

Article

Developing Statistical Population Datasets, England and Wales: 2021

Aggregate comparisons between the Statistical Population Dataset version 4.0 (v4.0) and Census 2021.

Contact:
Jake Argent, Josh Best, Karolina
Chalupniczak, Vicky Collison,
Emma Hamilton, Emily Rhodes
pop.info@ons.gov.uk
+44 3000 682506

Release date:
28 February 2023

Next release:
To be announced

Table of contents

1. [Main points](#)
2. [Overview of the SPD](#)
3. [Age and sex comparisons with the census](#)
4. [England and Wales](#)
5. [Local authorities](#)
6. [Output Areas](#)
7. [Developing Statistical Population Datasets data](#)
8. [Glossary](#)
9. [Data sources and quality](#)
10. [Future developments](#)
11. [Related links](#)
12. [Cite this article](#)

1 . Main points

- We have produced the Statistical Population Dataset (SPD) for 2021 using our most recent method; our analysis focuses on comparisons with Census 2021 to understand the quality and challenges of using administrative data for population estimates at aggregate level.
- SPD version 4.0 (v4.0) is broadly in line with Census 2021 estimates, yet we find challenges with overcoverage in younger working ages and undercoverage in older working ages.
- There are differences in coverage patterns between England and Wales, which reflect the data sources currently available for each.
- There are considerable differences in coverage patterns across local authorities (LAs); our analysis explores factors that may contribute to these differences, such as high volume and frequency of movement, or high numbers of self-employed people in that LA.
- Our low-level output area (OA) analysis shows that the presence of specific populations, such as university students, may present challenges in allocating individuals to the correct address.
- To support the delivery of high-quality admin-based population estimates from the dynamic population model (DPM) in future, we will develop the SPD by exploring new approaches and data sources, as well as use our learnings to inform the development of a coverage adjustment method.

These are not official statistics and should not be used for decision making. They are estimates from a new methodology, which is different from that currently used to produce official population and migration statistics. The information and research in this article should be read alongside the estimates to avoid misinterpretation. These outputs must not be reproduced without this warning.

2 . Overview of the SPD

The Statistical Population Dataset (SPD) aims to approximate the usually resident population down to small areas with admin data. The SPD is produced independently for each year and therefore any errors in one year are less likely to be rolled forward to the next. Our research has shown the need to include a coverage adjustment to the SPD to reduce the coverage error and measure its quality.

The SPD will support the delivery of high-quality admin-based population estimates from the dynamic population model (DPM), as described in our [Dynamic population model, improvements to data sources and methodology for local authorities, England and Wales: 2011 to 2022 methodology](#). The DPM uses statistical modelling techniques and demographic insights alongside a range of data sources to produce coherent and timely estimates of the population and population change. Comparisons between census-based and admin-based estimates for 2021 are discussed in our [Transforming population statistics, comparing 2021 population estimates in England and Wales article](#). This provides guidance on how best to interpret and use each of the estimates.

This article is part of a series examining the quality of the SPD through comparisons with Census 2021. An accompanying article looks at [record-level linkage between Census 2021 and the SPD](#). We collected Census 2021 data on 21 March 2021, and this remains our best estimate of the population for this time. The SPD version 4.0 (v4.0) reference date is 30 June 2021. Comparisons between the two sources provide a unique insight into the quality of our SPD methodology.

3 . Age and sex comparisons with the census

Age

Comparisons with Census 2021 show that the Statistical Population Dataset version 4.0 (SPD v4.0) is broadly in line with official estimates for England and Wales, with SPD v4.0 2021 being 1.1% lower than the census. However, this hides patterns of overcoverage (where SPD estimates are higher than Census 2021) and undercoverage (where SPD estimates are lower than Census 2021).

We see fewer records in the SPD for those aged under four years, which may reflect a lack of interaction with services. This pattern reverses for school ages, showing higher coverage in SPD v4.0 for those aged 5 to 16 years (Figure 1).

There is a fall in coverage for those aged 16 to 23 years. We suspect this is linked to patterns of interaction at this age, reflecting transitions between education and employment, as described in our [Understanding quality of the Statistical Population Dataset in England and Wales using the 2021 Census - Demographic Index linkage article](#).

We see steady and slight overcoverage for those aged 24 to 34 years. As our activity-based approach includes those who are active within the year prior to the SPD reference date, we need to explore further how this may be leading to the inclusion of short-term residents in the SPD.

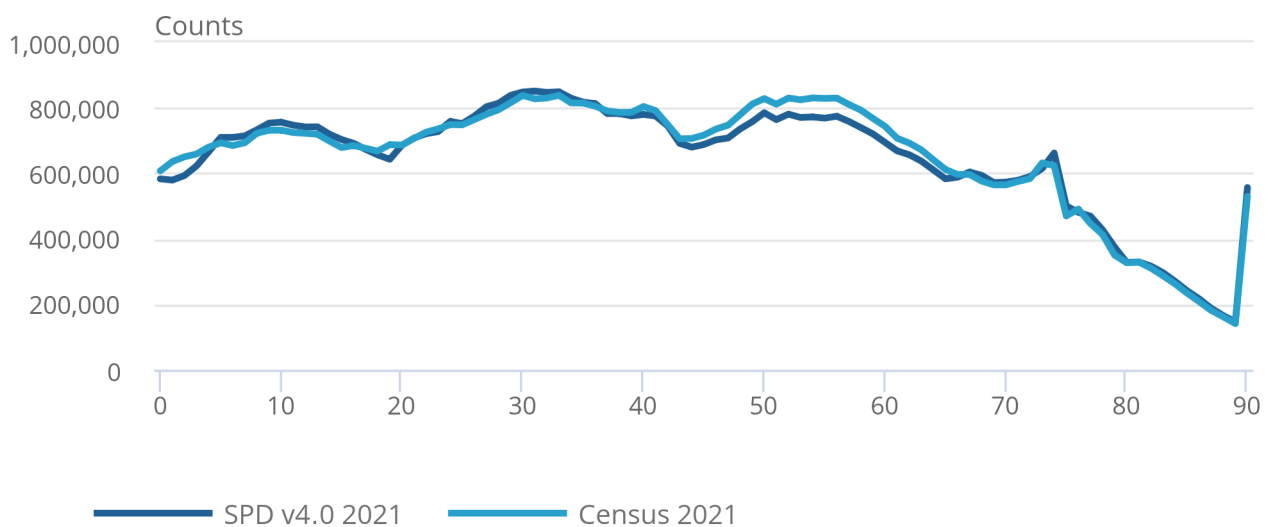
We tend to see lower counts in older working ages compared with the census. This may be because certain groups at this age are less likely to interact with services, such as those in early retirement, those living off a partner's income, or those who are self-employed (we do not currently include Self-assessment Tax data in the SPD).

Figure 1: There are age differences in coverage in the Statistical Population Dataset version 4.0

Population counts by age for the Statistical Population Dataset version 4.0 and Census 2021

Figure 1: There are age differences in coverage in the Statistical Population Dataset version 4.0

Population counts by age for the Statistical Population Dataset version 4.0 and Census 2021



Source: Office for National Statistics – Census 2021

Sex

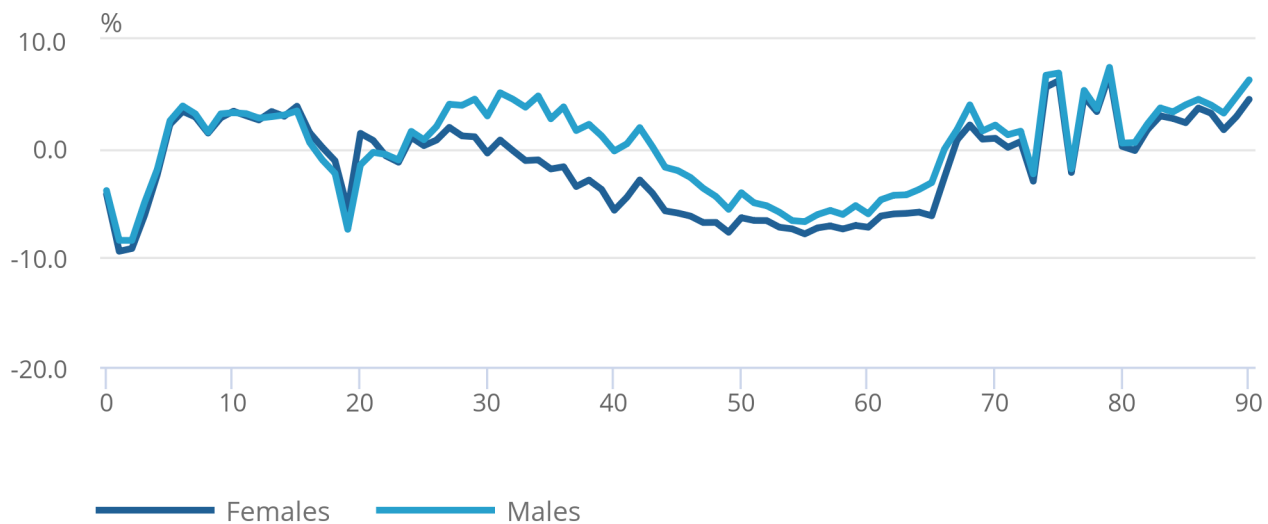
SPD v4.0 shows a tendency to count a higher proportion of males than females relative to Census 2021. This predominantly appears among working ages, and in particular younger working ages. An exception to this trend is for those aged 15 to 22 years, where SPD v4.0 shows higher counts for females relative to the census. This suggests a need to further understand the different interactions males and females have with services at different life stages.

Figure 2: There are differences in the coverage of males and females in the Statistical Population Dataset version 4.0

Statistical Population Dataset version 4.0 % difference from Census 2021 by sex and age, England and Wales, 2021

Figure 2: There are differences in the coverage of males and females in the Statistical Population Dataset version 4.0

Statistical Population Dataset version 4.0 % difference from Census 2021 by sex and age, England and Wales, 2021



Source: Office for National Statistics – Census 2021

4 . England and Wales

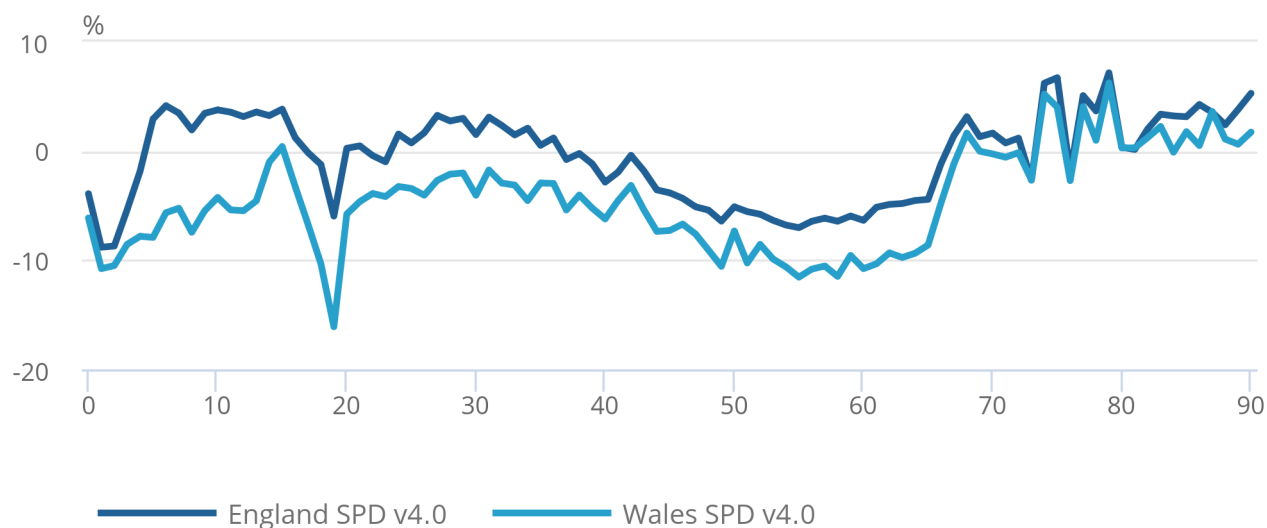
We compared the Statistical Population Dataset version 4.0 (SPD v4.0) for 2021 with Census 2021 for England and Wales to understand more about how coverage patterns differ by geography (Figure 3). SPD coverage for England is broadly in line with the census, being 0.9% lower than the census. The coverage for Wales is 5.2% lower than the census.

Figure 3: Statistical Population Dataset version 4.0 coverage for Wales shows more undercoverage compared with England

Statistical Population Dataset version 4.0 % difference from Census 2021 by age, England and Wales, 2021

Figure 3: Statistical Population Dataset version 4.0 coverage for Wales shows more undercoverage compared with England

Statistical Population Dataset version 4.0 % difference from Census 2021 by age, England and Wales, 2021



Source: Office for National Statistics – Census 2021

The differences between England and Wales may be linked to the inclusion of Hospital Episode Statistics (HES), the Emergency Care Data Set (ECDS), and the Individualised Learner Record (ILR), which only provide coverage for England. This may explain why the largest differences appear in those aged 17 to 19 years, which is a group more likely to attend further education and therefore appear on the ILR. We are in the process of obtaining Welsh equivalents of these data sources to use in future iterations of the SPD.

5 . Local authorities

Local authority-level analysis

We compared our 2021 Statistical Population Dataset version 4.0 (SPD v4.0) with Census 2021 for the 331 local authorities (LAs) in England and Wales to understand coverage patterns. We have plotted these against our P1 standard (Figure 4). The P1 quality standard is shown as the number of LAs whose population total is within plus or minus 3.8% of Census 2021. We are looking for 97% of all LAs to meet this quality standard.

Figure 4: 82 local authorities (24.8%) have Statistical Population Dataset version 4.0 counts 3.8% or less away from the census, and 19 local authorities (5.7%) have counts more than 3.8% higher than the census

Statistical Population Dataset version 4.0 local authority population count variations from Census 2021, England and Wales, 2021

Notes:

1. P1 quality standard is the maximum quality standard set out in evaluation criteria of the Beyond 2011 programme. At the local authority level, the P1 quality standard is to be within positive or negative 3.8% of the 2011 Census estimate (see [Beyond 2011: Options Report 2 \(PDF, 491 KB\)](#) for more information). We are planning to publish updates to the quality standards using Census 2021 later this year.

Download the data

[.xlsx](#)

At LA level in many areas, SPD v4.0 shows similar patterns to the national-level analysis, but we see considerable variation across LAs (Figure 5). While there are large numbers of LAs lower than Census 2021, there are also a few that are considerably higher. This highlights the importance of using our learning to inform the development of a coverage adjustment method.

The differences between the census and the SPD counts are likely to reflect a combination of true population change and features of the SPD. Census 2021 was carried out when England and Wales were placed under coronavirus (COVID-19) restrictions, and it collected information about people and their circumstances at that time. In comparison, these restrictions were removed by the SPD reference period in June 2021, and we do not know how interactions with services over the COVID-19 period may have been reflected in the admin data.

Figure 5: The coverage of the Statistical Population Dataset version 4.0 shows some variation at local-authority level

Statistical Population Dataset version 4.0 counts by local authority, age and sex, compared with Census 2021

Download the data

[.xlsx](#)

Counts higher than Census 2021

We usually find overcoverage in LAs that have demographic features associated with population churn, which can lead to time lags in the accuracy of administrative data. These demographic features include:

- high levels of migration
- high percentages of rental households
- urban areas
- younger population-size peaks

Analysis of Census 2021 information and LAs showing overcoverage found that LAs with more short-term residents and more rental accommodation have greater levels of overcoverage. For example, the SPD v4.0 for Manchester shows 3%; in this city, 62% of homes are rental properties compared with a national average of 25.3%. In Manchester, 3.3% of the population have been in the UK for under two years compared with a national average of 1%. Further information is available in our [Tenure dataset](#) and our [Length of residence dataset](#).

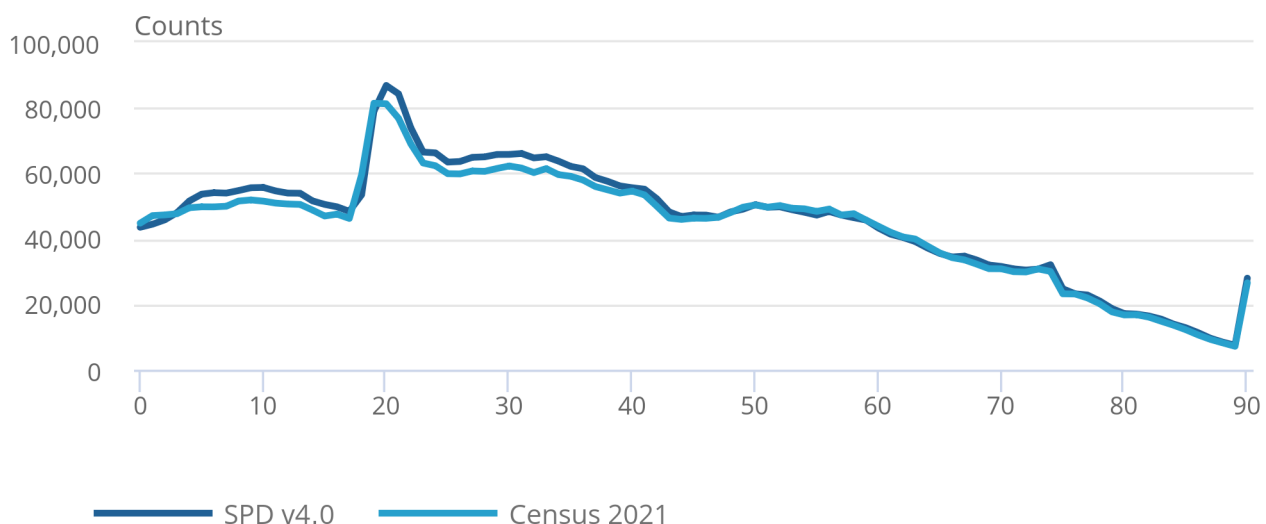
Outside of London, most of the 20 LAs with the highest overcoverage have large younger-working-age populations and/or larger-than-usual populations for those aged 18 to 22 years, which is an indication of universities in the LA (see Figure 6 and Table 1). We also see overcoverage in LAs associated with seasonal work and short-term migration, such as Boston. Most of the overcoverage in Boston is for males aged 20 to 40 years. There is a high proportion of people who have been in the UK less than two years (1.7%).

Figure 6: Local authorities outside London with Statistical Population Dataset version 4.0 overcoverage often have large university-age populations

Statistical Population Dataset version 4.0 population estimates compared with Census 2021, for 20 Welsh and English local authorities outside London with the highest levels of overcoverage, England and Wales, 2021

Figure 6: Local authorities outside London with Statistical Population Dataset version 4.0 overcoverage often have large university-age populations

Statistical Population Dataset version 4.0 population estimates compared with Census 2021, for 20 Welsh and English local authorities outside London with the highest levels of overcoverage, England and Wales, 2021



Source: Office for National Statistics – Census 2021

Table 1: English and Welsh local authorities outside of London with the highest Statistical Population Dataset version 4.0 overcoverage, 2021

Rank	Local authority	% difference from Census 2021
1	Boston	10.05
2	Richmondshire	5.11
3	Coventry	5.10
4	Nottingham	4.46
5	Liverpool	4.41
6	Wolverhampton	3.67
7	Peterborough	3.43
8	Middlesbrough	3.23
9	Manchester	3.04
10	Knowsley	2.34
11	Blackpool	2.10
12	Great Yarmouth	2.04
13	Kingston upon Hull, City of	1.97
14	Leicester	1.94
15	Luton	1.92
16	Halton	1.84
17	Stoke on Trent	1.79
18	Preston	1.70
19	Ipswich	1.52
20	Oxford	1.50

Source: Office for National Statistics – Census 2021

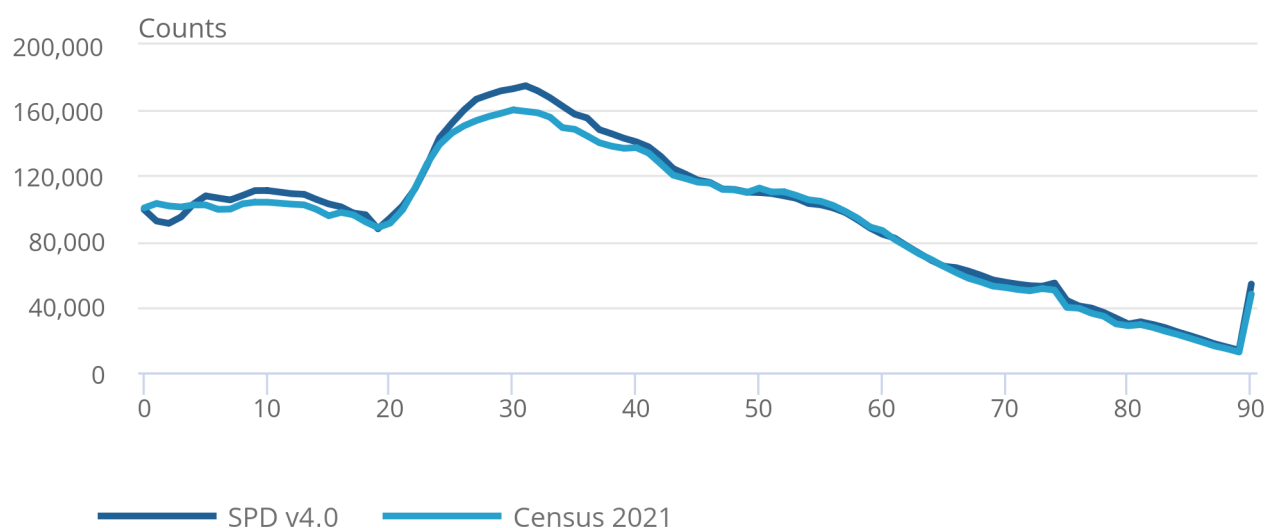
There are distinct population distribution patterns in LAs with SPD v4.0 overcoverage. In the 26 London LAs that show overcoverage, there are larger-than-usual populations aged between around 20 to 40 years, who are typically more mobile (Figure 7). In the seven London LAs with undercoverage, this peak is absent from the population distribution.

Figure 7: Local authorities in London have large younger-working-age populations

Statistical Population Dataset version 4.0 population estimates compared with Census 2021 for local authorities in London, England and Wales, 2021

Figure 7: Local authorities in London have large younger-working-age populations

Statistical Population Dataset version 4.0 population estimates compared with Census 2021 for local authorities in London, England and Wales, 2021



Source: Office for National Statistics – Census 2021

Counts lower than Census 2021

A high proportion of the LAs with undercoverage are in Wales (Table 2). This is likely to relate to the use of Hospital Episode Statistics (HES), the Emergency Care Data Set (ECDS), and the Individualised Learner Record (ILR), resulting in better coverage for England, but our analysis also shows some other patterns.

Analysis of the LAs with the most SPD v4.0 undercoverage tends to fall into the following categories:

- high levels of self-employment
- large proportions of people who are economically inactive
- mainly rural areas

Some of these patterns appear to reflect potential interactions with SPD data sources. Analysis of Census 2021 information and LAs showing undercoverage found that LAs with higher-than-average self-employed or economically inactive populations have greater levels of undercoverage. For example, the SPD v4.0 for South Hams has undercoverage of 5.8%. In South Hams, 37.2% of the population are economically inactive compared with a national average of 19%. Examples of people who are economically inactive include those who retire early or live off a partner's income. In South Hams, 13% of the population are self-employed compared with a national average of 8.1%. Further information is available in our [Economic activity status dataset](#).

Table 2: The 20 local authorities with the highest Statistical Population Dataset version 4.0 undercoverage, England and Wales, 2021

Rank	Local authorities	% Difference from Census 2021
1	Rutland	-9.15
2	Powys	-8.55
3	Pembrokeshire	-8.52
4	West Suffolk	-8.27
5	Eden	-7.83
6	Ceredigion	-7.75
7	Monmouthshire	-7.40
8	Carmarthenshire	-7.26
9	Harborough	-7.24
10	West Devon	-7.17
11	Gwynedd	-6.82
12	Bridgend	-6.67
13	Isle of Anglesey	-6.61
14	Conwy	-5.99
15	Vale of Glamorgan	-5.79
16	South Hams	-5.72
17	Wealden	-5.61
18	East Hampshire	-5.59
19	Derbyshire Dales	-5.51
20	Babergh	-5.39

Source: Office for National Statistics – Census 2021

Most LAs in England with the highest undercoverage have population distributions like those seen at the national level, although some do differ. Analysis of Census 2021 information and LAs showing undercoverage found that LAs with higher-than-average proportions of residents of certain types of communal establishments (CEs) – namely military bases, prisons, and care homes – generally have SPD undercoverage. For example, Rutland has undercoverage of 9.2%. In Rutland, 5.6% of the population are resident in one of the three types of CE, compared with a national average of 0.8%. Further information is available in our [Communal establishment residents by age and sex dataset](#).

6 . Output Areas

We compared population estimates between the Statistical Population Dataset version 4.0 (SPD v4.0) and Census 2021 at Output Area (OA) level. Because of the small size of OAs, this analysis can help us evaluate how SPD v4.0 performs at a more granular level of geography.

SPD v4.0 2021 and Census 2021 datasets refer to OA boundaries as they were in 2011. This is because the 2021 boundaries were not available at the time of our analysis.

Differences between SPD v4.0 and Census 2021 at OA level

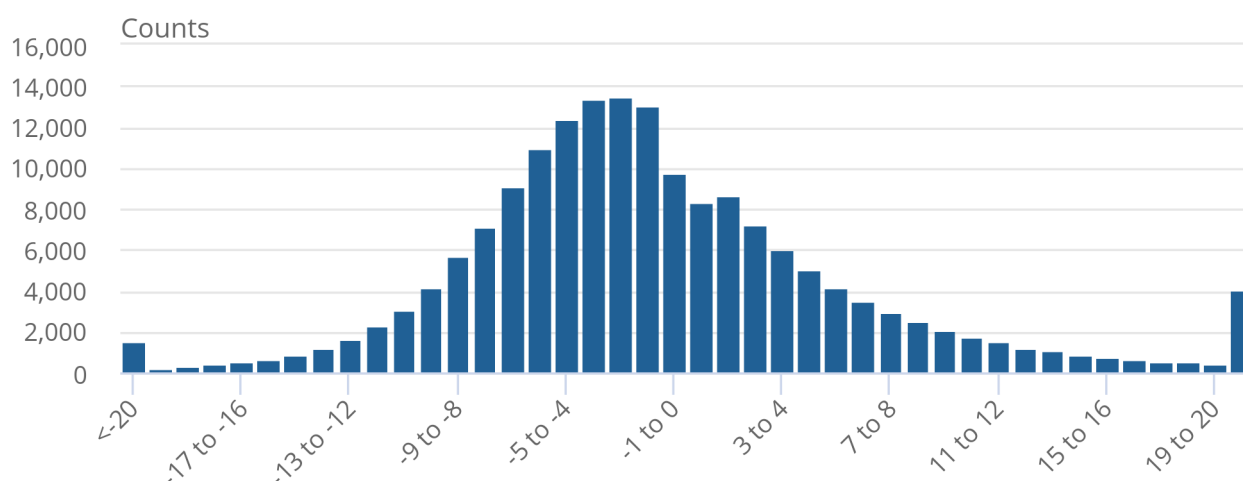
We compared population estimates for all 181,315 OAs. 62.3% show undercoverage in SPD v4.0, while 35.6% show overcoverage (Figure 8).

Figure 8: 85% of Output Area population counts show lower than plus or minus 10% coverage error

Percentage coverage error between Statistical Population Dataset version 4.0 2021 and Census 2021 by Output Area, England and Wales

Figure 8: 85% of Output Area population counts show lower than plus or minus 10% coverage error

Percentage coverage error between Statistical Population Dataset version 4.0 2021 and Census 2021 by Output Area, England and Wales



Source: Office for National Statistics – Census 2021

Outlier analysis

We used outlier detection to focus our analysis on the biggest differences between SPD v4.0 and the census. This included analysis of OAs themselves, in addition to each five-year age group in each OA.

We identified an OA, or an OA and age-group combination, as an outlier if the difference between the SPD and the census population estimates was more than five standard deviations away from the mean difference.

The number of outliers we found is in line with research previously published in our [Developing our approach for producing admin-based population estimates, subnational analysis for England and Wales: 2011 article](#). There were 999 (0.6%) outliers at OA level. Of these, 56.1% showed undercoverage in relation to the census, while 43.9% showed overcoverage. There were 9,322 (0.3%) OA and age-group combinations that were outliers. Of these, 62.7% showed undercoverage, while 37.3% showed overcoverage.

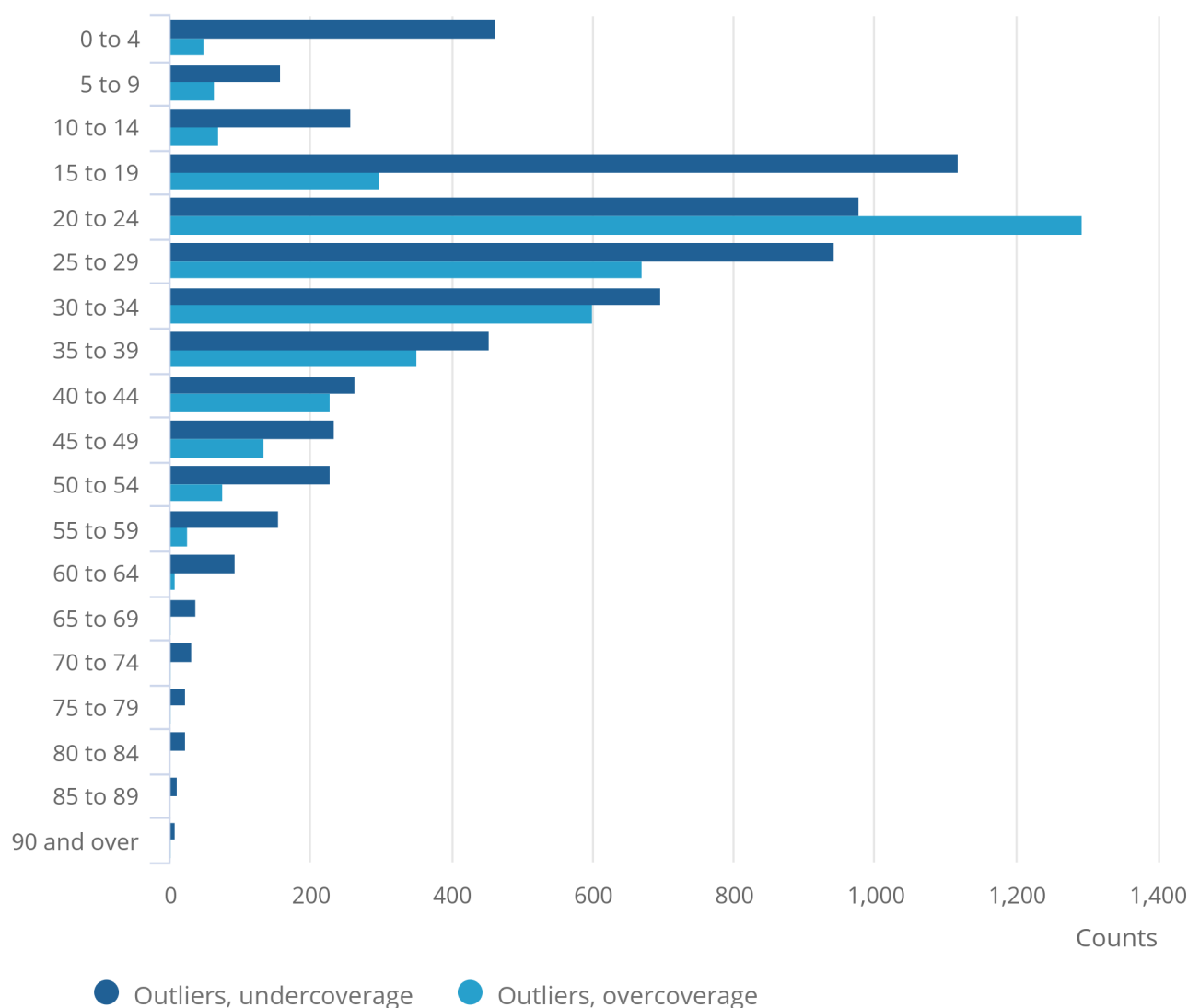
Figure 9 shows that we tend to see more outliers at younger ages in OAs with both undercoverage and overcoverage, suggesting that there are challenges in placing this age group in the correct address. To investigate this further, we looked into the presence of communal establishments (CEs).

Figure 9: Most age-Output Area outliers occur at younger ages

Count of Output Areas and age-group combination outliers, by age group and coverage type, England and Wales, 2021

Figure 9: Most age-Output Area outliers occur at younger ages

Count of Output Areas and age-group combination outliers, by age group and coverage type, England and Wales, 2021



Source: Office for National Statistics – Census 2021

Communal establishments

LA-level analysis and research published in our [Developing our approach for producing admin-based population estimates, subnational analysis for England and Wales: 2011 article](#) found that areas with large coverage error often contain CEs. These often indicate concentrations of specific populations, such as students or armed forces, who are less likely to interact with services in a typical way. Exploring the presence of CEs according to Census 2021 may help to explain the coverage error we see in our outliers.

Appearances of communal establishments in outliers

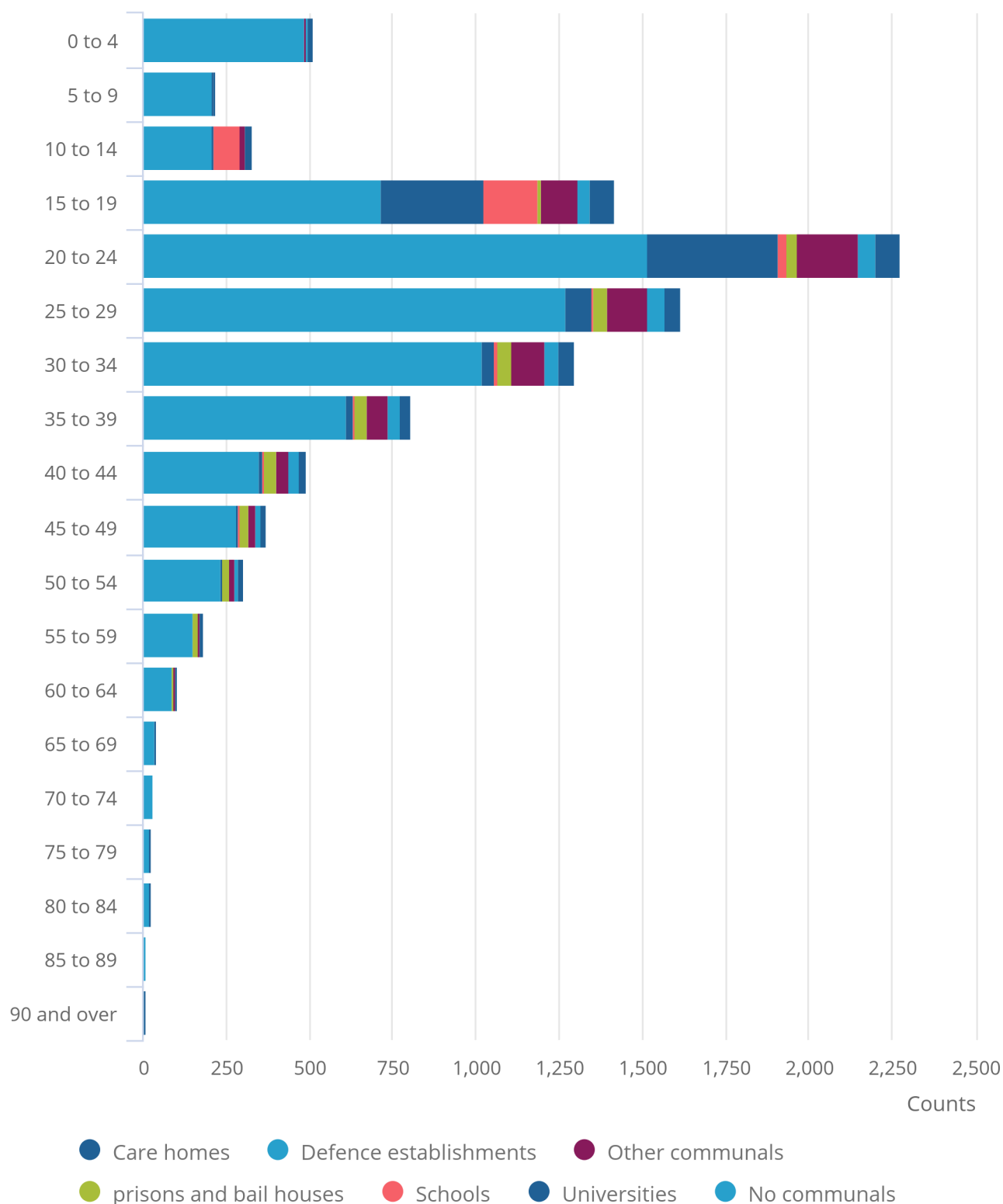
A CE is most likely to contribute to coverage error if it houses the same age group that we are seeing extreme coverage error for. We therefore focused this analysis on those outliers that occur in OA and age-group combinations (Figure 10).

Figure 10: Patterns of communal establishments by age-group outlier

Count of Output Area and age-group combination outliers, by age group and presence of communal establishments, England and Wales, 2021

Figure 10: Patterns of communal establishments by age-group outlier

Count of Output Area and age-group combination outliers, by age group and presence of communal establishments, England and Wales, 2021



Universities

There are universities in fewer than 1% of OAs, but in 10.7% of outliers, which is more than any other type of CE. They may provide a way of understanding the high number of outliers identified for typical student ages. For those aged 15 to 19 years, universities appear in 24.8% of outliers. For those aged 20 to 24 years, universities appear in 18.6% of outliers.

Because of the coronavirus (COVID-19) pandemic, the Higher Education Statistics Agency (HESA) issued [special guidance on how student data should be collected](#). This guidance places students unable to attend university because of the coronavirus pandemic in their intended university address. This may lead to the incorrect inclusion of records of those who were intending to come to university from outside England and Wales but who have chosen not to move because of coronavirus restrictions. This discrepancy highlights how the need to understand how specific populations interact with services is important to ensure we are placing them effectively in the SPD.

Residential care homes

Our research published in our [Developing our approach for producing admin-based population estimates, subnational analysis for England and Wales: 2011 article](#) demonstrated a high frequency of OA outliers for older age groups. This was associated with the presence of residential care homes. We did not observe the same association when comparing SPD v4.0 and Census 2021. This may suggest the inclusion of Hospital Episode Statistics (HES) and the Emergency Care Data Set (ECDS) are improving our coverage of those interacting with health services, but this conclusion may be premature given the impact of the coronavirus pandemic. During 2021, individuals needed to register with a General Practitioner (GP) to receive a vaccination. Registrations or recorded addresses may have been particularly up to date for this year and therefore not reflect the way individuals typically interact with administrative data. Consequently, we need to monitor how the quality of these data may change over time.

7 . Developing Statistical Population Datasets data

[Statistical Population Dataset version 4.0 2021](#)

Dataset | Released 28 February 2023

Statistical Population Dataset version 4.0 (SPD v4.0) counts by age, sex, and local authority.

8 . Glossary

Statistical Population Dataset (SPD)

Administrative data are used to approximate the usually resident population within England and Wales.

Administrative data

Collections of data maintained for administrative reasons, for example, registrations, transactions, or record-keeping. They are used for operational purposes, and their statistical use is secondary. These sources are typically managed by other government bodies.

Local authority

The general term for a body administering local government services. In England, local government is administered by either single-tier or two-tier local authorities. The single-tier authorities comprise unitary authorities, metropolitan districts, and London boroughs, though some services such as transport planning are carried out by the Greater London Authority. The two-tier authorities elsewhere comprise counties and non-metropolitan districts. In Wales, there are single-tier unitary authorities.

Output Area (OA)

Small geographical areas, typically with a population between 100 and 625 people and a minimum of 40 households.

Dynamic population model (DPM)

The SPD will be one of the core sources used in the DPM, which uses statistical modelling techniques and demographic insights alongside a range of data sources to produce coherent and timely estimates of the population and population change.

Activity

An individual interacting with an administrative system, for example, for National Insurance or tax purposes, when claiming a benefit, attending hospital or updating information on government systems in some other way. Only demographic information (such as name, date of birth and address) and dates of interaction are needed from such data sources to improve the coverage of our population estimates.

Population churn

A measure of migration into, out of, and within a geographical area.

Communal establishment (CE)

An establishment providing managed residential accommodation. “Managed” in this context means full-time or part-time supervision of the accommodation.

Outlier detection

In this article, outlier detection refers to identifying values in data that are significantly different to the majority. We identified an instance of coverage error as an outlier if its value was more than five standard deviations higher or lower than the mean average value.

Usually resident population

We are currently adopting the United Nations (UN) definition of “usually resident” – that is, the place at which a person has lived continuously for at least 12 months, not including temporary absences for holidays or work assignments, or intends to live for at least 12 months (United Nations, 2008).

9 . Data sources and quality

We have produced results using our most recent Statistical Population Dataset (SPD) method, referred to as SPD version 4.0 (v4.0). This builds on our previous SPD method and incorporates new data sources that contain activity information to help identify the usually resident population.

For the first time, we can make comparisons with Census 2021 data to analyse the SPD and understand how SPD v4.0 performs by age, sex, and geography. We collected Census 2021 data on 21 March 2021, and this remains our best estimate of the population for this time. The SPD v4.0 reference date is 30 June 2021. Comparisons between the two sources provide a unique insight into the quality of our SPD methodology.

SPD v4.0 builds on our previously published SPD v3.0 method, as described in our [Developing our approach for producing admin-based population estimates. England and Wales: 2011 and 2016 article](#). SPD v4.0 includes Hospital Episode Statistics (HES), the Emergency Care Data Set (ECDS), and the Individualised Learner Record (ILR).

HES and ECDS

HES contains information on those attending an NHS hospital and those accessing private healthcare in an NHS hospital in England. This is through an appointment, outpatient care, or accident and emergency admission. From 2020, ECDS superseded the accident and emergency data. The SPD includes records if they were active during the 12 months prior to the reference date. An activity can be defined as an individual interacting with a service, for example, attending hospital. The use of this source therefore provides additional indicators of who is usually resident in England.

ILR

The ILR is a dataset containing information on those participating in Further Education (FE) in England. The SPD includes records if they were studying during the academic year prior to the census reference date. As people typically attend FE between the ages of 16 and 18 years, the ILR helps to improve coverage in this age group.

10 . Future developments

The research in this article highlights the need for an appropriate coverage adjustment method, as well as further research into how we can effectively capture communal establishment groups. To improve the quality of the Statistical Population Dataset (SPD), as well as support the use of the SPD in the dynamic population model (DPM), future work will focus on:

- investigating new data sources to bring into the SPD
- examining model-based inclusion rules and household groupings to improve the quality of the SPD
- investigating methods of implementing a robust coverage adjustment method to provide the basis of an unbiased stock measure of the population, which will feed into the DPM
- exploring how best to ensure those in communal establishments are accurately reflected in the SPD

11 . Related links

[Developing our approach for producing admin-based population estimates, England and Wales: 2011 and 2016](#)

Article | Released 21 June 2019

Research into developing a new methodology to create population estimates from administrative data. These estimates are not official statistics on the population.

[Developing admin-based population estimates, England and Wales: 2016 to 2020](#)

Article | Released 26 November 2021

Progress made to extend our understanding of the stability and behaviour of two methods for producing admin-based population estimates.

[Understanding quality of the Statistical Population Dataset in England and Wales using the 2021 Census - Demographic Index linkage](#)

Article | Released 28 February 2023

Analysis of the Statistical Population Dataset Version 4.0 2021 using a linkage between Census 2021 and the Demographic Index.

[Transforming population statistics, comparing 2021 population estimates in England and Wales](#)

Article | Released 28 February 2023

Evaluating our progress towards a transformed population statistics system, using comparisons between census-based and admin-based population estimates and Census 2021.

[Dynamic population model, improvements to data sources and methodology for local authorities, England and Wales: 2011 to 2022](#)

Methodology | Released 28 February 2023

Developments of methods and data used in the dynamic population model.

[Admin-based population estimates: provisional estimates for local authorities in England and Wales, 2011 to 2022](#)

Article | Released 28 February 2023

Admin-based population estimates for all local authorities in England and Wales from the dynamic population model.

12 . Cite this article

Office for National Statistics (ONS), released 28 February 2023, ONS website, article, [Developing Statistical Population Datasets, England and Wales: 2021](#)