Quality assuring the 2011 Census population estimates

Jonathan Wroth-Smith, Owen Abbott, Garnett Compton and Peter Benton

Office for National Statistics

Abstract

The 2011 Census is a once in a decade opportunity to gain an accurate picture of the population. Quality assurance of the census population estimates is vital to ensure data robustness and that users have confidence in the results. This article provides a detailed summary of the methods being used to achieve this quality assurance, including the processes and adjustments.

Introduction

The 2011 Census is a once in a decade opportunity to get an accurate, comprehensive and consistent picture of the most valuable UK resource – its population – and a rich array of facts about it. The process of quality assuring the census population estimates is vital to ensure that they are robust and that users have confidence in the results.

This article 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. It focuses on the assurance of the census population estimates and the key population subgroups and housing characteristics associated with producing the census population estimates. Assurance of the other census characteristics, such as the labour market questions, will also be conducted.

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 Office for National Statistics (ONS) surveys. The CAA process estimates and adjusts for households and people missed in the census. More detail on the CAA process is set out in Population Trends (see Abbott, 2009)[1].

Building on 2001

The methods set out here build on the lessons learned from the Quality Assurance of the 2001 Census population estimates (see White, Abbott and Compton, 2006)[2] and the subsequent Local Authority (LA) Studies (ONS, 2004)[3]. The key lesson from the 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 is 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 include comparisons at geographic levels below LA.

Therefore, rather than have extensive discussions with LAs after the publication of the 2011 Census results, ONS is already engaging with a number of LAs to determine which locally held sources might be useful in the QA process. During 2010, ONS carried out QA studies with 40 LAs where census estimation is likely to be particularly challenging.

The data and intelligence provided as part of the 2010 QA studies have helped refine the QA process. The findings from these QA studies were published in December 2010 (ONS, 2010)[4]. All LAs have been asked to supply the administrative data sources identified as most useful in the QA studies' findings as at 27 March 2011 (census day).

Quality Assurance Overview

An overview of the QA process is shown in **Table 1**. The process is organised at three geographic levels (local authority, regional, and England and Wales) and by three stages (core checks,

supplementary checks, and options for subsequent improvements). Further information on the core and supplementary checks, and details on potential improvements are summarised in sections below.

Table 1 **Quality** **Assurance Process**

Level	Core Checks	Supplementary Checks	Improvements
Local Authority	• Checks against comparator sources • Operational Intelligence • LA provided evidence	• Low level aggregate comparisons • Additional characteristic checks • Comparison to alternative household count[5]	• Calibrate to external sources • Adjust CAA • Adjust using alternative household count
Regional	• Checks against comparator sources • Operational Intelligence • LA provided evidence	• Validation using visitor and second residence information • Assessment of Census Non-Response Link Study (CNRLS)	• Calibrate to external sources • Adjust CAA
England & Wales	• Checks against comparator sources • Operational Intelligence • LA provided evidence	• Validation using visitor and second residence information • Assessment of Census Non-Response Link Study (CNRLS) • Assessment of Longitudinal Study evidence	• Calibrate to external sources • Adjust CAA • Adjust using CNRLS • Adjust using Longitudinal Study

Core and Supplementary Quality Assurance

Core Quality Assurance Checks at the LA level

To quality assure census estimates it is necessary to draw comparisons with other data sources which provide alternative estimates of population groups and their structure and characteristics.

Quality issues with comparator sources mean that they will not provide a precise estimate of the population being measured by the census (see ONS, 2007)[6]. This imprecision will vary by LA as well as by age, sex, and other key characteristics. As a result the QA will not be checking that the census replicates the estimates from alternative sources, but will be using these to assess whether the census estimates are plausible using pre-defined tolerance ranges. These ranges build on the diagnostic range approach used to check age/sex estimates in the 2001 Census (see ONS 2001)[7].

Every LA will go through a set of core QA checks. The initial census population estimates for each LA will be compared with other existing sources, including mid-year population estimates and administrative datasets. Different sources will be used to assess the numbers and key characteristics of different population groups (see **Table 2**).

Table 2   **Sources used to assess the numbers and key characteristics of different population groups**

Core Check	Comparator Dataset	Check Approach
1. Under 1s	• Birth registration	• Number of births registered in the 12 months prior to census day compared with under 1s enumerated
2. Age and Sex	• Patient Register • Mid-year Population Estimates • School Census • Central Information System or Child Benefit/Pensions data	• Quinary age and sex • Absolute values and proportions • Diagnostic range tolerances capped to avoid implausibly large ranges
3. Number of Households and Average Size	• Council Tax • Address Register • Patient Register • Communities and Local Government household estimates	• Number of households by occupied and non-occupied. • Identification of dummy form responses. • Distribution of average household size. • Tolerance methods currently under development.
4. Ethnicity	• Population Estimates by Ethnic Group • Integrated Household Survey • School Census • Independent Schools data from Department for Education	• Broad and detailed ethnic group by sex and quinary age group • Diagnostic range where multiple comparators allow • Tolerances based on School Census data only are currently under development
5. Students	• Higher Education Statistics Agency (HESA)	• Students over 18 by age and sex • Students in communal establishments

	• Further Education Student Numbers from Business, Innovation and Skills (BIS)	• Tolerances relative to the quality of HESA data and the number of further education students as a proportion of the LA population
6. Armed Forces (Home/Foreign)	• Defence Analysis Statistics Agency (DASA) • US Armed Forces	• Age and sex based checks • Tolerance methods currently being developed
7. Migration (Internal)	• Patient Registration	• Age and sex based checks • Tolerance methods currently being developed
8. Migration (International)	• Patient Registration • International Passenger Survey • National Insurance Numbers	• Age and sex based checks • Tolerance methods currently being developed

In addition to checks against mid-year estimates and administrative sources, other analysis will be done as part of the core QA. An important element will be demographic analysis focusing on:

- comparing sex ratios by age group with administrative datasets
- comparing fertility and mortality rates, using the census estimates as the denominator, with rates derived using other sources (for example, the ONS mid-year estimates)

Additional key indicators have been built into the census data collection to help explain inconsistencies with administrative datasets. These will be assessed as part of the QA. They include information on second residences, short term-migrants and vacant properties.

As well as analyses at LA level, analyses on estimated population totals at Lower Super Output Area (LSOA) and Middle Super Output Area (MSOA) levels will be performed. These will involve:

- comparison with administrative data sources to identify outliers requiring further investigation, and
- comparison of the number of households in each MSOA with the total number of addresses on the address register

When analysing and reviewing the core checks it will also be important to understand key census field operation indicators and how the area has changed since the 2001 Census. These include:

- census and Census Coverage Survey (CCS) return rates in comparison with expected return rates and those achieved in 2001
- field information will be reviewed to identify any systematic coverage problems, including questionnaire tracking information and calls to the census contact centre
- background information will be collected for each LA and used within the QA. This includes a record of previous correspondence, reviews relating to the LA's population (including the post-2001 LA studies), the census QA studies, and information collected via census area managers about particularly challenging population or housing types

The main QA panel (see section on Publication of the results below) will consider a compiled set of core QA checks for every LA and findings from supplementary work undertaken. Ahead of estimates being referred to this panel, pre-defined guidance as set out below will be used to identify whether supplementary work is required.

Guidance on moving to supplementary analysis

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 the next section. Not all LA estimates will be subject to supplementary quality assurance.

Guidelines for moving to supplementary quality assurance are needed:

- to prioritise work and the effective targeting of further analysis
- to provide a more complete picture in complex cases, where there is a wider range of quality assurance information available in 2011 compared with 2001
- to provide greater transparency, and thus confidence, in the quality assurance process for users

The guidance for moving to supplementary analysis will be based on:

- the outcome of the core checks and where the census estimates fall outside the diagnostic or tolerance ranges for a number of key checks. These instances will be weighted to reflect the most important checks and where we have most confidence in the comparator sources (for example, birth registration or data on students from HESA)
- how well the census and CCS performed in an LA. This will take account of census return rates, CCS response rates and the size of the CCS sample

Supplementary Quality Assurance at the LA level

Where the core LA checks raise questions, supplementary more detailed analysis will be undertaken. The focus of this supplementary analysis will be an examination of the whole population at very low geographic levels or further investigation of particular population sub-groups at LA level.

Low level comparison will focus on the postcode clusters included in the sample for the Census Coverage Survey in each LA. It is in these postcodes that the Dual System Estimate (DSE) produced during the coverage assessment process will be most robust (the DSEs for these postcode clusters are the basis for estimating under-coverage for the whole Estimation Area[8]). Investigations 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 datasets for the same postcodes. Comparisons will be made for both individuals and households
- household size within these postcode clusters with administrative sources, to identify where large households may not have been adequately adjusted for in the estimates

Initial core checks may raise questions about estimates of population sub-groups. Examples of supplementary checks to be undertaken include investigation into the number of foreign or short-term migrant students to explore questions raised about total numbers of students.

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.

Should this supplementary analysis at LA level indicate that improvements to the census population estimates are required then improvements will be made. Options for improvements are summarised below.

Regional and National Quality Assurance

As well as core and supplementary LA level QA, regional and national analyses will also be undertaken. These include:

- analysis of sex ratios, fertility and mortality rates. A key issue here will be agreeing, ahead of the QA process, what we believe to be the most plausible sex ratios for each age group. Research is underway into national sex ratios from administrative data
- analysis of the ONS Longitudinal Study[9] (which will be linked to the 2011 Census data in time for use in census quality assurance) to identify people who have not been included in the 2011 Census (which could be for legitimate reasons) and also those who have been included more than once, at different locations
- matching of a sample of visitors on census questionnaires back to their usual residence to check for systematic under coverage
- modelling non-response to assess whether other census characteristics (for example, ethnicity and tenure) are more powerful predictors of non-response than age and sex that are currently used in the DSE
- matching responses and non-responses in the census and ONS social surveys. This work, referred to as the Census Non-Response Link Study (CNRLS), will help identify patterns in household and within household non-response

Assessment of the core and supplementary QA undertaken at regional and national levels will be made by a high-level QA panel. This panel sits above the main QA panel reviewing emerging trends and any LAs for which there are particular concerns.

Options available for making improvements to the census population estimates

Where the results of the supplementary analysis clearly indicate a problem with the estimates then improvements will be applied. The improvements 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 subsequent quality.

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

- revisit coverage estimation to post-stratify by different non-response characteristics
- revisit coverage estimation and post-stratify by other factors, for example, response rates as measured through the questionnaire tracking system
- calibrate to an alternative data source

- calibrate to an alternative by age/sex ratio
- adjust using the alternative household count

Publication of the results

Census population estimates will be considered by four panels prior to publication. A summary of these panels are:

- **Internal QA Working Group** – census experts will provide a daily steer on whether supplementary QA is required
- **Main QA Panel** – census and demographic experts from ONS (plus Welsh Assembly Government) will meet weekly to review all estimates at LA level
- **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, will be 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 LA once it is satisfied that the estimates have been through a thorough process of assurance. They will then be formally signed off by the Executive Panel and published.

An important element of the publication in July 2012 will be the information that accompanies the results. 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, and a summary of the improvements made during the quality assurance process.

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 LA according to the census definition. For example, the NHS patient 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, but through processing they will be excluded from the usual resident population estimate. Estimates of such populations (albeit lower quality than the main population estimates) will therefore be available to explain differences between the census and administrative sources.

Conclusions

The process for quality assuring the 2011 Census estimates builds on the lessons learned from the 2001 experience and the subsequent LA Studies programme. As a result, there are significant improvements to the process for 2011, which include:

- a greater understanding of local issues and intelligence, based on the QA studies work and the wider LA engagement process that the census has conducted in the lead up to the 2011 Census
- a clearer process for deciding where to undertake further supplementary analysis

- a clearer set of options for making improvements to the census population estimates should the quality assurance highlight deficiencies in the census population estimates

- the availability and planned use of more comprehensive information about the census operation, specifically the availability of detailed return rates to low levels of geography

- an improved address register for the census, and specific quality assurance processes that focus on household comparisons against the census address register and external sources such as council tax data

Notes and References

1   Abbott, O. (2009) '2011 UK Census Coverage Assessment and Adjustment Methodology', *Population Trends* 137: 25-32

2   White, N., Abbott, O., and Compton, G. (2006) Demographic analysis in the UK Census: a look back to 2001 and looking forward to 2011. 2006 Proceedings of the American Statistical Association, Survey Research Section [CD-ROM], American Statistical Association, Alexandria, VA.

3   See www.statistics.gov.uk/downloads/theme_population/LAStudy_FullReport.pdf

4   See www.ons.gov.uk/census/2011-census/process-info/data-quality-assurance/evaluation-of-the-census-2011-quality-assurance-studies-project.pdf

5   The alternative household count is compiled using the most up-to-date Census Address Register: returns received from each address indicating that an address is an occupied usual residence will be combined with information from dummy forms where no return was received but where an enumerator has indicated that an address was a usual residence.

6   Office for National Statistics (2007) A Review of the Potential Use of Administrative Sources in the Estimation of Population Statistics www.ons.gov.uk/about-statistics/methodology-and-quality/imps/archive-pre-08/updates/a-review-of-the-potential-use-of-administrative-sources-in-the-estimation-of-population-statistics.pdf

7   See www.statistics.gov.uk/census2001/pdfs/oncinfopaper.pdf

8   An Estimation Area is a single Local Authority or a group of neighbouring authorities.

9   The ONS Longitudinal Study (LS) links census and vital event information for 1 per cent of the population of England and Wales. Further information is available from: www.ons.gov.uk/about/who-we-are/our-services/longitudinal-study