Quality assuring the local authority census population estimates, England and Wales

Methodology for the validation of the Census 2021 population estimates for each local authority, as part of the wider quality assurance of the census data.

Table of contents

1 . Overview

This paper shows a practical illustration of the validation conducted on the Census 2021 population estimates for each local authority. This validation was part of the wider quality assurance (QA) of the census data described in [How we assured the quality of Census 2021 estimates](#).

Since some of the census statistics used in this validation were not published at the time this report was published, we have combined data for different areas to produce a case study for a fictional local authority – "Anytown". This means that the data discussed realistically reflect the issues encountered during our validation work, but ensures no census statistics are revealed without being part of a standard release.

The sections that follow are a condensed illustration of the full validation conducted on the Census 2021 estimates but provide a picture of the QA checks and comparisons completed on the Census estimates before publication.

2 . Validation of census population estimates: Anytown

Total population

Our first check compares Anytown's total population estimate with other available population estimates. Here, we will consider the two main comparator sources: the rolled-forward official mid-year population estimates (MYE), and the latest version of the 2020 admin-based population estimates v3.0 (ABPEs). We also used 2011 Census (on which the MYEs are based), the earlier ABPEv2.0, and the NHS Personal Demographic Service (PDS) as comparators. However, these were given less weight in our assessments for the reasons explained in [Administrative data used in Census 2021, England and Wales](#).

We don't expect the census estimate to be the same as an estimate from any other source. Definitions and the reference date may be different and all estimates, including the census estimate, are subject to a degree of statistical uncertainty. However, comparing estimates or counts from different sources can help us understand whether the census results are plausible or whether the evidence points towards a possible quality issue with them.

Table 1: Fictional population of Anytown

|	Usual residents (thousands)
2021 Census	139.1
2020 Mid-Year Estimates	160.3
2020 ABPE v3.0	140.8

Source: Office for National Statistics – Census 2021

Table 1 shows that the 2021 Census estimate is close to the ABPEs estimate but is further away from the MYE. The difference from the MYE is greater than that seen in most areas so we will want to understand the reasons for that difference. Our first step towards that is to look at the sex-age structure of the population.

Sex-age population profile
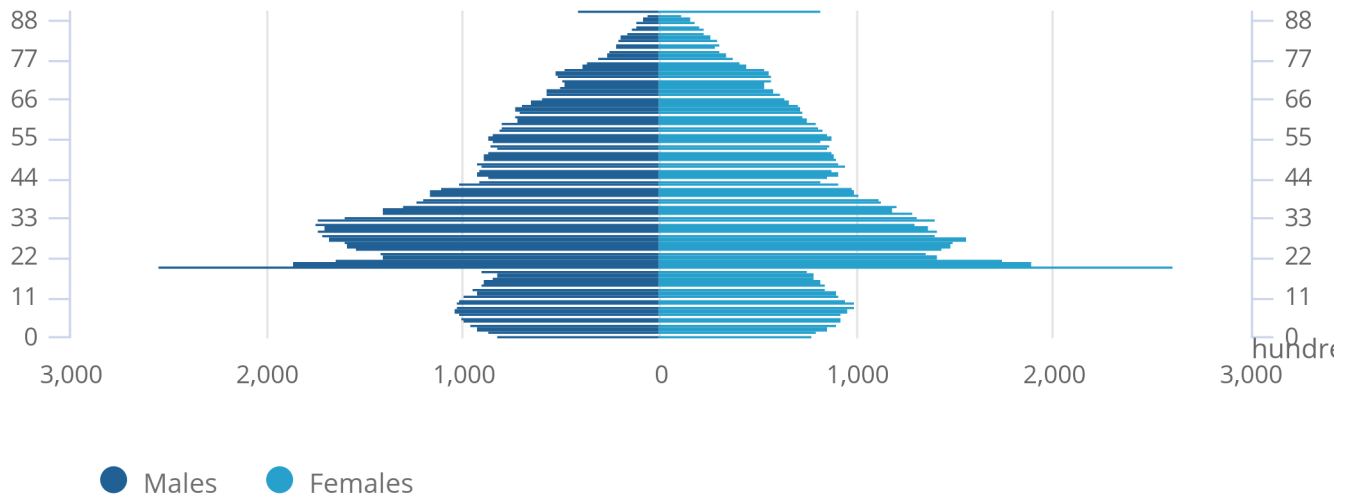
Figure 1a Population of Anytown: 2021 Census
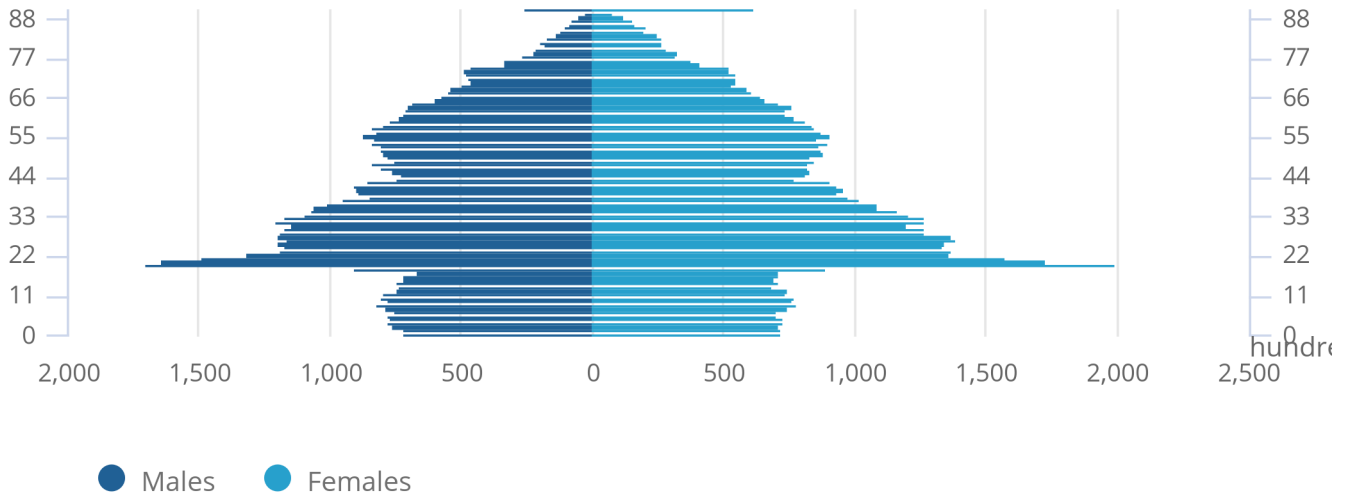
Source: Office for National Statistics, Census 2021

Figure 1b: Population of Anytown: 2021 Census and comparator sources

MYE 2020

● Males   ● Females

Source: Office for National Statistics

ABPE 3.0 2020

Figure 1c: Population of Anytown: 2021 Census and comparator sources

ABPE 3.0 2020

● Males   ● Females
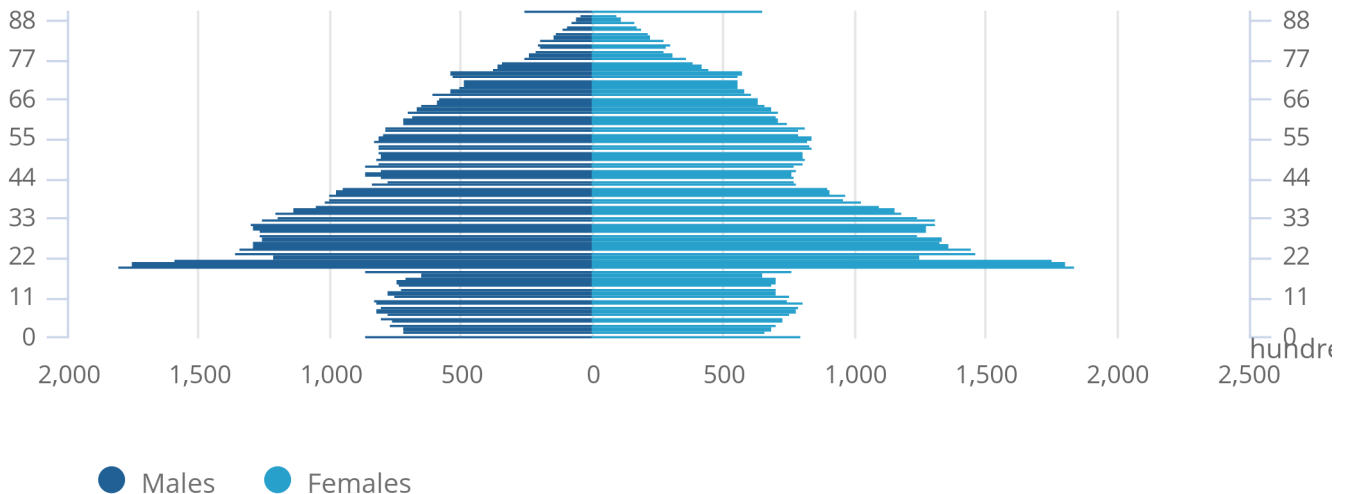
Source: Office for National Statistics

Figure 1a shows the sex and age structure of Anytown's population as estimated in the census. This can be compared with Figure 1b showing the same structure for MYE and Figure 1c for ABPE v3. While the profiles are broadly similar, there are interesting differences at some ages as described in the following paragraphs.

The census is lower than the MYE sources for children aged 0 to 15 years. It is also lower than the ABPE for very young children (aged under 3 years) but at broadly the same level for school-age children (aged 5 to 18 years). This difference for the very young is seen in many areas. Given the accuracy of the birth registration data this difference, while not large, was identified as an issue needing further investigation and, possibly, adjustment.

For older children, we place more weight on the ABPEs for Anytown than the MYEs. This is because MYEs will be subject to uncertainty related to the estimation of migration (such as, the estimation of the age distribution of international migration) while the ABPEs make use of other sources such as the NHS Personal Demographic Service (PDS), and School Census data sources. These have a very high coverage of this age group and aren't affected in the same way by the need to estimate migration. Though the ABPEs are only available for mid-2020, we have also checked the 2021 PDS and School Census data which show no major changes in this group. This provides reassurance that the estimates for children are plausible, other than the issue for children aged under three.

There is also a clear peak in the ABPEs and MYEs for those aged 18 to 22 years, which the Census estimates do not fully reflect. This pattern was seen in many student areas. It largely reflects lower numbers of students living in the area in 2021 than in previous years. This is because of the coronavirus (COVID-19) pandemic and students not travelling to live and study in the area. More evidence that this effect is because of a lower student population was provided by a check of the population estimates at Lower Super Output Area (LSOA) level. LSOAs showing the largest differences to the Small Area Population Estimates and the ABPEs were those containing concentrations of students on both HESA and census data.

The MYEs also show much higher numbers at those aged 28 to 40 years than the census or the ABPEs. This is likely to reflect inaccuracies in the MYEs rather than quality issues with the 2021 Census. While MYEs are likely to be most accurate immediately following a census (which provides the base population for those estimates), they can sometimes drift from the true values over the following decade. This is because of inaccuracies in the estimates of internal and international migration since that census. All three sources give very similar estimates for those aged 65 years and over, and this provides reassurance that the census has not missed a large cluster of addresses.

Students

Anytown's large student population meant that we carried out specific checks on this group. Our communal establishment checks had confirmed we had captured every hall of residence. We supplemented the census response data with data collected in our Student Hall Survey to provide the most reliable estimates of students living at those establishments. However, the total number of full-time students appearing in the adjusted census data was substantially below the number of students recorded on the HESA data as having a term-time address in Anytown. This difference was seen in other areas with large student numbers.

Further investigation suggested that a large part of the difference was because of international students who appeared on the HESA data but who were either not present (for example, because of the pandemic), or who were present but not "usually resident", specifically, not expecting to be in the UK for 12 months or more as may be the case for students on a one-year course.

Fertility and mortality rates

Further checks of the census estimates for some age groups are provided by calculating implied fertility and mortality rates. These checks combine census data with numbers of registered births and deaths in Anytown and calculate what the fertility and mortality rates would be if the census estimates were correct. These checks take advantage of the comprehensive nature of registration system for these life events and of our knowledge of expected shapes of fertility and mortality rate profiles based on previous demographic research.

First, we calculate Anytown's age-specific fertility rates calculated using registered births by age of mother and census estimates of women at those ages. Small numbers inevitably mean that the profile of these rates is more erratic than the national profile, but the broad pattern looks plausible. Summing those rates to find the total fertility rate returns a figure very close to the national figure.

Similarly, approximate age-specific mortality rates can be calculated using numbers of registered deaths at each age and sex and the census population estimates. These mortality rates showed the expected general pattern of rates gradually increasing as age increases with no spikes or dips, suggesting a substantial over or underestimate of the population in a particular age group. Overall, mortality rates for Anytown were about the national average, which is consistent with our expectations given our knowledge of Anytown's demography and history.

These fertility and mortality rates are only approximations to the true values as, for example, they take no account of migration and ageing in the year up to census day. However, they provide no reason to be concerned about the quality of the census population estimates.

Lower Super Output Area (LSOA) population

We extended the detailed quality assurance of the population structure across Anytown by checking the estimates for each LSOA in the area. These checks began with a standard comparison of the census estimates for LSOAs in Anytown against the:

- 2020 Small Area Population Estimates (SAPE)

- 2011 Census estimates

- 2020 Admin Based Population Estimates v3.0 (ABPEs)

- 2021 Personal Demographic Service (PDS) data

Anytown's LSOA population estimates broadly align with the main comparators. A comparison with the 2011 Census highlighted five LSOA outliers where the 2021 estimate was notably higher than the 2011 estimate, and the Valuation Office Agency data confirms a substantial increase in housing over the decade in those areas.

A scatterplot of the census estimates compared with the 2020 ABPE LSOA estimates suggested a further three LSOAs as needing investigation.

The largest LSOA outlier has a 2021 Census estimate significantly higher than the 2011 estimate, the 2020 ABPEs, and 2019 SAPE, but close to the 2021 PDS data. Most of this increase, is because of 400 new homes built in a new development scheme 'Anytown Park', and with the first residents moving in in January 2021. These new residents have been picked up by the census and PDS but are not reflected in the other estimates relating to earlier years.

The final two outliers seen in comparison with the ABPEs reflect where much of the student population reside in Anytown. Although the Census estimates are around 3% to 4% lower than the ABPEs estimates for these areas, this is not implausible. The evidence we have from the Student Halls Survey and elsewhere suggests that fewer university students were living in Anytown in 2021 than had usually been the case in recent years.

Households

After checking the population size and structure we checked the number and distribution of households in Anytown.

Figure 2: Households by LSOAs in Anytown: 2021 Census and Council Tax

Figure 2: Households by LSOAs in Anytown: 2021 Census and Council Tax
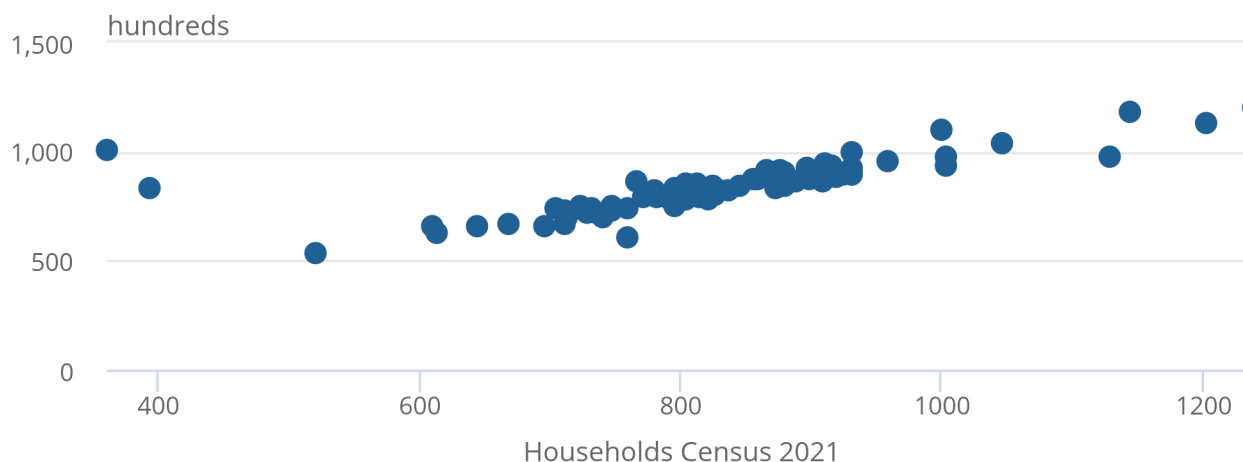
The total estimated number of households in Anytown is a few percent below the main comparator administrative source – the council tax data. Census estimates of households with at least one usual resident are generally different from counts of council tax records for various reasons explained in Administrative data used in Census 2021, England and Wales but the difference in Anytown is larger than average. Similar differences are seen when comparing the census estimates with data from the Electricity Central Online Enquiry Service (ECOES) data on electricity connections.

Though the census estimates are lower than those comparator sources, they are very close to the 'Alternative Household Estimate' (AHE). We calculate the AHE by combining counts of census responses with estimates of occupied household spaces from information provided by our field-staff during data-collection and from various administrative sources. This estimate provides a useful check of the standard census estimate of households calculated using our standard coverage adjustment methods.

The explanation for most of the differences between the census estimate of households and the figures from other sources is provided by an analysis of the figures at LSOA level. Figure 2 shows that the difference is largely concentrated in three LSOAs. These LSOAs are in the student quarter of Anytown and contain the main Halls of Residence in the area. A check of the record level Council Tax data confirms that units within some of these Halls have been included as individual records. In contrast, the census estimates would not include those units as household spaces but as part of a communal establishment.

After removing those records from the council tax counts, the census estimates are much closer, but still below, the council tax numbers. Part of this will be because of a slightly lower number of student households (for the reasons described in the Students section), but the difference is within the range of differences we expect based on differences in definition described in Administrative data used in Census 2021, England and Wales.

Anytown's household size profile looks plausible. As in 2011, the area contained more one and two person households than the national average which is typical for a largely urban area with a relatively young age-structure.

Communal Establishments

Our next stage of quality assurance was to check that we were not missing people from Communal Establishments (CEs). There were two parts to this. The first is checking that we received a response for the communal establishment (this is the 'CE1' form completed by the manager of the establishment). The second is checking that the population living within those establishments are accurately reflected in the census data (this relates to 'I' forms completed by individuals within the establishment and any adjustments made for undercoverage).

First, we conducted these checks for the large CEs (those with more than 50 expected residents). Our Address Frame (expected to be the best quality list of these establishments available on census day) records 15 large CEs in Anytown. Of these, 12 were directly matched with census responses and two of the remaining three appeared under different names. Desk research confirmed that the different names did not indicate a change in the nature or size of the establishments. The remaining CE was a large care home which had closed in late 2020.

We then checked how many usual residents responded in each of these large establishments. We compared this with the number reported by the establishment manager and against any capacity or occupancy figures available on the Address Frame or other administrative sources. For Anytown, these sources were the NHS Digital Capacity Tracker data for the care home population, Ministry of Justice data for the prisoner population and the ONS's Student Hall Survey, conducted shortly after the census for the student population in halls of residences. More detailed quality information around these sources can be found in Administrative data used in Census 2021, England and Wales.

After checking the large communal establishments, we also assessed coverage of smaller establishments. The census estimates that Anytown contained, in total, 160 CEs and 10,000 usual residents within them which aligns well with the Address Frame's 164 establishments and 10,500 recorded spaces (the Address Frame reflects the capacity of an establishment rather than the number of usual residents and so will be expected to be higher than the Census CE estimates). Of the four establishments not on census, we could confirm that one hostel which appeared on the Address Frame but not on census responses, had closed permanently in 2020. The three remaining establishments were guest houses with a listed total capacity of 30 people. We could not confirm whether these establishments were open or occupied on census day. The small possible element of undercoverage from these establishments is estimated and corrected for through our standard statistical methods and we took no further action on this.

Migration flows

We also used the census question on 'address one year ago' to check how the census measures of migration flows, by age and sex, in and out of Anytown compared to estimated flows in the MYEs.

This analysis allows us to check migration inflows from other countries and from elsewhere in the UK, and migration outflows to other areas in England and Wales.

The estimated international migration inflow to Anytown was substantially lower in the census than in the 2020 MYEs. There seemed to be two reasons for this.

The first is that the international migration estimates used in the 2020 MYEs would not have reflected the impact of the pandemic on migration, of international students. Anytown usually hosts a large number of international students, and we know from other data sources such as visa data and our Student Halls Survey that not all of the expected students have actually arrived in the UK. Comparing the sex-age profile of international migrants captured in the census with that of international migration estimated in the MYEs supports the idea that most of the difference is concentrated in those aged 19 to 24 years, which is the peak age for student migration.

The second is that the MYEs have tended to show higher international migration estimates than the census for urban areas like Anytown with correspondingly lower estimates for more rural areas. It's thought that this is likely to point to an inaccuracy in the methods used in the MYEs to allocate international migration to each local authority rather than to a quality issue with the census data.

The difference in these international migration estimates for Anytown was very similar to differences seen in similar LAs elsewhere in the country. We also found further corroborating evidence on Anytown's figures in the HMRC RAPID data on people born outside the UK and appearing on that data source within the previous 12 months. The evidence did not point to a quality issue with the census data on international migration.

Internal migration inflows and outflows were slightly lower than the figures used in the 2020 MYEs. This is consistent with the picture seen in most other areas, and is partly attributable to a difference in how migration is defined in the two sources. The MYE would count someone who moved in and out of an area within a year as both an in-migrant and an out-migrant while the census would not have captured them as living in the area during the period (it would only identify where they were resident on census day and one year before the census).

The age-sex profiles of internal migration flows do not display any cause for concern. As with all areas with large numbers of students, there is a peak in inflows for both sexes at ages 18 and 19 years with another, smaller, peak at aged 22 years, likely to largely reflect moves into Anytown of new graduates from other areas either for work, or to return to their family home. Similar, though smaller, peaks are seen in the age-profile for outflows. Comparison with the internal migration estimates used in the MYE shows a very similar profile for most ages, but differences at those student ages. This pattern is seen in all student areas and can be confidently attributed to the impact of the pandemic on students' place of usual residence. The estimates already reflect adjustments made to ensure that students were, as far as possible, included in the census data at their term-time address even if they were not physically present there at the time of the census.

Communities and Topics

Quality assurance checks for communities (defined with respect to concentrations of detailed country of birth, ethnicity, and religion) and topic data, such as broad ethnic groups and housing characteristics, did not reveal any concerning changes when comparing the 2021 Census with the 2011 Census data.

2011 Issues

We checked the quality assurance records from the 2011 Census and found no issues identified in 2011 that needed to be checked again in 2021.

LA Feedback

Anytown took part in our LA Insights initiative and provided very useful feedback on provisional Census estimates. This feedback became an integral part of our QA process as it provided confirmation on certain aspects of our QA and highlighted important aspects for us to investigate further.

The LA provided three points of feedback. First, the Census household estimate across the area seems low compared with their Local Land and Planning Gazetteer (LLPG) data for 2019, seeming to suggest that about 4000 properties may not have been identified in the census. Second, the household estimates for two LSOAs were much lower than the counts from council tax which had been provided for comparison with the provisional census numbers. Finally, we should check that a newly opened Halls of Residence was covered in Anytown's Census population estimates.

The difference between the census household estimate and the LLPG data was investigated but did not point to a quality issue with the census. The LLPG data was two years out of date and did not reflect the impact of redevelopment schemes which had demolished existing residential properties but not delivered new housing at time of the census. There was also a definitional difference in that the LLPG data quoted by the LA covered all properties rather than identifying only occupied household addresses. Finally, a current version of the LLPG data had been one of the data sources used in compiling the census Address Frame (used when sending out invitations to complete the census) so we could be confident that the data cited by the LA did not point to missed addresses.

The standard checks already described in the household section of this report provide further reassurance that the census estimate of households for Anytown is plausible.

The second concern related to two LSOAs where the census estimate of occupied households was very substantially below the council tax figures. These areas had already been flagged for investigation as part of the initial quality assurance.

The first LSOA of concern was in the student quarter of Anytown and contained several halls of residence. A check of the council tax data confirmed that units within these halls were included as separate records on that data source. However, they would be counted as part of communal establishments in the census data, rather than occupied households. Removing those records from the council tax numbers brought the two sources in line.

The second LSOA of concern was on the coastal area of Anytown. The census response data shows that there are cottages in this area used as second homes but that these are not described as such on the council tax records (Anytown does not offer any council tax discounts on second homes). There was also a small new block of flats which had been put on the market in January 2021. These had received invitations to complete the census, but the census field staff had found no signs of occupation, despite visits to the block. It is likely that the flats were appearing on the council tax data without being occupied at the time of the census in March 2021. These two factors combined accounted for the difference between the council tax data and the census estimate of occupied households.

The final concern raised by the local authority was that a new halls of residence, 'Anytown South Halls' might have been missed in the census. Initial investigation into census response data did not find any response record. However, further analysis into the address lines of each halls in the LSOA identified 'Anytown South' as the same building as 'Anytown House'. As 'Anytown House' has been well captured in the census, and its count of residents adjusted in line with the Student Hall Survey, there is no quality concern over this part of the population estimates.

This is a good example of how the LA Feedback process has proved to be highly useful throughout our QA process, through confirmation of discrepancies which need further investigation and providing snippets of up-to-date information for us to take on board. It allowed the QA team to further understand the census responses and ultimately led to a more comprehensive validation of the local authority population estimates.

Panel Discussion and Final Outcome

The evidence described above was taken to a LA QA Panel consisting of the head of a branch responsible for producing population statistics and a senior researcher from a demographic research team. The panel members were not involved in the production of the census estimates or in the quality assurance investigations so were able to take an independent view of the evidence presented to it. The panel discussed that evidence and checked that the feedback provided by the LA had been adequately addressed.

After a thorough discussion, the panel concluded the population estimates were generally of good enough quality to be published but asked for the discrepancy in the under-three age group to be escalated for further discussion at the Statistical Contingences Escalation Forum (SCEF).

QA Panels for other LAs had flagged the same issue for consideration by SCEF. The forum noted the comments from each Panel and considered the evidence on the estimation of those aged under 3 years and the findings of the national quality assurance for this age-group. The evidence clearly pointed to a slight national under-estimate of this group and the Forum agreed that the initial method for estimating this age-group should be strengthened by incorporating evidence from birth registrations and demographic analysis. Once that was done, Anytown's census estimates would be of good enough quality to publish.

3 . Related links

Administrative data used in Census 2021, England and Wales
Published: September 2022
The administrative data sources that were used in the statistical design for Census 2021, with information on their coverage, accuracy, and timeliness.

Quality and methodology information Census 2021
QMI | last updated 2 November 2022
Details the data strengths, limitations, uses, users and methods used for Census 2021, England and Wales.

How we assured the quality of Census 2021 estimates
Methodology | Released 7 November 2022
Methodology for the validation of Census 2021 population estimates for England and Wales, including the assurance of processes, assessment of estimates, and involvement of local authorities.

4 . Cite this methodology

Office for National Statistics (ONS), released 7 November 2022, ONS website, methodology, Quality assuring the local authority census population estimates, England and Wales