

Estimation and adjustment for Communal Establishments in the 2011 Census

Owen Abbott, Office for National Statistics

0. Summary

This paper outlines the proposed estimation and adjustment methods that will be applied for measuring coverage in Communal Establishments (CEs). This will differ between small and large CEs since we will only collect information in the Census Coverage Survey (CCS) from small (<100 bed spaces) CEs.

Within England and Wales, there are around one million residents within Communal Establishments, including Prisons, Hotels, Nursing Homes, Military Barracks, Halls of Residence and children's homes. In 2001, coverage was not measured in all types of CE and a total adjustment of 27,000 persons was made on an ad-hoc basis, and only in the largest CEs.

The proposed methodology is:

- a) for those CEs that have up to 100 bed spaces (which are included in the CCS), we will estimate the population using a dual system/ratio estimator to derive regional level estimates by broad type and broad age-sex. These estimates will be used to drive a simple imputation process to add people missed from small CEs onto the census database.
- b) those CEs that have 100 or more bed spaces (which are excluded from the CCS) will use the information gathered by the census special enumerators and any administrative data to make an establishment specific assessment of coverage, and adjustments will be made to these CEs if necessary using the same simple imputation approach as for small CEs.

1. Introduction

Within England and Wales, there are around one million residents within nearly fifty thousand Communal Establishments, which are defined as being 'managed' accommodation. Therefore, Prisons, Hotels, Nursing Homes, Military Barracks, Halls of Residence and children's homes are all types of Communal Establishment. The census treats large CEs (defined as large by the bed space capacity) differently to households, by using a different questionnaire and different enumeration methods. Because they are unlike the general household population, census non-response in CEs is different. This can be due to a number of factors, including the difference in living arrangements, the ability of the residents to respond (there are likely to be more people who require assistance), the level of security required to gain access etc.

This paper outlines the proposals for the estimation and adjustment methods that will be used in measuring coverage of persons within communal establishments.

2. Background

2.1 Scope

Communal establishments fall into 3 categories:

- a) **Small CEs with less than 10 bed spaces**

These are enumerated like normal households and thus will be included in the usual coverage assessment methodology for households (and therefore coverage of the small CEs will be assessed and adjusted). They are therefore out of scope for this paper.

b) Large CEs with 10 to 99 bed spaces

These are enumerated using a special questionnaire to define the type of CE in both the Census and the CCS. This paper defines the methodology for using this data to estimate and adjust for coverage for these types of CE.

c) Large CEs with 100 or more bed spaces

These are enumerated in the Census as above, but they are excluded from the CCS. Field and administrative data will be used to assess coverage within these large establishments, and specific adjustments will be made to adjust the database as was the case in 2001. This paper outlines how this will be achieved.

We will only be measuring coverage within the CEs, and not coverage of the CEs. This is because, particularly for larger CEs, it is not believed that the census will miss many of them – they are likely to be on the household frame and unlike a household have no reason to ‘hide’ from officialdom. The census process involved extensive list-checking of CEs in consultation with LAs and other relevant bodies (NHS; Universities; Armed forces etc.) and enumeration using specially trained staff.

Thus we are making the assumption that the Census frame will identify all CEs (with more than 10 bed spaces) and we therefore will only measure coverage within them (though we will use the information from sampled CEs that are in the Census only and in the CCS only), and impute residents into CEs that the census identifies. This does not rule out the creation of new CEs and residents within them, but this would have to be an ad-hoc manual process (the facility to do this will be available, but the creation of the records to be imputed will be manual much like a data file amendment).

2.2 The CE population

The 2001 Census results show that there were 46,431 Communal Establishments in England and Wales. Annex A shows the breakdown by type, and Annex B shows the type categories being used in the 2011 Census. The number of usual residents in the 2001 Census Communal Establishments was 934,256 – 448,554 in Medical Establishments, 485,702 in ‘Other’ types of establishment. Of all residents, 76,178 were resident staff and their families. Age and Sex breakdowns are available from the Census Results volumes. Note that the 2011 Census (and CCS) will **NOT** be collecting any individual level visitor data from CEs. Clearly in some types of CEs there are many more visitors than residents (e.g. Hotels).

3. Proposals for measuring coverage in CEs

3.1 Estimation for large (10 to 99 bed space) CEs

These CEs will be included in the CCS using a specific questionnaire, and therefore we will have matched data available for those that fall into the sample areas. Note that the CCS sample design does not take CEs into account, and therefore there is no control over the size of the sample that will be achieved.

The proposal is to use dual system and ratio estimation to estimate coverage within these CEs, by region, type and age-sex if the sample size is sufficient to support such estimates. This was the approach used in the evaluation of the 2001 CCS data, which estimated the population within CEs by type and age, estimating that 30,000 individuals were missed (see Annex C).

The main reason for aggregating across Estimation Areas is that the CCS sample is not optimised to measure this population at that level – but the sample size may be acceptable for making regional level estimates. Previous work suggested exploring whether particular ‘CE only’ postcodes could be included in the sample to help boost sample sizes – however, this was rejected for a number of reasons, including:

- a) This might require additional sample;
- b) there may be unwanted consequences for the household sample;
- c) there is no reliable way of identifying CEs that are mixed in with residential households (and thus there is not really any auxiliary data for helping to balance or anything that would form a sample frame).

In the 2001 CCS sample 374 CEs were interviewed, the majority containing less than 40 residents. They tended to be nursing homes, care homes, hotels, boarding houses, hostels and ‘other’ types. This sample size was enough for a national estimate by collapsed type, and possibly enough for direct regional estimates (there are 9 regions – so there would be about 40 per region), but these were not explored. Given the CCS sample size is similar for 2011, and the underlying CE population has not changed dramatically, we expect a similar achieved sample size for 2011.

The methodology proposed is to compute a set of regional level DSEs by collapsed age-sex groups and collapsed CE types (excluding some types of CE for which the population was very different and the sample size extremely small) to measure coverage within the CEs. The DSE is then used to derive the coverage ratio, and this is then applied to the Census counts for these types of CE.

The suggested CE types are the broad categories used in the 2011 Census as shown in Annex B. These are:

- Medical and Care
- Education
- Armed Forces
- Detention
- Travel and temporary accommodation
- Other

The collapsed age-sex groups are similar to those used in the 2001 study, with a split between children and young adults:

- Males and Females aged 0 to 15
- Males and Females aged 16 to 29
- Males and Females aged 30 to 59
- Males aged 60+
- Females aged 60- 84
- Females aged 85+

Despite the collapsing, these groups may still not contain enough sample within them to compute robust DSEs, and so further collapsing may be necessary to stabilise the DSEs (even with the use of the chapman correction). The actual groupings will thus depend on the data collected.

Regardless of the categories used to compute the DSEs, they can then be used to compute adjustment factors (i.e. the ratio) which can then be applied to the census counts for these groups to

obtain estimates of the population within these CEs. This means we can get estimates by LA, type and age-sex by using the adjustment factors and assuming that they apply at the lower level (i.e. a synthetic assumption).

Therefore the approach is to estimate the Local Authority by establishment type and age sex population \hat{T}_{lea} where l is the LA, e is the establishment type and a is the age sex group by:

$$\hat{T}_{lea} = \frac{\sum_{sample} \hat{y}_{rgc}}{\sum_{sample} x_{rgc}} X_{lea}$$

where the x_{rgc} are the census counts for the region r, grouped establishment type g and collapsed age sex group c, X_{lea} are the census totals for the local authority by establishment type and age sex, and \hat{y}_{rgc} are the DSEs estimated by region, grouped type and collapsed age-sex.

3.2 Estimation for special CEs

Special CEs, which have a capacity above 100 bedspaces are excluded from the CCS. This is the option adopted in the 2001 Census. For these, the proposal is to make adjustments based on a mix of the information collected by census enumerators and administrative based sources. This could also involve telephoning particular establishments to gather information where the field or admin data indicates there was a problem. The QA team will be obtaining the administrative data that will be used for this.

If these studies detect large discrepancies, we will make specific adjustments to the CE using a very simple imputation system that creates the required number of records within the CE with the required characteristics. Again this will use the same principles as the main imputation (i.e. impute the main characteristics and let CANCEIS impute the remainder).

Making specific adjustments is the approach used in 2001, but we will ensure in 2011 that there are systems built to assist with this analysis and imputation (in 2001 it was purely manual, and some clerical errors were made).

4. Adjustment methodology

4.1 Basic Principles

It is not proposed to use a modelling approach as is being used for the household population since there is less data on which to base any models (as the CCS sample size is quite small) and there is no data for the special CEs. Thus we will simply choose donors within any relevant impute classes to uprate the CE populations in line with the estimates described above.

We will impute skeleton persons from donor individuals within the same CE. This is simple to implement and replicates existing patterns within a CE (for instance, this will only replicate existing age-sex groups, ethnic groups or Activity last week status etc). This provides some protection against creating CEs with a mixture of individual characteristics that might not exist. It is also an identical approach to that used for the household imputation.

4.2 Imputation method for large (10 to 99 bed space) CEs

The estimation process results in a set of coverage weights by GOR, Broad CE Type and Broad age-sex. These are then fed into the following imputation process.

We firstly list all the CE persons for small CEs within each EA. We attach the relevant weights at individual level and then order this list by CE Type and weight. Then, as per the normal household imputation process, we compute a cumulative unweighted count and a weighted count down the list of individuals. When the difference between these counts becomes greater than 0.5, this indicates an imputation should be made, and the record at which this occurs is selected as the donor. A skeleton person with the same characteristics is imputed into the same CE as the donor. This whole process can use a modified version of the household imputation system already coded.

This will uprate the populations within broad type and broad age-sex group proportionately, so if the weight is 1.05, then we expect each CE to get bigger by about 5%.

This method should result in the correct number of imputations by CE type and broad age-sex group within the small CEs in the EA.

Imputation method for large CEs

The estimation process results in counts (by age-sex) of individuals required to be added in a particular CE.

We will use the same imputation procedure as for the small CEs. In order to do this we will calculate the coverage weights using the population totals (simply divide one by the other), and feed these in to the usual imputation process described above. This will then impute the correct number of persons from the counted individuals within the same CE by age-sex group.

This means that some individuals can be used as donors more than once, but only if the weights are 2 or greater (i.e. the estimates mean we are creating more records than exist already in the CE)

This method should result in the correct number of imputations within the CE type and will replicate the existing characteristics (e.g. ethnicity, after controlling for age-sex groups) within the specific CE.

Annex A – Numbers and Types of CE

Table UV70: All communal establishments: England and Wales

ALL COMMUNAL ESTABLISHMENTS	46,431
Medical and care establishments	23,435
NHS	1,907
Psychiatric hospital/home	530
Other hospital home	1,377
Local Authority	2,527
Children's home	393
Nursing home	51
Residential care home	1,906
Other home	177
Housing association	804
Home or hostel	804
Other	18,197
Nursing home	4,366
Residential care home	12,438
Children's home	349
Psychiatric hospital/home	352
Other hospital	82
Other medical and care home	610
Other establishments	22,996
Defence establishments (including ships)	291
Prison service establishments	150
Probation/bail hostel	92
Educational establishment (including halls of residence)	2,283
Hotel, boarding house, guest house	8,042
Hostel (including youth hostels, hostels for the homeless and people sleeping rough)	2,051
Civilian ship, boat or barge	6
Other	10,081

Annex B – CE Type categories in the 2011 Census

1 What is the nature of this establishment?

➤ Tick one box only

Medical and care

- General hospital
- Mental health hospital/unit (including secure units)
- Other hospital
- Care home without nursing
- Care home with nursing
- Sheltered housing only
- Children's home (including secure units)
- Other medical and care establishment

Education

- School
- University (for example halls of residence)
- Other educational establishment

Armed Forces

- Defence establishment (including ships)

Detention

- Prison service establishment
- Approved premises (probation/bail hostel)
- Detention centre
- Other detention establishment

Travel or temporary accommodation

- Hotel, guest house, B&B, youth hostel
- Holiday accommodation (for example holiday parks)
- Hostel/temporary shelter for the homeless
- Other travel or temporary accommodation

Other

- Religious establishment
- Staff/worker accommodation only
- Other establishment

Annex C – Estimates of undercount within 10-100 sized CEs in the 2001 Census

The 2001 evaluation study estimated that we missed some 5 per cent, or around 30,000 persons as shown in Table 1 below. The 460k counted in the census refers only to those types of CEs included in the analysis and which had less than 100 residents. Therefore it does not cover all CE types – the assumption was that in the others we achieved 100% coverage (in the absence of any robust information).

These estimates were not published or used for any subsequent adjustments – but they do highlight that the CCS did apparently pick up residents the census missed – no work has been done to examine who these people were.

Table 1 - Census counts and undercount estimates for collapsed type and agegroup for England and Wales.

Collapsed type	Collapsed ages	Census Count	DSE Undercount	
			Estimate	Adjustment
Care Establishments	Males and Females aged 0 to 29	13346	14074	728
Care Establishments	Males and Females aged 30 to 59	51361	53604	2243
Care Establishments	Males aged 60+	83529	88147	4618
Care Establishments	Females aged 60- 84	106914	113704	6790
Care Establishments	Females aged 85+	155601	163103	7502
Hostels and Hotels	Males and Females aged 0 to 29	23933	28959	5026
Hostels and Hotels	Males and Females aged 30 to 59	20650	22606	1956
Hostels and Hotels	Males aged 60+	3123	3346	223
Hostels and Hotels	Females aged 60- 84	1279	1279	0
Hostels and Hotels	Females aged 85+	199	199	0
TOTAL		459935	489021	29086