

Article

# Developing subnational multivariate housing by ethnicity statistics from administrative data, England and Wales: 2020

Feasibility research on producing housing by ethnicity statistics for England and Wales from administrative data.

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

Release date:  
16 February 2023

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 by ethnicity dataset for 2020](#)
4. [Population coverage for admin-based housing by ethnicity statistics](#)
5. [Exploring accommodation type by ethnic group](#)
6. [Developing subnational multivariate housing by ethnicity statistics from administrative data: England and Wales, 2020: Data](#)
7. [Glossary](#)
8. [Data sources and quality](#)
9. [Future developments](#)
10. [Related links](#)
11. [Cite this article](#)

# 1 . Main points

- By combining admin-based housing and admin-based ethnicity datasets, we established that the proportion of addresses that had at least one property characteristic, at least one person resident, and at least one person with a stated ethnicity, was 79.0% for England and 78.1% for Wales.
- Regionally, address-level coverage was lowest in London (69.4%) and highest in the North East (82.4%).
- The proportion of individuals with at least one property characteristic and a stated ethnicity in the Statistical Population Dataset version 3.0 (SPD V3.0) was 81.7% for England and 85.3% for Wales.
- Accommodation type varies by address-level ethnic group, both nationally and regionally, reflecting the different types of accommodation available in different regions.
- Official statistics on accommodation type by ethnic group are not routinely published between censuses; this research has the potential to provide more timely data, particularly for sub-regional geographies.
- Future work will include updating the admin-based housing by ethnicity dataset (ABHED) to a 2021 reference period, so we can make comparisons with Census 2021 housing by ethnicity data when it is released, as well as research to improve the methods that are used to produce the ABHED.

These are not official statistics and should not be used for policy or decision-making. They are published as feasibility research into a new method for producing subnational multivariate statistics on housing by ethnic group using administrative data.

## 2 . About our transformation research

At the Office for National Statistics (ONS), we are exploring the feasibility of producing statistics on a range of topics using administrative data sources. This might remove the need for us to collect data through a census or surveys.

This article presents feasibility research on the potential to produce subnational multivariate housing by ethnicity statistics to demonstrate our progress towards producing more frequent subnational multivariate statistics on population characteristics. We explore the coverage of our linked dataset across the housing and ethnicity variables at national and subnational levels.

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 by ethnicity dataset for 2020

To create the admin-based housing by ethnicity dataset (ABHED), we linked the [admin-based ethnicity dataset version 3.0](#) (ABED V3.0) for 2020, the admin-based household estimates version 3.0 (ABHE V3.0) for 2020, and the [admin-based housing stock version 1.0](#) (ABHS V1.0) dataset for 2020. The data sources are linked together using pseudonymised identifiers. In producing statistics using linked administrative data, particularly for small populations, we apply the same rigour in data security and privacy as with all official statistics. For further information about the security of these linked data, see our [Population and social statistics transformation: 2019 progress update](#).

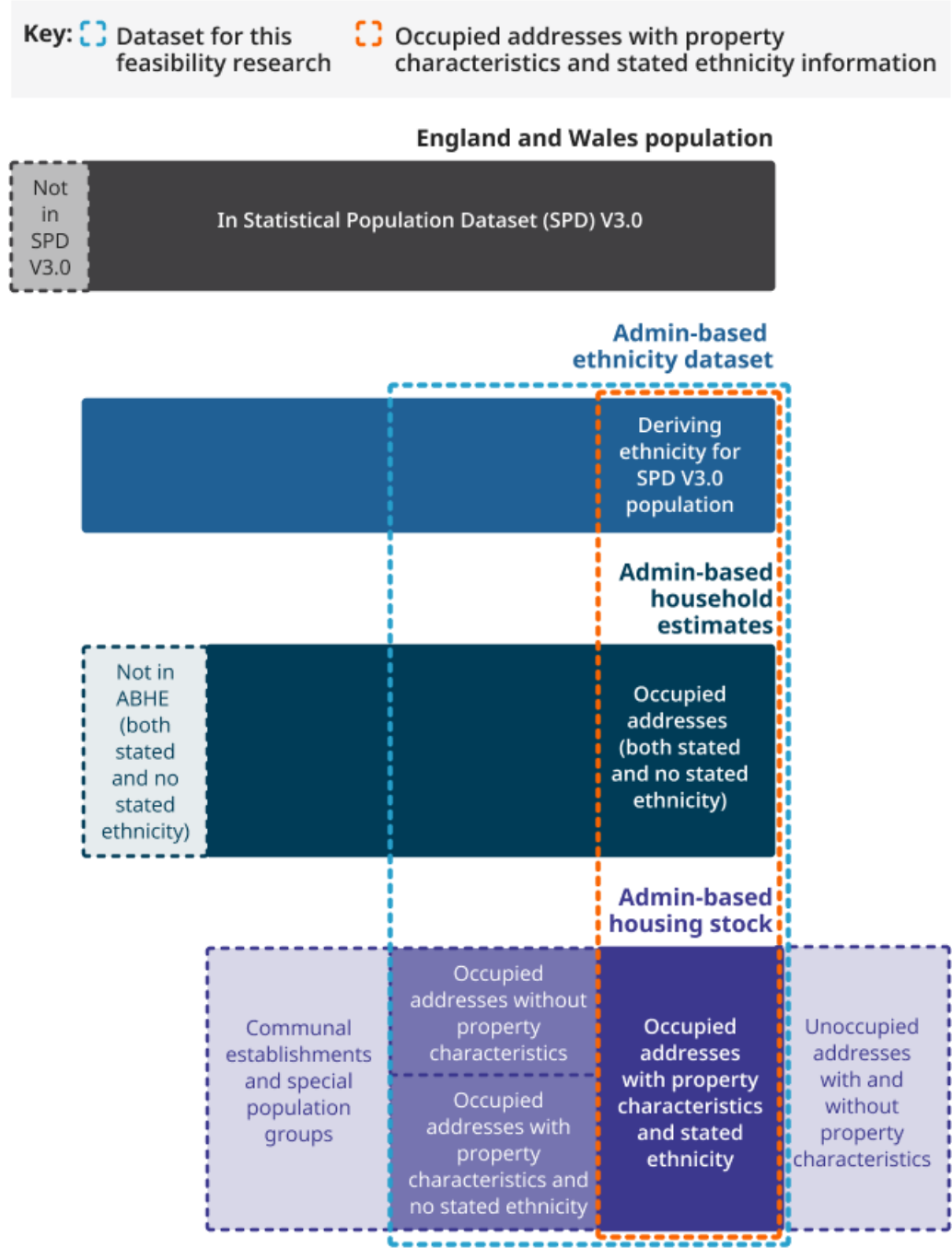
The ABHE V3.0 and ABED V3.0 are derived from multiple administrative data sources, and both use the [Statistical Population Dataset version 3.0](#) (SPD V3.0) as the population base. The SPD V3.0 is a record-level dataset, which includes individuals that meet one, or more, "activity-based" rules, meaning they were considered part of the usually resident population in England and Wales. Because they are both derived from the SPD V3.0, the ABHE V3.0 was linked to the ABED V3.0 using a unique identifier. This assigned a unique property reference number (UPRN) to as many individuals as possible in ABED V3.0. This then allowed individuals in the ABED V3.0 to be grouped into addresses using UPRN. To obtain property characteristics for each UPRN, the ABHS V1.0 was joined to the linked datasets using UPRN. The ABHS V1.0 provides information about residential addresses, so communal establishments and special population groups are removed from our dataset. This step removed 1.0% of records from the ABHED.

In this article, all the individuals at an address are considered to form one household; however, it should be noted that there may be more than one household at each UPRN. Because this is a different definition of household compared with the one used by the [Government Statistical Service \(GSS\) or 2011 Census and social surveys](#), we refer to "occupied addresses", rather than households, in this report. We have used the ethnicity variable from ABED V3.0, which makes use of 2011 Census data, in addition to multiple administrative data sources, to derive an individual's ethnic group. For more information on this method, see our [Changes to data and methods article](#) on this topic.

The ABHED for 2020 provides a dataset with multiple people per address, some with a stated ethnicity and some with no stated ethnicity. To produce an address-level analysis of housing by ethnicity, we have produced an initial approach which derives an ethnic group for an occupied address. Our research to develop this approach is outlined in our [Occupied address-level ethnicity measures for multivariate statistics, England and Wales: 2020](#) article.

Figure 1 provides a visual representation of how the ABHS V1.0 is linked to the ABHE V3.0 and ABED V3.0.

**Figure 1: Visual representation of the population coverage of the ABHED dataset, 2020**



Source: Office for National Statistics

**Notes:**

1. This diagram is not to scale.
2. 'Stated ethnicity' refers to those with a stated ethnicity and no refusal on their most recent administrative data record in 2020, in line with the methods used to derive an individual's ethnic group in the ABED V3.0 dataset.
3. 'No stated ethnicity' refers to the ethnicity being recorded as refused or unknown, in line with the methods used to derive an individual's ethnic group in the ABED V3.0 dataset. No stated ethnicity includes individuals who are in the SPD V3.0 but have not been linked to any sources of ethnicity data.
4. 'Property characteristics available' refers to properties that had information on at least one of accommodation type, number of rooms, number of bedrooms, and floor space in the ABHS V1.0.

The linkage of datasets with records at both the individual and the address (identified by UPRN) level means this article includes statistics on both.

## **4 . Population coverage for admin-based housing by ethnicity statistics**

### **Address-level coverage - national and regional**

Table 1 displays the coverage analysis of the admin-based housing by ethnicity dataset (ABHED) 2020. It shows that 79.0% of addresses in England and 78.1% in Wales had at least one property characteristic, at least one person resident, and at least one person with a stated ethnicity.

The ABHED data showed that 13.0% of addresses in England and 14.7% in Wales were identified as unoccupied. This was lower than the proportions identified by the admin-based housing stock version 1.0 (ABHS V1.0), where 13.6% of addresses in England and 15.2% in Wales were identified as unoccupied. This difference is because there were extra addresses in the ABHED that had been identified via the admin-based household estimates version 3.0 (ABHE V3.0) but not linked to ABHS V1.0.

Note that the proportion of unoccupied addresses in the ABHED was also higher than in the 2011 Census. This could have been caused by a combination of:

- the different reference periods
- undercoverage within the Statistical Population Dataset version 3.0 (SPD V3.0) that feeds into the ABHE V3.0
- assigning the incorrect address to individuals in the ABHE V3.0, like when an individual moves house but does not update their address on administrative data sources
- a failure to assign any address to an individual in the ABHE V3.0, so some properties identified as unoccupied within the ABHE V3.0 may be occupied with individuals who could not be assigned to an address.

Census 2021 estimates of unoccupied addresses have not yet been released; this analysis will be updated once the data are available.

Table 1 shows that at the regional level, the proportion of addresses that were occupied with at least one property characteristic and at least one stated ethnicity was lowest in London (69.4%) and highest in the North East (82.4%). These differences are partly caused by London having a higher proportion of flats compared with other regions. There are challenges involved in correctly assigning unique property reference numbers (UPRNs) to individuals within a block of flats, making it more difficult to identify occupied addresses. Further discussion of these trends can be found in our [Developing admin-based housing stock statistics for England and Wales: 2020 article](#).

Table 1: London has the lowest proportion of occupied addresses with at least one property characteristic and at least one stated ethnicity

Proportion of occupied and unoccupied addresses with at least one property characteristic and at least one stated ethnicity at the national and regional level, England and Wales, 2020

Region	Occupied addresses			Unoccupied addresses		
	Property characteristics and stated ethnicity	Property characteristics only	Stated ethnicity only	Property characteristics unknown and no stated ethnicity	Property characteristics available	Property characteristics unknown
England	79.0	5.0	2.6	0.3	12.3	0.7
North East	82.4	4.3	1.3	0.1	11.7	0.2
North West	82.0	3.8	2.0	0.2	11.4	0.7
Yorkshire and The Humber	80.8	4.3	2.2	0.3	10.9	1.5
East Midlands	82.3	4.2	2.5	0.2	10.1	0.7
West Midlands	81.9	4.3	2.3	0.3	10.7	0.5
East of England	82.1	3.9	2.1	0.2	11.1	0.5
London	69.4	6.7	4.7	0.8	17.4	1.0
South East	78.1	6.7	2.3	0.4	12.0	0.6
South West	78.6	4.5	3.0	0.3	13.0	0.6
Wales	78.1	4.4	2.4	0.3	14.0	0.7

Source: Office for National Statistics

#### Notes

1. Proportions may not sum to 100.0% because of rounding.
2. 'Stated ethnicity' refers to those with a stated ethnicity and no refusal on their most recent administrative data record in 2020.
3. 'No stated ethnicity' refers to the ethnicity being recorded as refused or unknown, in line with the methods used to derive an individual's ethnic group in the ABED V3.0. This includes individuals who are in the SPD V3.0 but have not been linked to any sources of ethnicity data.
4. 'Property characteristics available' refers to properties that had information on at least one of accommodation type, number of rooms, number of bedrooms and floor space in the ABHS V1.0.

## Local authority (LA) coverage

As displayed in Figure 2, coverage at LA level varies widely. The proportion of occupied addresses that had at least one property characteristic and at least one stated ethnicity within the ABHED linked dataset in England was lowest in City of London (37.9%), followed by Westminster (39.7%), and Kensington and Chelsea (39.8%), and highest in Rochford (88.5%).

There was less variation in coverage at LA level in Wales. Figure 2 shows that the proportion of occupied addresses with at least one property characteristic and at least one stated ethnicity ranged from 65.9% in Gwynedd to 84.5% in Flintshire.

## Figure 2: Address-level coverage of housing and ethnicity varies by local authority

Local authorities by proportion of occupied addresses with at least one property characteristic and at least one stated ethnicity, England and Wales, 2020

### Notes:

1. 'Occupied addresses' refers to addresses with a person living there.
2. 'Stated ethnicity' refers to those with a stated ethnicity and no refusal on their most recent administrative data record in 2020.
3. 'No stated ethnicity' refers to the ethnicity being recorded as refused or unknown, in line with the methods used to derive an individual's ethnic group in the ABED V3.0. This includes individuals who are in the SPD V3.0 but have not been linked to any sources of ethnicity data.
4. 'Property characteristics' refers to properties that had information on at least one of accommodation type, number of rooms, number of bedrooms, and floor space in the ABHS V1.0.

### Download the data

[.xlsx](#)



## Lower layer super output area (LSOA) coverage

Table 2 shows that the proportion of addresses that had at least one person living there, at least one property characteristic, and at least one person with a stated ethnicity varied widely across LSOAs within England and Wales. Just over half of LSOAs in both England (53.7%) and Wales (53.6%) had between 80% and 89% of addresses as occupied, with at least one property characteristic and at least one person with a stated ethnicity.

Table 2: Just over half of LSOAs in England and Wales have between 80% and 89% of addresses with at least one person living there, at least one property characteristic, and at least one person with a stated ethnicity  
Counts and proportions of LSOAs within each coverage bracket, England and Wales, 2020

Coverage bracket	England		Wales	
	Counts	Percentage	Counts	Percentage
49% or less	730	2.2	28	1.5
50 - 59%	942	2.9	56	2.9
60 - 69%	2243	6.8	172	9.0
70 - 79%	6958	21.2	519	27.2
80 - 89%	17636	53.7	1023	53.6
90% or more	4335	13.2	111	5.8

Source: Office for National Statistics

### Notes

1. 'Coverage bracket' refers to the range of percentages for the address-level coverage. For example an LSOA that had 72% of addresses with at least one person living there, at least one property characteristic and at least one person with a stated ethnicity would fall within the 70 to 79% bracket.
2. 'Stated ethnicity' refers to those with a stated ethnicity and no refusal on their most recent administrative data record in 2020.
3. 'No stated ethnicity' refers to the ethnicity being recorded as refused or unknown, in line with the methods used to derive an individual's ethnic group in the ABED V3.0. This includes individuals who are in the SPD V3.0 but have not been linked to any sources of ethnicity data.
4. 'Property characteristics available' refers to properties that had information on at least one of accommodation type, number of rooms, number of bedrooms and floor space in the ABHS V1.0.
5. The LSOAs in this analysis were defined according to the 2011 boundaries set out by the Office for National Statistics (ONS) Geography portal. Findings may differ slightly depending on when the boundaries are drawn.

## Individual level coverage

Coverage analysis of the ABHED linked dataset at an individual level showed that in England, 81.7% of individuals in the SPD V3.0 had a stated ethnicity and at least one property characteristic. In Wales, the equivalent figure was 85.3% of individuals. Within the SPD V3.0, the proportion of individuals with a stated ethnicity and at least one property characteristic was highest for children and lowest for adults of working age. This trend was mainly driven by the availability of ethnicity data by age, which is described in our [Developing admin-based ethnicity statistics for England and Wales: 2020 article](#). Further breakdowns of individual level coverage are included in our [accompanying dataset](#).

## 5 . Exploring accommodation type by ethnic group

As part of our feasibility research on the potential to produce subnational multivariate housing by ethnicity statistics using administrative data, we explored three approaches to measuring occupied address-level ethnicity. For this initial analysis, we have used an admin-based version of the household reference person (HRP) from the census, which we refer to as the address reference person (ARP). We recognise that using the ethnic group of one person to represent an occupied address is problematic, as it conceals information particularly where there are two, or more, ethnic groups within the household. This approach allows a comparison with statistics from the 2011 Census; this comparison will be updated once the equivalent statistics from Census 2021 have been released. Official statistics on accommodation type by ethnic group are not routinely published between censuses. This research has the potential to provide more timely data, particularly for sub-regional geographies. A full discussion of our research to produce a measure of address-level ethnic group can be found in our [Occupied address-level ethnicity measures for multivariate statistics, England and Wales: 2020](#) article.

In this article, we present initial analysis of Valuation Office Agency (VOA) accommodation type (2011 Census definition, see [Section 7. Glossary](#)) by five-category ethnic group of the admin-based ARP. This is done as an initial case study of the potential for using admin data sources to produce housing by ethnicity statistics. The analysis is based on occupied addresses. Note that our admin-based dataset does not have complete coverage. These are not official statistics and should not be used for policy or decision-making.

## England and Wales

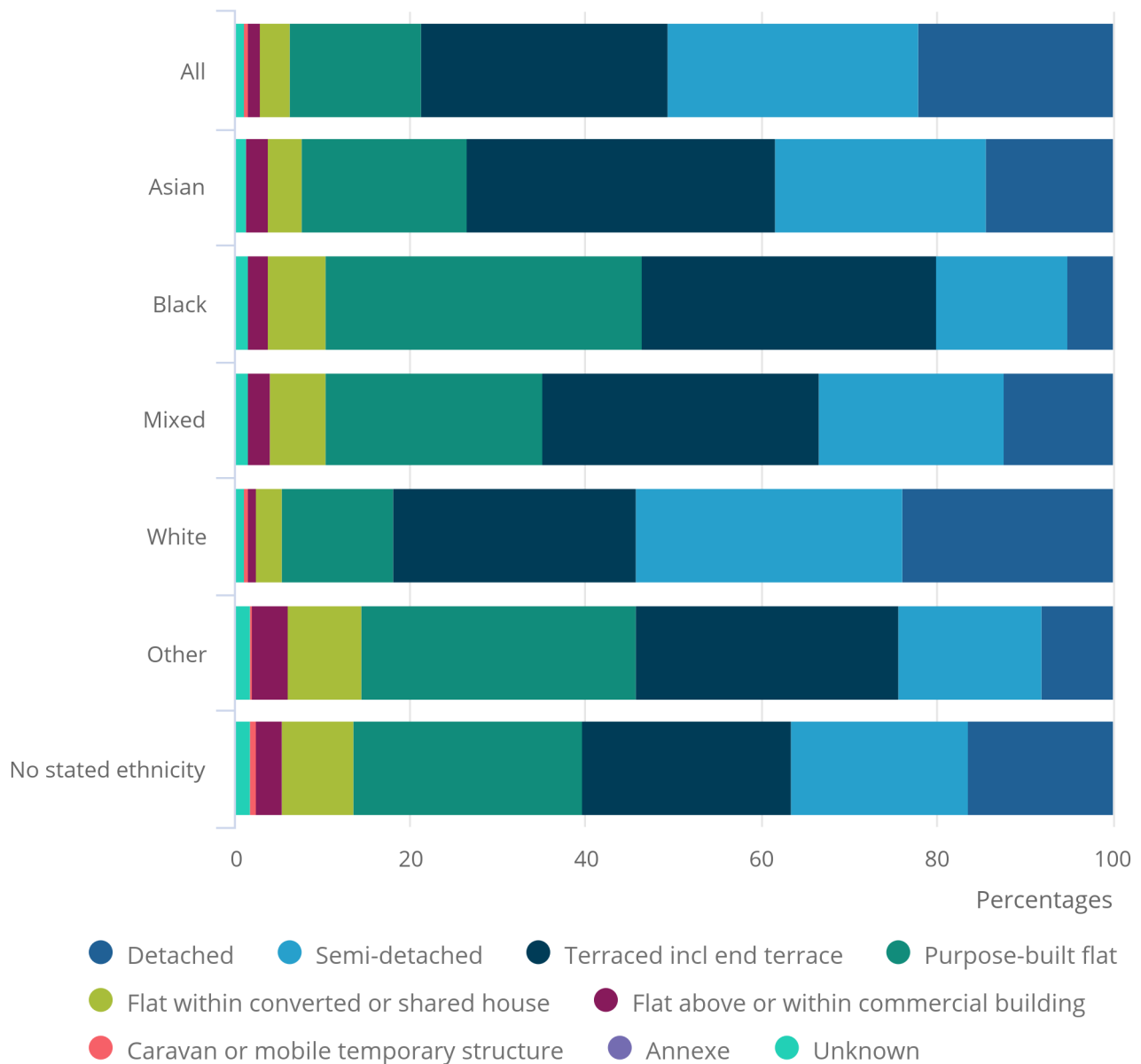
Figure 3 shows that across England, the most common accommodation types for Asian ARPs and White ARPs were semi-detached and terraced houses. For Black and Mixed ARPs and ARPs from the Other ethnic group, the most common types of accommodation were terraced houses and purpose-built flats. Of all the groups, White ARPs were most likely to live in detached houses. Figures for accommodation type by ethnic group of the HRP from the 2011 Census (as seen in Figure 4) showed broadly similar patterns for all ethnic groups, apart from the Other ethnic group. HRPs from the Other ethnic group in the 2011 Census were more likely to live in purpose-built flats than ARPs from the Other ethnic group in the 2020 admin-based housing by ethnicity dataset (ABHED).

**Figure 3: Accommodation type varies by ethnic group of the address reference person (ARP) in the admin-based housing by ethnicity dataset (ABHED)**

Proportion of occupied addresses by accommodation type and five-category ethnic group of the admin-based address reference person, England, 2020

Figure 3: Accommodation type varies by ethnic group of the address reference person (ARP) in the admin-based housing by ethnicity dataset (ABHED)

Proportion of occupied addresses by accommodation type and five-category ethnic group of the admin-based address reference person, England, 2020



Source: Office for National Statistics

Notes:

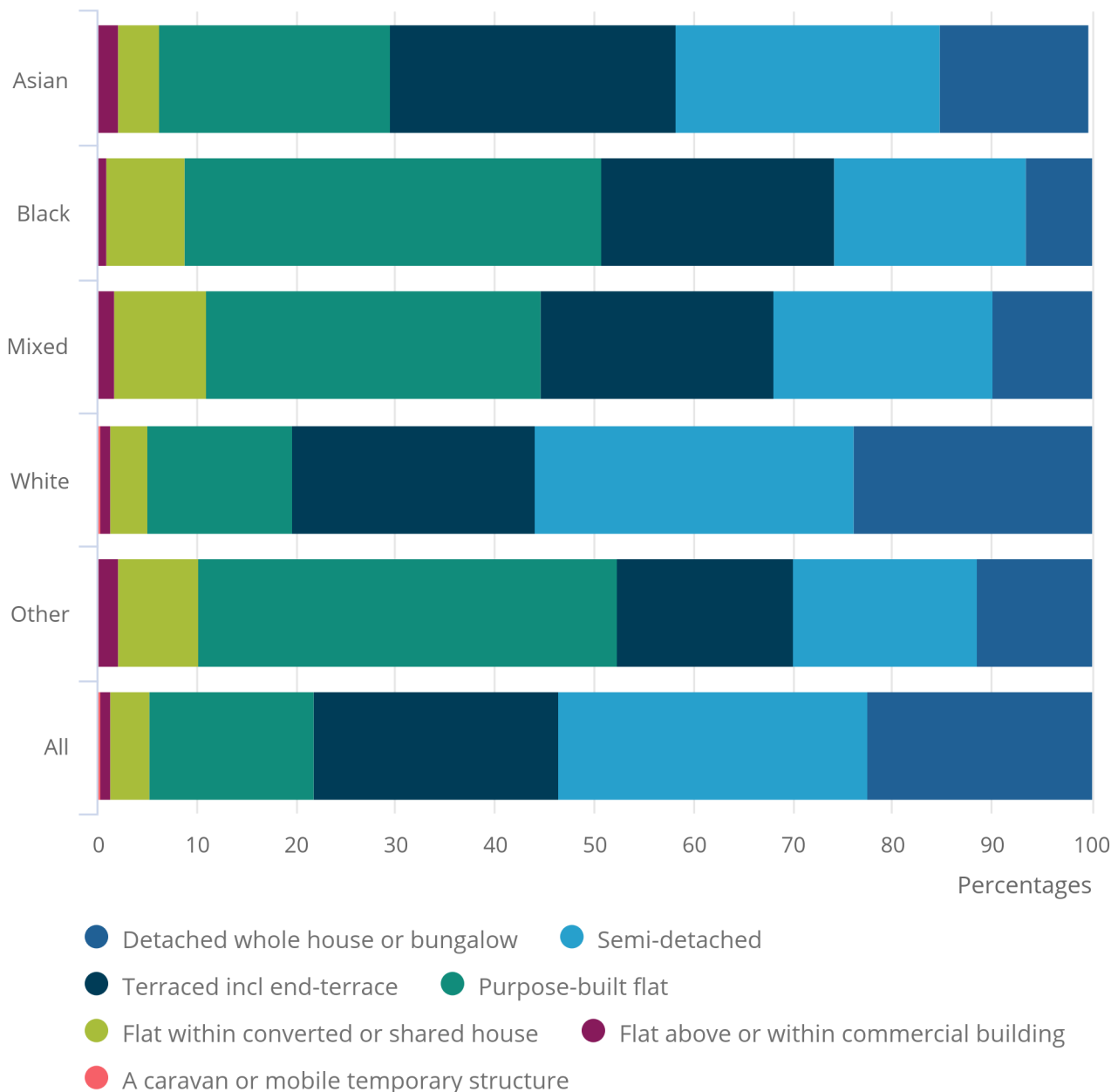
1. Proportions may not sum to 100.0% because of rounding.
2. 'No stated ethnicity' refers to the ethnicity being recorded as refused or unknown, in line with the methods used to derive an individual's ethnic group in the ABED V3.0. This includes individuals who are in the SPD V3.0 but have not been linked to any sources of ethnicity data.
3. 'Property characteristics available' refers to properties that had information on at least one of accommodation type, number of rooms, number of bedrooms, and floor space in the ABHS V1.0.

**Figure 4: Accommodation type varies by the ethnic group of the 2011 Census household reference person**

Ethnic group of the household reference person by accommodation type from the 2011 Census, England

Figure 4: Accommodation type varies by the ethnic group of the 2011 Census household reference person

Ethnic group of the household reference person by accommodation type from the 2011 Census, England



Source: Office for National Statistics

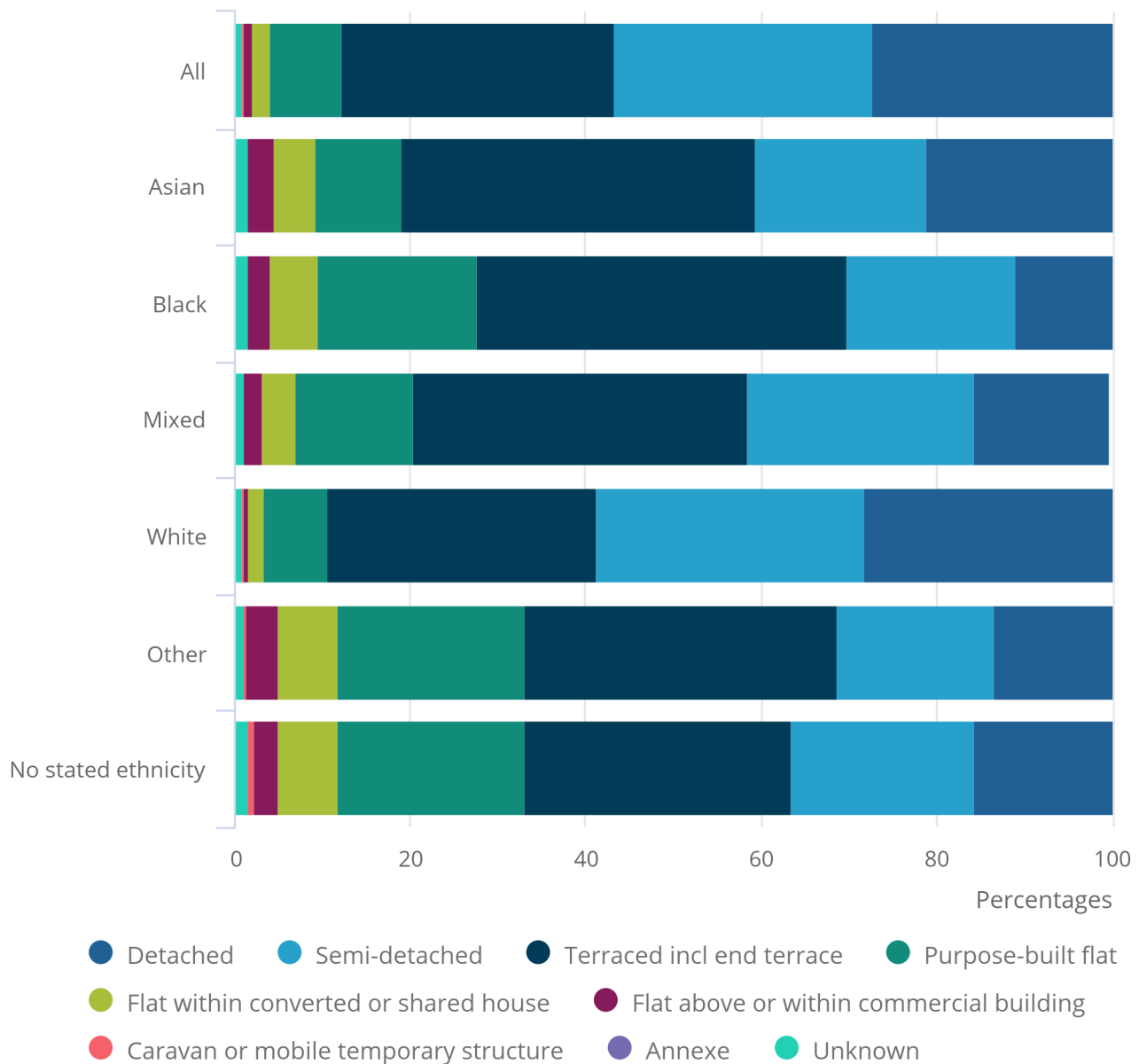
Figure 5 shows that across Wales, the most common accommodation type for admin-based ARPs of all ethnic groups was terraced houses. White ARPs were most likely to live in detached houses. Comparison with figures for Wales from the 2011 Census (seen in Figure 6) showed higher proportions of admin-based ARPs in the ABHED living in terraced houses, compared with HRPs in the 2011 Census. This was true for all ethnic groups.

**Figure 5: Terraced houses are the most common accommodation type across all ethnic groups of the admin-based ARP**

Proportion of occupied addresses by accommodation type and five-category ethnic group of the admin-based ARP, Wales, 2020

Figure 5: Terraced houses are the most common accommodation type across all ethnic groups of the admin-based ARP

Proportion of occupied addresses by accommodation type and five-category ethnic group of the admin-based ARP, Wales, 2020



Source: Office for National Statistics

Notes:

1. Proportions may not sum to 100.0% because of rounding.
2. 'No stated ethnicity' refers to the ethnicity being recorded as refused or unknown, in line with the methods used to derive an individual's ethnic group in the ABED V3.0. This includes individuals who are in the SPD V3.0 but have not been linked to any sources of ethnicity data.
3. 'Property characteristics available' refers to properties that had information on at least one of accommodation type, number of rooms, number of bedrooms, and floor space in the ABHS V1.0.

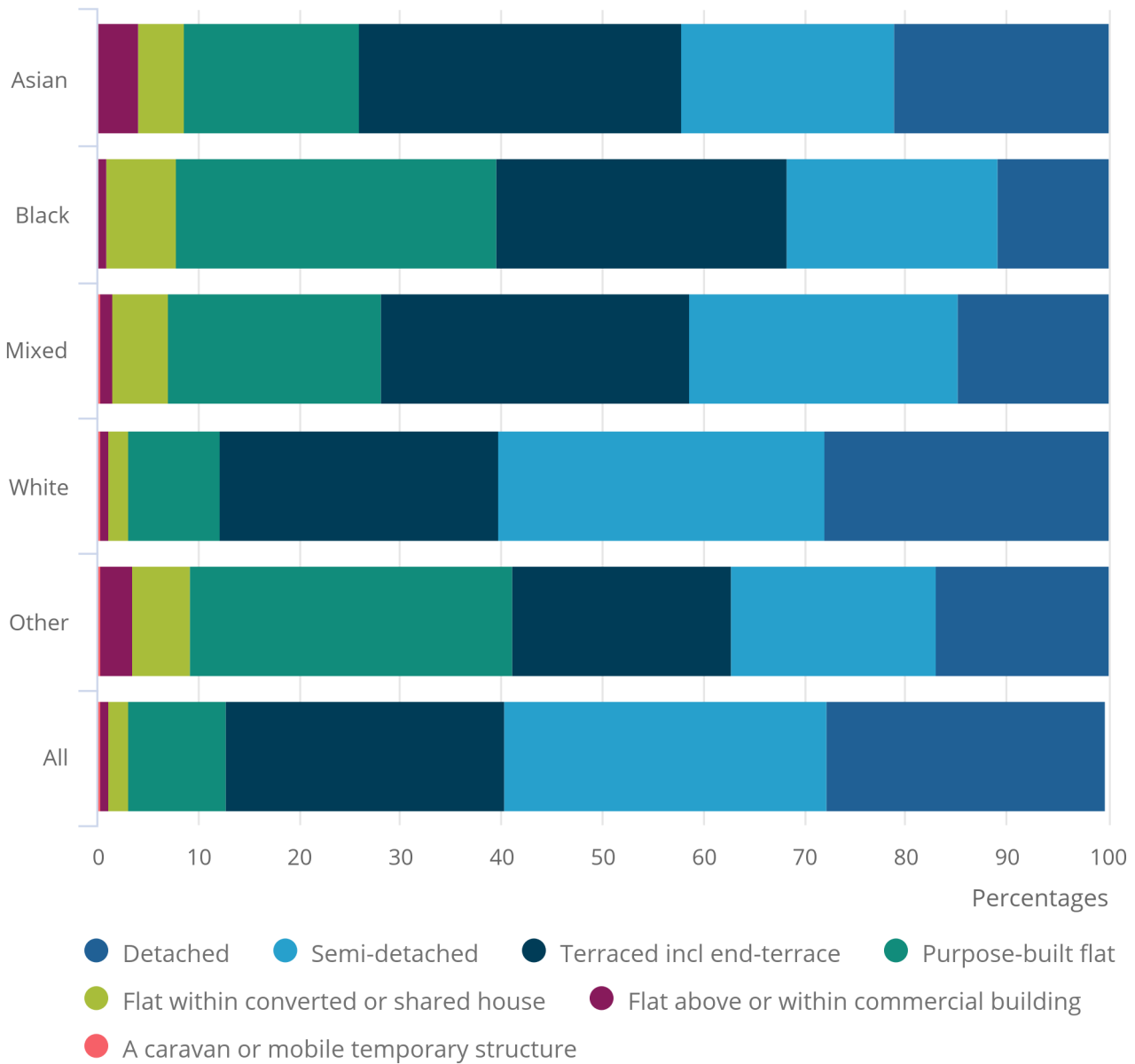


**Figure 6: Accommodation type varies by ethnic group of the 2011 Census household reference person (HRP)**

Ethnic group of HRP by accommodation type from the 2011 Census, Wales

Figure 6: Accommodation type varies by ethnic group of the 2011 Census household reference person (HRP)

Ethnic group of HRP by accommodation type from the 2011 Census, Wales



Source: Office for National Statistics

## Regional statistics

Figure 7 shows that London was the region with the highest proportions of admin-based ARPs living in purpose-built flats across all ethnic groups, ranging from 31.7% for White ARPs to 47.8% for Black ARPs. White ARPs were most likely of all ethnic groups to live in detached houses in all regions. Black ARPs were most likely of all ethnic groups to live in purpose-built flats in all regions apart from the North West, where it was ARPs from the Other ethnic group. The 2011 Census (figures not shown) showed broadly similar patterns for HRPs, apart from for purpose-built flats. HRPs from the Other ethnic group were most likely in five of the regions, and Black HRPs were most likely in the remaining four regions, to live in purpose-built flats. Some of these differences can be explained by the different average (median) ages of admin-based ARPs across the different ethnic groups. [Census 2021](#) provides data about average (median) age by ethnic group for the entire usually resident population.

### Figure 7: Accommodation type by ethnic group of the admin-based ARP varies by region

Proportion of occupied addresses by accommodation type and five-category ethnic group of the admin-based ARP, regions of England, 2020

#### Notes:

1. Proportions may not sum to 100.0% because of rounding.
2. 'No stated ethnicity' refers to the ethnicity being recorded as refused or unknown, in line with the methods used to derive an individual's ethnic group in the ABED V3.0. This includes individuals who are in the SPD V3.0 but have not been linked to any sources of ethnicity data.
3. 'Other flat' includes flats within converted or shared houses and flats above or within commercial buildings.
4. 'All Other' includes caravans or mobile or temporary structures, annexes, and unknown property characteristics.

#### Download the data

[.xlsx](#)

## Sub-regional statistics

Accommodation type by five-category ethnic group of admin-based ARP tables have been produced for Local Authorities (LAs) and lower layer super output areas (LSOAs) in England and Wales. These can be found in the [accompanying dataset](#). Small population sizes in some geographical areas, and potential localised factors, mean that the ABHED data for LAs and LSOAs are likely to be affected more by coverage issues and accuracy of the administrative data. Where an ethnic group is small in an LA or LSOA, the value in the accompanying table has been suppressed for statistical disclosure control (SDC). This is in accordance with the approaches currently required by administrative data suppliers, which vary from those used for other statistics (for example, Census 2021). A full explanation of the SDC approach used can be found in the notes of the accompanying datasets.

In England, the LA with the highest proportion of admin-based ARPs of all ethnic groups living in purpose-built flats was Tower Hamlets, ranging from 72.0% for White ARPs to 83.0% for Black ARPs. The LA with the highest proportion of ARPs of all ethnic groups living in detached houses was North Kesteven (in the East Midlands), ranging from 50.0% for ARPs from the Other ethnic group to 59.6% for Asian ARPs. The LA with the highest proportion of ARPs of all ethnic groups living in terraced houses was Pendle (in the North West), ranging from 54.0% for White ARPs to 78.3% for ARPs from the Other ethnic group.

In Wales, no LAs had purpose-built flats as the most common accommodation type for admin-based ARPs. Blaenau Gwent had the highest proportion of ARPs of all ethnic groups living in terraced houses, ranging from 41.7% for Asian ARPs to 59.3% for Mixed ARPs. The Isle of Anglesey had the highest proportion of ARPs of all ethnic groups living in detached houses, ranging from 24.0% for Mixed ARPs to 46.1% for White ARPs.

## 6 . Developing subnational multivariate housing by ethnicity statistics from administrative data: England and Wales, 2020: Data

[Developing subnational multivariate housing by ethnicity statistics from administrative data, England and Wales: 2020 data](#)

Dataset | Released 16 February 2023

Data for feasibility research on producing housing by ethnicity statistics for England and Wales from administrative data.

## 7 . Glossary

### Accommodation type

The accommodation type variable is derived from the Valuation Office Agency (VOA) property type and VOA dwelling code variables. These are to resemble the census accommodation type 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, that has been constructed or adapted for use as separate living accommodation.

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](#).

### Admin-based address reference person (ARP)

The admin-based address reference person (ARP) is a variable derived from administrative data. It is designed to serve the same purpose as the census-based household reference person (HRP) variable. An individual within an occupied address acts as a reference point for producing further derived statistics and for characterising a whole occupied address according to the characteristics of the chosen reference person. We have created this variable as an admin data-based equivalent of the census-based household reference person. It is used for the purpose of this feasibility research to identify an individual whose stated ethnicity can represent each occupied address. In the analysis presented within this publication, we used the ARP2 approach, which is explained in our [Occupied address-level ethnicity measures for multivariate statistics, England and Wales: 2020](#) article.

### Communal establishments

A communal establishment is an establishment providing managed residential accommodation. "Managed", in this context, means full-time or part-time supervision of the accommodation. Communal establishments include:

- sheltered accommodation units (including homeless temporary shelter)
- hotels
- guest houses
- B&Bs, inns, and pubs
- all accommodation provided solely for students (during term-time)

More information is available in our [2011 Census glossary](#).

## Ethnic group

The self-reported ethnic group of the individual, according to their own perceived ethnic group and cultural background.

Five categories are presented in this article. The ethnic groups included in each category are:

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

## Household reference person (HRP)

In the census, household reference persons provide an individual person within a household to act as a reference point for producing further derived statistics. They also characterise a whole household according to characteristics of the chosen reference person. The full definition used can be found in our [2011 Census Glossary](#). In this analysis, we have derived an ARP variable from administrative data designed to serve the same purpose as the HRP variable from the census, as described in the ARP glossary entry.

## No stated ethnicity

No stated ethnicity refers to the ethnicity being recorded as refused or unknown, in line with the [methods used to derive an individual's ethnic group in the admin-based ethnicity dataset version 3.0](#) (ABED V3.0). No stated ethnicity also includes individuals who are in the Statistical Population Dataset version 3.0 (SPD V3.0) but have not been linked to any sources of ethnicity data.

## Occupied address

For this research, an occupied address is a unique property reference number (UPRN) on the Address Frame which has been successfully linked to at least one individual in the [Statistical Population Dataset version 3.0 \(SPD V3.0\)](#). It is different to the concept of a household, which uses a definition based on shared facilities. More information on the differences between a traditional "household" and an "occupied address" is available in our [Occupied address \(household\) estimates from Administrative Data: 2011 and 2015 article](#).

## Special population groups

Special population groups include armed forces personnel and dependants stationed in the UK, foreign armed forces based in the UK (mainly US Air Force personnel and dependants, and the prison population).

## Stated ethnicity

Stated ethnicity refers to the ethnicity being recorded as a specific ethnic group and not refused or unknown on their most recent administrative data record in 2020, in line with the [methods used to derive an individual's ethnic group in the admin-based ethnicity dataset version 3.0](#) (ABED V3.0).

## Unique property reference number (UPRN)

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

## Notes for: Glossary

1. The Gypsy, Roma and Irish Traveller ethnic groups have been aggregated because of differences in response options across data sources, meaning that it is not possible to separate them. Hospital Episode Statistics (HES) and Improving Access to Psychological Therapies (IAPT) do not include any Gypsy, Roma or Irish Traveller response options.

2. The Higher Education Statistics Agency (HESA) data for England and Wales only have categories for White and Gypsy or Traveller within the higher-level White ethnic group. Those with a sub-category ethnic group of White in HESA were recoded as White not specified.

## 8 . Data sources and quality

### Population base

The [Statistical Population Dataset version 3.0](#) (SPD V3.0) was used as the population base for the [admin-based ethnicity dataset version 3.0](#) (ABED V3.0) for 2020, and the admin-based household estimates version 3.0 (ABHE V3.0) for 2020. The SPD V3.0 is a record-level dataset, which includes individuals that meet one or more "activity-based" rules. This means they were deemed to be part of the usually resident population in England and Wales as of 30 June 2020. The quality of the population base, as explained in our [Admin-based population estimates and statistical uncertainty: July 2020 article](#), will have an impact on the quality of the ABED V3.0 and ABHE V3.0. More information about the coverage of the population base can be found in our [article on Population and migration statistics system transformation](#).

### Admin-based household estimate (ABHE)

The ABHEs are derived from the Statistical Population Datasets (SPDs). 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. To create the ABHE V3.0, the SPD V3.0 successfully assigns a UPRN to 98.3% of usual residents directly from the personal demographic service (PDS) data.

Our ability to accurately identify occupied addresses depends on the quality and coverage of the SPD V3.0. It also depends on 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 dataset

The [admin-based housing stock version 1.0](#) (ABHS V1.0) dataset for 2020 brings together data from several administrative sources. The aim is to develop a new method for producing more regular census-like statistics for occupied residential addresses (down to small geographies) across England and Wales. The ABHS V1.0 2020 was produced by linking a residential Address Frame from June 2020 to Valuation Office Agency (VOA) data from June 2020 and the ABHE versions 2.0 and 3.0 (ABHE V2.0 and V3.0). To align more closely with the 2011 Census definition of a household, communal establishments were removed from the Address Frame. A more detailed description of how we developed the ABHS V1.0 and assessed its quality can be found in our [Developing admin-based housing stock statistics for England and Wales: 2020 article](#).

### Admin-based ethnicity dataset

The ABED V3.0 was produced using the following administrative data sources:

- [English School Census \(ESC\), 2011 to 2020](#)
- [Hospital Episode Statistics \(HES\), 2009 to 2020](#)
- [Emergency Care Data Set \(ECDS\), 2020](#)
- [Improving Access to Psychological Therapies \(IAPT\), 2012 to 2018](#)
- [Higher Education Statistics Agency \(HESA\), 2010 to 2020](#)
- [Birth Notifications, 2006 to 2020](#)
- [Individualised Learner Record \(ILR\), 2008 to 2020](#)
- [Welsh School Census \(WSC\), 2011 to 2020](#)

For more information about these data sources, see our [Developing admin-based ethnicity statistics for England and Wales: 2020 article](#).

Ethnicity records from these data sources were linked to the 2020 SPD V3.0, using unique identifiers. A method to select a final ethnicity per person was then implemented, as described in our [Producing admin-based ethnicity statistics for England: changes to data and methods methodology](#).

## Feedback

We welcome feedback on this research and our planned future developments. Please email your feedback to [admin.based.characteristics@ons.gov.uk](mailto:admin.based.characteristics@ons.gov.uk), including "Housing by ethnicity" in the subject line.

## 9 . Future developments

This feasibility study of developing an admin-based housing and ethnicity dataset (ABHED) for 2020 shows the potential to produce outputs in intercensal years at lower levels of geography and for smaller population groups than is currently possible from surveys alone.

Further work is required to explore how we can develop and improve upon subnational multivariate ABHED. Future work will include:

- producing the ABHED for 2021 and conducting comparisons to Census 2021
- improving the underlying ABHS dataset, including the way we identify occupied addresses
- exploring the potential for including admin-based income estimates and admin-based labour market statistics in the development of address reference person methods, with a view to aligning more closely to existing census household reference person definitions
- exploring methods for producing multivariate population statistics using administrative and survey sources, building on research set out in our research paper [Methods for producing multivariate population statistics using administrative and survey sources](#).

## 10 . Related links

### [Developing admin-based housing stock statistics for England and Wales: 2020](#)

Article | Released 5 December 2022

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

### [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.

### [Developing subnational multivariate income by ethnicity statistics from administrative data, England: tax year ending 2016](#)

Article | Released 31 August 2022

Feasibility research on an initial case study producing income by ethnic group statistics for England from administrative data.

### [Developing admin-based ethnicity statistics for England and Wales: 2020](#)

Article | Released 7 February 2023

Research updates on producing population statistics by ethnic group for England and Wales from administrative data, with comparisons with Census 2021 estimates.

### [Change over time in admin-based ethnicity statistics, England: 2016 to 2020](#)

Article | Released 23 May 2022

Feasibility research on producing statistics on the population by ethnic group for England for 2016 to 2020 from administrative data, exploring change over time. This builds on work to improve the 2016 admin-based ethnicity statistics. These research outputs are not official statistics.

### [Admin-based population and migration estimates: research update](#)

Article | 26 November 2021

Provides an overview of all the latest research into the transformation of population and migration statistics to an administrative data first methodology.

## 11 . Cite this article

Office for National Statistics (ONS), released 16 February 2023, ONS website, article, [Developing subnational multivariate housing by ethnicity statistics from administrative data, England and Wales: 2020](#)