Estimating Populations in Large Communal Establishments (CEs)

Greg Tilsar, Clare Farmer. 2021 Census Statistical Design of Processing, April 2021

Main Points

Estimation strategy for each type of communal establishment (CE), which contains at least one large CE in the 2021 Census:

CE Type	Estimation strategy
Halls of Residences	We will use a variety of measures to maximise response from large CE student halls. This includes: • Collecting aggregate age/sex distributions through the Student Hall Survey • Using administrative data supplied by private halls of residence providers • Copying individuals who responded at their out of term address but identified a term-time address
Care Homes	We will use administrative data supplied by the NHS
UK Military bases	We will contact individual base personnel to sense check results from the Census
US Military bases	We will use administrative data provided by the United States Armed Forces
Prisons	We will use administrative data provided by the Ministry of Justice
Approved Premises	We will use administrative data provided by Her Majesty's Prison Service
Boarding Schools	We will implement a variety of measures to increase response in this type of establishment. This includes: • Copying individuals who responded at their out of term address but identified a term-time address • Contacting establishments asking for aggregate age/sex distributions
Children's Homes	We will accept the results from the Census
Religious Establishments	We will contact establishments asking for aggregate age/sex distributions of residents.
Hostels for the Homeless	We will accept the results from the Census
High Security Mental Health Unit	We will accept the results from the Census
Low/ Medium Secure Mental Health Unit	We will accept the results from the Census
Staff Accommodation	We will contact establishments asking for aggregate age/sex distributions of residents.
Immigration Removal Centres	We will use administrative data provided by the Home Office

Education Other	We will use a variety of different strategies to improve response rates in this type of establishment. This includes: • Copying individuals who responded at their out of term address but identified a term-time address • Using administrative data supplied by the Department for Education

Introduction

Large CE's are defined as establishments with 50 or more usual residents. According to the 2021 Census CE/Special Population Group (SPG) address frame, more than a million people could be living in accommodation classified as a large CE in 2021.[1] Whilst this figure will only make up a small percentage of the actual population, living in England and Wales, these populations are likely to be clustered in particular locations and will share certain characteristics.[2] Therefore, it is important that we maximise response from these establishments.

Whilst we hope that everyone in large CEs fills in their Census forms, previous experience from past Censuses has shown that this is not always the case and that we often see under coverage in establishments such as student halls and prisons. Unlike households or small CEs (establishments with less than 50 usual residents), the Census Coverage Survey will not be used to correct this undercount.[3] Instead, a separate estimation strategy will exist for these establishments.

This paper will firstly outline the overall strategy that existed for the estimation of populations living in large CEs in 2011. Then, it will highlight the differences between the strategy pursued in 2011 and the one that will be pursued in 2021. Finally, the paper will outline the estimation strategy that exists for each type of CE, which has at least one establishment with an expected capacity of 50 or more, according to the CE/SPG address frame.

Large CE estimation strategy in 2011

In 2011, large CEs were defined as any establishment with 100+ bed spaces, which provided managed residential accommodation. During the 2011 Census, the Census count for large CEs was checked against two sources:[4]

---

[1] This figure is based on the overall expected capacity number within the first version of the CE/SPG address frame. The expected capacity exists to enable Census planning and does not guarantee that this number of people will be in any given establishment. In 2021, we will get the CE managers' assessment of actual number of residents to measure against.

[2] Ghee, C and Fraser, O 'EAP141 – Design of Address Frame, Collection and Coverage Assessment and Adjustment of Communal Establishments in 2021 Census', 13th January 2021, https://uksa.statisticsauthority.gov.uk/publication/eap141-design-of-address-frame-collection-and-coverage-assessment-and-adjustment-of-communal-establishments-in-2021-census/

[3] Detailed reasoning as to why the Census Coverage Survey will not be used for this purpose can be found in Ghee, C and Fraser, O 'EAP141 – Design of Address Frame, Collection and Coverage Assessment and Adjustment of Communal Establishments in 2021 Census', 13th January 2021,

[4] ONS, 'Estimation and Adjustment for Communal Establishments', September 2012, https://webarchive.nationalarchives.gov.uk/20160108085304/http://www.ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-data/2011-first-release/first-release--quality-assurance-and-methodology-papers/coverage-within-communal-establishments.pdf

1. Questionnaire Tracking system (QT)[5]
2. Administrative data

The QT was used to check the number of responses received against the number of residents at each large CE.[6] Administrative data was then used to quality assure the figures received from certain establishments e.g., Ministry of Justice data was used to quality assure the figures that were received from prisons.[7] Large CEs were considered for estimation, if they met one of the following criteria:

1. The overall return rate for the establishment was less than 75%
2. Actual non-response was 50 or more
3. Administrative data had a count, which was 50 higher than the number of Census returns received.

In 2011, 719 large CEs were considered for estimation and adjustment based on failing one of the three thresholds listed above.[8] Estimation and adjustment was conducted for 69% of these establishments, which equated to roughly 56,800 people being added into large CEs.[9]

Large CE estimation strategy 2021

In 2021, large CEs will be treated differently than they were previously. The key difference between the strategy, which existed in 2011 and the one that exists in 2021 is that the definition of a large CE has changed.

Originally, we changed the definition of a large CE from an establishment with 99+ bed spaces, which provided managed residential accommodation to an establishment with 50+ bed spaces, which provides managed residential accommodation.[10] However, this definition has had to be amended for a couple of reasons:

1. The Census does not collect information on the number of bed spaces that an establishment has.
2. The CE expected capacity number, as identified in the CE/SPG address frame, does not always specify an establishment's number of bed spaces and instead states the number of usual residents expected.

The number of expected usual residents is sometimes used because in certain instances it is more useful for census planning to use this number instead. A key example of this is with hotels. Hotels are likely to have a significant number of beds but very few usual residents due to the nature of these establishments. It is unnecessary to have a form sent to each room at this type of establishment as most of the individuals staying in this type of accommodation will only be visitors and should complete the Census at their place of usual residence instead. Therefore, a default capacity of 1 was given to each hotel.

Consequently, a different metric was needed to separate large and small CEs. It was decided that the number of usual residents recorded on the CE Manager form was the most suitable number to be

---

[5] The QT was based on the number of usual residents collected by the enumerators
[6] ONS, 'Estimation and Adjustment for Communal Establishments', September 2012,
[7] ONS, 'Estimation and Adjustment for Communal Establishments', September 2012,
[8] ONS, 'Estimation and Adjustment for Communal Establishments', September 2012
[9] ONS, 'Estimation and Adjustment for Communal Establishments', September 2012,
[10] Ghee, C and Fraser, O 'EAP141 – Design of Address Frame, Collection and Coverage Assessment and Adjustment of Communal Establishments in 2021 Census', 13th January 2021,

used to divide large and small CEs. Therefore, an establishment was decided to be a large CE if it had 50 or more usual residents, according to the CE Manager form.

Unlike in the 2011 Census, estimation work may be conducted for any large CE, which has a return rate of less than 100%. However, establishments will be prioritised for estimation if they meet certain thresholds.

These are:

(a) The return rate of the establishment being less than 75%
(b) The actual non-response rate of the establishment being 50 or more

The amount of estimation work that will be conducted will be dependent on the time and resources available. However, we are hoping to have the ability to make more adjustments to large CEs than we did in 2011.

The current plan for each type of CE, if an adequate level of response is not received is as follows:

US Military Bases

The current plan is to use administrative data to estimate the populations living in United States (U.S.) Armed Forces bases, if an adequate level of response is not received. We asked the United States Armed Forces to provide data on the individuals who are living on their military bases in England and Wales. This request includes an aggregate age/sex breakdown of all individuals living on each of these bases (including dependents). This data was delivered to us in March 2021.

Prisons[11]

Presently, the strategy is to use administrative data to estimate figures for people living in prisons, if we fail to receive a sufficient number of responses from prisoners. The Ministry of Justice have agreed to supply us with data on the individuals who live in this type of establishment. This data will provide us with information on the prisoner's sentence length, their custody status as well as providing an aggregate age/sex breakdown of all the individuals living in each establishment.[12] This data was sent to us in April 2021.

Approved Premises[13]

The current plan is to use administrative data to estimate the Census counts for these establishments, if we do not receive enough returns. We asked Her Majesty's Prison Service (HMPS) to provide an aggregate age/sex breakdown of all the people residing within this type of accommodation. Furthermore, we also requested information on the amount of time that these individuals have spent in custody. This data was delivered to us in April 2021.

---

[11] Young Offender Institutions are classified under the same definition as prisons and therefore the plan for the estimation of these establishments is the same as prisons as they fall under the same category.
[12] The custody status and sentence length variables have been requested to ensure that any estimation complies with the 2021 Census definition of who should be classified as usual resident at these establishments,
[13] The CE type approved premises also includes bail hostels

UK Military Bases

The plan is to meet with individual base personnel to sense check results from the Census, if we fail to receive a sufficient level of response from these bases. We did look into whether administrative data could be used to correct Census counts. However, this was unable to be conducted for a variety of reasons. Firstly, the Ministry of Defence were unable to supply this data owing to the lack of a centralised database, which contains such information. In addition, research into Personal Demographic Service (PDS) data showed that it was unsuitable for this purpose.[14]

Halls of Residences

The current plan is to use a variety of different strategies to correct our counts for student halls, if an insufficient number of Census returns have been received. We considered whether Higher Education Statistics Agency (HESA) data or PDS data could be used to estimate the populations living in student halls. Our desk-based research found some statistical quality issues with HESA and PDS data, which made it unsuitable for either of these sources to be used for this specific purpose[15]. Instead, a variety of other measures have been put in place to ensure that the correct estimates are achieved for the student population living in large CE student halls.

Firstly, we will be running a post-Census quality survey for student halls between May and June 2021. Within this survey, we will contact the providers of halls of residences, which have failed to provide a sufficient number of Census returns and will ask them to provide an aggregate age/sex distribution for all the individuals who had a valid tenancy agreement, at a student hall, on 21st March 2021. If the provider is unable to provide such information, then we will ask broader questions about the age and sex of the individuals residing in this type of establishment.

Furthermore, we are also planning to use administrative data sent by some private hall of residence providers. For example, Unite Students provide accommodation for nearly 70,000 students in England and Wales and have agreed to supply us with information on the individuals residing in their properties on Census day. This data will include an age/sex breakdown of everyone who had a valid tenancy agreement at a Unite Students property.

Finally, we have also developed methods to copy students, who responded to the Census only from their out of term-time accommodation address e.g., parental/ guardian home and place them into their term-time residence. We will identify individuals who have stated, on a household form, that they are in full time education and that they have a term-time residence. We will check that this individual has not responded to the Census from their term-time address as well. If the student has not responded at their term-time address, then a record for that student will be created at their term-time address, as this is where the individual should be considered usually resident.[16]

Boarding Schools

We are planning to use a variety of different approaches to maximise response from large CE boarding schools. We did conduct research to see whether administrative datasets such as the English School Census, the Weekly Attendance Record or the School Level Annual School Census could be used for this purpose.

---

[14] See Annex A for further details on the research conducted into PDS
[15] See Annex B for further details on the research conducted into HESA and PDS
[16] The record of the individual at their out of term address would be removed to avoid duplication

However, a key issue with all of these sources is that none of them comply with the Census definition of who should be counted at a boarding school. In the 2021 Census, only individuals who board four or more nights a week, should complete the Census at a boarding school.[17] Whilst most of these administrative datasets are able to provide figures on the number of people who board at a boarding school, none of them can be filtered on individuals who board four or more nights a week. It is for this reason that we have decided that these sources will not be the sole way in which the populations living in large CE boarding schools are estimated. Instead, a variety of different measures have been put in place.

Firstly, in a similar manner to student halls, we have developed methods, which will allow us to copy students, who have responded to the Census only from their out of term accommodation e.g., parental/ guardian home and place them into their term-time address. We will identify individuals who have stated, on a household form, that they are in full time education and that they have a term-time residence. We will then check that this individual has not responded to the Census from their term-time address. If a student has not responded to the Census from their boarding school as well, then a new record will be created for the student at their term-time address, as this is where the individual should be considered usually resident.[18]

We may also contact boarding schools directly asking them for an aggregate age/sex breakdown of all pupils, who board four or more nights a week, if an insufficient level of response is received.

Hostels for the Homeless

The current plan is to conduct no estimation work for individuals living in this type of establishment. Due to the transient nature of this population group, they are very unlikely to appear on administrative datasets. In addition, even if they do appear on administrative datasets, they are unlikely to have accurate address information. We did look into whether the Homeless Case Level Information Collection, which is a dataset compiled by the Ministry for Housing, Communities and Local Government (MHCLG), could be used for the estimation of large CE hostels for the homeless. However, this dataset will not be available on time for the estimation work. Consequently, owing to the lack of any administrative datasets, which would be able to validate our responses from these establishments, the decision was made to not make any alterations to the results received from Census.

Children's Care Homes

Currently, the plan is to not conduct any type of estimation work with this particular CE type. There are a couple of reasons for this with the primary reason being that a 100% response rate was achieved for residential children's homes in the 2011 Census.[19] Therefore, a high-level of response is expected once again in 2021. Moreover, internal research has shown that we do not have access to any administrative datasets, which would be able to validate the collected Census response from these establishments.

---

[17] ONS, 'Boarding School or Children's Home Managers' https://census.gov.uk/help/types-of-household-or-accommodation/communal-establishments/i-manage-an-establishment/boarding-school-or-childrens-home-managers
[18] The record of the individual at their out of term address would be removed to avoid duplication
[19] ONS, 'Estimation and Adjustment for Communal Establishments'

Religious Establishments

Currently, the plan is to contact these establishments if they fail to provide us with enough returns to the 2021 Census. Owing to the fact that a 100% response rate was received for large CE religious establishments in 2011, we believe that a similar result will be achieved in 2021.[20] However, if an establishment does fail to provide a sufficient number of returns, then the plan will be to contact the establishment asking for an aggregate age/sex breakdown of all the usual residents in this accommodation.

Care Homes

The plan is to use administrative data, if an insufficient level of response is received from care homes. Research conducted in 2019 revealed that PDS data is able to provide effective coverage of people residing in care homes within England and Wales. This research uncovered that "PDS 2017 counts of people in care homes [were] of broadly similar magnitude to the 2011 Census" count for numbers of people in care homes.[21] Further research was conducted into PDS 2019 data to understand its effectiveness in terms of coverage of care homes. This research uncovered that over 90% of care homes in England and Wales appear in PDS data and that the number of residents in most establishments is similar to the expected capacity number registered for each establishment in the CE/ SPG address frame. This data has been delivered to the ONS.

Staff Accommodation

We will ask for the aggregate age/sex distribution of all the usual residents in these establishments, if an adequate level of response to the 2021 Census is not achieved. This is because we do not have access to any administrative data sources, which can provide this information.

High Secure Mental Health units

In a similar manner to residential children's homes, no estimation work will be conducted for this type of large CE. This conclusion has been reached for a variety of reasons. Firstly, in the 2011 Census, we received a 100% response rate from these establishments, and this provides us with confidence that we will be able to achieve a similar result in 2021.[22] In addition, we do not have access to any administrative datasets, which would be able to validate the collected data for this type of establishment.

Low/Medium Secure Mental Health units

In a similar manner to high secure mental health units, no estimation work will be conducted for this CE type. This decision has been reached for the same reasons that we have decided that no estimation strategy will exist for high security mental health units. Given the historically high response rates for this CE type and confidence around response rates for 2021, we did not think that the benefits of estimation work for this CE type outweighed the additional burden that the request for age/sex data would have put on the managers, who had already worked hard to provide/enable responses for the CE residents.

---

[20] ONS, 'Estimation and Adjustment for Communal Establishments'
[21] ONS 'Phase 2 Communal Establishment Sprint Final write-up', 2019. See Annex C for further detail
[22] ONS, 'Estimation and Adjustment for Communal Establishments'

Immigration Removal Centres

Our plan is to use administrative data to estimate the populations residing in immigration removal centres (IRC), if we fail to obtain enough Census returns. The Home Office have agreed to provide us with basic demographic information on the individuals residing in this type of establishment. This includes their age, sex and nationality as well as the centre in which people are currently living in. This data was delivered to the ONS in May 2021.

Education Other

Currently, the plan is to use a variety of different measures to ensure that the correct number of individuals are stated to be living in establishments categorised as education other.

Owing to the third lockdown in England, a number of students may not have been in their term-time accommodation on Census day and may have only filled in the Census at their out of term-time accommodation e.g., parental/ guardian home. In these instances, where an individual has stated, on a household form, that they are in full-time education and that they have a term-time residence, then we will create a record for this individual at their term-time address as this is where they should be considered usually resident.[23]

We may also use administrative data to try to correct these counts. Individualised Leaner Record (ILR) data provides information on individuals living in this type of accommodation including their demographic characteristics such as their age and sex. The latest version of this dataset was collected in January 2021 and was delivered to us in May 2021.[24]

---

[23] The record of the individual at their out of term address would be removed to avoid duplication

[24] The ONS are currently not pursuing the Lifelong Learner Wales Record, which is the Welsh equivalent of this dataset for large CE estimation. This is because there are currently no education other establishments located in Wales with an expected capacity of 50 or more on the CE Frame

Annex A

This research involved comparing the number of people registered as living at a UK armed forces base in PDS 2020 data against the CE expected capacity number, which exists for every UK armed forces base on the CE/SPG Address Frame.[25] If the two figures were similar, it could have indicated that PDS may be able to accurately state the number of individuals living at this type of accommodation. The results of this analysis were not promising, and they uncovered that the majority of armed forces personnel in PDS data had an address registered as a medical practice rather than an actual UK armed forces base.

Annex B

Firstly, the current timing of HESA data extracts means that this data will not be available for the estimation work. HESA figures are collected at the end of the academic year and consequently, even if only raw data were available, it would be received too late to be used in the estimation work being conducted for large CE student halls.

In addition, we need to better understand the statistical quality of the addressing information supplied by students to HESA. Whilst we have confidence in the ability of HESA data to accurately state the location of first year students, our research suggests that students are less likely to update their term-time postcode after first year.[26] Consequently, in HESA data, students appear to still be living at halls of residences, when, in all likelihood, they have probably moved address.

We conducted research into PDS 2019 data to see where students were recorded as living in this administrative dataset compared to other sources. If PDS had them registered as living at a different postcode, which was in the same geographical area as the hall of residence they were registered as living in, according to HESA data, this could indicate that they have moved to a student house.

We chose to look specifically at full-time students who met the following criteria:

1. Contained an administrative data record in both HESA 2018/19 data and PDS 2019 data
2. Did not have a valid 'reason for removal code' in PDS 2019 data.
3. Registered as living at a university owned hall of residence in HESA 2018/19 data
4. Not in their first year at university[27]
5. Had a term-time postcode in HESA 2018/19 data, which did not match to the postcode registered in PDS 2019 data
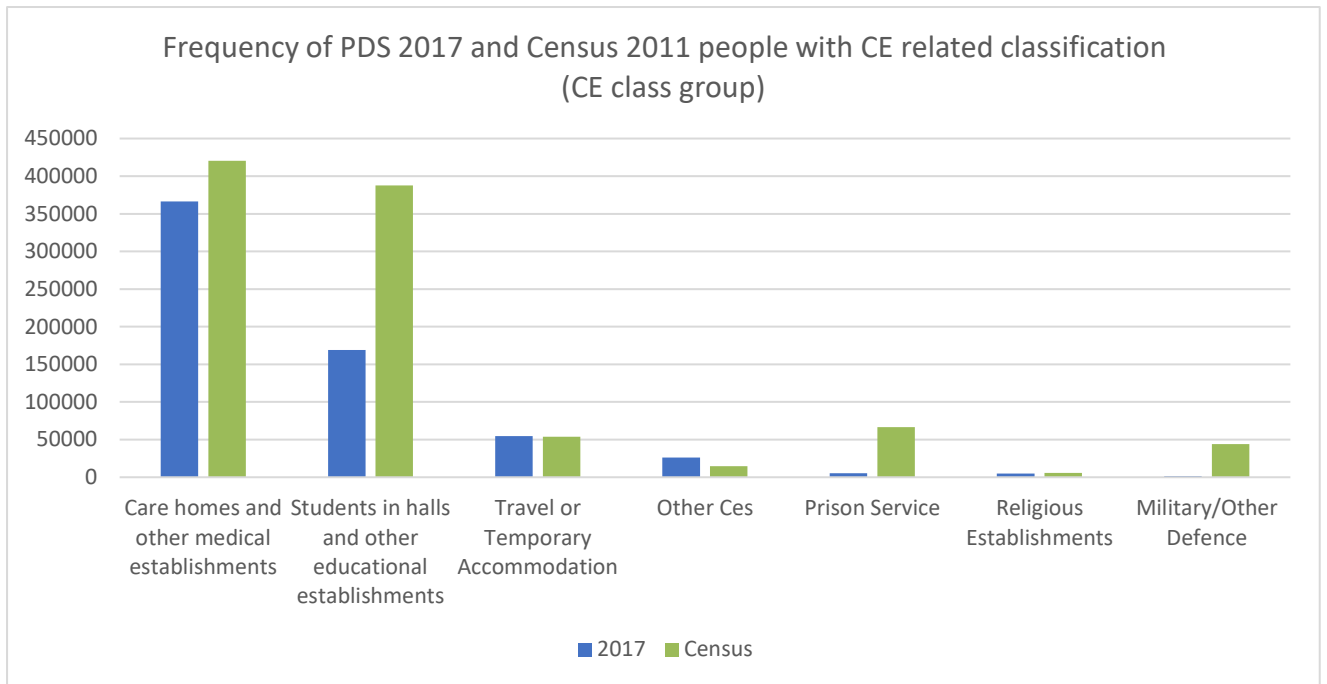
Our analysis found that out of the roughly 45,000 students that met this criteria, roughly 46% of these individuals had a postcode, which did not match between both sources because they were recorded on PDS as living at their home address rather than their term-time address. Moreover, another roughly 6% did not match between both sources because they did not have a valid postcode in PDS 2019 data. Therefore, even though we know there are issues with students not updating their term-time address information, it does not seem that we will be able to address this issue for our statistical use by comparing addressing information with another administrative data source.

---

[25] CE Expected Capacity number for UK Armed Forces bases were mostly provided by the Ministry of Defence

[26] This was highlighted in research conducted by the ONS in 2019, which found that between the years 2013-2018, nearly 75% of the student population present in this dataset had the same term-time postcode throughout their studies, ONS, 'Phase 2 Communal Establishment Sprint Final write-up', 2019.

[27] Included anyone who had been studying at university for 2 or more years and included students, whose year of study was classified as unknown.

Annex C[28]

Frequency of PDS 2017 and Census 2011 people with CE related classification (CE class group)

---

[28] ONS 'Phase 2 Communal Establishment Sprint Final write-up', 2019

References

- Ghee, C and Fraser, O 'EAP141 – Design of Address Frame, Collection and Coverage Assessment and Adjustment of Communal Establishments in 2021 Census', 13th January 2021, https://uksa.statisticsauthority.gov.uk/publication/eap141-design-of-address-frame-collection-and-coverage-assessment-and-adjustment-of-communal-establishments-in-2021-census/

- ONS, 'Boarding School or Children's Home Managers' https://census.gov.uk/help/types-of-household-or-accommodation/communal-establishments/i-manage-an-establishment/boarding-school-or-childrens-home-managers

- ONS, 'Estimation and Adjustment for Communal Establishments', September 2012, https://webarchive.nationalarchives.gov.uk/20160108085304/http:/www.ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-data/2011-first-release/first-release--quality-assurance-and-methodology-papers/coverage-within-communal-establishments.pdf

- ONS 'Phase 2 Communal Establishment Sprint Final write-up', 2019