2011 Census - Methodology for Quality Assuring the Census Population Estimates

Contents

1 Introduction

This paper provides an overview of the methods for the quality assurance (QA) of census population estimates and the process by which the estimates will be adjusted (if required) and finally agreed.

The focus of this paper is the assurance of the census population estimates and the key population subgroups or housing characteristics associated with producing the census population estimates. Assurance of the other census characteristics, such as the labour market questions undergo a similar process to that set out in section 3.

The quality assurance of the census estimates occurs after the coverage assessment and adjustment (CAA) process, which itself includes a number of quality assurance checks – for example against the census address register and ONS surveys. The CAA process estimates and adjusts for missed households and people in the census (see Abbott, 2010).

The methods set out in this paper build on the lessons learned from the Quality Assurance of the 2001 Census population estimates (see White, Abbott and Compton, 2006) and the subsequent Local Authority Studies (ONS, 2004).

The key lesson from the Local Authority (LA) studies was that the 2001 Census address register and field information systems were inadequate. This has already been addressed by development of a national address register and questionnaire tracking system to underpin the entire census operation.

Beyond that, it was also clear that ONS should do more to explore the use of local knowledge and locally available data sources in the quality assurance of the census results. This will potentially include matching individual census records with other data sources to really understand discrepancies that might be found at the aggregate level.

Therefore, rather than having extensive discussions with LAs after the publication of the 2011 Census results, ONS has already engaged with a number of LAs to determine which locally held sources might be useful in the Quality Assurance process. The findings from these QA studies, were published in December 2010 (ONS, 2010). Those sources identified as most useful in the QA studies have subsequently been requested from all LAs with a reference date of March 27th 2011.

A further lesson learnt from the 2001 QA was the need to maintain a focus on the emerging picture across the country as local areas estimates are put through the QA process. This has been directly addressed with the establishment of a high level panel to review emerging regional and national estimates. This panel, which includes independent experts, sits alongside a main quality assurance panel tasked with reviewing all LA estimates.

To support the quality assurance at local, regional, and national levels this paper sets out what will be checked at each of these levels. It also sets out the adjustment options at each level.

Publication of this overview will be supplemented by additional detailed papers in July 2011. These will set out:

(i) the criteria to be used to determine when more detailed investigation will be carried out,
(ii) additional information on what the more detailed 'supplementary' investigations will include,
(iii) further detail on improvement methods, including the distribution of any national level improvement to Local Authorities.

2 Overview

This section provides a brief overview of the process for quality assuring the census population estimates. It outlines some of the key checks that will be undertaken and reflects the areas of risk associated with the estimation process. A fuller explanation of the QA process is set out in sections 3 to 6.

As set out in the introduction, quality assurance will be carried out at local, regional and England & Wales levels. The work to be carried out is summarised in table 1, along with options for improving estimates at each level if required.

A distinction is made between those checks which are core and those which are supplementary. Core checks will be routinely assessed for all areas. Supplementary checks will be carried out to investigate inconsistencies found with the core checks.

Table 1 – Local, Regional and England & Wales overview of quality assurance and improvement options

Level	Core Checks	Supplementary Checks	Improvements
Local Authority	Checks against comparator sources Assessment of the distribution of population below LA Demographic analysis Operational Intelligence Area profiles LA provided evidence	Analysis of visitor and second residence Additional checks on population sub-groups Low level aggregate comparisons Data matching to administrative data	Calibrate to external sources Adjust Coverage Assessment/Adjustment (CAA) Adjust using alternative household count
Regional	Aggregate checks against comparator sources Demographic analysis Modelling non-response Assessment of Census Non-Response Link Study (CNRLS)	Analysis of visitor and second residence	Calibrate to external sources Adjust CAA Adjust using CNRLS
England & Wales	Aggregate checks against comparator sources Demographic analysis Assessment of CNRLS Assessment of Longitudinal Study	Analysis of visitor and second residence	Calibrate to external sources Adjust CAA Adjust using CNRLS Adjust using Longitudinal Study

2.1 Core Assurance Checks at the LA level

All LA estimates will be subject to the six types of core check set out in the table above. A fuller specification of the checks and sources used is provided in section 3.

The census population estimates by age and sex for each LA will be compared with other existing sources, including mid year population estimates and a range of administrative datasets. Different sources will be used to assess the numbers and key characteristics of different population groups.

Census estimates of the distribution of the population within each LA will also be checked. The assessment will be made at both Lower and Middle Super Output Areas, again in comparison to administrative data.

The 2011 QA process will routinely use demographic analysis to assess the census population estimates. For each LA this will include sex ratios (the ratio of men to women) and fertility/mortality rates using the census estimates as the denominator. Census based fertility/mortality rates will be compared to rates derived from other sources (e.g. the ONS mid-year estimates).

When analysing and reviewing the core checks it will be important to understand key census field operation indicators. This includes overall return rates (set against expectations), variation of return rates within an LA as well questionnaire tracking information and calls to the census contact centre.

To provide valuable context about each LA an area profile has been compiled for the QA process. This includes a record of previous correspondence and the findings from reviews of the LA's population (including the post-2001 LA studies). It also provides a regional/national context for the LA population estimates and insights into change since the 2001 Census using the mid-year population estimates and other administrative sources.

Finally the quality assurance process will be able to draw on evidence provided from each LA. This has either come through correspondence with ONS on the 2001 Census or on the subsequent mid-year populations. As set out in the introduction, users have also been asked to be provide locally held data and relevant research as part of the QA studies (and subsequent call for information).

All of this information will be assessed using published guidance on whether the census estimates should go to the Quality Assurance panel for approval or be referred for supplementary analysis. This guidance approach is set out in section 3.

2.2 Supplementary Quality Assurance at the LA level

Where the core LA checks raise questions, supplementary, more detailed, analysis will be undertaken. As set out in section 4, this work may assess the population as a whole at very low geographic levels or particular sub-groups within the LA.

An analysis of specific small geographic areas will focus on the postcode clusters included in the sample for the Census Coverage Survey in each Local Authority. It is in these postcodes that the Dual System Estimate (DSE) produced during the coverage assessment process will be most robust. The DSEs are particularly important as they are the basis for estimating and adjusting for under-coverage across the wider LA(s). This work will include:

- comparing population estimates for each of these postcode clusters, and the distributions of key characteristics such as age and sex, with aggregate statistics from the available administrative dataset for the same postcodes. Comparisons will be made for both individuals and households.

- An analysis of household size within these postcode clusters with administrative sources to identify where large households may not have been adequately adjusted for in the estimates.

- Matching between the census records for these postcode clusters and administrative records depending on the patterns seen in the aggregate comparisons, to further understand differences.

- Matching between administrative records for these postcodes and census records from addresses outside these postcodes (including outside the Local Authority) where individuals from other areas have indicated in their census return either that their address one year ago or their second residence address is in the CCS postcode, indicating potential over-count in the administrative datasets.

Population sub-groups where questions have been raised will be analysed using the additional detail provided by the sources used in the core quality assurance. Examples of these supplementary checks include investigation into the number of foreign or short-term migrant students to explore questions raised about total numbers of students.

Supplementary analysis will also include investigations into the information built into the census data collection to help explain inconsistencies with administrative datasets. This includes information collected on second residences, short term migrants, vacant properties and visitors.

It is likely that the analysis may need to be completed for a number of LAs before firm conclusions can be drawn about the existence of systematic

coverage problems. Where supplementary QA indicates that improvements to the census population estimates are required then improvements will be made either to a single LA, to a number of specific LAs or across all LAs.

2.3 Core Assurance Checks at Regional and National Levels

As at the LA level, a range of core checks will be routinely run at both the regional and England & Wales levels. These include:

- Repeating checks against comparators at these higher geographic levels. Some comparators, such as survey based estimates, will be more robust at higher levels.
- Repeating demographic analysis of sex ratios, fertility and mortality rates. Recent work has highlighted the potential usefulness of comparing sex ratios from administrative data at the England & Wales level (Smallwood, 2009). As an extension to this work ONS are in the process of repeating this analysis for 2011 using more administrative datasets.
- Analysis of the ONS longitudinal study to identify people who were missed in the 2011 Census (which could be for legitimate reasons) and also those who have been included more than once, at different locations.
- Analysis of the Census Non-Response Link Study (CNRLS). The CNRLS matches census returns to recent returns from ONS social surveys and will be used to identify where individuals within households were missed in the census but were identified in social surveys. Regional and national level patterns of missed census respondents will be assessed.
- Modelling non-response to assess whether other census characteristics (e.g. ethnicity and tenure) are more powerful predictors of non-response than age and sex that are currently used in the DSE.
- Assessment of the census population estimate for England & Wales against the equivalent mid-year population estimate rolled forward from the 2001 census. Statistical variation and biases in the sources used to roll forward population estimates will be quantified (by sex) to help understand differences between the two estimates.

Evidence from these core checks will be considered by the high-level QA panel to either approve, request supplementary analysis or improvement.

2.4 Supplementary quality assurance at the regional and national levels

As at LA level, supplementary analysis will be carried out to further investigate potential issues with the core checks.

This will include extending some of the supplementary checks initially carried out at LA level. For example this may include extending the outputs of the low

level comparisons and the matching done during the LA supplementary analysis to understand the extent of systematic bias across areas.

Across England & Wales, census forms will be returned stating that an address is a second residence or that an individual is only a visitor at that address. In such cases individuals should record that their usual residence is at another address, potentially in a different LA. When census forms have been processed across the country, individuals will be matched back to their usual residence to check for systematic under-coverage.

2.5 Options available for making improvements to the census population estimates

Where the results of the supplementary, regional or national analysis clearly indicates a problem with the estimates then improvements will be applied. The adjustment made will depend on the nature of the problem and the intelligence gained during the supplementary analysis, particularly our understanding of the different data sources and their quality.

One, or a combination, of the following improvements is available:

- Revisit coverage estimation (post-stratification by alternative characteristics, post-stratification by re-grouping estimation areas, borrowing strength from other LAs).
- Adjust coverage estimation to account for between and within household bias.
- Calibrate using external data (national adjustments, direct calibration for small geographic areas/population sub-groups).

Section 5 has more detail on improvement options at national, regional and local levels.

2.6 Publication of the results

Census population estimates will be considered by four panels prior to publication. As detailed in section 6, these panels are:

- **Internal QA Working Group**
  Census experts providing a daily check on whether the QA checks should lead to supplementary QA, prioritise the work analysis required, and give a steer on whether improvements are needed.

- **Main QA Panel**
  Census and demographic experts from ONS (plus Welsh Assembly Government) meeting weekly to review all estimates at LA and higher geographies.

- **High Level QA Panel**

  A group of internal, UK and independent external reviewers who will provide expertise and guidance on the emerging national picture.

- **Executive Panel**

  An executive panel, chaired by the National Statistician, accountable for the final sign-off of the national and local census population estimates ahead of publication.

The High level QA Panel will approve and endorse the estimates for each Local Authority once it is satisfied that the estimates have been through a thorough process of assurance. They will then be formally signed off by the National Statistician and the executive panel ahead of publication.

An important element of the publication in July 2012 will be the information that accompanies the results. Alongside the census estimates, a number of the key comparisons will be shown. In order to give users confidence in the results, it is imperative that the results include a description of how the estimates for each LA relate to other sources, including previous mid year population estimates, and administrative datasets. A summary of the improvements made during the quality assurance process will also be published.

When comparing with other datasets, it is important that definitional issues are described and explained. Many administrative datasets will include people in LAs who are not 'usually resident' in a Local Authority according to the census definition. For example, the NHS register and National Insurance registers will include people who are living in the UK for less than 12 months, or whose family home is in another part of the UK. The census questionnaire has been designed to explicitly identify such people and, through processing, they will be excluded from the usual resident population estimate. Estimates of such populations will be available to help explain differences between the census and administrative sources.

3. Core Assurance Checks

3.1 Core Checks and Data Sources at LA level

This section provides a more detailed specification of those core checks identified in the overview. A series of tolerances have been derived for each of the core checks to help distinguish between potential errors in the census estimates and definitional/coverage issues with the comparators. Further information is provided on how these tolerances have been devised and how they will be used during the operation to guide when supplementary analysis is required.

It is noted that all core checks will be based on the usual resident population intending to stay in the UK for twelve months or more. This is consistent with the definition used in mid-year population estimates. Those census respondents identifying themselves as not usually resident in the UK (and those imputed as not usually resident through the coverage adjustment process) will be checked separately.

The results of the core checks are compiled together into a quality assurance pack available to the quality assurance panel. Other information considered by the panel includes information about the conduct of the census field operation in the area ('operational intelligence'), area profiles, and local authority provided intelligence.

A glossary of the comparator sources referenced in the following section is provided as appendix A to this paper.

Checks Against Comparator Sources

a. Age and Sex

The total number and distribution of the population by age and sex will be compared against a number of comparator sources. For the most part, the checks will be undertaken at 5 year age groups as this is the level at which coverage estimation is first assessed. Checks for single year age groups will be undertaken where this is supported by the comparator sources.

Comparator sources:

- Birth registration
- ONS mid-year population estimates
- NHS patient register
- Department for Education (DfE) and Welsh Assembly Government (WAG) school census

- HM Revenue and Customs (HMRC) child benefit data[1]
- Department for Work and Pensions (DWP) pension claimants data[1]
- Customer Information Service data from DWP/HMRC on interactions with benefits and taxations systems[1]

b. Total Population Below LA Level

The distribution of the census population estimate at Lower-Layer Super Output Area (LSOA) and Middle-Layer Super Output Area (MSOA) levels will be assessed across each LA. Differences between the census and comparator sources across LSOAs and MSOAs will be reviewed to identify any systematic differences, for example where the census estimates tend to be generally higher or lower than comparator sources, and to identify outliers.

Comparator sources:

- NHS patient register
- ONS mid-year population estimates (small area population estimates)

c. Household Numbers

The estimated number of occupied and un-occupied households will be checked based on census household returns. This can only be done where the comparator sources indicate whether an address is occupied (for example on the patient register).

Included in the check will be a household count made up of the number of occupied census household returns and 'dummy forms' (which are completed by census field staff for non-responding addresses) which indicate that an address is a main residence. It will also include addresses where no information was collected - with an assumption made about which of these addresses are a main residence.

Other sources which do not identify occupancy will also be included on the check for completeness. This includes both the Census Address Register and the Valuation Office Agency (VOA) Council Tax data.

Comparator sources:

- Valuation Office Agency (VOA) Council tax data
- NHS patient register
- Census Address Register

[1] Child benefit, pensions data and data from the Customer Information Service (CIS) is currently being assessed to understand which of these sources should be used in the QA process. Central to this work is the understanding of the CIS which has only recently become available but will contain claimants of child benefit, state pensions but also people of any other age who claim benefit or pay tax.

- Department for Communities and Local Government (DCLG) household projections
- Alternative Household Count (further information provided in section 5.3.1)

d. Household Size

Census estimates of household size will be checked by assessing average household size and the frequency distribution of household size. The frequency distribution will be particularly useful to identify whether census household continuation forms were used (or were not used as intended). The main household census form only has space for six individuals, a sudden drop between the number of six and seven households might indicate that large households often did not request a continuation questionnaire. This would be validated against comparator data.

Comparator sources:

- NHS patient register
- Integrated Household Survey
- ONS population projections and Department for Communities and Local Government (DCLG) household projections

e. Large Communal Establishments

Large communal establishments (with more than 100 bed spaces) will be checked separately from the total census population estimates. As large communal establishments are not covered by the Census Coverage Survey (CCS) there is no systematic adjustment through Dual System Estimation for where individuals did not complete a return.

By default comparator datasets will be used to adjust the results when the census estimates for that communal establishment are significantly lower than the expected number. If comparator data are not available, direct contact will be made with the establishment.

An assessment of the number of census forms returned will also be made in relation to the number of forms which were issued.

Comparator sources include:

- Higher Education Statistics Agency
- Census Communal Establishment list – expected number of residents
- Ministry of Justice Prisons data
- Department for Education data on school boarders
- Patient Register (nursing home residents)

f. Ethnicity

Census population estimates by broad and fine ethnic group as well as by broad age and sex will be checked against comparator sources:

- ONS Population Estimates by Ethnic Group
- DfE & WAG School Census
- Integrated Household Survey

g. Students

Census estimates of students in full-time education aged 18 or over will be checked by single year of age and sex. This includes students in further and higher education. The check covers all students (regardless of whether they live in communal establishments or in households), students living in communal establishments and students living in households.

Students tend to be geographically clustered and so checks will be carried out within an LA at MSOA level.

Comparator sources:

- Higher Education Statistics Agency data
- Department for Business Innovation & Skills (BIS) and Welsh Assembly Government (WAG) Further Education data

h. Armed Forces (Home and Foreign)

Armed forces personnel are highly concentrated in local authorities with (or near) an armed forces base. The check will assess the census estimate of the number of armed forces personnel by age and sex by both base and residence using comparator data from the UK and US armed forces. This includes both personnel resident on the base and in private accommodation.

The comparator data for both UK and US armed forces is by base only. There is no information about where individuals are resident. It has been necessary to apportion armed forces personnel by base to their LA of residence using the 2001 Census. The process of apportioning is used in the production of mid-year population estimates but is likely to be out of date in some areas.

Alongside comparator data from the UK/US armed forces, the number of census returns received from personnel living at each barracks will be checked against the number of forms which were initially distributed.

Comparator sources:

- Defence Analytical Services Agency
- US Armed Forces data

- Census Communal Establishment list – expected number of residents

i. Migration (Internal)

Census estimates of migration within England & Wales and from Northern Ireland and Scotland are based on the address one year ago question. Internal migration to each local authority will be checked by age and sex. It is not possible to check internal migration from each local authority until all areas have been processed.

Comparator sources:

- NHS patient register

j. Migration (International)

The address one year ago question will be used to identify international migrants who previously lived outside the UK. The number of short and long-term migrants are both checked by age and sex. Short-term migrants are distinguished from long-term migrants using the date of arrival and intention to stay questions in the census questionnaire for people whose country of birth was outside the UK.

Comparator sources:

- NHS patient register
- International Passenger Survey
- Migrant Worker Scan
- ONS estimates of international long and short-term migration

Demographic Analysis Checks

a. Fertility Rates

Fertility rates will be calculated based on census population estimates of women aged 15-44. Calculating and analysing rates of fertility and mortality can be a useful alternative way of assessing census population estimates given the accuracy of compulsory birth and death registration.

These rates will be compared to those calculated using previous mid-year population estimates and those referring to mid-2011. Unusually high or low fertility rates (calculated using the mid-year estimates) will be assessed to check whether they become more or less unusual when calculated using the 2011 Census. Similarly, any unusually high or low rates calculated using 2011 census estimates will be checked against existing MYEs and other comparator sources. Analysis of fertility will also assess variation in fertility rates within each LA.

Comparator sources:

- Births registrations data
- ONS mid-year population estimates

b. Mortality Rates

Similar comparisons will be made for mortality rates as set out for fertility. Age-standardised mortality rates will be calculated using census population estimates and compared to the equivalent rates calculated using mid-2011 population estimates. Analysis will also be carried out on age-specific mortality rates calculated using the census population estimates.

Comparator sources:

- Deaths registrations data
- ONS mid-year population estimates

c. Sex Ratios

The ratio of the number of men per 100 women in the census population estimates will be checked by single year of age, to see whether the census population estimates result in implausible values.

Operational Checks

Operational intelligence also includes core checks on the Census return rates including:

- The overall LA return rate
- The LA return rate compared to the expected return rate
- The variability in the response rate across LSOAs
- The 2011 return rate against the return rate for the 2001 Census

Where potential problems are identified a range of additional intelligence can be considered. Additional intelligence includes the findings from debriefings with Area Managers and Local Authorities on any problems with the field operation. It will also draw on the detailed information provided by the Questionnaire Tracking system.

3.2 Core Checks at Region/National level

a. Aggregate checks against comparator sources

All of the core checks listed in section 3.2 can be aggregated to regional and national levels. The age/sex and household checks will be routinely repeated at these levels as these are central to the quality assurance process.

At regional level the number of armed forces personnel by base will be routinely checked. It is assumed that region of residence (from the census) is directly comparable to workplace (from the comparator data). As described in section 3.1, no comparator data is available which directly identifies where armed forces personnel are resident, only on where they are based. Evidence from the 2001 Census suggested that personnel did not always live in the local authority in which they were based.

b. Demographic Analysis

Analysis of mortality/fertility rates and sex ratios will be repeated at both regional and national levels. Experience from 2001 suggests that sex ratio analysis will be particularly valuable.

For 2011, the ONS Centre for Demography (ONSCD) has been undertaking research to determine plausible national age/sex ratios before the census QA process. Currently work is focusing on sex ratios based on an analysis of the Department for Work and Pensions' (DWP) Lifetime Labour Market Database (known as L2). The L2 is a one per cent sample of National Insurance numbers. For these individuals, interactions with both Her Majesty's Revenue and Customs (HMRC) and DWP are linked to provide a longitudinal picture of economic activity (including registration for benefits).

c. Analysis of the ONS Longitudinal Study (LS)

The LS contains linked census and vital events for 1 per cent of the population of England & Wales. A discrepancy was found in the LS after matching between the 1991 and 2001 Census, even following adjustments made for levels of expected attrition. This provided evidence for an adjustment to the number of males aged 25-34 at the England & Wales level.

Recognising the potential value in analysis of the LS for quality assurance, ONS has brought forward the process of LS matching to the 2011 Census. As a result of bringing forward this matching the evidence will be checked during the quality assurance process.

d. Census Non-Response Link Study (CNRLS)

The CNRLS will match census respondents to relevant ONS social surveys, including the Integrated Household Survey. Matching will identify those people who responded to a survey but not to the census. This will include individuals missed from whole households where no census return was received, as well as individuals missed from households where a census return was returned.

Results from the matching will be particularly useful to assess whether there is systematic under-coverage of individuals from households where a census form was returned.

e. Modelling non-response

The coverage adjustment process is based on the estimation of non-response using age by sex distributions within hard-to-count areas. During census data processing, predictors of non-response will be monitored as they emerge in the accumulating data. This modelling will assess whether potential alternative variables could be better predictors of non-response. Ethnicity, tenure or a combination of variables could be used as alternatives to age by sex for improving coverage estimation/adjustment.

f. Assessment against mid-year population estimates

At England & Wales level the census population estimates will be compared to mid-year population estimates rolled forward from the 2001 census. Any difference between these estimates (by sex) will be better understood by quantifying the statistical variation and biases in the administrative and survey sources used to roll forward the mid-year population estimates. This is being undertaken as part of the preparations for the quality assurance.

3.3 Devising Tolerances at LA level

Each of the core comparator checks will have a pre-defined tolerance range which will be used to identify potential discrepancies between census estimates and the comparator data. As set out in the guidance section of this paper, tolerances are used to identify where more detailed 'supplementary' quality assurance is needed.

Using tolerances recognises that although comparator data provide a good estimate of the indicator in question, the comparators themselves will contain error and variance. There are also definitional and coverage differences. For instance, the School Census does not include children in independent schools.

Understanding of the census and comparator sources will improve during the quality assurance process. Thus, during the quality assurance process tolerances (and the associated guidance) will be reviewed to assess how effectively they identify census estimates requiring further investigation.

Tolerance Methodology

There are four types of method used to set tolerances:

- Diagnostic range (multiple comparators)
- Quality assessment (single comparator)
- Set percentage (single comparator)
- Change over time (single comparator)

A. Diagnostic Range

The diagnostic range approach was first used in the quality assurance of the 2001 census estimates of age and sex. The same basic approach will be used again in 2011 for a number of checks.

The diagnostic range is defined as:

the midpoint of the range of different comparators plus or minus the range itself

By defining the diagnostic range in this way the range is broadly between 2 and 2.5 standard deviations beyond the extreme values of the range of comparators.

For the 2011 census quality assurance, diagnostic ranges are calculated for each local authority and each category checked. In the case of age/sex for instance, different diagnostic ranges are calculated for each five-year age group. Differences in the width of the diagnostic ranges reflect variation in comparator sources between local authorities and between age/sex groups.

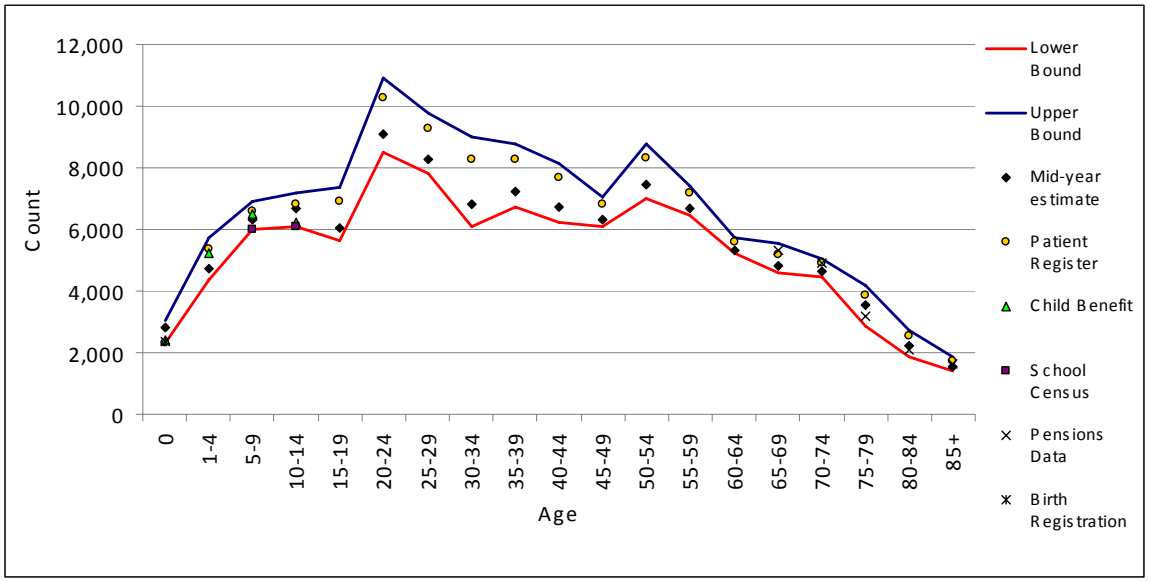
Worked example

The diagnostic range can be demonstrated when applied to 20-24 year old males for a hypothetical LA with two comparator data sources. If for this age group there were 9,102 men in the mid-year population estimates and 10,293 in the patient register this would give a range (or difference) of 1,193.

Taking the mid-point of the two estimates gives 9,102 + (1,193 / 2) = 9,697. Using this the mid-point +/- the range (9,697 +/- 1,193) gives a diagnostic range of 8,507 to 10,889.

A diagnostic range is then calculated for each age/sex group using the sources available. Figure 1 shows the diagnostic range for all ages based on the hypothetical example above. The lower bound for under one year olds and children aged 5 to 14 have been constrained to the Patient Register and School Census respectively as described below.

Figure 1 – Worked Example of Age/Sex Diagnostic Range



Core QA Checks using the Diagnostic Range are:

- Age/sex
- Household number

There are two refinements made to the diagnostic range approach when applied to age/sex. These are:

(i) Comparator data used as a lower bound rather than the diagnostic range where sources of sufficient quality exist. Patient register data is used as the lower bound for under 1 year olds and the school census is used for 5 to 15 year olds.

Replacing the lower bound of the diagnostic range with these two sources the quality assurance process will be in effect controlling to these values. This amendment was introduced following a recommendation made by the Independent Review of Coverage Assessment/Adjustment and Quality Assurance (Plewis, 2011).

Development of the quality assurance approach showed that using the sources in this way narrowed the diagnostic range in almost all cases.

The school census is a high quality source which records children's residential address (as well as the location of their school). However as it only covers children in state maintained schools it only replaces the diagnostic range lower bound.

The Patient Register count of babies under one year of age is likely to be particularly accurate because of the regular contact with the health system for

new born babies. At other ages there is greater uncertainty in the Patient Register because individuals are not required to register when they move (either within the UK or when emigrating overseas), there is less likely to be contact with doctors, and because (for young adults in particular) more moves take place. Compulsory birth registration data is of very high quality but is not updated to include changes of address.

(ii) Where the diagnostic ranges vary greatly, diagnostic ranges are amended to ensure that ranges are not implausibly wide or implausibly narrow.

Amending the ranges involved using cut-off values to remove the very widest and narrowest ranges based on the latest available data, mid-2009. These cut-offs were based on the width of the diagnostic range in an age/sex group for all local authorities, expressed as a percentage of the mid-point. They were identified as the 10^{th} and 90^{th} percentiles.

A similar amendment was made in preparations for the 2001 census quality assurance. However rather than using broad age group (0-19, 20-24, 25-29, 30-79, 80-84, and 85+) as in 2001, cut-offs will be calculated and applied in five-year age groups. This refinement is in line with how the diagnostic range will be used, by five year age group rather than in broad age groups.

B. Quality Assessment

Where only a single comparator dataset is available, it is not possible to use the diagnostic range approach. For instance, the number of higher education students will only be checked against data from the Higher Education Statistics Agency (HESA) (combined with data on further education students).

In such cases tolerances have been set after an assessment has been made of the quality and coverage of the comparator data. Quantifying quality and coverage is inevitably difficult in the absence of a comprehensive independent source (such the census) against which to compare.

Local authority users have been consulted in the process of identifying and quantifying issues of quality and coverage. As part of the Quality Assurance Studies, summaries of eight of the key administrative data sources were sent to the 30 local authorities involved. The authorities were asked to comment on the issues identified and provide evidence of the size of their impacts (ONS, 2010).

As an example of how the quality assessment approach developed, the check of the number of students identified that on HESA data for some records students' term-time address was missing. Tolerances were set for this check taking into account the uncertainty around where these students lived.

Core QA Checks using the Quality Assessment:

- Students
- Ethnicity
- International Migration

C. Set Percentage

In some cases it proved particularly difficult to set tolerances in relation to the quality and coverage of the comparator. Typically these issues were not only difficult to measure but also varied between areas. Where this was the case, it has been necessary to use set percentages above and below the comparator.

Armed forces (home and foreign) checks both use set percentages. Comparator data are available on the number of armed forces personnel stationed in each local authority, but not where personnel are resident. Although an adjustment is made to the comparator to indicate residence, this is based on the 2001 Census and so will inevitably be out of date. Tolerances have been arbitrarily set at +/- 5% around the comparator but will be reviewed during the quality assurance process.

Core QA Checks using Set Percentage:

- Household size
- Home Armed Forces
- Foreign Armed Forces
- Operational checks (Census return rates)

D. Change Over Time

The fourth approach sets tolerances in relation to past trends in comparator data. The approach is used for checks where there is a single comparator data source and a time series of data available.

To use the change over time approach, there was a requirement for the comparator to be particularly well understood. Mortality and fertility rates are both routinely produced by ONS at local authority level, as are estimates of internal migration based on patient registration data.

There was also a requirement for the comparator data to be of high quality. Births and deaths registration data are both of very high quality as registration is compulsory.

Core QA Checks using Change Over Time:

- Mortality Rates
- Fertility Rates

3.4 Additional Intelligence

Alongside the core checks a range of additional information will be available on which to assess the validity of the LA level estimates. This includes additional operational intelligence, area profiles and local authority provided evidence.

Additional operational intelligence

Information from the census and census coverage survey field operation will be reviewed to identify systematic coverage issues which may be reflected in the census estimates. In addition to information on return rates discussed above this may include:

- Notes from field staff debriefs
- Information about field incidents
- Information from the Question Tracking (QT) system

Diagnostics from the data processing stages will also be available such as the amount of data editing and imputation which was carried out.

Area Profiles

Background information has been compiled for each LA to help understand recent demographic, social and economic change. This includes how official estimates of population have changed since the 2001 Census with detail on births, deaths and migration over this period. This is set alongside change in other sources such as patient registers and housing stock. To provide context, the information provided for the LA is set against the wider region and against estimates for England & Wales as a whole.

A summary of correspondence between each LA and the ONS will be available for reference. This includes correspondence on the 2001 Census, mid-year population estimates and other relevant population products such as ethnic population estimates.

Local Authority Provided Intelligence

Local Authority users have been directly involved in the preparation for census quality assurance. This has resulted in LAs providing local intelligence to be used in the process.

To identify local data sources and intelligence, ONS undertook Quality Assurance Studies in 40 LAs where census estimation is likely to be most challenging. Many of the sources identified by users had already been built into the quality assurance process. Two sources not previously identified were Council Tax data with exemption and discount information (relating to second

residences, vacant properties, and properties with only a single adult resident) and Patient Registration data by country of birth. Council Tax and Patient Registration data were important sources ONS had planned to use but the additional locally held detail (exemptions/discounts and country of birth respectively) provided new insights.

Many of the research reports users had undertaken/commissioned and had provided as part of the QA Studies were also recognised as being of potential use.

A final report from the studies was published in February 2011 and listed the local sources and intelligence that were most useful. ONS have subsequently requested this information from all 348 LAs, referring to March 27th 2011.

All LAs have also compiled a Census Local Partnership Plan (CLPP) in conjunction with the census Area Managers which includes data and intelligence of potential use in quality assurance.

3.5 Guidance on Moving to Supplementary QA or Sign-Off

All LA level estimates will be subject to the 'core' quality assurance checks outlined above. Where these comparisons raise questions about the validity of the census population estimates, further 'supplementary' quality assurance will be carried out as detailed in section 4.

Not all LA estimates will be subject to supplementary quality assurance. This section sets how guidance will be set to identify where more detailed quality assurance investigation is required. A more detailed paper on the guidance to be used will be published in July.

Initially, the guidance will be relatively conservative, ensuring that initially a larger proportion of LAs are referred for supplementary QA, than is likely to ultimately be necessary. This will enable ONS to learn more about the process and the comparator datasets. Guidance will be subject to review at a number of points in the process, after supplementary quality assurance checks have been completed an initial set of LAs' estimates. The high-level quality assurance panel will be asked to consider proposals for changes to this guidance, if required.

There will be instances when supplementary quality assurance is carried out for LAs in addition to those identified by the guidance. The main quality assurance panel will be responsible for reviewing the core quality assurance carried out for every LA. During this review the panel may conclude that an estimate requires further analysis.

Why are tolerances by themselves not the most effective way of guiding where supplementary quality assurance is required?

Tolerances will continue to be a useful way of raising questions about the validity of census estimates. However there are a number of reasons why guidance is needed:

- In the prioritisation of work and in the effective targeting of further analysis.
- Some aspects of a check might be more important than others. It might be more important that the number of young men have been identified as having fallen outside tolerance than men aged over 60.[2]
- There is a wider range of quality assurance information available in 2011 compared to 2001. As a result, while a more complete picture is presented, the picture is more complex.
- Guidance would be of value in providing greater transparency in the quality assurance process for users. As already noted, not all LAs will be subject to the more detailed quality assurance.
- Guidance would include aspects of the quality assurance checks (including LA profiles for example), not just those based on the census estimates themselves.

Proposed Guidance Approach

Guidance will initially be set based on whether each of the core checks is highlighting a potential difference.

The three stages below illustrate how the guidance will be used to decide whether supplementary work will be undertaken.

Stage 1 – Identify whether a check is highlighting a potential problem
Stage 2 – Summarise the scores from all checks
Stage 3 – Conclusion on whether supplementary QA is required

Stage 1 – Identify whether a check is highlighting a potential problem

The first stage involves assessing whether the LA census estimate falls inside tolerances. This is done separately for each check.

Confidence intervals around the age/sex census estimates will be used in assessing whether a tolerance has been exceeded. The LA census estimate

[2] Not only do tolerances tend to be wider for young adults (reflecting variation in other comparator sources) but also estimates of the number of people missed by the census tend to be greater for young adults.

will be judged to fall inside tolerance if the upper or lower limit of the confidence interval is within tolerance.

Some checks will be part of a 'family' of checks, whose individual assessments may be combined. For example, checks of the population estimates for each individual age and sex group will be combined, with weighting to reflect the relative importance of each component.

This is illustrated in table 1, where weights have been assigned to the age groups from 0-4 to 25-29. Higher weights have been assigned to the 20-24 and 25-29 age groups.

Having assigned scores for the components, the scores across the components are summed for the LA. Where this score exceeds a set level for the LA, this is flagged as requiring supplementary quality assurance.

Table 1 – Age/Sex Tolerance Weighting

Age Group	Weight	Inside Tolerance (Score) LA 1	Inside Tolerance (Score) LA 2	Inside Tolerance (Score) LA 3
0-4	5	Yes (0)	Yes (0)	Yes (0)
5-9	5	Yes (0)	Yes (0)	No (5)
10-14	5	Yes (0)	No (5)	No (5)
15-19	5	Yes (0)	Yes (0)	Yes (0)
20-24	10	No (10)	Yes (0)	Yes (0)
25-29	10	Yes (0)	Yes (0)	Yes (0)
...				
Total		10	5	10

Three example LAs are shown in table 1. In LA 1 only one age group falls outside tolerance (20-24). As this age group has weight 10, the total score for that check for LA 1 is 10. In LA 2 only one tolerance has been exceeded (10-14) but as this age group has weight 5, the total score is lower at 5. Finally in LA 3, two tolerances have been exceeded (5-9 and 10-14). As these age groups are only weighted 5 the total score for the LA is 10.

Whether the check for the LA is judged to have fallen outside threshold is dependent on the level set for that particular check or family of checks. For this example, if the threshold is set at 10, LAs 1 and 3 would both be flagged as requiring for supplementary quality assurance.

Stage 2 – Summarise the scores from each check

Using this guidance approach all checks can be easily summarised for an LA. This would indicate across all checks where there was evidence to trigger supplementary QA.

An example summary for an LA is provided in table 2. This shows that potential problems are flagged for two checks in LA 1 (the red boxes) and one for LA 3.

The table 2 summary shows that LA 2 has no red boxes, but it does have an amber box for the age/sex check.

Table 2 – Summary of Six Checks in Two LAs

	Checks Against Comparator Data			Demographic Analysis		Operational Intelligence
LA	Age/Sex	Students	Armed Forces	Mortality	Fertility	Return Patterns
1	Red	Green	Green	Green	Green	Red
2	Amber	Green	Green	Green	Green	Green
3	Green	Red	Green	Green	Green	Green

Stage 3 – Conclusion on whether supplementary QA is required

The final stage will be to combine the results of all indicators to decide whether supplementary QA is required.

Initially, a conservative approach will be taken, so that more LAs than necessary will have supplementary QA, so that the process and sources and can be evaluated more fully. The detailed guidance is yet to be finalised, but initially it might be along the lines that any LA with at least one red check or with multiple amber checks will be go forward to supplementary QA for example.

After an initial set of LAs have been through supplementary quality assurance, the guidance process will be reviewed, and a refinement is likely to be made at this point.

4. Supplementary Quality Assurance

Where the core checks at LA level have identified that further analysis is required, more detailed supplementary quality assurance will be carried out. The further analysis to be undertaken will depend on which of the core checks were highlighting potential problems. It will include analysis of specific geographic areas within an LA and analysis of specific population subgroups.

This section provides an overview of supplementary analysis. A more detailed paper on methods and sources, particularly involving the use of data matching will be published in July 2011.

4.1 Analysis of Population Subgroups

During the development of the core checks, additional analysis was identified which could be used to help further explain why questions were raised. Examples of the supplementary analysis identified are provided below. Further analysis will inevitably be identified as the preparation for the census continues and during the QA operation.

Households & Population by Age/Sex

- Number, type and distribution of second residences identified on returned census questionnaires (where no-one is usually resident) and on dummy forms.
- Number, type and distribution of second residences identified by individuals who state that they have a second residence within the LA, based on census questionnaires completed at a different usual residence. (Not all second residences identified in this way will be available until all questionnaires have been processed.)
- The above information collected from census forms will be matched to assess over-count, the accuracy of dummy form information and the accuracy of questionnaires returned from second residences. Accuracy will also be assessed against second residences and vacant properties identified on local authority provided Council Tax data.

Students

- International students (including short-term migrants) and domestic students reviewed separately (Higher Education Statistics Agency)
- Second (i.e. term time) address information for domestic students completed at their parents' home address

Fertility

- Male fertility rates
- Fertility rates by country of birth

Ethnicity

- Language spoken (School Census)
- Country of birth (School Census, Annual Population Survey, Patient Register)
- Short-term migration (International Passenger Survey)

Armed Forces

- School age children with parents in the armed forces (School Census)
- Households including armed forces personnel (Council Tax)

Internal Migration & Population by Age/Sex

- Address twelve months ago identified on census returns in comparison to current address on administrative data – indicating that administrative data have not been updated.

4.2 Analysis of Specific Geographic Areas

The accuracy of census population estimates at local authority level will have previously been assessed as part of the core check work carried out for all areas. Estimates will also be checked at LSOA and MSOA as part of this work. To fully understand differences between census estimates and comparator data it will in some cases be necessary to undertake supplementary analysis at even lower geographic levels. In some cases this will involve matching administrative data to census records.

Following the 2001 Census a programme of detailed LA studies were carried out on 32 census population estimates. This involved matching studies in Manchester and Westminster. Findings from the 2001 Census LA studies were published in 2004 alongside resulting adjustments to mid-year population estimates. Building on what was done in 2001, the 2011 QA process will incorporate detailed studies ahead of the planned first release in July 2012.

Analysis of specific geographic areas involves two stage, firstly aggregate analysis for Census Coverage Survey (CCS) postcode clusters, and secondly unit record matching of individuals and households.
a. Analysis of CCS postcode clusters

Census estimates for the postcode clusters included in the sample for the CCS will be subject to rigorous quality assurance. It is in these postcode clusters that Dual System Estimates (DSEs) will be most robust. The estimates are particularly important as they are used to adjust for under-coverage across the Estimation Area (EA).

The census population estimates for each of these postcode clusters, and the distributions of key characteristics such as age and sex, will be compared with aggregate statistics from each administrative dataset for the same postcode cluster. This analysis, carried out for individuals and households, will make use of confidence intervals around the census estimates.

A fuller explanation of the CCS postcode cluster analysis will be published as part of a more detailed paper on use of low-level aggregate comparison and data matching to be published in July 2011.

b. Unit Record Level Matching

Record-level matching between Census and administrative sources will help to validate Census population estimates. Initially this work will focus on CCS postcode clusters, for which ONS has the most comprehensive snapshot of the Census-day population and which are used to adjust for undercount across estimation areas. However, the same approach may be applied to non-CCS areas where Census counts pose QA challenges.

Address matching between Valuation Office Agency data, Electoral Rolls, the Patient Register, School Census, CCS address lists and the Address Register will identify addresses that are held in common or are unique. This will involve both automatic matching and the clerical resolution of potential matches and unmatched residuals. A set of rules has been devised to identify addresses that administrative sources suggest are occupied but which have been missed by Census processes. These rules will be reviewed in the light of analysis of matched records that include both administrative and Census information.

Individual person matching between Census, CCS, Patient Register, School Census and Higher Education Statistics Authority data will create additional evidence based on administrative sources for validating Census/ CCS matches and dual system estimates. This analysis will also draw on newly-collected information in the 2011 Census on second residents, visitors, students' usual residence and usual address one year ago to help to account for records in administrative sources that have no corresponding Census record at the same address. Adjustments will be made for short-terms migrants. Analysis of the linked data may also provide evidence of within-household under-enumeration.

5 Improvements to the census population estimates

5.1 Introduction

Section 4 outlined the further, or supplementary, analysis that will be undertaken should the guidance indicate that additional investigation is required. This section provides an overview of the improvement options that are available should this analysis conclude that the census estimates are too low or high and that improvements are required.

One, or a combination, of the following improvements is available:

- Revisit coverage estimation (post-stratification by alternative characteristics, post-stratification by re-grouping estimation areas, borrowing strength from other LAs).
- Adjust coverage estimation to account for between- and within-household bias.
- Calibrate using external data (national adjustments, direct calibration for small geographic areas/population sub-groups).

The improvement to be applied will depend on the nature of the problem and will be heavily informed by the intelligence gained during the core and supplementary analysis, particularly our understanding of the different data sources and their quality.

Some of the improvements described in this section are possible as a consequence of Census methodology that is new in 2011, such as the capture of second addresses and visitor information. Others draw on the lessons learned from 2001 Census QA and the subsequent Local Authority Studies, such as the use of demographic indicators, particularly calibration to an alternative national age/sex ratio.

5.2 Revisit coverage estimation

Where local authority estimates cannot be approved following QA, depending on the nature of the issue(s) to be resolved, one option is to assess whether the DSE can be improved.

5.2.1 Dual system estimate post-stratification by alternative characteristics

Estimation of non-response using age by sex distributions within hard-to-count areas may leave residual bias that the dual system estimation does not correct. Other characteristics such as ethnicity or tenure, or a combination of variables, may be more powerful predictors of response and thus give a better dual system estimate.

During census data processing, predictors of non-response will be monitored as they emerge in the accumulating data (as described in section 3.2). This modelling will identify potential alternative variables for different estimation area types, which could be applied if the approach based on estimating for age/sex groups generates anomalous results.

5.2.2 Estimation area post-stratification

Estimation areas consist of either single LAs or groups of contiguous LAs. Patterns of undercount could vary significantly between local authorities in an estimation area, which could generate under- or over-estimation for an LA. Alternative indicators could be used to group LAs into estimation areas, such as return rates or the ONS LA classification. The estimation methodology would than be applied to the re-formed estimation areas.

5.2.3 Borrow strength from other local authorities

In some areas, the Census Coverage Survey (CCS) may not collect sufficient data to enable a robust estimate of coverage. This might be due to low response rates in the CCS, insufficient capture of people missed by the Census or as a result of data processing problems.

In these instances, it may be appropriate to borrow strength from other areas in place of the CCS information collected. This can involve either borrowing strength for a single LA within an Estimation Area (where it consists of more than one LA) or borrowing strength for an entire Estimation Area. The

method was used to adjust the 2001 census estimates in a number of LAs. (ONS, 2003).

Where it is necessary to borrow strength for an LA within an Estimation Area, this will be taken from the other LAs in that Estimation Area. In effect this will adjust the census count for the LA in line with adjustments made for surrounding LAs. If it is necessary adjust for an entire Estimation Area, strength will be borrowed from similar LAs to each LA in question. LAs are similar based on their socio-economic and demographic characteristics.

It is the ratio estimator which is itself adjusted when borrowing strength. The ratio estimator is essentially the adjustment factor used to calculate the census estimate from the census count in each Estimation Area. If the original ratio estimator for an LA was 1.2 and the census count was 4,000 then the census estimate would be 4,400. By borrowing strength across the five most similar LAs, a ratio estimator might be calculated as 1.3. Hence the LA estimate of interest becomes 4,600 (1.3*4,000).

5.3 Adjust Coverage Estimation for Bias

Dual System Estimation (DSE) used in coverage assessment and adjustment is based on a series of assumptions which if violated may result in biased estimates of the population. One of the critical assumptions made is that the probability of being counted in the CCS is independent of the probability within each stratum of being counted in the Census. Independence is assumed both for households and individuals within households.

5.3.1 Between Household Bias Adjustment Using the Alternative Household Estimate

To correct for any potential between household bias an Alternative Household Count will be produced and routinely used to adjust the DSEs in each Estimation Area.

The alternative count of the number of households is compiled from:

- The number of returned household census questionnaires which are occupied.
- An estimate of the number of occupied household questionnaires which did not respond to the census. This includes households identified as occupied (through a dummy form or where a blank household questionnaire was returned) and an estimate of how many households were occupied where no return or dummy was received.

The alternative household count will be produced for each CCS postcode cluster so that DSE adjustments can be applied. In addition the alternative household count will be produced for the total number of occupied

households for the LA as a whole. This will be used in the core QA process as set out in section 3.1.

5.3.2 Within Household Bias Adjustment Using the Census Non-Response Link Study (CNRLS)

As described in section 3.2, the Census Non-Response Link Study (CNRLS) will be used to assess where census returns missed individuals. This is done by matching census returns to recent returns from ONS social surveys.

Where there is evidence at either the regional or national levels of systematic under-coverage this will be adjusted through the DSEs.

5.4 Calibrate the estimates using external data.

Estimates of the total population, population sub-groups (e.g. students) and demographic rates all exist in sources independent of the census. There are quality and comparability issues with alternative sources to the census population estimates that prohibit a comprehensive direct calibration to these sources. However, there will be instances, such as localised problems, where alternative sources (or combinations of alternative sources developed through data matching) clearly show a real discrepancy with the census estimates.

A number of different options for a calibration improvement are set out in the following sections.

5.4.1 Apply national sex ratios by age

Evidence of a systematic under or over estimation in men or women would be needed in order to implement an adjustment that would bring the sex ratios in line with the independent source(s). In addition, there would also need to be evidence that indicated whether the adjustment was equally dispersed across England & Wales or whether it was geographically clustered in some LAs only.

Methods for distributing any England & Wales level improvement (including any Longitudinal Study based improvement as set out in 5.4.2) would be assessed once the size of such an adjustment was identified. This would need to take into account the extent to which any systematic under or over estimation was concentrated in certain types of area. Research on how a national distribution should be applied has explored distributing in relation to population size and in relation to the level of local coverage adjustment. A further paper will be published describing the distribution method.

Once apportioned to LA level, improvements would be introduced through the ratio estimation process element of coverage adjustment. The ratio estimator would be adjusted (or calibrated) for the appropriate age/sex strata so that a rerun of the estimates would be adjusted by the appropriate amount.

5.4.2 Revise coverage adjustment using evidence from the ONS Longitudinal Study

Following the 2001 Census, evidence from the ONS Longitudinal Study (LS) identified a shortfall in the Census estimation of young men. This was subsequently used to adjust 2003 mid-year population estimates. The method will also be available to assess and, if necessary, adjust the 2011 Census population estimates. Section 3.1 outlined how the LS would be used as a core check for missing or double counted people at the national level.

5.4.3 Use external sources for direct adjustment

5.4.3.1 Direct adjustment for small geographic areas

Some population sub-groups could be partially or completely missed as a result of deliberate non-compliance in both the census and the CCS. Where omissions are identified, survey or administrative sources may be used either to add to the population estimates or to change imputed characteristics.

As identified in section 3, there are coverage and definitional differences between administrative sources and Census. Similarly there are inconsistencies between the administrative sources themselves.

In most instances a single administrative source cannot be used in isolation to replace a census estimate at any geographic level without introducing greater uncertainty.

During development of the Census QA methodology ONS have been carrying out a series of administrative matching studies. This work is reviewing issues in coverage and synchronicity in a number of administrative sources at both individual and household level. As set out in section 4.2 the work will be extended to include 2011 Census counts and estimates and a further paper will be published.

By matching across sources it may be possible to compile counts from across administrative based sources against which the census can be calibrated if required for small geographic areas. Data matching (and its interpretation) is inevitably resource intensive and so limits how the approach can be used across all LAs.

5.4.3.2 Direct adjustment for population sub-groups

A small number of administrative datasets exist against which it is possible to calibrate sub-groups of the census population estimates. The independent review of quality assurance and coverage adjustment recommended that these sources should be identified and improvements routinely made where required following quality assurance.

Direct adjustments will be made using:

(i) School Census data to adjust estimates of school-age children.

School Census data is recognised as a particularly accurate source against which to potentially improve census estimates. While only children in the state school sector are included (children in independent schools and those who are home-schooled, for example, are not covered) the source provides a lower bound for 5 to 15 year olds against which census estimates can be calibrated.

(ii) Patient Register data to adjust estimates of babies under one year of age.

The Patient Register count of babies under one year of age is likely to be accurate and therefore may be used in calibration. At other ages there is greater uncertainty in the Patient Register as individuals are not required to register when they move (either within the UK or when emigrating overseas) and because (for young adults in particular) more moves take place.

Compulsory birth registration data is of very high quality but will not be updated to include changes of address. This data will be used in the quality assurance process and so would have been assessed prior to any adjustment being made.

(iii) Administrative Data on Residents of Individuals in Large Communal Establishments

Census counts for large communal establishments (with more than 100 bed spaces) will be calibrated to those administrative sources used in their quality assurance (section 3.1). Where there is evidence of census under-enumeration and no administrative data is available, direct contact will be made with the establishment.

Large communal establishments are not covered by the CCS and so are not included within the coverage estimation process. As a result the large communal establishment population is not routinely adjusted for under-enumeration as is the case for the household population (and residents in smaller communal establishments).

A number of administrative sources are available with which to quality assure the different types of large communal establishments (student hall of residence, armed forces base, prison, boarding school, nursing home). Quality assurance will also take into account the number of census forms given out and the number returned.

6 Signing off the 2011 Census Population Estimates

6.1 Introduction

The production of the 2011 census population estimates includes a number of automatic adjustments for bias and much more focus on identifying, and improving, shortfalls in the estimation process than in 2001. As a result it is essential that the process for approving these estimates is robust, transparent and has credibility with users.

The approach outlined below builds on the 2001 approach and in particular introduces additional layers of accountability and review. Although the basic approach for 2001 worked well, there was clearly a need to have more external involvement and a separate assurance process that focussed on emerging issues at pre-defined periods before publication.

6.2 Approach to governance and sign-off

The approach for 2011 has been developed based on a number of key principles:

- clear accountability for signing off the census estimates for LAs and at higher geographies
- clear accountability for identifying emerging national issues or systematic biases affecting a large number of areas
- clear accountability for commissioning additional analysis and invoking options for improvement
- Involvement of external experts to ensure that decisions on the estimates are robust and credible
- Involvement of senior managers and the National Statistician in governance and final sign off

Considering these principles, the 2011 approach utilises a hierarchy of four groups, which review and endorse the 2011 Census population estimates, resulting ultimately in sign-off by the National Statistician.

1. **Internal QA working group** - A working group, meeting daily, which will: review the outcome of the core QA; using the checks described in section 3, decide whether supplementary QA is required prior to estimates going to the main QA panel; and, if necessary, recommend any options for improving the estimates. This will reduce the burden on the main QA panel and streamline the agreement of the estimates.

2. **Main QA panel** - an internal ONS Panel (including Welsh Assembly Government representation) which will focus on assessing census population estimates at LA level. The group will meet on a weekly

basis and be presented with a suite of material and evidence needed to formally review the census estimates for that LA. The panel will either:

- Endorse the estimate and recommend approval to the High Level QA panel.
- Request further research for areas where the panel does not endorse the estimates.

3. **High level panel** - a group of internal, UK (i.e. from census offices in Scotland and Northern Ireland) and independent external reviewers which will provide expertise and guidance on the emerging national picture. The panel will meet every six to eight weeks and will: comment on the emerging national picture; advise on emerging issues; and review/endorse key decisions made by the QA panel.

   A key focus of the group will be to look at areas where estimates were rejected by the main QA panel and subject to subsequent research. They will challenge any areas of concern and commission additional analysis from the census QA team and invoke options for improvement where necessary. The group will report progress and interim findings to the Executive panel on a three monthly basis.

   At the end of the process the high level panel will make a recommendation to the Executive panel to sign off the regional and national estimates. It will also be asked to endorse the recommendation made by the main QA panel to sign off each of the LA level estimates.

4. **Executive panel –** An executive panel, chaired by the National Statistician, accountable for the final sign-off of the national and local census population estimates ahead of publication. The Executive panel will consider the recommendations of the high level panel, explicitly considering any areas of disagreement highlighted between high level panel members.

6.3 Membership of the panels

The table below gives an indication of the membership of each panel and the different levels of expertise that this brings to the QA.

Membership	QA working group	Main QA panel	High level QA panel	Executive panel
Census	✓	✓	✓	✓
ONS Population and Demography Division	✓	✓	✓	✓
ONS Methodology Division	✓	✓	✓	✓
ONS Centre for Regional and Local Statistics	×	✓	✓	×
Welsh Assembly Government	×	✓	✓	×
ONS Surveys and Administrative Sources Division	×	×	✓	✓
National Records of Scotland (formerly GROS)	✓	×	✓	✓
Northern Ireland Statistics and Research Agency	×	×	✓	✓
Subject matter experts (as required)	✓	✓	✓	✓
External members	×	×	✓	×
National Statistician	×	×	×	✓

The role of each of the members is to bring a different aspect of experience and expertise to the assurance and sign-off of the estimates.

2011 Census

- Expertise in 2011 census operation and quality issues identified in the field.
- Expertise in 2011 census processing and quality issues identified with each processing stage.
- Expertise in 2011 methodologies used in processing and contingency options (specifically with coverage estimation/adjustment).
- Expertise in working with comparator datasets at aggregate level and in data matching.
- Experience from 2001 quality assurance process.

ONS Population and Demography Division

- Expertise in population and migration estimation.
- Expertise in working with administrative and survey data sets used to measure population and migration.
- Expertise in demographic analysis specifically mortality and fertility analysis.
- Experience from 2001 quality assurance process.

ONS Methodology

- Expertise in 2011 statistical methodologies for edit and imputation, coverage estimation/adjustment and disclosure control.

- Expertise in available adjustment methodologies (specifically with coverage estimation/adjustment)
- Experience from 2001 quality assurance process.
- Expertise gained from providing input to the development of the 2011 QA methods.

ONS Centre for Regional and Local Statistics

- Expertise in local and regional area analysis.
- Experience from 2001 quality assurance process.

Welsh Assembly Government

- Expertise in the demographic issues relevant to Welsh LAs.
- Expertise in administrative, surveys and other population products in Wales.
- Expertise gained from providing input to the development of the 2011 QA methods.

ONS Surveys and Administrative Sources

- Expertise in survey and administrative sources used in the QA process.

National Records of Scotland (formerly GROS)

- Experience of the 2011 Census operation in Scotland (including the QA process).
- Expertise in administrative, surveys and other population products in Scotland.
- Expertise gained from providing input to the development of the 2011 QA methods.

Northern Ireland Statistics and Research Agency

- Experience of the 2011 Census operation in Northern Ireland (including the QA process).
- Expertise in administrative, surveys and other population products in Northern Ireland.
- Expertise gained from providing input to the development of the 2011 QA methods.

Subject matter experts

Additional subject matter experts will be called to attend a panel where their expertise would help illuminate particular problems, either the main QA panel or the high level. This could include, for example:

- census staff with particular insight into particular operational issues, e.g. field ops team
- statistical or demographic consultants from Southampton University Centre for Population Change
- experts with a particular geographic expertise, including a particular LA where there may be significant issues that are difficult to resolve.

6.4 Selection of external members of high level panel

In addition to the ONS employees and UK representatives, external experts will also be appointed to sit on the high level panel, providing both expertise and an element of external assurance to users.

External membership will be required to cover expert perspectives on the following topics that are critical to the assurance of the census population estimates:

1. An understanding of the statistical methodology for coverage assessment and adjustment, and wider census methods
2. National demography, including mortality, fertility and sex ratios
3. Migration and cultural diversity
4. London's population, including individual LA perspectives and urban issues more widely
5. Good understanding of the aims and objectives of the census design, including the enumeration and output population bases
6. Understanding of wider census user perspectives

In selecting and appointing external members it will be important that they are well known within their field, highly regarded by census users and that they demonstrate the necessary impartial, independent expertise required for the process.

Census will work with relevant external bodies to identify individuals that cover these areas of expertise as candidates for external membership.

References

No	Paper	Weblink
1	Abbott, O. (2010) '2011 Census Coverage Assessment and Adjustment Methodology'	http://www.ons.gov.uk/census/2011-census/process-info/statistical-meth/coverage-assessment-guide.pdf
2	Smallwood, S and de Broe, S (2009) 'Sex ratio patterns in population estimates'	http://www.statistics.gov.uk/downloads/theme_population/PopTrends137web.pdf
3	Abbott, O. (2009) 'Visitor Sampling and Matching', version 0.2 presented to the UKCAAWG 25.03.2009, ONS.	Available on request
4	Abbott, O. (2007) 2011 UK Census Coverage assessment and adjustment strategy. Population Trends, 127, 7-14.	www.statistics.gov.uk/downloads/theme_population/PopulationTrends127.pdf
5	Abbott, O. and Large, A. (2009)	Available on request

	'Measuring the level of duplicates in the 2011 Census', paper presented to the 15th meeting of the GSS Methodology Advisory Committee.	
6	White, N., Abbott, O., and Compton, G. (2006) 'Demographic analysis in the UK Census: a look back to 2001 and looking forward to 2011'. Proceedings of the American Statistical Association	Available on request
7	Abbott, O., McLeod, P., Baillie, M. and Brown, J. (2010) 'Correcting for bias in the DSE-Constructing an alternative household estimate', Titchfield: ONS.	Available on request
8	Blackwell, L., Lynch, K., Smith, J. and Goldblatt, P. (2003) 'Longitudinal Study 1971-2001: Completeness of Census Linkage', London: ONS.	www.statistics.gov.uk/downloads/theme_population/LS_no10.pdf
9	Brown, J. and Abbott, O. (2008) Response to Discussion at NSMAC on Estimation of Over-Count. Note for 14th Meeting of the National Statistics Methodology Advisory Committee.	www.statistics.gov.uk/methods_quality/downloads/Estimation_overcount_followup.doc
10	Brown, J., Taylor A. and Abbott O. (2009) Overcount in the 2011 Census: Estimation Issues. Paper presented to 17th Meeting of the GSS Methodology Advisory Committee.	www.statistics.gov.uk/methods_quality/downloads/GSSMAC17.pdf
9	Cruddas, M. and Diamond, I. (2000) 'A Quality Assurance and Contingency Strategy for the One Number Census', London: ONS.	See also: www.statistics.gov.uk/census2001/pdfs/oncinfopaper.pdf
11	Office for National Statistics (2004) 'Local Authority Population Studies: Full report' London: ONS.	www.statistics.gov.uk/downloads/theme_population/LAStudy_FullReport.pdf
13	Office for National Statistics	www.ons.gov.uk/census/2011-

	(2010) 'Evaluation of the Census 2011 Quality Assurance Studies project'	census/process-info/data-quality-assurance/
14	Office for National Statistics (2003) 'One Number Census Quality Assurance Information: Quality Assurance Themes Borrowing Strength'	http://www.statistics.gov.uk/census2001/pdfs/borrowing_strength.pdf

Appendix A - Glossary of Sources Used in the QA Process

Compulsory Registration

- Birth registration

Compulsory registration of births in England & Wales available at record level. QA will also use age of mother (to calculate age-specific fertility), age of father (to calculate male fertility), and mother's country of birth.

- Deaths registrations

Compulsory registration of deaths in England & Wales at record level. Age-specific mortality rates use age at death.

Administrative Sources

- NHS patient register

Registration with a GP in England & Wales available at record level by age and sex. Anyone resident in the UK for longer than three months can register but registration is not compulsory, nor is it compulsory to re-register when changing address. Where individuals do change their address, it is assumed re-registration occurs a month later.

- NHS patient register (provided by LAs)

A small number of LAs have been able to provide patient registration data by country of birth (not at record level). These data were received in response to a request made to all LAs in December 2010 following the QA studies.

- Department for Education (DfE) and Welsh Assembly Government (WAG) school census

Children in maintained schools aged 5 to 15 by their usual residence, single year of age and sex. Record level data for England provided by DfE, data for Wales provided by WAG (pending establishment of legal gateway). Excludes children in independent schools and children who are home educated.

- Higher Education Statistics Agency data

Students at higher education institutions by term-time address. Available at record level by single year of age and sex. Separately identifies students in halls of residence and residential accommodation, overseas students, and students on courses lasting less than a year. Does not include students in further education, language schools or private universities.

- Department for Business Innovation & Skills (BIS) and Welsh Assembly Government (WAG) Further Education data

Students at further education institutions by home address. Available at Middle Layer Super Output area by single year of age and sex. Aggregate level data for England provided by BIS, data for Wales provided by WAG

- HM Revenue and Customs (HMRC) child benefit data

Children whose parents are in receipt of child benefit by residential address. Available at LA level and by single year of age and sex.

- Department for Work and Pensions (DWP) pension claimants data

Individuals in receipt of state pension by residential address. Available at LA level and by single year of age (over 65 for women and 60 for men) and sex.

- DWP Lifetime Labour Market Database

A one per cent sample of national insurance numbers together with all activity from DWP and HMRC systems. This allows the resident population and migrant flows to be inferred from activity. ONS only has access to aggregate information from this source.

- DWP/HMRC Customer Information System

The Customer Information System (CIS) holds information on both DWP and HMRC clients. Updated from a range of sources including the Child Benefit System, National Insurance & PAYE, DWP Benefit Systems & State Pensions, Tax Credits. Available by single year of age and sex. Aggregate data (postcode sector data) supplied in November 2010 and is currently being assessed for use in Census QA. Access to record level data dependent on the establishment of a legal gateway.

- Ministry of Justice Prisons data

Number of prisoners by establishment and by single year of age and sex.

- Defence Analytical Services Agency

Number of UK armed forces personnel by base rather than residence. Available by single year of age and sex at local authority level.

- US Armed Forces data

Number of United States armed forces by base rather than residence. Available by single year of age and sex at local authority level.

- Council tax data (provided by Valuations Office Agency)

List of addresses maintained for the purpose of council tax bandings and non-domestic rates. Excludes commercial only addresses and contains no discount or exemption information for the addresses listed. Available at record level.

- Council Tax data (provided by LAs)

List of residential addresses including discount and exemption information. Requested following QA studies from all LAs at postcode level.

DCLG/ONS Population Products

- DCLG household projections

Projected number of households and average household size at LA level consistent with ONS sub-national population projections.

- ONS mid-year population estimates

Annual population estimates at LA level by single year of age and sex. Previous census is aged on annually and adjusted for the number of births/deaths and net migration. Lower and Middle Layer Super Output area estimates also available by age and sex.

- ONS Population Estimates by Ethnic Group

Annual population estimates at LA level by broad and fine ethnic groups. Ethnicity distribution based on 2001 Census distribution. Available at LA level by age and sex.

Survey Sources

- Integrated Household Survey (IHS)

The IHS is a composite household survey combining the answers from a number of ONS household surveys to produce an experimental dataset of core variables. Estimates on a range of population sub-groups are available.

- International Passenger Survey (IPS)

The IPS is a survey of a random sample of passengers entering and leaving the UK by air, sea or the Channel Tunnel. Identifies international immigrants and emigrants by age and sex.

2011 Census Sources

- Addresses identified as second residences / holiday home

The 2011 Census asks (question H1) whether an address is a second residence or holiday home that no-one usually lives in. It also asks (question H4) whether individuals recorded on the form are staying at that address because it is their second address.

- Census Address Register

Underpinning the 2011 Census is an up-to-date address register compiled from a range of administrative address sources. The register will be updated up to and beyond census day with addresses identified during the operation.

- Information recorded on dummy forms about second residences

Where a completed census is not returned, enumerators will attempt to make contact. If contact cannot be made (or if an individual refuses) a dummy form is completed by the enumerator who will assess whether the address is occupied, is vacant or is a second residence at which no one is usually resident.

- Census Communal Establishment list – expected number of residents

To identify how many forms to send, preparations for the census included establishing an expected number of residents for each communal establishment.