Administrative data used in Census 2021, England and Wales
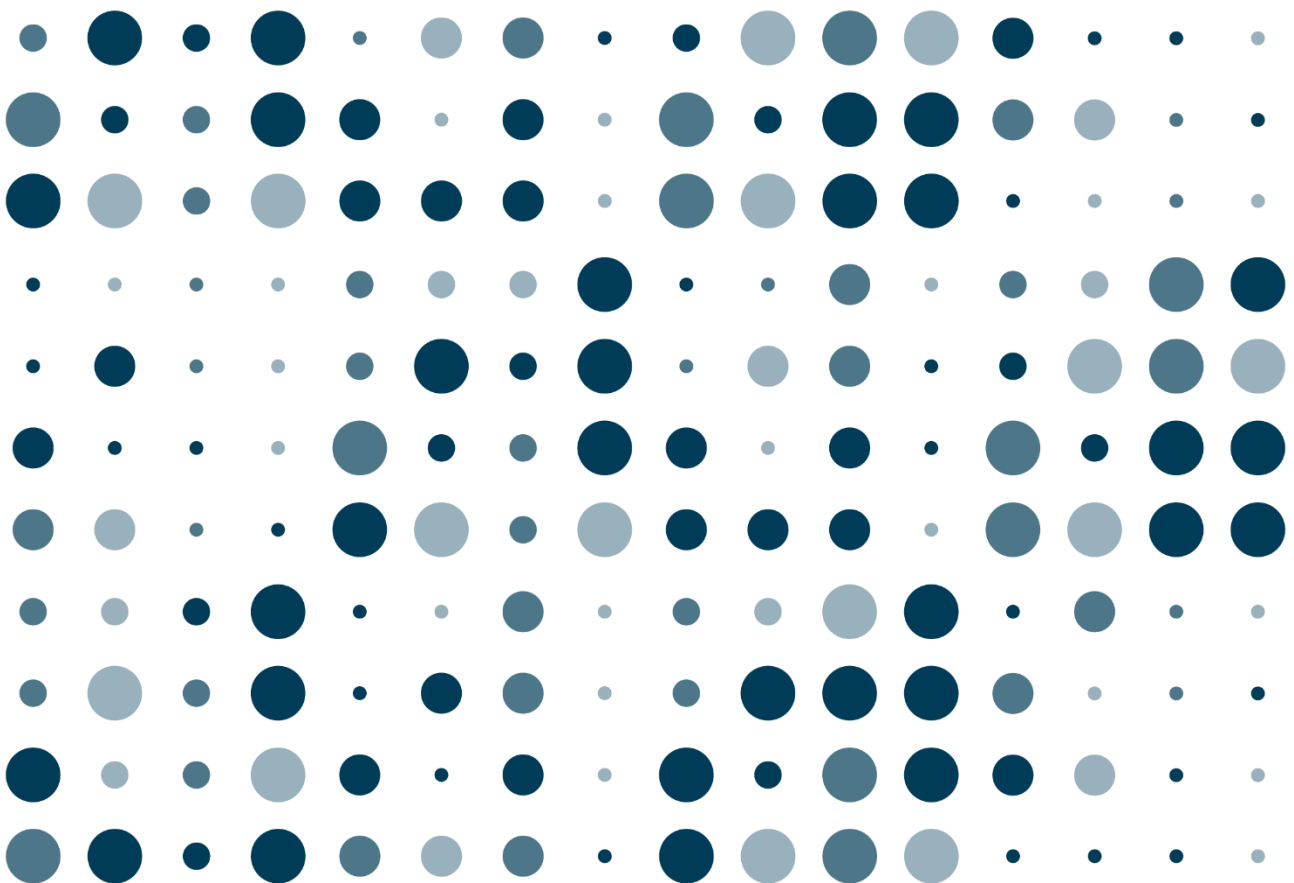
Census 2021 Statistical Design, Census 2021 Data Transformation Division

Date 15/09/2022

1 Introduction

Census 2021 made greater use than ever before of administrative data collected by public and some private organisations; recognising the importance of utilising the best available sources to deliver a high-quality census for all. This paper provides a comprehensive description of each of the sources that the Office for National Statistics (ONS) used in the statistical design of the census (as referenced in the contents section), covering population, households, and housing. Details are provided about how each data source was used within the Design for Census 2021, along with information about the coverage, accuracy and timeliness of the source set against the requirements of census.

The order of the sections of the report broadly follows the census process, covering the administrative data sources that were used to:

- Construct and quality assure the Census 2021 Address Frame (the list of residential addresses in England and Wales used to contact people to take part in census)
- Predict how and when households were likely to respond, to ensure the right resources were in place to support those who needed help
- Quality assure the census results, while delivering the information needed for the processing and estimation stages.

The use of administrative data to support the census aligns with international best practice and the 'Quality' pillar of the Code of Practice for Statistics on the use of Suitable data sources. Data sources were chosen based on their coverage and accuracy against the needs of Census 2021, including how well the data aligned to the census population and definitions. This involved working closely with data suppliers to understand the quality of the data sources, while drawing on ONS's experiences of using many of the sources in existing National Statistics and as part of ONS's wider transformation programme.

As with all census data, administrative data have only been used for statistical purposes in the public interest. The data are subject to robust controls to ensure that individuals cannot be identified. Furthermore, ONS complies with all data protection legislation, including the General Data Protection Regulation, the Data Protection Act 2018, the Statistics and Registration Act 2007 and the Digital Economy Act. Further information, including ONS's privacy statement and data protection policy can be found on our Data protection page.

Although this paper focuses on how we have used administrative data to support Census 2021, the insights gained from the census research into administrative sources has and will continue to inform our work to improve our future population, migration and social statistics. Details about this can be found in the Future of population and social statistics.

2 Census Address Frame

The Census 2021 Address Frame (the frame) is a comprehensive list of all residential addresses in England and Wales, including households and communal establishments (CEs). It is used as the basis for contacting people to take part in the census. The frame is constructed using multiple administrative data sources to ensure a high degree of accuracy, as outlined below.

The primary source for the frame is the AddressBase Premium product (ABP), which contains information relating to every address in Great Britain, from its provisional designation and throughout its life cycle. The ABP is widely used by the public and private sectors and is continually updated by Geoplace. The data come from the Local Land & Property Gazetteers (LLPGs) and other sources, such as the Valuation Office Agency (VOA), Royal Mail and Ordnance Survey.

The initial frame was created from both ABP and additional administrative and commercial data sources. These additional sources were particularly useful to address the challenges in enumerating CEs. For more information see the Design of Address Frame, Collection and Coverage Assessment and Adjustment of Communal Establishments in 2021 Census and 2021 Census Address Checking (PDF, 251 KB). These sources helped to:

- identify missed or misclassified addresses in the ABP
- validate addresses where there was a lower confidence in their accuracy
- remove invalid addresses from the final frame

The sources used (covered in more detail below) include:

- Ministry of Defence Armed Forces accommodation
- US Visiting Forces accommodation
- Ministry of Justice (prisons)
- Cushman and Wakefield (student halls)
- Department for Levelling Up, Housing and Communities and Statistics for Wales Traveller caravan data

- [Care Quality Commission](#) and [Care Inspectorate Wales](#) (care homes)

The initial frame was produced in August 2020, to allow for the printing of initial contact letters. A further version which accounted for possible new addresses or case type changes (for example, from a household to CE or vice-versa) was produced in January 2021. Further deliveries of address data from GeoPlace were also used to supplement the frame up until Census Day, 21 March 2021.

Further administrative data sources were utilised to create addresses for the frame for some establishment types that cover smaller sub-groups of the population, like boarding schools and homeless shelters. These include:

- [Probation Finder](#) - formerly called UK Probation Directory (approved premises)
- Department for Education's [Get Information About Schools](#) (GIAS) - formerly called Edubase (boarding schools)
- [English School Census](#) (boarding schools) (covered in [Section 16](#) of the report)
- [Care Inspectorate Wales](#) and GIAS (children's homes)
- [Homeless Link](#) (hostels for the homeless and homeless day shelters)
- [Independent Hostel Guide](#) web scraped data (hostels)
- [Care Quality Commission](#) (low or medium secure mental health hospitals)
- [NHS](#) (hospitals)
- [Hospice UK](#) web scraped data (hospices)
- [Government Diplomatic List](#) (embassies)

2.1 Ministry of Defence Armed Forces accommodation

The UK Armed Forces and their families are entitled to live in subsidised accommodation in the UK, either at or within an appropriate distance from their place of work. There are two main types of accommodation provided by the Ministry of Defence (MoD) for the armed forces; [Single Living Accommodation (PDF, 1,555KB)](#) (SLA) and [Service Family Accommodation](#) (SFA).

SLA is accommodation provided to those who are single, married or in a civil partnership but unaccompanied service personnel, generally within barracks at, or close to, their duty station. SFA is accommodation provided for the authorized, eligible, or entitled use of military personnel and their dependants. SFA can be either on or off base but always close to it. Officers are allocated SFA primarily by rank; however, other factors such as family size, appointment, representational responsibilities, and personal choice may influence the final allocation.

MoD provided data on both SLA and SFA accommodation to the ONS to support the creation of the Census 2021 Address Frame (the frame). The data helped ensure the completeness and accuracy of the frame for those living in the different accommodation types, ensuring the correct Census 2021 questionnaire form was sent to all addresses (a household form for SFA or a communal establishment one for SLA).

Coverage

Although the MoD hold data covering the whole of the UK, in line with census-taking responsibilities, only data for England and Wales were required by the ONS. The data provided good coverage for the armed forces living in England and Wales, but some missing SFA and SLA were identified during collection operations that had not been supplied in the MoD.

Accuracy

The MoD data were accurate for both SLA and SFA accommodation types, supporting the quality assurance of the frame constructed from ABP, and updates to the frame, where address data were inaccurate or missing from ABP.

Timeliness

A snapshot of the MoD data taken in July 2020 was supplied to the ONS. This was used for the first version of the frame produced in August 2020. The data were then checked again via desk-based research as part of the production of the final frame in January 2021. This was to ensure that the MoD supplied data were still valid and accurate close to Census Day on 21 March 2021.

2.2 United States Visiting Forces accommodation

Similar to the UK Armed Forces discussed in Section 2.1, United States Visiting Forces (USVF) and their families are eligible to live in Single Living Accommodation (SLA) or Service Family Accommodation (SFA). The associated data were provided to the ONS to validate the Census 2021 Address Frame (the frame) and where possible to add any missing addresses. In a similar way to the UK Armed Forces data, the USVF data were also used to identify the correct census form type to send (a household form for SFA or a communal establishment one for SLA).

Coverage

USVF data only cover England since there are no bases with USVF presence in Wales. The data provided comprehensive coverage of both SLA and SFA accommodation.

Accuracy

The USVF data were accurate for both SLA and SFA accommodation types, which supported both the quality assurance of the frame constructed from AddressBase Premium (ABP), as well as updates to the frame, where address data were inaccurate or missing from ABP.

Timeliness

The data were provided to the ONS in July 2020 and were used for the first version of the frame in August 2020. The addresses were checked again close to Census Day to ensure they were still accurate.

2.3 Ministry of Justice (prisons)

Her Majesty's Prison and Probation Service (HMPPS) is an executive agency, sponsored by the Ministry of Justice, supported by HM Prison Service, Probation Service and Youth Custody Service. It is responsible for carrying out sentences given by the courts, either in custody or the community, and for rehabilitating people in their care through education and employment.

The HMPPS list was used to support the creation of the Census 2021 Address Frame (the frame). Namely, to confirm the existing prisons' list in the AddressBase Premium (ABP) and add any additional prisons that may have been missed. In addition, Youth custody data, provided by the Youth Custody Service, were used to capture Secure Training Centres for young people in custody.

Coverage

The HMPPS list and Secure Training Centres data include all the relevant establishments in England and Wales. There were 115 prisons in England and five in Wales, and three Secure Training Centres overall.

Accuracy

The HMPPS data provide a list of prisons and Secure Training Centres presenting high accuracy of address level information as was needed for the creation of the frame.

Timeliness

An extract of all prisons and Secure Training Centre addresses was taken in May 2020 and used for the first version of the frame (August 2020). The addresses were checked again close to Census Day to ensure they were still accurate.

2.4 Cushman and Wakefield (student halls)

Cushman and Wakefield (C&W) are a global real-estate services firm with an insight into commercial properties, including student accommodation in the UK.

The ONS used the Student Accommodation Tracker (SAT), which is a C&W database holding a list of addresses for university-owned and privately owned student halls in the UK. The data were used to help create a list of student halls of residence for the Census 2021 Address Frame (the frame), which formed the basis for both the census collection (through the communal establishment questionnaire) and the Student Occupancy Survey collection operations (see Section 13).

Coverage

Rather than providing comprehensive coverage, C&W data were used to add 1,600 halls of residence to the frame. Whilst there are more halls of residence in the C&W dataset and within England and Wales as a whole, only a subsection was added as these data were used to supplement halls added through a separate piece of engagement work. This involved directly contacting universities and private hall of residence providers and asking for unit-level information on some of their halls of residence, where the ONS did not have this information, and covered approximately 2,000 other halls of residence.

Accuracy

The accuracy of the C&W data was high in terms of correctly identifying establishments which were student halls of residence. There were some initial difficulties with punctuation within the C&W addresses which affected the ONS's ability to automatically match the data to the AddressBase Premium (ABP). These were resolved through a manual resolution of room level C&W addresses.

In addition, on occasion, multiple halls of residence (or blocks of a hall of residence) were combined in C&W data under one entry which referred to the colloquial name for the complex. This caused an issue during live operations as, whilst the correct number of forms were delivered, students were all captured under a single hall of residence,

where it may have been more appropriate to capture these students across multiple separate halls in these complexes.

Timeliness

The ONS received two data extracts from C&W to support the census. One in July 2020, which was used for the first version of the frame (August 2020), and one in October 2020, which was used for the final version of the frame (January 2021).

2.5 Department for Levelling Up, Housing and Communities and Statistics for Wales Traveller caravan data

Provided by the Department for Levelling Up, Housing and Communities (DLUHC) (formerly the Ministry of Housing, Communities and Local Government) and Welsh Government's Stats Wales, Traveller caravan data are part of a series of official statistics on the count of Traveller caravans in England and Wales.

The counts are carried out to provide local insights on the number and seasonal movement of caravans. Local authorities carry out the counts twice a year, in January and July, and include caravans on both authorised and unauthorised sites. Authorised sites are provided either by local authorities, private registered providers of social housing, or on land owned by Gypsies and Travellers and which have a (non-expired) planning permission (either temporary or permanent). Unauthorised sites are those without planning permission (which can be on either land owned or not owned by Travellers).

The data were used by the ONS to update the list of authorised caravan site addresses in the Census 2021 Address Frame (the frame). This was to ensure that those living in caravans were not missed from the census enumeration.

Coverage

The data extracts used to support Census 2021 (see Timeliness Section) covered 266 authorised Traveller caravan sites in England, and 94 in Wales. For England, this covered all socially rented caravan sites included in the official Traveller caravan count:

January 2020[1]. For Wales, the official count of January 2020 included 102 authorised sites; however, the extract that was available to the ONS for use in the frame was that of July 2019, when the official count for Wales was 94 authorised sites. The official counts for Wales are available at: Number of sites by authorisation and local authority.

Accuracy

Caravan sites addresses were linked to AddressBase Premium (ABP) to validate the list of addresses in the frame. Given some inconsistencies in the structure and the type of existing ABP addresses, that is, whether the addresses referred to caravan sites or units (caravans), the ONS decided to link the data clerically to ensure matching to the right address. Therefore, there was not a reliance on high accuracy of the received data to achieve strong automatic matches. The ONS was able to manually assign a Unique Property Reference Number (UPRN) to each record in the administrative data and include them all in the final version of the frame.

Timeliness

Traveller caravan data are collected in January and July every year. However, in line with official guidance from the UK Statistics Authority and the Office for Statistics Regulation, the collection of the summer 2020 and winter 2021 Traveller caravan count for England was suspended due to the coronavirus (COVID-19) outbreak. Therefore, the extract used to support Census 2021 was the Traveller caravan count: January 2020. For Wales, the collection of the summer 2020 Gypsy and Traveller caravan count was also suspended due to the coronavirus outbreak; but the closest extract to Census Day (21 March 2021) that was available to the ONS for use in the frame was that of July 2019. However, the ONS had the facility to include any additional Traveller sites which were made aware of during the collection through local authorities or members of the public.

---

[1] DLUHC does not hold a definitive list of sites across England. The data received covered socially rented Traveller sites provided by local authorities and private registered providers of social housing. Although address level data for privately funded sites are not published, the quality of the frame is unlikely to be affected as private caravan sites are captured through the main census enumeration (as households).

2.6 Care Quality Commission and Care Inspectorate Wales (care homes)

The administrative data sources used to enhance the list of care homes in the Census 2021 Address Frame (the frame) are the Care Quality Commission (CQC) dataset for care homes in England and the Care Inspectorate Wales (CIW) for those in Wales. The data provided hold a list of care homes' names and addresses. The addresses on the administrative data were matched to the AddressBase Premium (ABP) in order to verify that all care homes had the correct ABP classification code before including them in the frame.

Coverage

The ONS accepted all matches made between the administrative data and the ABP addresses which had the correct classification code in ABP, carrying out further checks on the records that did not have the correct classification. There were around 15,300 addresses which were classified as care homes in the frame. Since every care home must be registered with either the CQC or the CIW, both sources had a good coverage of all the care homes found in England and Wales.

Accuracy

Both the CQC and CIW data provided detailed and highly accurate information on all establishments registered as care homes. Moreover, the datasets provided very reliable and accurate address information for all care homes, which enabled the ONS to easily link the dataset to the ABP product and ensured they were able to be added to the frame. One challenge faced with the CQC data was the Unique Property Reference Numbers (UPRNs) supplied with these data for each property. It was discovered that a number of these were inaccurate (that is, the UPRN provided did not correspond to a valid UPRN in ABP), and this led to a number of establishments not being included on the final frame[2]. However, these establishments were subsequently accounted for and corrected as part of large communal establishment estimate work conducted following enumeration.

---

[2] CQC are aware of the issue and are investigating what is needed within their systems to keep the data up to date.

Timeliness

The ONS received the first extract of both CQC and CIW in July 2020 in time for the first version of the frame (August 2020). Second extracts of both datasets were received in November 2020, which helped identify new additions and deletions from the lists, in time for the final version of the frame (January 2021). Any new care homes that became active between January 2021 and Census Day (21 March) were identified either during the census collection operation or the census estimation process.

3 Driver and Vehicle Licensing Agency

The Driver and Vehicle Licensing Agency (DVLA) is an executive agency of the Department for Transport, responsible for holding more than 49 million driver records and more than 40 million vehicle records. The aim being to maintain the registration and licensing of drivers in Great Britain. DVLA also maintain the registration and licensing of vehicles, together with the collection and enforcement of Vehicle Excise Duty (VED), in the UK.

To support Census 2021, the ONS used aggregate DVLA data on driver transactions, such as when applying for a new licence or renewing an existing licence. The ONS used these data to update the methodology for the digital domain of the Hard to Count index (PDF, 833KB) (Office of Communication data were also used in the index, which is covered in Section 4). That is, as an indication of peoples' use of internet to interact with a government website. The index in turn was used to identify areas (lower-layer super output areas (LSOAs)) which were provided with a paper questionnaire as first contact rather than simply an invitation letter and online access code.

Coverage

The data provided to the ONS on the number of transactions to apply for or renew a driving licence in England and Wales are aggregated to LSOA level of geography. The transactions are for people aged 16 and over by single year of age and sex and split by mode of transaction (paper or online). The first year of data received was 2016, followed by 2017 (see Timeliness Section).

In 2016 the data covered 16% of the population in England and Wales aged 16 and over, and 7.6 million transactions, of which 55% were made online. Increasing to 17% of the same population in 2017, and to 7.9 million transactions made in 2017, of which 59% online. This is consistent with research conducted in 2019 by DVLA, which suggests

there has continued to be an increase in the uptake of digital services for DVLA transactions.

DVLA data have partial population coverage, as they include only people aged 16 and over, and among these only those who apply for or renew a driving licence. People who never applied or renewed a driving licence will not be captured in the data. In addition, online transactions exclude anyone not holding a UK passport and those whose name or title have changed (for example, married), as these groups are not permitted to apply via the online route.

Accuracy

DVLA data helped the ONS identify areas and groups that were more likely to need digital assistance when completing their census questionnaire, which informed the census field collection and follow-up activities. The data showed that the percentage of online transactions decreased with age, with the oldest age group (aged 75 years and over) having the lowest percentage of online transactions. It is possible that a younger individual will apply on behalf of an older applicant, which may cause inaccuracies when using the DVLA data as a proxy for a person's digital uptake. However, this is unlikely to affect area level digital uptake, which is how the data were used to support Census 2021. Approximately 1% of transactions in 2016 were missing age and sex information, which reduced to 0.5% in 2017.

Timeliness

The 2016 and 2017 extracts of DVLA data were used to develop the Hard to Count index methodology. The final version of the index that fed into the Design for Census 2021 used 2017 DVLA data.

4 Office of Communications data

The Office of Communications (Ofcom) is the UK's communications regulator and has responsibility for tracking the coverage and availability of broadband and mobile services.

Similarly to DVLA data (as described in Section 3), data from Ofcom for 2016 and 2017 were used to develop the digital domain of the Hard to Count index. Whereas DVLA data were used to measure whether households used the internet to interact with a government website, Ofcom data measured whether households had the digital infrastructure to respond to an online census.

The ONS used the Ofcom data to measure the percentage of households in an area that had access to the internet, estimated using the number of fixed line broadband connections to premises from Ofcom, and the number of occupied households from the 2011 Census. Ofcom and census data were matched at postcode level and then aggregated to lower-layer super output areas (LSOA) level of geography. Details about this can be found in [Hard to Count index for the 2021 Census (docx, 1.24MB)](#).

Coverage

The Ofcom data cover broadband connections to premises in England and Wales, which can include both residential properties and small business premises. This means that the number of connected premises will, in general, be higher than the number of residential properties with a broadband connection within an LSOA (see Accuracy Section). Nonetheless, the data provided a useful proxy for levels of household connectivity by local area.

Accuracy

For disclosure reasons, Ofcom data were not available for several postcodes in England and Wales where there were fewer than four broadband connections (in 2016 this was around 180,000 postcodes). In such cases, for the Hard to Count index, an assumption was made that there were at least two broadband connections in each of the postcodes.

As mentioned, the data included both households and small business premises, as a consequence around 2.5% of LSOAs (878 in 2016) had a higher number of premises with a broadband connection than there were occupied households in an area (based on the 2011 Census). In these cases, an assumption was made that the percentage of households with access to the internet was 100%.

Timeliness

Ofcom annually publishes open data on the number of fixed line broadband connections at postcode level. The ONS used Ofcom data from 2017 to construct the final Hard to Count index.

5 Utilities (gas and electricity)

There are several companies that support the gas and electricity industry by providing a single point of contact and data transfer solution for energy suppliers and distributors. As such, they hold a record of most properties with an active gas or electricity supply,

including data on the supplies based on meter readers and smart meter data. These data were used for the census to provide insight into the number of properties that were likely to be occupied within a local area. This supported the quality assurance of census household estimates. There are three commercial organisations that provided data to the ONS:

- Xoserve is the Central Data Service Provider (CDSP) for Great Britain's gas market. They are responsible for providing information to gas transportation companies from their central register
- ElectraLink is responsible for operating the data hub that underpins the UK energy market, facilitating the development of a more efficient and competitive market
- the Electricity Central Online Enquiry Service (ECOES) assists suppliers in the electricity customer switching process by validating meter point registration numbers and addresses

Coverage

Xoserve hold information on more than 24 million gas premises covering all properties which have an active gas contract. This therefore excludes households that do not use gas, which constitute 14% (approximately 4 million) of all domestic properties in Great Britain. For information about this at different geographies see Sub-national estimates of properties not connected to the gas network.

ElectraLink covers properties with both gas and electricity. However, only data on electricity supplies were used for the census since Xoserve already provided adequate data on gas supplies.

ECOES cover properties with an active electricity contract. Among the three sources, ECOES has the highest coverage of residential addresses in England and Wales covering approximately 26 million addresses compared to 23 million properties provided with electricity by ElectraLink. ECOES total address counts closely align with those from Council Tax and Valuation Office Agency data. In combination, the utility companies cover most residential addresses in the country.

Accuracy

The addresses on the utilities data compared well against those recorded on the Census 2021 Address Frame for England and Wales. However, the address on utilities data refers to the location of the meter, which in certain cases, might not match the exact

address of a household. For example, where a single meter point covers multiple households in blocks of flats and households of multiple occupancy.

An indicator of occupancy was created by the data suppliers working with the ONS, using information on whether there was a live meter point and the level of energy usage, including amount and dates of readings for domestic properties. This provided a useful indication of the number of occupied households within a local area. However, the use as an indicator is complex as there were some notable limitations, whereby gas and electricity might have been used at a property that was not occupied as a main residence under the census definition. For example, properties that are used frequently as short-term lets are not generally considered anyone's usual residence but will be using utilities at a similar level to an address that is someone's home. In addition, owners of second homes, or otherwise vacant properties, may keep some level of heating on during winter to avoid pipes freezing, have cleaners in the property using appliances, or may keep lighting on to deter burglars. Due to these limitations, utilities data were used with other sources of information, such as local authority supplied Council Tax data (see Section 7).

Timeliness

Xoserve data were extracted on 22 March 2021, the day after Census Day, and contained information on occupancy based on meter readings submitted in the last two years. ElectraLink data were supplied in early April and provided information on occupancy based on estimated annual consumption and whether there had been meter readings close to Census Day, that is, within six weeks before, or three weeks after this date. ECOES data are based on monthly extracts and the extract used for census had 31 March 2021 as the reference date. ECOES data provided information on occupancy based on whether there had been any events, such as a customer switching electricity supplier, or the property being disconnected, and the "effective from date" for these.

6 Valuation Office Agency property attributes

The Valuation Office Agency (VOA) is an executive agency of Her Majesty's Revenue and Customs (HMRC). Since the 1990s, it has been responsible for banding dwellings liable for Council Tax (CT) in England and Wales. To fulfil this function, VOA collect data on property attributes for residential properties. Further information about the VOA data can be found in the Valuation Office Agency data source overview.

For Census 2021, the ONS used VOA data to:

- assess the total number of dwellings by local area and provide insight into new-build developments
- impute for type of accommodation in instances of non-response to the census
- provide data on number of rooms, in place of directly collecting this on the census questionnaire as referenced in the Census 2021 White Paper (PDF, 967KB)

Aggregate statistics published in Table CTSOP3.0 (xlsx, 25.2MB) of the Council Tax: stock of properties, 2021 were also used, to validate census statistics on the number of bedrooms and accommodation type (for example, house, flat).

Coverage

The VOA data extract used for Census 2021 had approximately 26 million records, nearly 25 million in England and more than 1 million in Wales. The data have excellent coverage of the census address frame for households. The ONS successfully linked 96% of Census 2021 households to VOA property data.

Both census and VOA data do not include properties that are in disrepair, uninhabitable or derelict, as such properties will be removed from the CT valuation list. The VOA provide detailed information on the process of valuing domestic properties in How domestic properties are assessed for Council Tax bands.

Accuracy

The property attribute data are updated when a new residential property is built and thus a new entry is created in the database, or when new information about the property comes to the attention of the VOA, either when a challenge (or band review) is raised by a taxpayer or a VOA List Officer or when the property is sold. However, changes to a property can occur without a trigger to VOA. For example, homeowners can change internal partition walls without the need for planning permissions or make structural changes to the property without legal planning permission, and the information has not fed through to VOA (as a result, the number of rooms may be inaccurate).

Thus, while VOA makes every effort to maintain an accurate list, the accuracy of VOA data vary by geography and property type, because data updates are mostly linked to the sale of properties. Nonetheless, previous research by the ONS based on the 2011 Census showed a high level of consistency between VOA and census accommodation type and that VOA number of rooms is a suitable alternative to the census number of rooms questions, despite definitional differences.

Timeliness

A snapshot of the VOA data as of 12 April was used for Census 2021. This ensured that the reference period covered Census Day (21 March 2021). The aggregate statistics used from the Council Tax: stock of properties are published on an annual basis and refer to the stock of all dwellings existing between 1 April 2020 and 31 March 2021 in England and Wales, by Council Tax band and property attributes.

7 Local Authority supplied Council Tax data

Each local authority (LA) in England and Wales is responsible for the collection of Council Tax (CT), a yearly charge for all domestic properties. CT data are collected monthly at dwelling level and include information about exemptions, discounts, and premiums applied to certain types of properties.

CT data were used for Census 2021 to assess the number of properties likely to be occupied or unoccupied within a local area. They were also used to enhance the Census 2021 Address Frame (the frame) (see Section 2) by identifying potential additional domestic addresses.

Coverage

As with Valuation Office Agency (VOA) property attribute data (see Section 6), locally provided CT data cover all dwellings in England and Wales that are either habitable or capable of being repaired.

Information on properties that receive an exemption, discount or pay a premium due to being vacant or second homes was used to identify dwellings that were likely to be occupied or unoccupied for census purposes. However, the rules around exemptions, discounts and premiums vary by local authority and over time. For example, not all LAs offer a second home discount, and if they do, the discount could be withdrawn at any time. This makes it difficult to make comparisons across LAs and compare to census data (see Accuracy Section).

Currently in England, empty homes for CT purposes are dwellings that are unoccupied and substantially unfurnished, as detailed in Council Tax base 2021 in England (PDF, 326KB). At the LA's discretion, empty homes can be subject to a discount of between 0% (no discount) and 100% (exemption), unless they have been empty for more than two years. If this is the case, they may be subject to a premium in addition to the full amount of tax due. Second homes, for CT purposes, are dwellings that may be empty for part of the year and are furnished and may be eligible for a discount of between 0% (no discount) and 50%. The rules are slightly different in Wales where most LAs no

longer give discounts to long-term empty or second homes and some charge a premium for second homes. For more information about this see Council Tax Dwellings in Wales, 2020 to 2021(PDF, 475KB).

Accuracy

While billing authorities make every effort to ensure that the correct level of CT is paid by all chargeable dwellings, the accuracy of CT data to assess the numbers of occupied and unoccupied addresses varies across different areas. This is driven by the differing rules around exemptions and discounts between LAs and by different levels of take-up of discounts and exemptions. For example, in areas where full CT rates are already low, owners of second homes may not claim a second home discount, making the property appear as their main home (or occupied for census purposes) when it is not.

Some homeowners may consider their second home (by the census definition) as their usual residence. This is further complicated by the possible effect of the coronavirus (COVID-19) pandemic, whereby the increase in home working meant some may have chosen to use their second homes as the main residence during the census period. This would cause discrepancies between the CT data and census counts, particularly at local area level.

In addition, CT data may include addresses that for census purposes would not count as occupied household spaces. For example, some CT records will relate to communal establishments (CEs), or constituent units of a CE, such as, individual rooms or flats in a student hall. Newly built houses may also be liable for CT before they are occupied as may rental properties when briefly unoccupied between tenancies. All these factors may cause discrepancies between the number of addresses that are likely to be occupied based on CT data and Census 2021 household estimates.

Timeliness

Locally provided CT data are supplied monthly to the ONS. The April 2021 extract of CT data was used to support the census, which covered the census reference date. A small number of LAs were not able to provide data as of April 2021, but slightly later. Back series data were also used where possible to supplement the April 2021 extract. However, there may have been lags in the data possibly due to delays in people updating their CT records or communicating changes in their circumstances, such as moving home, or claiming an exemption or discount.

8 Immigration Detention Dataset

Immigration removal centres and short-term holding facilities are holding centres for foreign nationals who are typically awaiting decisions on an asylum or other type of protection claim or waiting to be returned. The Immigration Detention Estate dataset, produced by the Home Office, provides information on those detained in the UK solely under Immigration Act powers in immigration removal centres and short-term holding facilities. Data on detentions come from Home Office administrative data systems, specifically the Case Information Database (CID). Case workers enter information onto the system for operational purposes. For more information on the Home Office dataset see the User Guide to Immigration Statistics.

The ONS received aggregate data on age, sex, nationality, place of detention and length of detention at the immigration detention estate. The dataset was mainly used for two purposes. Firstly, to understand whether the centre had 50 or more usual residents and so whether the establishment should be considered a large communal establishment (CE) or not. Secondly, the dataset was used to correct for possible under-coverage in the census of those living in large CEs (immigration detention centres) [3].

Coverage

The dataset has full coverage of all eight immigration removal centres and short-term holding facilities in England and Wales. There are explicit rules on who is held in these centres and who appears in the dataset. Individuals are only detained under immigration powers in certain circumstances, while case workers enter information for all individuals detained into a centralised database. This means the dataset should cover almost everyone in an immigration removal centre or in a short-term holding facility.

Accuracy

As stated above, the accuracy of the Immigration Detention Estate dataset is believed to be high. There are minor definitional differences in the data compared with census. That is, for census purposes, people who reside in an immigration detention centre should be enumerated only if they have been or intended to stay in the UK for three months or more. In the Immigration Detention Estate dataset, there is only information about how

---

[3] In Census 2021, large (as opposed to small) communal establishments are establishments providing managed residential accommodation for 50 or more usual residents. This distinction is made because the ONS uses a different method to correct for under-coverage in large CEs - see Estimation and Adjustment of Large CEs (PDF, 208KB). A future publication will explain in more detail the large CE estimation process and how under-coverage was corrected in this type of establishment.

long an individual has been in detention and not about length of stay or intention to stay in the UK. However, it is likely that a person who is claiming asylum or other type of protection in the UK, has the intention to stay for three months or more. Additionally, those in detention who have not made an asylum or other protection claim are likely to have already been in the UK for three months or more before their detention. Therefore, the ONS made this assumption and included all those who were in an immigration removal centre or in a short-term holding facility in England and Wales to compare to the numbers of usual residents of these establishments.

Timeliness

The dataset referenced 31 March 2021, which provided data close to Census Day (21 March 2021). There may be a small number of cases where there is a delay or failure to notify the caseworker that an individual is being transferred into immigration detention, which could lead to some individuals not being included in the dataset.

9   Ministry of Defence Armed Forces counts

The Ministry of Defence (MoD) Armed Forces personnel data include the number of serving UK Armed Forces personnel and civilian personnel with a Defence Medical Services (DMS) registration. Personnel with a DMS registration have their primary care (GP services) provided by the MoD rather than by the NHS.

The MoD data are compiled by Defence Statistics Health from the Defence Medical Information Capability Programme (DMICP) and are based on the number of DMS registrations in DMICP. This includes UK Armed Forces (including Regulars, Gurkhas, Officer Designates and Full Time Reserve Service personnel) and entitled civilians (that is, family dependants of service personnel and MoD employed civilian personnel who are entitled to care). The dataset contains DMS registrations by location, age and sex. The data are open source and released bi-annually. For more information see Defence personnel NHS commissioning bi-annual statistics.

The MoD data were used as a main comparator source to quality assure the census UK Armed Forces population estimates, alongside the Personal Demographic Service (PDS) (see Section 18).

Coverage

The MoD data include all UK Armed Forces and civilians entitled to care provided by the MoD. The numbers presented are not representative of the number of MoD employed

civilians or military dependants associated with the MoD, as most MoD civilian employees are not entitled to military health care and the MoD do not publish data on those without a DMS registration.

Accuracy

Individual MoD medical centres are responsible for ensuring the accuracy of the patient registration information. The data are collected by many different medical and administrative staff and are subject to quality limitations. However, demographic data for UK Armed Forces are cross referenced with the Joint Personnel Administration (JPA) system which is the MoD's reference for personnel information. For more information on the quality of the data see the Bi-annual NHS Commissioning Population Statistics Background Quality Report (PDF, 119KB).

The data were aggregated into clinical commissioning groups (CCGs), which were NHS bodies responsible for the planning and commissioning of health care services for their local area[4]. This makes it difficult to compare the data to census estimates which are aggregated using census geographies. As the data was not disaggregated by accommodation type, the ONS was unable to use this information to separately quality assure the counts for military personnel in households and communal establishments (CEs). Therefore, a post-census follow-up survey was conducted which allowed the ONS to quality assure the numbers of CE residents and to support with coverage adjustments[5].

Timeliness

The data provided by the MoD are entered into the electronic patient record which is updated regularly providing timely data. Defence Statistics Health extract the data and publish NHS Commissioning figures on a bi-annual basis. The MoD aggregate data used by the ONS were a snapshot taken on 1 April 2021, providing information close to the census reference period.

---

[4] CCGs replaced primary care trusts on 1 April 2013. As of 1 April 2021, there were 106 CCGs in England. From July 2022 they were dissolved, and their duties taken on by the new integrated care systems.
[5] The aim of the post-census follow-up survey was to obtain administrative data from specific establishment types. This involved contacting establishments after census collection, and requesting data on the age and sex of usual residents in the establishment as of Census Day, supporting the Estimation and Adjustment of Large CEs (PDF, 208KB).

10 United States Visiting Forces counts

Data for foreign armed forces resident in England and Wales are collected by the United States Visiting Forces (USVF). They collate data from several sources to include liaison officers, service Headquarters, and local reporting units within the UK.

UK Host Nation Coordination Cell (HNCC) are the data controller for the USVF personnel data they hold. The data are taken from six sources and include data from US Army, US Navy, US Air Force, US Space Force, US Coast Guard, US Marine Corps, as well as US Civilian workers. The ONS received aggregate counts of US Air Force personnel and dependants (partners and children) located in the UK broken down by base, single year of age and sex.

To support Census 2021, USVF data were used to quality assure the US Air Force population counts on the census. The USVF data were the only source the ONS could use as a comparator for the US Air Force population. However, the use of the data for census was limited, as described below.

Coverage

The USVF data include all USVF personnel and their dependants in England (there are no USVF bases in Wales) that are permanently assigned to the UK for three months and longer, which also met the census definition of a usual resident. Foreign armed forces personnel who are not from the US are not accounted for as part of the special populations as there are no data currently available, however, these are considered very small in number.

Accuracy

USVF data are at establishment level and refer to the base in which personnel are stationed rather than the local authority of residence. Inevitably there will be personnel living off base, potentially in a different local authority, which means they may not be counted in the right area when comparing to the census. Similar to the MoD Armed Forces personnel data (see Section 9), USVF personnel data do not distinguish between those usually resident in communal establishments (CEs) or households. The ONS were, therefore, unable to use the data to quality assure the census data by different residence types.

Timeliness

The USVF data were extracted between 1 September and 1 November 2021, which meant that the data reference period was six to eight months after Census Day.

11 Ministry of Justice Prisoners (Prison-NOMIS)

Ministry of Justice (MoJ) prisoner data are taken from the Prison-National Offender Management Information System (Prison-NOMIS) operational database which is used in prisons for the management of offenders. Prison establishments record details for individual inmates on the prison IT system, known as Prison-NOMIS. For more information on MoJ data on prisoners see Prisoners data, quality assurance of administrative data used in population statistics, May 2019.

The ONS received aggregated data on the population of prison establishments in England and Wales, broken down by single year of age, sex, ethnicity and custody status. It also included whether the individual was sentenced to 6 to 12 months, or 12 months and more. MoJ data were used to validate the Census 2021 population estimates for prison communal establishments (CEs) in England and Wales and to correct for under-coverage of the prison's population if present.

Coverage

The MoJ dataset is comprehensive as it covers all prisons across England and Wales with a requirement that all prisoners be included, since prison establishments must record details on the prison IT system (Prison-NOMIS).

For census purposes, prisoners sentenced for fewer than 12 months should be counted as usually resident at their permanent or family home unless they have no other place of usual residence, while those sentenced for 12 months or more should be counted as usual residents at the prison. Information on how prisoners are defined in Census 2021 is detailed in Output and enumeration bases: residential address and population definitions of Census 2021.

Additionally, if a person has been convicted and is in prison awaiting sentencing (that is, they are on convicted unsentenced remand), they should be counted as usual residents at the prison. The MoJ dataset provides information about the length of sentence and allows identification of those who are on convicted unsentenced remand as opposed to those on pre-trial remand. This means there is alignment with the census definition of usual residence. The inclusion of prisoners in the data received who may leave prison before the end of their sentence and not reside there on Census Day, may lead to over-

coverage in MoJ data when used for census purposes. However, in terms of the overall prison population, it is likely that this affected very few, if any, prisoners.

Accuracy

MoJ data were useful in providing detailed information on individuals serving a sentence of 12 months or more in prison. The only aspect of the census definition not covered in the MoJ data is around capturing individuals who should have been enumerated in the census at the prison, despite having a sentence of less than 12 months as they had no other place of usual residence. This is because MoJ data does not contain information on where people serving a sentence of under 12 months were living before detention or what they would consider as their place of usual residence if not the prison. This meant the ONS were unable to identify those who were serving a sentence of fewer than 12 months who should have been enumerated in the census at the prison. However, the number of individuals falling into this group is likely to be low and will not have had a major impact on the quality of MoJ data used to support the census.

Timeliness

The reference period for the data was 19 March 2021, which closely aligns with Census Day (21 March 2021).

12 Capacity Tracker data for care homes

Capacity Tracker is a web-based digital tool built by the North of England Commissioning Support Unit (NECS) in partnership with NHS England. The tool is a single data capture platform, providing insight into what support is needed and where in the health and social care sector across England. Establishments such as care homes, substance misuse facilities, NHS community rehabilitation and hospices are asked to upload bed vacancy counts to the platform daily, alongside other information such as the number of coronavirus (COVID-19) cases and vaccinations, Personal Protective Equipment (PPE), and workforce information. They are not legally obliged to do this, but since the start of the coronavirus pandemic the use of Capacity Tracker has been encouraged.

Capacity Tracker combines data from the Care Quality Commission (CQC) with data collected within the platform. The CQC data include basic information about the establishments, such as their location, service provided and their capacity, in terms of total number of beds, while the data collected via the platform provides real-time information on number of vacant beds.

To support Census 2021, the ONS received aggregate data from all establishment types, the vast majority being care homes. The data included name and postcode of the establishment, bed vacancies and total resident count. The data were used to assess census estimates of the care home communal establishments (CEs) population and in large CE estimation[6].

Coverage

Capacity Tracker data have a very high coverage compared with the census, with 99.8% of all health and social care sector establishments in England included in the data, and all local authorities covered. Most of the establishments captured in the data are care homes. The data source is only available for England, so all Welsh establishments are excluded.

In terms of coverage of the residents, Capacity Tracker provides information on total number of beds and number of vacant beds, which were used to derive number of occupied beds at a point in time. However, not all beds may be occupied by people who meet the census definition of a usual resident of a CE. This definitional difference was considered when conducting large CE estimation and Capacity Tracker data were used in conjunction with other sources of information. For example, the CE expected capacity stated by the census field officer, the number of residents as stated by the CE manager on the census form, and data from the Personal Demographic Service were also used, to determine the correct number of usual residents, or those who should have responded to the census at each establishment on Census Day.

Additionally, demographic information is not collected in Capacity Tracker data and therefore this source did not provide an age-sex breakdown to use as a census comparator.

Accuracy

In terms of the accuracy of the information on the Capacity Tracker, establishments have a vested interest to keep the data as accurate as possible for operational purposes (for example, to ensure that bed vacancies are appropriately filled). Additionally, NECS

---

[6] In Census 2021, large (as opposed to small) communal establishments are establishments providing managed residential accommodation for 50 or more usual residents. The ONS uses a different method to correct for under-coverage in large CEs - see Estimation and Adjustment of Large CEs (PDF, 208KB). A future publication will explain in more detail the large CE estimation process and how under-coverage was corrected in this type of establishment.

data analysis team validates the data by checking the vacancy numbers provided do not exceed the total capacity, as provided by CQC. Although establishments are encouraged to upload information on the platform daily, this is not a mandatory requirement so the number of responses provided can fluctuate. During the coronavirus pandemic only 80% of the establishments registered on the platform completed the tracker regularly. This may have caused inaccuracies in the reported number of vacant beds due to delays in updating the tracker.

Timeliness

The data were delivered to the ONS on the 30 April 2021 and comprised of a weekly collection of data through the period 5 April 2020 to 4 April 2021. Receiving a back series of data allowed the ONS to assess vacancy rates over time, both in the lead up to Census Day and the following week.

13 Student Occupancy Survey

Student occupancy survey data refer to administrative data collected by the ONS on students who were contracted to live in communal establishments (CEs) at the time of the census. These data were collected to assist in correcting possible under-coverage in large CEs (those with 50 or more usual residents)[7].

All potential large CE boarding schools and most student hall providers were sent an electronic questionnaire, by the ONS, via either the newly created Student Hall Survey or Boarding School Survey. Each provider was asked to submit the number of usual residents contracted to live at their establishment on 21 March 2021[8]. In addition, providers were asked to break the number down by age, sex, and international student status. An international student was classified as an individual with a non-term time address that was outside of the UK.

---

[7] In Census 2021, large CEs are establishments providing managed residential accommodation for 50 or more usual residents. A different method is used to correct for under-coverage in large compared to small CEs (Estimation and Adjustment of Large CEs (PDF, 208KB)). A future publication will explain in more detail the large CE estimation process and how under-coverage was corrected in this type of establishment.

[8] Detailed information on the enumeration of students during Census 2021 and the usual resident definition for students living in halls of residence can be found in this paper on the Methodology for accurately enumerating students in Census 2021 (PDF, 315).

A different data collection method was used for the 10 largest private student accommodation providers, as it was deemed too burdensome to ask a single provider to potentially complete over 100 electronic questionnaires. Consequently, separate engagement was done with these providers and a different mode of data delivery was conducted.

Coverage

The list of establishments, included in the survey, came primarily from the Census 2021 Address Frame (the frame) (see Section 2). The ONS supplemented this list with desk-based research, establishments discovered during census collection and discussions with suppliers. Overall, a total of 682 providers were contacted across England and Wales, with 629 valid providers found. Of these, 364 were boarding schools and 265 were university or privately owned halls of residence providers, covering a combined total of more than 2,000 student halls.

A total of 557 providers responded, with 92% of boarding schools, 90% of universities, 20% of the 10 largest private student accommodation providers and 75% of the remaining private providers supplying a full response. Some student hall providers were only able to provide a usual resident count and no demographic information, while others could supply some demographic information but not complete data (when these partial responses were included, the return rate increased to 92% for universities, 60% for the 10 largest private providers and 81% for the remaining private providers (boarding schools remained at 92%)).

Accuracy

The ONS found discrepancies, on occasion, between an establishment's usual resident count and the age and sex count of that establishment. In instances where it was not possible to resolve this discrepancy through direct contact with the provider, the figure that was most aligned to the census communal establishment form (the Census 2021 questionnaire that CE managers were asked to complete to count those living at the establishment on Census Day) was often chosen.

Some establishments informed the ONS that they did not capture non-term-time address for students and therefore in response to the international student question, had used nationality as a proxy for this variable.

ONS analysis also revealed that the number of census responses sometimes exceeded the survey figure. As the administrative data had a lower count than the census in these cases, no new individuals were added to the census counts in estimation.

Finally, in a small number of cases:

- the figure stated in the data exceeded the capacity of the establishment
- the information collected did not conform to the census definition of usual residence

In the latter two situations, the ONS removed the administrative data from large CE estimation to prevent potential overestimation in the final census results.

Timeliness

The providers were contacted between May and October 2021. The ONS requested data on students with a contract to live at a student accommodation establishment on Census Day (21 March 2021). The age information was not always as of Census Day as some organisations could not retrospectively go back into their systems and calculate this information. Therefore, some providers supplied age information as of the date the survey was completed.

14 Higher Education Statistics Agency student data

The Higher Education Statistics Agency (HESA) is a not-for-profit private limited company owned by Universities UK and GuildHE. As the current designated data body for England, HESA has a statutory role in England under the Higher Education and Research Act. HESA also helps Higher Education (HE) providers in Wales, Scotland, and Northern Ireland to fulfil their statutory role to provide information to their funding and regulatory bodies.

HE student data are collected by HESA and cover students enrolled at publicly funded HE providers in the UK in an academic year. It is optional for privately funded universities to submit their data to HESA, consequently, the dataset also includes information from a small number of private providers.

For Census 2021, the ONS used HE student data along with Individualised Learner Record data (see Section 15), to:

- assess the numbers of students aged 18 years and over in full-time education - at local area and national level, as well as by age and sex

- validate the statistics on qualification questions

HE student data were also used as one of the sources used to generate admin-based population estimates, which were used in census quality assurance (see Section 24).

Coverage

HE student data cover students at:

- HE providers in England registered with the Office for Students (OfS) in the Approved (fee cap) or Approved categories
- publicly funded HE providers in Wales, Scotland and Northern Ireland
- Further Education (FE) colleges in Wales (see Student coverage historical changes)

Students who are studying their entire course outside of the UK and those not funded for study by distance learning overseas are excluded at individual level but are collected in the aggregate offshore record. For more information see HESA Student record. There are a small number of distance learning students studying outside the UK who are funded (for example, Crown servants overseas and the Services); these are also included in the HESA student record.

Most students covered by HE student data are aged between 18 and 30, but there are students of ages outside of this range. In the data used to support the census, records referring to the following sub-groups were removed, as so to capture the census population of interest:

- students younger than 18 years old
- those that had ended their studies by Census Day
- those staying overseas and who intended to be out of the UK for a year or more, including students who had started the course in the UK but had gone abroad to continue their course (for example, for an exchange year)
- part-time students (those registered on a course of less than 24 weeks per academic year)

Accuracy

HE student data are considered a good quality data source, however, there are limitations in its use for census as student addresses become less accurate over their years of study as they move between accommodation. It is a legal requirement for HE providers to accurately return their data, and providers carry out their own quality and validation checks. HESA run extensive, additional validation checks on the data. If at any stage the data fail the checks, they get returned to the HE provider to re-submit.

Despite this, the accuracy of the data is also affected by the student updating the HE provider after the initial collection in August and September (see Timeliness Section), creating individual variation on how up to date each record will be.

HE student data on the number of students living in a local authority (LA) will differ from that under the census definition for several reasons. Research carried out by the ONS in 2021 comparing HE student data and other data sources, such as the Personal Demographic Service (see Section 18), has shown that the address information in HE student data becomes less accurate as students move from one year of study to the next. This suggests that the address information may not always be updated after a student's first year. Therefore, HE student data were not used to correct for under-coverage of students living in large communal establishments (CEs) - the Student Occupancy Survey data discussed in Section 13 were used for that purpose instead. The timeliness could also result in HE student data providing different counts to Census 2021, with the reference period for Census Day being seven months after the initial collection, during which students could have moved and not told the HE provider.

Due to the coronavirus (COVID-19) pandemic, special guidance was issued by HESA on how HE providers should collect the required student data. For the 2019 to 2020 academic year, HE providers were advised the term-time postcode should be defined as where the student was for most of their time during teaching periods, although it should not be changed if the change was a direct result of the pandemic. For the 2020 to 2021 academic year, the guidance included further instruction that postcode did not need to be returned for remote learners if it was not possible or too burdensome to obtain. This and the wider challenges of the pandemic resulted in a significant increase in the number of students without a term-time postcode recorded (the number of "unknowns" doubled for the 2020 to 2021 academic year, compared with previous years).

The challenges of the pandemic also increased the likelihood that a student's term-time postcode reflected where they intended to have been, rather than where they were. This is particularly pertinent for international students, whereby the HE student data might in certain instances record a student as residing at a postcode in England and Wales, when they were in fact living abroad remote learning, given their circumstances had changed due to the pandemic. This caused discrepancies between the number of students derived from HE student data and census estimates of the student population in some areas (see Quality and methodology information (QMI) for Census 2021).

Timeliness

The HE student data extract used to validate census statistics on the qualification questions, included all the academic years from 2010 to 2011 until 2020 to 2021, since

the aim was to look at qualifications gained since the 2011 Census. The extract used to quality assure the numbers of students aged 18 years and over in full-time education, covered the academic year 2020 to 2021, because the aim was to understand where students were on Census Day. The data do not contain a variable indicating when a record was last updated, meaning it cannot be determined whether any records were updated after Census Day. Also, the bulk of data collection is largely done at the start of the academic year, in the first few weeks of September. After this point updates are made through the accommodation office or the individual student choosing to update the HE provider of a change.

15 Individualised Learner Record

The Individualised Learner Record (ILR) is an ongoing collection of data about learners and the learning undertaken by them from learning providers in the Further Education (FE) and Skills sector in England. This includes FE colleges, sixth form colleges, training organisations, local authorities, academies, and voluntary and community organisations. For more details see the ILR data submission guidance.

The data are collected by the Education and Skills Funding Agency (ESFA), an arm's length body of the Department for Education, which is responsible for:

- allocating education and skills funding across the education and training sector
- providing assurance that public funds are properly spent, achieving value for money for the taxpayer and delivering the policies and priorities set by the Secretary of State
- providing, where necessary, financial support for learning providers

ILR data were used, alongside the Welsh Government's Lifelong Learning Wales Record (LLWR) and HESA data (see Section 14), to validate Census 2021 estimates of the number of full-time students aged 18 and over in England and Wales. The aim of using ILR and LLWR data was to capture students in further education. The ILR data were also used to validate the statistics on qualification questions, again in conjunction with the HESA data.

Coverage

Learning providers must submit ILR data if they receive funding through one or more of the following funding models:

- 16-19 (excluding apprenticeships)
- adult skills

- other 16-19 and Other Adult skills funding
- apprenticeships
- community Learning
- skills Bootcamps
- European Social Funding (ESF)

The ILR also covers learning that is not funded by the ESFA, as follows:

- all providers must return an ILR for Office for Students (OfS) funded learners
- FE colleges must send data for all learners, including those that are not funded by the ESFA, and those where the full cost of the learning is met by the learner or their employer
- apprenticeships not funded by the ESFA will appear on the ILR if the training is delivered within the terms of an ESFA contract, or on a voluntary basis in other cases

The ILR data also include all learners and learning aims financed by an [Advanced Learner Loan](). Some learning aims financed by an Advanced Learner Loan may require Higher Education data to be returned on the ILR.

Students aged 16 and 17 years old were filtered out by the ONS as only students aged 18 and over were in scope for the Census 2021 quality assurance using ILR data.

Learners are removed from the system if they withdraw without completing one episode of learning. ILR data may include learners that have not completed an episode of learning in a calendar year but have done so previously and their record is flagged as "withdrawn" in the data rather than deleted.

Accuracy

The ILR data are expected to be very accurate since they must be submitted in accordance with ESFA [validation rules and schema](). Providers will not receive funding for learners who have invalid or incomplete entries. ILR files are validated at the point of transmission to ESFA's data submission portal or through the Funding Information Service, with incomplete returns being returned to the provider. Providers are expected to make monthly returns and thus can update their entries up until the end of an academic year, which runs from 1 August to 31 July. For more details see [ILR standard file specifications and reference data](). Therefore, errors can be identified and fixed ahead of delivery to the ONS. However, previous years' files cannot be amended, so any errors identified after the year has concluded remain incorrect.

One limitation of the ILR data for use in Census 2021 was that the data did not distinguish between part-time and full-time students. Therefore, when the ONS used the data to assess the census numbers of full-time students aged 18 and over, the ILR-based numbers of full-time students were derived using the proportions of full-time students in the LLWR instead.

Timeliness

For every academic year, which runs from 1 August to 31 July, the ESFA publishes 14 different snapshots of the ILR, as described in the ILR standard file specifications and reference data. The extract used to validate Census 2021 estimates of the number of full-time students aged 18 and over was snapshot 14 of the 2020 to 2021 ILR (PDF, 296KB), which had a reference date of 31 July 2021. The extract used to validate the statistics on qualification questions covered all the academic years from 2010 to 2011 until 2020 to 2021; however, for the academic year 2020 to 2021, only learners who had completed their studies prior to Census Day (21 March 2021) were included in the analysis.

16 English School Census

The English School Census (ESC) is carried out by the Department for Education (DfE) and collects information on all pupils attending state funded schools in England. This includes local authority-maintained schools (where funding goes to schools through the local authority) and those where funding goes directly to schools (for example, academies and free schools). This includes nursery, primary, secondary, and special educational needs schools.

The submission of school census data is a statutory requirement of schools under Section 537A of the Education Act 1996. The ESC data are collected three times a year (in January, May, and October), to ensure appropriate school funding and to evaluate, or assess changes in, education policy.

For Census 2021, ESC data were used at an aggregate level to assess census estimates on the number of children of school age, that is, total number of children aged 5 to 15 years old, from national level down to local authority level, and by age and sex. For this purpose, ESC data were used both as a direct comparator source in the census quality assurance, and indirectly as one of the sources used to generate the admin-based population estimates (see Section 24). ESC data were also used in household adjustments, to provide an indication of whether an address (using a Unique Property Reference Number) had signs of activity on Census Day and was likely to be occupied.

Finally, ESC data were used to supplement the Census 2021 Address Frame (see Section 2), to aid with identification of boarding schools.

Coverage

In the academic year 2020 to 2021, the ESC covered 22,047 state-funded schools and 8,342,521 children, including all pupils enrolled in an English state school from the start of nursery to end of secondary. For more information see Schools, pupils and their characteristics, Academic Year 2020/21. The ESC has high coverage of the general population of children aged 5 to 15 years, when education is compulsory, which aligns with the need to quality assure census estimates for these ages. The coverage is lower for children younger than 5 years old because fewer children attend early years education, while some children older than 15 are also included due to the provision of special educational needs. However, pupils outside the 5 to 15 years old range were excluded from the data used to support the census.

The ESC excludes children attending private schools (also known as 'independent schools') and home-schooled children. Ministry of Defence schools based overseas are encouraged to participate in the ESC on a voluntary basis, whilst registered independent schools and general hospital schools instead complete the school-level annual school census (SLASC). Unique circumstances such as permanent exclusion may also mean some pupils are not covered. While these exclusions did not affect the overall high coverage of ESC data for the census population of interest (children aged 5 to 15 years), in cases where the census estimates were found to be noticeably higher than ESC the estimates were also assessed against aggregate data combining ESC and SLASC. This was to account for children attending independent schools as well.

Accuracy

The accuracy of ESC data is high. The headline statistics on Schools, pupils and their characteristics produced by the DfE using ESC data are designated as National Statistics, which means they are compliant with the Code of Practice for Statistics.

The ESC data are collected under a statutory requirement for all schools that need to submit them and are used by DfE to finalise funding allocations, thus there is a vested interest from schools to ensure the accuracy of the data. Data are collected throughout the year via the school management information system and submitted to DfE via their own system called COLLECT. The latter runs automatic validation checks on the data, including checks for unexpected characters, out of range values and missing data, and guidance is provided on how to investigate and correct errors. A summary report with comparisons to the previous year is approved by the head teacher and, for local

authority-maintained schools, sense checked by the local authority, who are responsible for resolving duplications. For more information about validation carried out on ESC data see Check your data - Complete the school census.

Timeliness

The extract used to support Census 2021 quality assurance is the 2020 to 2021 ESC collected on Thursday 21 January 2021. Thus, there was a lag of two months between the ESC reference period and Census Day of 21 March 2021. Due to the nature of the quality checks undertaken by schools and local authorities, the reference period is firmly set and schools cannot submit out of date data. This means that the location of pupils on Census Day (as reported in ESC data) may differ from Census 2021 if they changed school or dropped out of school between January and March 2021.

17 Welsh School Census

The Welsh School Census (WSC), or Pupil Level Annual School Census (PLASC), is an electronic collection of pupil and school level data provided by all state funded, or local authority maintained, schools in Wales. The data are collected every year, usually in January. However, due to the coronavirus (COVID-19) pandemic and the resulting school closures between December 2020 and March 2021, the WSC was delayed until 20 April 2021. For more information about this see Schools' census results: April 2021.

Similar to the ESC (see Section 16), completion of the WSC is a statutory requirement on all state funded schools in Wales. The data are used for resource allocation in the Welsh Local Government Finance Settlement and the Pupil Development Grant, as well as to inform education policy and aid research in educational attainment. The data include information on schools and pupils' characteristics, such as ethnicity, free school meal eligibility and class sizes.

For Census 2021, WSC data were used to assess census estimates for children of school age (those aged 5 to 15 years), from national level down to local authority level, and by age and sex. The data were used both as a direct comparator source in the census quality assurance, and indirectly through their use in the admin-based population estimates (see Section 24).

Coverage

The WSC 2020 to 2021 data included 1,473 local authority-maintained schools and 474,724 pupils – see Schools' census results: April 2021. This includes all Welsh pupils

attending a state school from the start of nursery to end of secondary school. Thus, there is high coverage of the population of children aged 5 to 15 years, which aligns with the need to use the data to quality assure census estimates for these ages. Pupils excluded at the time of the WSC will not appear on that academic year's collection, although those that are only temporarily excluded will appear in the following year's collection.

Similar to ESC, the WSC excludes children attending independent schools, namely private schools that may also be financed by individuals, companies, or charitable organisations. Thus, if there were a considerable number of privately funded schools in a specific local authority (LA), the Census 2021 numbers of children by age and sex in that LA would appear higher than the WSC. In these cases, the census estimates were assessed against aggregates from WSC and Welsh Independent Schools (WIS) data.

Children educated outside of school will not be captured in the WSC and the WIS. Data from the Educated Other Than at School Census (EOTAS) show that there were 1,792 pupils (of any age) educated outside of school in April 2021. For more information about pupils receiving education outside of school in Wales see Pupils educated other than at school.

Accuracy

Similar to ESC, the accuracy of WSC data is high. The data form the basis for an annual statistical release, Schools' census results: April 2021, following validation as part of the Local Government Finance settlement, and are designated as National Statistics.

Data are submitted to the Welsh Government on the date of the school census, having been collected throughout the year via the school's management information system. Data can be updated for between one to two months after submission although this is mainly done to correct errors. After generating the PLASC return, schools will receive reports of errors and queries – the former being unexpected characters or values or logical inconsistency, and the latter being unusual features of the data or particular high levels of omission. These must be resolved before the head teacher signs off on the completed return. The Welsh Government's statistical service also carries out checks on duplicates.

Timeliness

As stated previously, the 2020 to 2021 WSC was conducted on 20 April 2021, and this extract was used to support Census 2021 quality assurance. Thus, if a pupil had changed school between Census Day (21 March) and 20 April there may be a

discrepancy between the census address and the WSC address on Census Day, which would affect aggregate comparisons between the two sources.

18 Personal Demographic Service stocks

The Personal Demographic Service (PDS) is the master demographics database for the NHS in England, Wales, and the Isle of Man. The PDS is used by NHS organisations and enables a patient to be readily identified by a healthcare professional to obtain their correct medical details quickly and accurately.

The PDS data were used to derive total population estimates by single year of age and sex at all geography levels. These aggregates are a central comparator for the validation of local authority (LA), regional and national census population estimates. It has been used as a comparator data source alongside mid-year estimates and the admin-based population estimates (see Section 24). It was also used, along with MoD Armed Forces personnel data (see Section 9), to quality assure the armed forces special population estimates, and to estimate and adjust the populations in care homes with and without nursing.

Coverage

The master database contains approximately 80 million patient records, some of which will relate to patients who no longer use the healthcare system due to death or emigration but may not have been removed. Records are created for newborns or when a patient contacts an NHS service, primarily by registering with a GP practice, but also through accessing A&E or attending hospital.

In general, the PDS has been found to have high coverage of the total population at the national level (England and Wales), with only limited evidence of under-coverage for those people who have private health care and new migrants. The PDS is dependent on members of the public interacting with the NHS and updating their address. Therefore, there is a typical trend for the PDS to show some over-coverage, particularly for more urban LAs where there is a higher population churn. For example, people visiting an LA may temporarily interact with an NHS service and then move out of the area but remain on the PDS (for example, short-term migrants). Similarly, people may not notify their GP when they move until they need to see a doctor again. At local area level, this can also cause over-coverage in one area and under-coverage in another area.

Accuracy

The PDS tends to be more accurate for younger ages (school children and particularly those aged under one year) and older ages (those aged 55 years and over), whereas student ages and young men tend to show more divergence from the census. As mentioned above, this is likely due to the amount of interaction these groups have with NHS services, as it is known that students and younger men are typically less likely to interact with healthcare services or update their details, such as failing to update their address when moving out of the parental home.

With regards to the accuracy of the PDS for specific regions across England and Wales, it tends to vary across LAs. For more rural areas it is generally a better source where typically fewer people aged 18 to 40 years are expected to live, whilst for the more urban areas, and in particular London, it is less accurate. This again, is likely due to the high population churn.

The PDS was found to be an effective source for estimating the populations residing in large care home communal establishments (CEs). A few instances were noticed where PDS either had too many individuals registered at an establishment based on its capacity or too few when compared to information captured in Capacity Tracker (see Section 12) or the CE Manager form. When these situations were encountered, the data were not used to estimate the population residing in the establishment and a different estimation strategy was used.

The coronavirus (COVID-19) pandemic appears to have affected interactions with the PDS system. There was an observed drop in GP registrations and address changes from April 2020 (the start of the pandemic) and an increase in the first half of 2021, rising above pre-pandemic levels. This may be due to a catching up effect of people returning to the GP after lockdown, updating their details because of the vaccination programme, but also people moving house and changing their address as the property market reopened (in May 2020 in England and in June 2020 in Wales).

Timeliness

The ONS were supplied an extract of the PDS referenced on the 21 April 2021. This extract, one month after Census Day, allowed for lags in updating the system. This was the best option available, however, there is always a concern with time relevance in that people may not have updated their details with the NHS in a timely way (for example after a move).

19 Registration and Population Interaction Database

The Registration and Population Interaction Database (RAPID) is created by the Department for Work and Pensions (DWP) to provide a single coherent view of citizens' interactions across the breadth of systems in the DWP, HM Revenue and Customs (HMRC) and local authorities (LAs) via Housing Benefit. RAPID covers everyone with a National Insurance number (NINo) and for each person, the interactions within these systems are summarised in each tax year from the one that ended in 2011 to the most recent available (year ending 5 April 2021). RAPID is based on various DWP benefit systems, supplemented by data extracts from HMRC systems. For further information about RAPID, including its coverage and timeliness, refer to Methods for measuring international migration using RAPID and Children in low income families: local area statistics.

RAPID was used to derive aggregate counts of the population interacting with the tax and benefits system during the tax year that Census Day fell within, with breakdowns by UK nationals and foreign nationals. These data supported both the quality assurance and the assessment of whether a census Residual bias adjustment (PDF, 613KB) was required.

Coverage

RAPID provides excellent coverage of the population of England and Wales, given that the majority of people interact with the tax or benefits systems, including children who are included in a child benefit claim. However, it does include people who are no longer part of the resident population for England and Wales, either because they have moved in the UK outside of England and Wales, or because they have left the UK and have not informed the DWP or HMRC. Rules were applied in the construction of the aggregates for census to try and account for the over-coverage based on whether people had interacted with DWP and HMRC in the tax year ending 2021 (see Accuracy Section).

The application of the activity rules also results in under-coverage in some areas, whereby people are resident in England and Wales, but have not interacted with DWP, HMRC, and LA Housing Benefit, so would not be counted. Under-coverage affected the following groups:

- students may be excluded if they are not working or claiming benefits. This is possible for both UK and international students and is more likely among the latter

- older working age men and women, who have stopped working but not yet reached retirement age and are not in receipt of an occupational pension. This occurs more often for females than males

- non-UK nationals may become inactive in the dataset whilst still living in the UK, or people who have not yet registered for a National Insurance number

- children will not be active in the dataset if child benefit has not been claimed

Accuracy

The standard measure of residency derived by DWP in RAPID is defined differently to that of census. In RAPID, if someone has a registered address in the UK and has at least one week of activity, they will be considered a resident. This means that the dataset may include people who would not be considered usually resident by the census. The ONS produced outputs from RAPID that had two alternative definitions: one required activity in the month before the census reference date, the other required activity in every month in the tax year.

Timeliness

The RAPID dataset used to support Census 2021 contained activity up to the end of the 2020 to 2021 tax year (5 April 2021), which covered Census Day. However, there are always time relevance limitations as people may not have updated their address details with the DWP or HMRC in a timely way (for example after a move). There will be a bias around working age males in the 2020 to 2021 tax year data, because this group are more likely to be self-employed and their tax returns were not available at the time the census quality assurance work took place. For more information see Trends in self-employment in the UK. This means that for the RAPID data the ONS used to complete the census quality assurance work, those who had activities on DWP or HMRC systems because of their Self-Assessment (SA) returns, were only given a proxy status on their record if they had a SA record from the previous tax year. RAPID will not record people as resident who are newly self-employed until the SA record is updated, which is usually the year after starting self-employment due to the SA submission timetable.

20 Births and deaths data

Birth and death registrations are routinely published by the ONS as part of Births, Deaths and Marriages statistics. The two sources were used as comparator data in Census 2021 quality assurance and are described below.

20.1 Births data

The ONS produces birth statistics in England and Wales from weekly received birth registrations, a service carried out by the Local Registration Service in partnership with the General Register Office (GRO), as well as birth notifications from NHS Digital. Following a birth notification, which is generally done soon after a baby is born by a midwife or doctor, the parents have a legal duty to register the birth within 42 days at the local register office (over 95% of births are usually registered in that timeframe). However, this was affected by the coronavirus (COVID-19) pandemic where birth registration services were suspended in March 2020, before being reinstated in June 2020, where it was safe to do so. This had an impact in 2020 and 2021, where a significant number of registrations came in after 42 days – 42% in 2020 and 26% in 2021 (see sections below on how the delays were accounted for).

At national level, birth registrations were primarily used for calculating age-specific fertility rates, sex ratios and to demonstrate areas which have missing babies and women of fertile ages. Birth notification data were also used to check whether babies had been missed on the census return. Babies aged under one who appeared in the Personal Demographic Service (PDS) data (see Section 18) were also used in quality assurance of Census 2021. This provided an additional insight into the number of births occurring, particularly in local authorities (LAs) where the effect of registration delays was greater due to the coronavirus pandemic.

Coverage

The birth registration extract used for census covered births occurring in England and Wales from April 2020 to near Census Day. Usually, an extract would be taken six weeks after the end of the reference period to allow for the registration period to elapse. However, registration delays due to the coronavirus pandemic meant birth registration extracts were taken later (see Timeliness Section). In contrast, the birth notifications data were not subject to the same delays. Therefore, the two sources were used together to ensure adequate coverage to support the census quality assurance and the checks on whether babies had been missed on the census returns.

Accuracy

Before the pandemic, the completeness of birth registrations was higher compared to birth notifications. However, the delay to birth registrations in the pandemic meant the presence of a notification and a mother of childbearing age was used as evidence of a birth for the purposes of calculating rates and assessing census counts.

It is a legal requirement to register a birth under the Births and Deaths Registration Act 1836 meaning that every birth registered in England and Wales should be included in the data. In addition, as part of the birth registration process, validation checks are conducted to help the registrar ensure all data are accurate. A number of checks are also carried out on records once the ONS receive the data to ensure they are valid and any inconsistent records are queried with the GRO. For more information on the accuracy of the births data see Births Quality and Methodology Information.

Timeliness

The birth registration extract used represents births that occurred in England and Wales between April 2020 and near to Census Day (21 March 2021). The extract for this was taken in August 2021 to allow extra time for the delay to birth registrations to catch up. The birth notifications arrive daily at the ONS and were therefore a timely resource to use for LA assessment. The combination of using birth registrations with notifications and the PDS extract meant that interruption to the census assessment was minor.

20.2 Death registrations

The registration of deaths occurring in England and Wales is a service carried out by the Local Registration Service in partnership with the General Register Office (GRO). A death must be recorded within five days of the occurrence, except for stillbirths (which can be registered at any time). This information is submitted onto the Registration Online (RON) system (implemented in 2009), which checks that the information provided has been entered accurately. The data are collected under the Births and Deaths Registration Act 1836, which made it a legal requirement for all deaths (and births) to be registered from 1 July 1837.

The ONS used deaths data and census estimates by local authority to derive age-specific mortality rates which helped identify potential inaccuracies with census estimates in the underlying population. In addition, estimates of the very old (which are also derived from death registration data) were used at the national level as a comparator of high ages in the census.

Coverage

The death registrations source achieves almost complete coverage of deaths occurring in England and Wales to UK residents, making this a highly valuable source for use in census. Deaths which occur outside the UK to residents of England and Wales are not included in the death registrations; these may include those taken ill in another country,

those serving overseas in armed forces, or travel accidents. For information on the quality of death registrations data see Mortality statistics in England and Wales QMI

Accuracy

The accuracy of the death registrations source is considered high. In a small number of cases death registrations can be delayed, and in some cases can take years to register (see Impact of registration delays on mortality statistics in England and Wales). This delay occurs when deaths are referred to a coroner to ascertain the cause of death and the death is only registered once the cause is confirmed. Therefore, deaths occurring in the census reference period may be missing, having not yet been registered.

Timeliness

The registration of deaths was prioritised during the coronavirus (COVID-19) pandemic and therefore the number of deaths used for the census were not affected by delays in the same way birth registrations were (see Section 20.1 on Timeliness). The ONS used the 2020 deaths extract with the addition of weekly death registrations into 2021 to help understand the population as of Census Day.

21 Electoral Register

The Electoral Register (ER), sometimes called the 'electoral roll', lists the names, and addresses of everyone who is registered to vote in the UK. The local Electoral Registration Offices (EROs) are responsible for maintaining the register for those living in England, Scotland, and Wales. The Electoral Office for Northern Ireland is the responsible office for those living in Northern Ireland.

It is mandatory to register when asked to do so and when meeting the conditions for registering. The main use of the register is to enable people to vote in elections or referendums. It is also used for the purpose of prevention and detection of crime, and for identity checking when applying for credit. To support Census 2021, ER data were used to quality assure the census population estimates by local authority (LA) in England and Wales. However, it was only used as an extra comparator source due to the exclusion of important sub-groups, such as children. In addition, sex is not collected on the ER and therefore it could not give an age-sex breakdown to use as a census comparator.

Coverage

People must be aged 16 years and over to register to vote (or aged 14 years and over in Scotland and Wales) [9]. They must also be either a British citizen, or an Irish or EU citizen living in the UK, or a Commonwealth citizen who has permission to enter or stay in the UK, or one who does not need permission.

The vast majority of those on the ER will be registered to vote for both parliamentary and local government elections. However, British citizens living abroad who have been registered to vote in the UK in the last 15 years can vote in UK Parliament elections but not in local government elections, while EU citizens and peers in the House of Lords can vote in local elections but not in UK Parliament elections. In Scotland and Wales, people as young as 14 can be included on local government registers (as attainers) but not on the parliamentary register.

The copies of the registers used to support Census 2021 contain all those registered in an area for any type of election, except from those [registered to vote anonymously](#) and British citizens living abroad. However, the usefulness of ER data is constrained by the eligibility criteria on age and nationality as the data exclude children under registration age and some foreign citizens without UK citizenship. Many EU citizens will be included because they can vote in local elections, whilst in Wales and Scotland most foreign nationals can now be registered to vote in some elections. Some people are legally excluded from voting in all types of elections and so will not appear in the ER, such as prisoners serving a prison sentence and those who have been convicted of committing certain electoral crimes.

Furthermore, there will be some under-coverage caused by people who are eligible to vote but do not register. Equally there may be over-coverage due to people moving home, being registered in more than one LA (students can register to vote both at their term-time address and their home address) or leaving the country and their record only being removed at the time of the next annual canvass. However, individuals can register to vote even if they do not have a permanent address, so the dataset can include those that are homeless or living in a communal establishment.

---

[9] In England, the entitlement to register to vote does not relate to being aged 16 years and over, but it is for someone who turns 18 during the lifetime of the annual register. Therefore, it is mainly 17-year-olds and some 16-year-olds that register as "attainers". In Scotland and Wales, where people aged 16 and over can vote in some elections, the entitlement to register to vote relates to being aged 14 years old or over, therefore, 15-year-olds and some 14-year-olds will register as attainers.

Accuracy

The ONS uses the data supplied by the local Electoral Registration Officers (EROs) to produce Electoral statistics for the UK. These represent the most accurate count possible of the number of people on electoral registers each year. However, electoral registrations are subject to error, mostly due to entries for individuals who no longer live at an address, as detailed in Electoral Commission's Accuracy in Great Britain.

People normally only need to register once and not for every election. However, they do need to register again if they change name, address, or nationality. From July each year EROs contact all the households in their areas to check if the details on the register are correct, called the annual canvass. Contact is made by post, email, phone or by someone visiting the property. The register is then continuously updated through a 'rolling registration' system of monthly revisions, where electors are added and removed. The registration form includes a request to provide National Insurance number and date of birth. This information is used by the EROs to verify against other administrative data and ensure the registration is valid. Checks against other local and national sources are carried out at various points throughout the year.

Date of birth has been a mandatory field on the registration form since 2014, so there may be some missingness in the date of birth information caused by those who have not needed to update their ER record since then. The ONS received date of birth information only for those electors who are either aged 16 or 17 years and therefore approaching voting age. This further limited the use of the ER to support census.

Timeliness

The ER extract used to support the census refers to 1 December 2020, four months before Census Day of 21 March 2021. Anyone who has registered to vote, de-registered due to death or living aboard or changed their personal details during this time may therefore have been missed, incorrectly included, or not captured at the address where they were on Census Day.

22 Energy Performance Certificate

The Energy Performance of Buildings Register holds all Energy Performance Certificates (EPCs) for England and Wales. EPCs are valid for 10 years and published by the Department for Levelling Up, Housing and Communities (DLUHC) on their website. EPCs indicate the energy efficiency of a building to prospective tenants or buyers, with the intention to improve it and consequently to tackle climate change. EPC

data were used as part of Census 2021 quality assurance, to validate the statistics on central heating type (for example, electric or gas) and accommodation type (for example, house, or flat) questions.

Coverage

Each record in EPC data refers to a single EPC when this has been issued for a dwelling, therefore the data may include multiple EPCs per dwelling if more than one has been issued over time. There were around 21 million EPC records up to 31 March 2021, for 16 million unique properties.

Since their introduction in August 2007, EPCs have been required by law when a property is constructed, offered for sale, or let, as detailed in EPC Action Plan (PDF, 253KB). Therefore, the register does not hold data for every domestic dwelling in England and Wales since if a property was constructed prior to the introduction date of EPCs, and not offered for sale or rented since then, the property may not appear in the data. For the same reason, EPC data tend to have a higher coverage of rental as opposed to owner-occupied properties due to more frequent moves between rental properties. Also, older dwellings are less likely to be covered in comparison with newly built dwellings. It was important to consider these inherent biases in EPC data when making comparisons with census data, as some types of properties are less likely to be included.

In line with Census 2021, vacant buildings suitable for demolition are exempt from providing an EPC and therefore not included in the data.

Accuracy

Primary responsibility for the accuracy of the data in relation to individual buildings lies with the energy assessor who carried out the energy assessment. For more information on the validation rules that the register applies to assessor input and further quality assurance work see the EPC Action Plan (PDF, 253KB).

EPC data were linked to census addresses via the Unique Property Reference Number (UPRN) that uniquely identifies an address. Where there were multiple EPC records associated with the same address, the most recent record was retained. However, EPC data only provide a snapshot of the energy efficiency features of a dwelling at the point of the assessment, usually prior to a sale or rental if the property didn't already have a valid EPC. There is no obligation for homeowners who are trying to sell, or landlords who are trying to let, to update the EPC if they make improvements to the energy efficiency of the property and a pre-existing certificate is within date. Consequently,

EPCs may not necessarily capture the most up-to-date features of a dwelling if changes have been undertaken to the property, which may affect the accuracy of EPC data in relation to the type of central heating.

Timeliness

The Energy Performance of Buildings Register published by DLUHC on the Open Communities website is updated quarterly. For quality assurance of Census 2021, all EPC records lodged up to the end of March 2021 were used. Some records will be timelier than others depending on when an EPC was issued and whether any changes have been made to the property since March 2021. For example, if the latest EPC for a property is dated 2007, this will still be used (even if the property has undergone changes since) as it is the best information available for that property.

23 Ministry of Defence Joint Personnel Administration System

Census 2021 collected, for the first time, data on veterans of the UK Armed Forces. The data will be used by service providers such as central government departments, local authorities and third-sector organisations to support veterans in line with the Armed Forces Covenant.

The veterans' question in Census 2021 was designed to collect information on veterans who had previously served in the UK Armed Forces, but who were not currently a serving member. However, in certain instances, it is possible that currently serving personnel may have recorded that they were a veteran on the census questionnaire. For example, if they were a veteran of a particular war or campaign but still currently serving or if they had previously left the armed forces but had re-joined and were currently serving (see Armed forces community (veterans) question development for Census 2021).

In order to ensure the accuracy of the census data, the Ministry of Defence worked with the ONS to provide data to correct for such instances where serving personnel had identified themselves as veterans.

Coverage

The MoD data used to quality assure and correct the census responses covered serving armed forces personnel including those in the Royal Navy, Royal Marines, Army, and Royal Air Force as well as both regulars and reservists. A very small number of

reservists might be missing from the data that were used; however, this negligible amount would not affect the ability of census to identify veterans. For more information see MoD Background Quality Report for UK Armed Forces Personnel Statistics (PDF, 349KB).

Accuracy

The MoD data used is extremely accurate and is taken from the MoD's Joint Personnel Administration (JPA) System. There were only a small number of instances where a census record could not be linked to the MoD data, generally due to insufficient geographic information.

Timeliness

The data referenced 1 April 2021; the reason for the delay from Census Day was to ensure the data extract fell within the MoD validation schedule and therefore was the most complete and accurate it could be. However, in a very small number of cases late reporting of exits from the armed forces can occur, which may affect whether someone is counted in the MoD data as currently serving, when they were in effect a veteran as per Census Day.

24 Admin-based population estimates

The admin-based population estimates (ABPE) are research-based estimates of the usually resident population in England and Wales, derived from linked administrative data. The ONS produced ABPEs for years 2016-2020 based on two methods, referred to as ABPE v2.0 and ABPE v3.0. Each method uses a different set of rules and data to decide whether to include an individual in the usually resident population. The data used in ABPE v2.0 are from the NHS Patient Register, Customer Information System, Higher Education Statistics Agency and the English and Welsh School Censuses. The data used in ABPE v3.0 included additional sources such as data on benefits and income (for example, Pay As You Earn, Tax Credits, Housing Benefit and Child Benefit), and the NHS Personal Demographic Service.

Estimates were produced for the usually resident population by age and sex, at national, local authority (LA), lower-layer super output area (LSOA) and output area (OA) level. Both versions of ABPE were used in Census 2021 quality assurance to compare against the census estimates. This is the first time a combined administrative dataset has been available to provide this comparator.

Coverage

The ABPE coverage was evaluated against the official mid-year population estimates (MYEs) and 2011 Census by age and sex. The results from the comparison of ABPE with Census 2021 will be published in a subsequent release. Caution should be exercised in interpreting differences from the official MYEs as these accumulate higher uncertainty the further away from the census they are.

Previous research shows that ABPE v2.0 typically over-estimates the population compared with official estimates, particularly for working age males, because people remain present on administrative data sources after they have moved or left the country. In ABPE v3.0, stricter activity-based inclusion rules were applied to target the over-coverage seen in v2.0 and produce an ABPE with under-coverage that would allow dual system estimation to be applied to address the coverage issues. Overall, this approach was successful as research published in 2020 found ABPE v3.0 generally showed lower estimates than the 2011 Census, although they were higher for some populations. This under-coverage is likely to include self-employed people, full-time students who were not in higher education, and people not in work due to looking after family but not receiving any Carer's Allowance.

The differences in coverage between ABPE v3.0 and the official estimates remain stable over time, unlike those for ABPE v2.0, which suggests that activity-based rules are more likely to remove people who have left England and Wales.

Accuracy

The aggregate comparisons made between the ABPEs and official estimates suggest that both ABPE methods produce estimates that broadly align with the official population estimates at the national level. Discrepancies are seen at local area level. LAs which have a high proportion of students, a young and mobile population group, as well as communal establishments (CE).

Differences between the estimates are likely due to a number of factors including discrepancies in the data collection and processing procedures in the underlying data, and definitional differences in the concept of usual residence; for example, presence on a data source or claiming a benefit does not necessarily reflect usual residence. Furthermore, aggregate comparisons give us the net effect of different coverage patterns, they do not tell us whether we are including the right people in the right place.

Comparisons of ABPE v3.0 for mid-2020 with the Census 2021 estimates showed that they aligned well at the England and Wales level and across a range of LAs, although

we observed discrepancies when comparing these estimates for some age and sex groups and local areas. Further analysis of the accuracy of the ABPE will be undertaken by comparison with Census 2021. This will include analysis of a detailed linkage exercise between the ABPE and the census to fully understand the coverage patterns by area, age and sex.

Timeliness

The ABPEs used to support Census 2021 are annual estimates covering the years from 2016 to 2020 and with a reference date of 30 June. This is in line with the reference date of the ONS's annual MYEs. Consequentially, there is a lag of nine months between the last year of available ABPEs and Census Day. The lack of timely interaction with admin data sources can also create time lags affecting the accuracy of estimates, especially in areas with high population churn such as London boroughs. The effect of such lags can also be seen in some LAs with large universities. Those of graduate age may be overestimated in the ABPEs, as when graduates move out of an area, the admin data may still show them as living at their student address. The acquisition of data sources with more regular interactions (such as HMRC PAYE data) along with more frequent updates of data (such as DWP and HMRC benefits data), is likely to improve both the timeliness and accuracy of the ABPEs.

25 Contact

Census 2021 Statistical Design, Census 2021 Data Transformation Division

census.quality.assurance@ons.gov.uk

Room 4200E
Office for National Statistics
Segensworth Road
Titchfield
Fareham
PO15 5RR

26 Acknowledgements

The ONS would like to thank:

- all the data suppliers for allowing us to use their data to support Census 2021 and for providing us with feedback and technical details about their data: Care Quality Commission, Care Inspectorate Wales, Cushman and Wakefield, Department for Education, Department for Levelling Up, Housing and Communities, Department for Work and Pensions, Driver and Vehicle Licensing Agency, Electralink, Electricity Central Online Enquiry Service, Electoral Commission, Her Majesty's Revenue and Customs, Higher Education Statistics Agency, Home Office, Ministry of Defence, Ministry of Justice, NHS Digital, Office of Communications, Statistics for Wales, United States Visiting Forces, Valuation Office Agency, Welsh Government, Xoserve
- all educational establishments who completed surveys on students occupying their accommodation
- all local authorities who provided Council Tax data
- all local Electoral Registration Offices who provided electoral registrations data