This is one of a series of reports published to support the release of results from the 2011 Census. This series of methods and quality reports provides information on the different methods used to collect, process, clean, adjust and protect the census results. The series also reports on the quality assurance of the results and provides quality indicators.

Terms used in the series are explained in the 2011 Census glossary.7

Table of Contents

Executive Summary ................................................................. 3

1. Introduction ........................................................................... 5

2. Background ........................................................................... 6
   2.1 Quality Assurance in 2001...................................................... 6
   2.2 Aims and objectives of the quality assurance ......................... 7
   2.3 Scope of 2011 Methodology Quality Assurance ..................... 8

3. Project overview .................................................................... 10
   3.1 Development phase ............................................................ 10
   3.2 Implementation and operation phase ................................. 10
   3.3 Evaluation and close down ............................................... 11

4. Engagement – development and outcomes .......................... 12
   4.1 Engagement with Users ..................................................... 12
      4.1.1 2011 Census Advisory Groups .................................... 12
      4.1.2 2011 Census Roadshows – 2008 to 2011 ................... 12
      4.1.3 2011 Census Methodology tutorials – November 2010 ... 13
      4.1.4 2011 Census Methodology workshops – June/July 2012 .... 13
   4.2 Local authority QA Studies ............................................... 14
   4.3 Independent review of methods ......................................... 15
Executive Summary

Although every national census aims to capture the total population, no census is perfect and people are missed. For the 2011 Census, ONS developed a set of methodologies to identify and adjust both for the number of people and households not counted in the census and those that were counted more than once, with the aim of producing robust and widely accepted population estimates. The methodology for adjusting for under/over enumeration is described by ONS (2012a). To ensure that the results were both robust and fit for purpose, a quality assurance (QA) process was developed. This 2011 Census QA evaluation report describes the methods and processes used to quality assure the 2011 Census estimates.

Quality assurance is primarily about the final product, the tabulated results, and assuring that the results conform to expectations. However, it is also about assuring the quality of individual level data. That is, it is necessary to ensure that changes or actions (e.g. edit and imputation) to the data do not result in systematic errors that are unlikely to be detectable when quality assuring the main results but may affect the quality of more detailed cross-tabulations.

Principal features of the 2011 Census quality assurance process:

- Significant changes introduced which built on the experience from the 2001 Census quality assurance process including:
  - the introduction of quality assurance checks of household estimates and characteristics
  - the insertion of additional checks of sub-local authority level person and household estimates
  - the inclusion of a wider range of comparator data, including the use of local data sources where available
  - the introduction of additional methods for making adjustments for estimates that require resolution
  - the inclusion of a process for making a national adjustment
  - the expansion of QA panels (in number and membership), including independent external experts, to provide assurance of the estimates and eventually recommend sign-off and publication of the census estimates

- Extensive activities to engage with users and external experts. These activities were focused on improving the development of the methods and processes, and to build also user confidence in the methods, and therefore the census results.

- Central to the QA process was a hierarchical set of four QA panels which reviewed the evidence provided through the quality checks and comparators sources; and, ultimately were responsible for recommending to the National Statistician that national and local census population estimates should be published. These panels had a membership of internal and external, independent reviewers to provide challenge and subsequent assurance both internally in the sign-off of the estimates and externally to users.
Overall the process to quality assure the results was highly successful and met its primary objectives. Most importantly the methods and data sources used were transparent and gave users confidence in the process and therefore the census estimates.

Summary of the main lessons from the evaluation of the quality assurance:

- Frequent and open engagement with users was a major contributory factor in improving the QA methods and building user trust in the results
- Independent assessment of the methods was an important tool in providing assurance to ONS and users that the quality assurance methodology would meet the aims of identifying and resolving issues
- Rehearsal of the QA process should take place as early in the census process as possible
- Earlier understanding of the impacts and interdependencies of all processes – such as edit and imputation – will enable a better understanding of the potential impacts on the data and results
- More consideration should be given to the output content when developing quality assurance checks
- Use of external experts worked really well and early engagement with them paid dividends
- The use of the QA panels worked well, with the panels providing a rich source of challenge and subsequent assurance both internally in the sign-off of the estimates and externally to users

Given the recommendation by the National Statistician about the need for a 2021 Census (predominantly online, supplemented by further use of administrative and survey data), this quality assurance evaluation report will be an invaluable reference source when considering the design and development of a quality assurance methodology for that census.
1. Introduction

To achieve the aim of producing robust and widely accepted population estimates, ONS developed a set of methodologies for the 2011 Census to identify and adjust both for the number of people and households not counted in the census and those that were counted more than once (ONS (2012a)\(^8\)).

This report describes the methods and processes used to quality assure the 2011 Census estimates. It summarises how the 2011 Census quality assurance methodology was designed and applied, evaluates the outcomes, outlines the lessons learned and provides suggestions to improve QA processes for future censuses and other large-scale population estimation projects.

Given the recommendation by the National Statistician about the need for a 2021 Census (predominantly online, supplemented by further use of administrative and survey data), this report will be an invaluable reference source when considering the design and development of a quality assurance methodology for the 2021 census.

The methods and processes (developed and implemented) for the 2011 Quality Assurance were built using the framework developed for 2001 (evaluation of the 2001 Census is available from the ONS website\(^6\), taking into account the lessons learnt and the outcomes of the subsequent Local Authority Studies (ONS, 2004)\(^9\). In particular, the main changes made were:

- quality assurance of household estimates and characteristics
- additional checks of sub-local authority household and person estimates
- inclusion of a wider range of sources, including local data
- additional methods for making adjustments
- expansion of the quality assurance panels to include internal and external experts
- the potential for making a national adjustment was included in the process

Quality assurance was carried out at local, regional and England and Wales levels. The main QA process was a series of ‘core’ and ‘supplementary’ checks at these geographical levels. These included checks against comparator sources; area profile checking; demographic analyses; and data matching to administrative data. The checks are laid out in sections 6.1 and 6.2. In parallel with this main QA process was a topic quality assurance process which checked important characteristics that were not considered as part of the core and supplementary checks (for example information on ethnicity or general health) to ensure no systematic errors (bias) were introduced as a result of statistical processing.

Notable features of the QA process were:

- a range of quality assurance panels reviewed estimates at varying levels of detail, including different geographic levels
- a range of evidence was used, including comparisons against administrative data sources
- census estimates used in the QA included persons and their key characteristics (students, armed forces, ethnicity)
- topic quality assurance made sure no systematic errors (bias) were introduced
- estimates of households occupied by usual residents were quality assured
- identification of issues which ONS was able to adjust for in the data processing
- analysis which explained inconsistencies between comparator data and census data


2. Background

2.1 Quality Assurance in 2001

The quality assurance of the 2001 Census population estimates is described by ONS (2001)\(^{10}\). The process and outcomes were reviewed by the Local Government Association (LGA, 2003)\(^{11}\), who concluded that internal evidence from the census regarding its own completeness is not convincing and external validation is essential. In addition, planning for production of timely metadata and information about the QA process was an important lesson.

White et al (2006)\(^{12}\) outline the following lessons from the 2001 quality assurance process:

- More information about census data collection would have helped the quality assurance process. The lack of reliable quantitative information from the Field Management Information System hampered some judgments.

- The matching studies highlighted differences between the census address lists and those provided by Manchester and Westminster. Address lists used for 2011 need to be more up-to-date, with clear definitions for vacant, derelict, communal establishments and commercial properties.

- Local data and knowledge should have been used in the quality assurance process. More robust arguments and reasons may have been formed if such data had been used.

- Delays delivering the data from the suppliers used up all of the three-month contingency period. This time would have been valuable for better understanding and quality assuring the census estimates, and it should be protected in future.

- A wide range of possible actions should be in place to act upon where census estimates were identified as implausible.

- Contingency plans are necessary, particularly in the event of delayed data or the requirement for a national adjustment.
2.2 Aims and objectives of the quality assurance

The broad strategic aims of the 2011 Census were:

- to give the highest priority to getting the national and local population counts right
- to build effective partnerships with other organisations, particularly local authorities, in planning and executing the field operation
- to provide high quality, value-for-money, fit-for-purpose statistics that meet both user needs and inspire user confidence, and which are as consistent, comparable and accessible across the UK as is possible
- to maximise overall response rates and minimise differences in response rates in specific areas and among particular population sub-groups
- to protect, and be seen to protect, confidential personal census information

A 2011 Census quality strategy was developed to deliver these aims.

The development of a 2011 Census Data Quality Assurance Strategy guided the QA work by outlining the quality assurance objectives, the methodology and the mechanisms (i.e. the key principles and deliverables to be developed to achieve the objectives) and formed part of the overall 2011 strategy.

The production of the 2011 census population estimates was much more focused on identifying, and improving, shortfalls in the estimation process than in 2001. It was essential that the process for approving these estimates was robust, transparent and had credibility with users. Although the basic approach for 2001 worked well, it was identified that there was a clear need to have more external involvement and a separate assurance process that focused on emerging issues at pre-defined periods before publication.

To ensure that the development of the methodology was a transparent process, a consultation mechanism was agreed that had two main aims: to gain a level of input from users and stakeholders into the development of the methodology, and to manage user expectations so that users were very clear about the process ahead of the actual operation.

The methodology included identifying, establishing and developing links with relevant experts in ONS and other organisations to ensure that a variety of experts with relevant topic knowledge were involved in the QA. It also covered the procedures to be implemented if the data did not conform to expectations.

Quality assurance is primarily about the final product, the tabulated results, and assuring that the results conform to expectations. However, it is also about assuring the quality of individual level data. That is, it is necessary to ensure that changes or actions (e.g. edit and imputation) to the data do not result in systematic errors that are unlikely to be detectable when quality assuring the main results but may affect the quality of more detailed cross-tabulations.
2.3 Scope of 2011 Methodology Quality Assurance

The Downstream Processing Project (DSP) provided a set of IT systems capable of carrying out the subsequent processing of all 2011 Census and Census Coverage Survey data. The project was responsible for the live running of the data through the downstream process and providing operational support during live running. The whole process started with the loading of the data and ended with a disclosure control process, before the production of outputs. Figure 2.1 shows the main steps and how quality assurance fitted into the overall process.

Figure 2.1 - Quality Assurance in data processing

Quality assurance checks were carried out at most of the major processes (as can be seen in Figure 2.1) to ensure that no systematic error was being introduced. The main focus of the quality assurance work, and hence this report, was on the quality assurance of the census results. This took place after the coverage assessment and adjustment (CAA) process, which itself included a number of quality assurance checks (see ONS, 2012c)\(^3\).

After the coverage assessment and adjustment process, census population estimates for all 348 local authorities were compared with upper and lower tolerance bounds derived from administrative and survey data sources. These gave a range of plausible values within which census estimates were expected to fall. The tolerance bounds were designed to reflect known differences between alternative sources and census estimates in terms of definitions, accuracy and timing.
QA was also carried out at a topic level on a number of important characteristics collected in the census (for example information on ethnicity or general health) that were not considered as part of the main checks. This topic QA was designed to identify and where possible correct for systematic errors, thus helping to ensure the published results were both accurate and plausible.

The QA was the final process in agreeing the census population estimates that were published on 16 July 2012. Information from both the CAA and the quality assurance process was published alongside the census population estimates to help users understand the quality of the estimates (such as how much adjustment was applied by area) and place the estimates in the context of other administrative data sources. This package of supporting information included:

- response rates by local authority and by age and sex
- 95 per cent confidence intervals by local authority
- the size of the household bias, overcount, communal establishment and national adjustments
- census estimates against other sources for each local authority, such as patient register, school census and child benefit, but also showed the tolerance bounds for each area
- the components of the estimates, quantifying the effects of the various processing steps

Evaluation papers for other 2011 Census processes may be found on the census page “How did we do in 2011?”.
3. Project overview

The project was structured into three overlapping phases:

- Development phase - 2005 to 2011
- Implementation and operation - 2009 to 2013
- Evaluation and close down - 2013 to 2014

3.1 Development phase

The first phase began in 2005 with a review of the lessons from the 2001 Census and the accompanying reviews from external organisations [including the Local Government Association11]. The outcome of these reviews was summarized in White et al (2006)12 and highlighted a number of areas to focus on. After further research and development, a 2011 Census Data Quality Assurance Strategy1 was published which set out our strategic plans for QA. This covered the methods for quality assuring the estimates and the topic QA, user engagement and quality reporting to accompany the publication of results.

An internal working group was established to review and guide the design work, including the approach to topic QA, the local authority QA studies, and the specification of demographic indicators and made recommendations through the formal census governance arrangements. This working group included representatives from the census programme, subject experts from other areas of ONS, such as methodology and the population statistics division, and representatives from the Northern Ireland Statistics and Research Agency (NISRA), National Records of Scotland (NRS) and the Welsh Assembly Government (WAG). In addition, further advice was sought through external events, such as a 2011 Census Methods Day hosted at the RSS, the Census Advisory Groups and local authorities through the 2011 Census local authority QA studies work (see section 4.2). Lastly, the plans and methods were subject to an independent review.

3.2 Implementation and operation phase

The implementation and operation phase commenced in 2009 when this project began working with the census processing project which was developing the environment and systems for processing the captured data. This phase operated in parallel with the remaining methodological development and therefore a flexible approach to the development and testing of final systems was undertaken.

Main activities of this phase:

- Specifying and acceptance testing the design, and, presentation of the quality assurance checks and user interfaces
- Designing and developing the reporting procedures and interactive tools to be used by the quality assurance panels and the topic QA experts as part of their review of the results
- Resource planning and recruitment – including recruitment of the quality assurers and QA panels
- Developing operational management procedures to manage and report on the results and progress of the quality assurance
- Training of the analysts undertaking the quality assurance and the members of the quality assurance panels
- Live running – managing and operating the quality assurance processes and the movement of areas through the quality assurance panels throughout the live running period
- Production of material to support outputs – including metadata and methodology guidance for users
- Responding to user queries with specific questions about the quality of the census results
The production of the 2011 Census population estimates was much more focused on identifying, and improving, shortfalls in the estimation process than in 2001. It was essential that the process for approving these estimates was robust, transparent and had credibility with users. Although the basic approach for 2001 worked well, there was a clear need to have more external involvement and a separate assurance process that focused on emerging issues at pre-defined periods before publication.

### 3.3 Evaluation and close down

The evaluation and close down phase began shortly after the publication of the Key and Quick Statistics in February 2013.

The evaluation (of which this report is a part) includes the following activities:

- overall project evaluation
- evaluation of the methods and processes used to quality assure the results
- further analysis of the utility and performance of the quality assurance checks during the operation and from subsequent processes (for example where data amendments had to be applied as a result of a processing error)
- evaluation of the processing diagnostics to understand the timing in each of the quality assurance processes to inform future plans for resourcing and processing times
4. Engagement – development and outcomes

4.1 Engagement with Users

The quality assurance methodology was the subject of extensive consultations with a variety of stakeholders, including academics, statisticians, demographers and expert census users. The process was designed to:

- ensure 2011 Census estimates were fit for purpose
- use comparator sources to identify discrepancies with census estimates
- use contingencies, where required, to improve census estimates
- ensure census population characteristics were accurate
- build user confidence through transparency in the methods

4.1.1 2011 Census Advisory Groups

A number of 2011 Census advisory groups met regularly (generally every six months) throughout the whole process to give advice on all areas of the 2011 Census. These groups comprised experts in local government, central government, health, academia, business, the Welsh assembly and local authorities, diversity groups and faith groups. 2011 Census advisory group papers are available from the ONS website. A number of papers specifically relating to data quality were presented to the groups. Minutes and actions from these meetings record the advice and suggestions from the groups and these are also available from the ONS website.

4.1.2 2011 Census Roadshows – 2008 to 2011

ONS took to the road in 2008, twice in 2009 and again in 2011 to not only to inform and update those with an interest in the census, but also to answer questions, learn lessons and help in the further development and enhancement of the methods. Quality assurance was one of many topics covered in London, Leicester, Birmingham, Durham, Manchester, Newcastle, Cardiff, Nottingham, Peterborough and Crewe.

Evaluation from each roadshow was provided on feedback forms. Questions were recorded and answers fed back to the attendees. The questions and answers from the 2011 Census roadshows are available from the ONS website.
4.1.3 2011 Census Methodology tutorials – November 2010

The 2011 Census team ran a series of national methodology tutorials, “Spotlight on completing the count”, in November 2010. The tutorials provided census users with an accessible explanation of the methodology involved in coverage estimation (the process of estimating the number and characteristics of those who did not complete a census questionnaire) and quality assurance (the process of checking the plausibility of the census data, with the priority being the population estimates for local authorities by age and sex). They were aimed at non-technical users, with the main target audience being representatives from local authorities with a need to understand and explain to others how estimates are produced and quality assured.

The broad aims of the tutorial sessions were:

- to educate users on the methodology for coverage adjustment and quality assurance
- to highlight the improvements to the methodology since the 2001 Census
- to give confidence in the census results by demonstrating that the process to produce the census population estimates is based upon sound methodology
- to allow a forum for sharing questions and concerns among local authority users

Attendees were led through working examples to help explain the methodology behind the generation of census estimates. There was also the opportunity for attendees to get involved with interactive group exercises. Feedback was given in a selection of questions and answers and is available from the ONS website19.

4.1.4 2011 Census Methodology workshops – June/July 2012

In June/July 2012 the Office for National Statistics put on an additional series of workshops, “From process to publication – understanding your census estimates”, to outline the methods used to arrive at census estimates. These were timed to take place just before the release of the first census estimates to strengthen user awareness and their understanding of the methods and metadata that would accompany the release. The one day workshops were designed to enable the delegates to:

- understand how local authority estimates are created and the key components in the methods including estimation methods, adjustments for communal establishments and adjustments for bias and overcount
- see the quality assurance processes, data sources used, and examples of the investigations conducted
- discover what would be available for local authorities

Using anonymised real examples, the workshops took delegates through processes and aimed to give them the confidence needed to explain the census estimates to key policy makers and finance leaders within their organisations. They covered the background on the methods, to build confidence that local authority’s census estimates would be robust, fit for purpose and accurate. Further information on these workshops is available from the ONS website20.
4.2 Local authority QA studies

Experience suggested that the census would produce population estimates that were fit for purpose in the vast majority of local authorities (LAs). However in some LAs the demographic composition of the area, or societal changes since 2001, suggested that additional research to understand the complex issues in these areas was required. ONS undertook a series of QA studies working closely with certain LAs to further understand their concerns and how the QA process could be improved for all LAs. It was also intended to use the QA studies to identify data sources held by LAs which would be of benefit to the QA process and where appropriate to gain access to these datasets.

On 27 October 2009, the Census Regional Champions (as described in Evaluation of local authority engagement37) endorsed a proposal for ONS to undertake a pilot for the QA studies project to trial the approach. This pilot involved ten LAs and ran between February and April 2010. Following the successful completion of these studies a wider QA study was undertaken between July and October 2010. A further 30 LAs were involved in this work.

The local authority QA studies project had several aims:

- to identify LA data sources that could serve as comparators during QA analysis
- to improve ONS’s understanding of the quality of comparator data sources for QA
- to provide softer intelligence on population sub-groups, unique population characteristics or any other issue that could help ONS’s understanding of the census results
- to establish relationships between ONS and LAs to support the QA work
- to inform process / engagement required for LAs not involved in the pilot or QA studies

Some notable outcomes were instrumental in the development of the methodology including:

- improved understanding within ONS of the data sources held by LAs
- secure data transfer and storage options refined
- an early insight into sources and checks ONS had already planned to use
- opportunity to improve data analysis skills
- analytical methods, in particular the use of scatter plots to compare datasets and identify outliers
- the value of council tax data including exemptions and discount codes confirmed
- national collection of HMO data initiated
- useful feedback on the quality issues of key comparator datasets received
- importance of engagement with LAs for knowledge-sharing and better understanding of data sources recognised

Recommendations that were taken forward from these QA studies included:

- request for all local authorities to provide council tax data including exemption and discount information
- patient register data by country of birth to be requested from local authorities
- ONS to review information submitted through the Census Local Partnership Plans to identify other sources of supporting evidence which would be useful for census quality assurance
- council tax exemption and discount data to be used as part of supplementary QA to investigate cases where census household estimates fall outside expected ranges, and to reconcile specific population sub-groups, for example UK and foreign armed forces
- check on Average Household Size to be included as part of the quality assurance process
- electoral registration data held by ONS to be used as part of the quality assurance process
More information on the QA studies and the evaluation of the QA Studies is available from the ONS website\textsuperscript{21,22}.

### 4.3 Independent review of methods


The independent reviewers were:

- Ian Plewis, Professor of Social Statistics, University of Manchester
- Ludi Simpson, Honorary Professor of Population Studies, University of Manchester
- Paul Williamson, Senior Lecturer in Population Geography, University of Liverpool

The review made a number of recommendations, which ONS accepted and addressed, and the review team concluded that:

‘We would like to put on record our belief that many lessons have been learned from the Census in 2001 (which was itself a considerable improvement over the 1991 Census). We have been impressed by the scope and depth of the methodological investigations initiated by ONS, by their willingness to discuss with a wide range of interest groups concerns about coverage and Quality Assurance (QA), and by the procedures that are in place to use field staff flexibly. We are reasonably optimistic that, having taken account of our recommendations to develop, document and consult on specific aspects of methodology, the 2011 Census in England and Wales will provide population estimates that can guide resource allocation and social policy in the right direction for the next ten years. It must, however, be recognised that the target 95 per cent confidence intervals set by ONS for the population counts – a maximum interval of ± 3 per cent for all LAs – are entirely contingent on achieving local as well as national targets for non-response.’

‘We are delighted to learn from ONS that early indicators of response to the Census suggest that the targets will be met. If confirmed, this will be a considerable achievement at a time of falling response rates to official enquiries.’

The full review and the ONS response to the recommendations can be found on the ONS website\textsuperscript{23}.
5. The process for quality assuring 2011 Census estimates

The material published prior to the 2011 Census set out the proposed approach for the QA of the census population estimates. Central to this process were the QA panels, a hierarchy of four groups, which reviewed and endorsed the 2011 Census population estimates, using processes that ultimately required the National Statistician to sign off national and local census population estimates.

The four panels which considered 2011 Census population estimates were:

- **Internal QA Working Group**
  ONS census experts provided a daily check on whether the QA checks should lead to supplementary QA, prioritised the work analysis required, and gave a steer on whether improvements were needed.

- **Main QA Panel**
  Census and demographic experts from ONS (plus Welsh Assembly Government) met to review all estimates at local authority (LA) and higher geographies. On average this panel met on a weekly basis.

- **High Level QA Panel**
  A group of internal, UK and independent external reviewers provided expertise and guidance on the emerging national picture. On average the group met every six weeks.

- **Executive Panel**
  An executive panel, chaired by the National Statistician, was accountable for the final sign-off of the national and local census population estimates ahead of publication. It was convened when all initial estimation was completed. The High level QA Panel approved and endorsed estimates for each Local Authority once it was satisfied that the estimates had been thoroughly assured. They were then formally signed off by the National Statistician and the executive panel ahead of publication.

The document Quality Assurance of 2011 Census Population Estimates describes in detail the process used by ONS to quality assure the 2011 Census population estimates, including the principal outcomes of each of the four QA panels.
6. Quality Assurance Methodology

This section provides a summary of this process used to quality assure the 2011 Census population estimates. It outlines the main checks, including both core checks (which were routinely assessed for all areas) and supplementary checks (which were carried out to investigate inconsistencies found with the core checks). It also covers topic QA which helped to ensure that no systematic errors (bias) were introduced as a result of processes such as imputation, filter rules or derived variables. Quality assurance was carried out, in the main, at local, regional and England and Wales levels.

In preparation for the QA process, ONS published details of the evidence that would be considered as part of the process in the document 2011 Census - Methodology for Quality Assuring the Census Population Estimates25.

An overview of the census data quality assurance strategy is provided in Wroth-Smith et al (2011)26. This is summarised in figure 6.1. The quality assurance process was organised into three stages: core checks on the data, supplementary checks and, for data anomalies that cannot be understood using aggregate-level data, record matching. Core and supplementary checks were carried out at local authority and lower levels of geography. There were also regional and national-level checks, using data comparators and demographic indicators, to ensure that the emerging and final regional and national estimates were accurate.

In parallel with the core and supplementary checks, topic QA checked important characteristics collected in the census that were not considered as part of the core and supplementary checks, for example information on ethnicity or general health. The primary purpose of this topic QA was to identify and where possible correct for systematic errors. Checks and validation of data were carried out at various processing stages to ensure that errors did not appear at the final output stage.

Figure 6.1 - Quality Assurance Overview

The QA checking was an iterative process. Each test may have been carried out a number of times depending on the supporting intelligence. Where changes were made, the QA checks were carried out again to ensure the changes had no unintended impact. Table 6.1 summarises the core and supplementary checks at each level.

<table>
<thead>
<tr>
<th>Comparator data</th>
<th>Level 1</th>
<th>Core checks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operational intelligence</td>
<td>Level 2</td>
<td>Supplementary checks</td>
</tr>
<tr>
<td>Local authority provided intelligence</td>
<td>Level 3</td>
<td>Detailed aggregate analysis of local areas and population sub-groups, Data matching</td>
</tr>
</tbody>
</table>

- Administrative data sources
- ONS survey sources
- Return rates
- Issues identified by questionnaire tracker
- Comparator data
- Background data on each LA
- Correspondence
- Post 2001 LA studies
Table 6.1 – Summary of Local, Regional and England and Wales quality assurance checks

<table>
<thead>
<tr>
<th>Level</th>
<th>Core Checks</th>
<th>Supplementary Checks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local Authority</td>
<td>• Checks against comparator sources</td>
<td>• Analysis of visitor and second residence</td>
</tr>
<tr>
<td></td>
<td>• Assessment of the distribution of population below LA</td>
<td>• Additional checks on population sub-groups</td>
</tr>
<tr>
<td></td>
<td>• Demographic analysis</td>
<td>• Low level aggregate comparisons</td>
</tr>
<tr>
<td></td>
<td>• Operational intelligence</td>
<td>• Data matching to administrative data</td>
</tr>
<tr>
<td></td>
<td>• Area profiles</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• LA provided evidence</td>
<td></td>
</tr>
<tr>
<td>Regional</td>
<td>• Aggregate checks against comparator sources</td>
<td>• Analysis of visitor and second residence</td>
</tr>
<tr>
<td></td>
<td>• Demographic analysis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Modelling non-response</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Assessment of Census Non-Response Link Study (CNRLS)</td>
<td></td>
</tr>
<tr>
<td>England and Wales</td>
<td>• Aggregate checks against comparator sources</td>
<td>• Analysis of visitor and second residence</td>
</tr>
<tr>
<td></td>
<td>• Demographic analysis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Assessment of CNRLS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Assessment of Longitudinal Study</td>
<td></td>
</tr>
</tbody>
</table>

### 6.1 Core checks

All core checks were based on the usual resident population (those who had been born in, or were intending to stay in, the UK for twelve months or more). This was 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) were checked separately, but also used to help explain differences with administrative data.

The results of the core checks were compiled into a quality assurance pack, made available to the main quality assurance panel. Other factors considered by the panel included information about the conduct of the census field operation in the area (‘operational intelligence’), area profiles, intelligence provided by local authority.
A series of tolerances were established for each of the core checks to help distinguish between potential errors in the census estimates and definitional/coverage issues with the comparators. These tolerances helped to identify the occasions when supplementary analysis was required.

A full set of core checks can be found at Annex A.

### 6.1.1 Core assurance checks at the local authority level

All LA estimates were subject to six types of core check:

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

- Census estimates of the distribution of the population within each local authority were checked. The assessment was made at both Lower and Middle Super Output Areas, again in comparison to administrative data.

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

- When analysing and reviewing the core checks it was important to understand key census field operation indicators. This included overall return rates (set against expectations), variation of return rates within an LA, and questionnaire tracking information.

- To provide valuable context about each local authority, an area profile was compiled for the QA process. This included a record of previous correspondence and the findings from reviews of the local authority’s population (including the post-2001 LA studies). It also provided 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.

- The quality assurance process was able to draw on evidence provided from each local authority. This evidence either came through correspondence with ONS on the 2001 Census or on the subsequent mid-year populations. Users were also asked to provide locally held data and relevant research as part of the QA studies.

Information from the core checks was assessed using published guidance on whether the census estimates should go to the main quality assurance panel for approval or be referred for supplementary analysis. It was not necessary for all the core checks to be completed before requesting supplementary checks.
6.1.2 Core assurance checks at regional and national levels

As at the LA level, a range of core checks was routinely run at both the regional and England and Wales levels, including:

- repetition of the checks against comparators at these higher geographic levels
- repetition of demographic analysis of sex ratios, fertility and mortality rates
- 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) which matches census returns to recent returns from ONS social surveys and was used to identify where individuals within households were missed in the census but were identified in social surveys (ONS, 2012d)\(^7\).
- assessment of the census population estimate for England and Wales against the equivalent mid-year population estimate rolled forward from the 2001 census

Evidence from these core checks and questions and issues escalated by the main QA panel were considered by the high-level QA panel to either approve, request supplementary analysis or request improvement.

6.2 Supplementary checks

6.2.1 Supplementary assurance checks below and at the local authority level

Where the core LA checks raised questions, more detailed analysis was undertaken. This work assessed the population as a whole at very low geographic levels and/or particular sub-groups within an area.

An analysis of specific small geographic areas focused on postcode clusters included in the sample for the Census Coverage Survey (CCS) in each LA. The information within these clusters was particularly important as they were the basis for estimating and adjusting for under-coverage across the wider LA(s). This work included:

- comparing population estimates (for both individuals and households) for each selected postcode cluster, and the distributions of key characteristics such as age and sex, with aggregate statistics from the available administrative dataset for the same postcodes
- 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 LA) 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

No adjustments were made on the basis of matching, but a greater understanding of the underlying data and local situations was gained. The document Results from using routinely-collected government information for 2011 Census quality assurance\textsuperscript{39}, describes valuable new insights gained on the statistical quality of the administrative sources that ONS used for the 2011 Census.

Where questions were raised, population sub-groups were analysed using the additional detail provided by the sources used in the core quality assurance. Supplementary analysis also included investigations into the information built into the census data collection to help explain inconsistencies with administrative datasets. This included information collected on second residences, short term migrants, vacant properties and visitors.

It was anticipated that it would be necessary to complete analysis for a number of LAs before firm conclusions could be drawn about the existence of systematic coverage problems, and this indeed was the case. Where supplementary QA indicated that improvements to the census population estimates were required, improvements were made either to a single LA, to a number of specific LAs or across all LAs. For example, there were cases where this led to revised communal establishment adjustment.

\textbf{6.2.2 Supplementary assurance checks at regional and national levels}

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

This included extending some of the supplementary checks initially carried out at LA level. An example was 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.
6.3 Improvements to the 2011 Census population estimates

As expected, there were instances where estimates fell outside the bounds set. For the majority of cases further investigation and analysis was able to explain differences between census and comparator sources. For a small number of cases issues were identified which resulted in adjustments to the data prior to publication. This included the correction of a small number of communal establishments that had been misclassified as households. Some communal establishments had been enumerated correctly but needed to be moved to the correct geographical area in the census data. Single year of age ‘spikes’ (which occurred when a particular donor was used several times during the imputation process), were also identified and resolved.

The available options to deal with instances where there were questions about the estimates were:

**Revisit coverage estimation**

Where local authority estimates could not be approved following QA, depending on the nature of the issue(s) to be resolved, one option was to assess whether the estimates could be improved by using one of the following methods:

- Dual system estimate post-stratification by alternative characteristics
- Estimation area post-stratification
- Borrow strength from other local authorities

In practice, these methods did not need to be used as the other adjustments we had in place (household bias adjustment, sample balance, national adjustments and some of our tuning approaches) were sufficient for producing robust estimates.

**Adjust Coverage Estimation for Bias**

Dual System Estimation (DSE) used in coverage assessment and adjustment was based on a series of assumptions which if violated could have resulted in biased estimates of the population. One of the critical assumptions made was that the probability of being counted in the CCS was independent of the probability of being counted in the census. Independence is assumed both for households and individuals within households. The methods used for adjusting coverage estimation for bias were set out in the documents:

- ONS, 2012d - Between Household Bias Adjustment Using the Alternative Household Estimate
- ONS, 2012e - Within Household Bias Adjustment Using the Census Non-Response Link Study (CNRLS)

Any bias not picked up by the these types of adjustments was designed to be addressed with the national adjustment.
Apply national sex ratios by age (national adjustment)

Evidence of a systematic under or over estimation in men or women was 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 and Wales or whether it was geographically clustered in some LAs only. Once apportioned to LA level, improvements could 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.

The approach and the processes we followed are outlined in the document ONS (2012f)\textsuperscript{29}, which covers the options of:

- revising estimates using evidence from the ONS Longitudinal Study
- using external sources for direct adjustment (direct adjustment for small geographic areas or direct adjustment for population sub-groups)

The document Making a National Adjustment to the 2011 Census (July 2012)\textsuperscript{30} sets out the evidence for making a national adjustment in 2011, the method used for making the adjustment and for cascading it to local authority level. The overall effect of the national adjustment on the national population was to add 303,400 male usual residents, increasing the population by 0.5 per cent.

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. For this reason this option was only used for adjusting census population estimates in large communal establishments (in particular student halls of residence). The paper Estimation and Adjustment for Communal Establishments\textsuperscript{31} describes this in detail.
6.4 Topic Quality Assurance

Checks were also developed to validate census estimates by topic against data such as the 2001 Census and ONS survey data. The broad topic areas covered included:

- demography
- ethnicity, identity, language and religion
- health
- education
- labour market
- travel/transport, and
- households/housing

ONS and other topic experts were periodically invited to review the checks for their specialist subject areas and identify instances where census estimates deviated from expectations. Anomalies were investigated and the process refined as appropriate.

The primary purpose of topic QA was to identify and where possible correct for systematic errors. Checks and validation of data were carried out at various processing stages to ensure that errors did not appear at the final output stage.

Checks were carried out by comparing the 2011 Census data with comparator data. 2001 Census data was primarily used along with additional comparators such as the Annual Population Survey (APS) or Integrated Household Survey (IHS).

The QA was conducted at LA, regional and national level with sub LA analysis conducted if an issue was found that required looking at smaller geographies.

A topic was considered a higher priority if it was:

- used for population estimates
- used in the coverage estimation/imputation process
- a key output variable e.g. Eurostat variable
- part of a new question or there are concerns over response e.g. civil partnerships

Due to time constraints, efforts were initially focused on the high and medium priority checks, with a selection of low priority checks conducted when possible. The topic quality assurance checks are summarised in the table 6.2.
Table 6.2 – Summary of topic quality assurance checks

<table>
<thead>
<tr>
<th>Check Name</th>
<th>Individual Checks</th>
<th>Household Checks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marital Status, Second Address, Country of Birth, Passports held,</td>
<td>General Health, Long term illness/disability, Carers, Religion, National identity,</td>
<td>Tenure and type of landlord, Accommodation type, Self-contained, Household</td>
</tr>
<tr>
<td>Economic activity, Ethnicity (write in only ), Main language, English</td>
<td>Occupation, Industry, NS-SEC</td>
<td>composition/family type (PDD)</td>
</tr>
<tr>
<td>language ability</td>
<td>Address one year ago, Hours worked, Welsh language, Qualifications, Self-</td>
<td>Number of bedrooms, Number of rooms, Central heating</td>
</tr>
<tr>
<td>employed/employee, Supervisor status, Ever worked</td>
<td>employed/employee, Supervisor status, Ever worked</td>
<td></td>
</tr>
</tbody>
</table>

Checks were carried out on the data after it had passed through various processing stages. Table 6.3 shows the level of check carried out at each stage.

Table 6.3 – Topic quality assurance checks at each stage

<table>
<thead>
<tr>
<th>Processing Stage</th>
<th>Check type</th>
<th>Geographic level of check</th>
</tr>
</thead>
<tbody>
<tr>
<td>First loaded</td>
<td>Proportional frequency distributions of census data checked against</td>
<td>Delivery Groups (Groups of local authorities)</td>
</tr>
<tr>
<td></td>
<td>comparator data.</td>
<td></td>
</tr>
<tr>
<td>After item imputation</td>
<td>Proportional frequency distributions of census data checked against</td>
<td>Delivery Groups</td>
</tr>
<tr>
<td></td>
<td>comparator data.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cross checks – comparing differences between counts at Filters and Derived</td>
<td>Local authority</td>
</tr>
<tr>
<td></td>
<td>Variables and Item Imputation.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Implausible Values – check to identify any implausible values such as</td>
<td>National initially, then lower geographies to investigate.</td>
</tr>
<tr>
<td></td>
<td>married babies, young children who are working.</td>
<td></td>
</tr>
<tr>
<td>Prior to output</td>
<td>Cross checks – comparing differences between counts at Item Imputation</td>
<td>Local authority by 5 year age groups and sex.</td>
</tr>
<tr>
<td></td>
<td>and Post Coverage Item Imputation.</td>
<td></td>
</tr>
</tbody>
</table>

Topic leads, considered experts in their respective fields, were consulted with on a periodic basis during the topic QA process. The topic leads provided assistance in interpreting results and advice on emerging issues in the data such as implausible distributions. Table 6.4 summarises the issues that were identified during the topic checks.
Table 6.4 – Summary of issues identified during topic quality assurance

<table>
<thead>
<tr>
<th>Issue identified</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implausible values such as married children.</td>
<td>Data amendments applied to ensure implausible values, such as married children, were no longer present in the data.</td>
</tr>
<tr>
<td>Higher than expected increase of 16-29 year olds divorced and widowed due to imputation.</td>
<td>This issue was raised at Census Data Quality Forum (DQF) where it was decided that this issue would be difficult to resolve. It was agreed that these records would be left in, as they were based on observed data and the Census Outputs Group (COG) had previously made decisions to adhere to the principles in the Census Edit Strategy.</td>
</tr>
<tr>
<td>Higher numbers of dissolved civil partnerships and surviving civil partners when checked against comparator data at Item Imputation stage.</td>
<td>No significant change between stages (when compared to Filters and Derived Variables stage) so not a result of imputation. It was decided that these figures should remain in the data as there was no clear indication of error and most likely explained by respondents misunderstanding the question.</td>
</tr>
<tr>
<td>Single year age spikes found in the data making the distribution of data inconsistent with those who responded. For example, figure 6.1 shows how the number of 38 year old widowed females in one LA had increased as a result of Item Imputation.</td>
<td>Data amendment applied to smooth the data by adjusting the way donor records were selected. This helped to ensure the Census data reflected the distribution of those who had responded.</td>
</tr>
<tr>
<td>Religion variable was not being imputed correctly at the Post Coverage Item Imputation stage.</td>
<td>New records imputed as part of the Post Coverage Item Imputation stage were being assigned as having ‘no religion’. A Data File Amendment (DFA) was raised to correct this issue.</td>
</tr>
</tbody>
</table>

As an example of the types of issue identified by topic quality assurance checks, figure 6.1 shows a single year of age spike of widowed females for one of the local authorities checked.
Figure 6.1 – Example of single year of age spike of widowed females for one of the LAs checked
7. Assessment and lessons learnt

This section evaluates components of the quality assurance from a methodological, operational and project perspective.

7.1 Methodological assessment

Generally the statistical solution was a success and a significant improvement on 2001. In addressing the lessons from 2001, the tools available allowed the QA team and the panels access to a wider set of checks, allowing them to have more confidence in the local authority age-by-sex population estimates. Whilst these were the key estimates, additional variables and lower level geography checks were included. Naturally, this resulted in the ability to consider issues in smaller domains, such as patterns across single years of age or coverage rates across ethnic groups. The following points highlight lessons learnt and recommendations for any future exercise.

Engagement

Frequent and open engagement with users was a major contributory factor in improving the QA methods and building user trust in the results (as detailed in Chapter 4). Engaging with users on the proposed 2021 Census will continue to be a critical exercise.

The need for improved engagement with local authorities and community organisations was one of the main lessons learned from the 2001 Census. Their understanding of their local areas and communities was an important contributor to the success of the census and they promoted it through their communication channels and ran over 6,000 events to help promote the importance of the census.

It was particularly important that these groups had an understanding of the quality assurance methods to give them a high degree of confidence in the census results.

Independent Review

The Independent Review of Coverage Assessment, Adjustment and Quality Assurance (described in section 4.3) produced a number of recommendations which ONS responded to. The quality of the review panel and the detail of the reports that they produced not only gave ONS confidence in the process and the quality assurance, but led to improvements in the methodology for 2011, which should provide a good baseline for a census in 2021. The final reports from the Independent Review and the responses from ONS are available from the ONS website.

The Independent Review was important in reassuring census users (in particular local authorities) that the methodology and the data assurance would be instrumental in producing robust population estimates. Consideration should be given to a similar review of the methods ahead of the 2021 Census.
Planning

Evaluation of other components of the processing system produced similar comments about the need to fully rehearse the systems ahead of the live operation. This is also true of the quality assurance systems. Delays in earlier processing and late fixes to the quality assurance components put some pressure on the time available to undertake the full quality assurance process that was planned. Although much of this was mitigated though the use of additional resources, it did have some impact on the implementation of some checks which required additional fixes and delays in the output systems. Therefore, for the 2021 Census, it is recommended that the rehearsal of quality assurance methodology, systems and processing takes place well in advance of live operations.

Method

Quality assurance of the 2011 Census was more rigorous and comprehensive than in previous censuses.

The independent review of the coverage assessment and adjustment methodology that took place in 2011 stated that “the methods give confidence that the resulting final census population estimates will be better than any other method and will be suitable for use in resource allocation and planning”.

The methods, procedures and participation in the quality assurance process were very successful in 2011, with some innovative use of people and systems to aid the quality assurance process. However not all aspects of the results were quality assured, and some issues arose after outputs were released and experienced users began detailed analyses; these highlighted areas for future consideration. This will always be the case given the sheer volume of information to process and check (more than 8 billion census statistics have been published), and the difficulties with trying to automate the interpretation of rare events or respondent circumstances.

- Earlier understanding of the impacts of all process – such as Edit and Imputation – would enable better understanding of their possible impacts on the data. For instance, topic QA was used to highlight implausible numbers that were a direct result of some of the earlier processes.

- Checks need to be reviewed to ensure that diagnostics are plausible – e.g. response rates – can very high or low response rates be sufficiently explained?

- More consideration of the output content when developing checks would be beneficial. For example, checks on workplace should also consider the plausibility of the flows to that workplace – this would highlight implausible flows which may not affect the results, but can impact on user confidence in the overall quality of the census.

- There were several issues with single year of age patterns which were not anticipated. This was as a result of not considering them as a part of the design of earlier processes. Future census designs should consider this. Key checks need to be in place (for example for the very old) and checks need to be done to the highest single age possible.

- Whilst the comparisons at local authority level were the focus, and worked well, comparison with the previous census, particularly at lower levels, is recommended as this may highlight some small area issues – e.g. the placement of communal establishments.
• Initially the focus of the topic QA was on using comparators; unfortunately the use of social survey comparator data proved to be unreliable when looking at LA level and below (mainly due to the smaller sampling size of the social surveys). As the QA progressed it was realised that cross checks of data between processing stages was proving more useful and so the focus of the project switched to this. If applicable, use of the 2001 Census seemed to be the most reliable comparator to use for checking the distributions, along with topic lead knowledge to identify expected shifts in the distributions.

• During topic QA, cross checks between the different processing stages proved to be very useful and helped to identify issues found in the data such as married babies and single year of age spikes.

Data

• In particular, local authority administrative data were of great value. It would be beneficial to have local authority Council Tax (CT) data well in advance of operations. It would also be valuable to ensure that all data and sources used for operational purposes (for example for the construction and validation of the address register, for management of the field operations, operational data from census including management information, and all associated census data such as images), are also available for quality assurance.

System

• An outstanding success was using electronic packs at QA panels allowing panel members to explore, analyse and visualise the data during meetings. Maximising the use of available technical innovations generally had a positive impact on the process.

• It is recommended that a flexible, more integrated approach be taken to system development. Analysts and developers should work more closely to develop systems that are flexible and provide change facilities that more readily respond to the complexities of the data being processed.
7.2 Operational assessment

The operational solution was considered to be very successful, particularly the way the panels were used and the additional, fresh level of challenge and subsequent assurance they provided.

The most notable operational aspect of the QA process was the way that the QA panels functioned. Overall the panels worked very well and gave confidence in the final estimates. The panels provided a rich source of challenge and subsequent assurance both internally in the sign-off of the estimates and externally to users.

The main QA panels required considerable training and briefing around estimation issues during the meetings so that they understood what could be achieved and what could not. A lesson for the 2021 Census is that members of the QA panel should be trained and briefed using actual data and QA materials, to help their understanding and to clarify expectations around the likely outputs and communications.

The mix of assurance from internal and external colleagues was very valuable and a similar approach should be considered for the next census in 2021. However, the structure, timing and membership of the panels should be reviewed in the context of the methods and timetable for the processing and publication of the 2021 Census.

7.3 Output assessment

The supporting material that was published with the first release of population estimates (available from the ONS Website\textsuperscript{38}) was considered to be highly successful. Providing this information alongside the population estimates helped users to understand the quality of the estimates (such as how much adjustment was applied by area or response rates) and place the estimates in the context of other administrative sources.

A similar approach should be considered for the census in 2021, but with consideration to the relevant content and delivery mechanisms appropriate to the methods employed in the next census.
8. Conclusions

It was a significant achievement to design, implement and apply a quality assurance methodology that contributed to the production of census estimates that are of extremely high quality, are widely accepted, and that were delivered to time. The 2011 QA methodology used the framework laid down in 2001 and improved upon it, to deliver a more robust methodology. This provides a way of developing new methodology and accompanying systems that can be harnessed for future censuses (including the census in 2021) and large scale population statistics projects.

Overall, the process to quality assure the results was highly successful and met its primary objectives. Most importantly, the methods and data sources used were transparent and gave users confidence in the process and therefore the census estimates.

There are some important lessons to learn from the 2011 Census quality assurance methodology; elements that went well and should be considered for the proposed 2021 Census and others that did not go so well, but are just as useful as input into future planning.

In summary, the key lessons learned from the quality assurance methodology assessment are:

- frequent and open engagement with users was a major contributory factor in improving the QA methods (from 2001) and building user trust in the results
- independent assessment was an important tool in providing assurance to ONS and users that the methodology and the quality assurance would produce robust estimates
- rehearsal of the QA methodology should take place as early in the process as possible
- earlier understanding of the impacts and interdependencies of all processes - such as Edit and Imputation – will enable a better understanding of the possible impacts on the data
- more consideration should be given to the output content when developing quality assurance checks
- use of external experts worked very well, and early engagement with them paid dividends
- the use of QA panels worked well, with the panels providing a rich source of challenge and subsequent assurance both internally in the sign-off of the estimates and externally to users
9. References

1. 2011 Census quality strategy

2. Data quality assurance

3. 2011 Census General Report

4. Independent assessments

5. 2001 (One Number) Census quality assurance process


7. 2011 Census glossary

8. ONS (2012a) 2011 Census Coverage assessment and adjustment process

   ONS (2012b) 2011 Census Coverage Survey summary

9. Local Authority population studies 2004

10. A Quality Assurance and Contingency Strategy for the 2001 (One Number) Census
11 2001 process and outcomes reviewed by the Local Government Association

12 Lessons from the 2001 Quality Assurance process

13 ONS (2012c) Tuning the Coverage Estimation Process

14 How did we do in 2011? - Evaluation papers for the other 2011 Census processes

15 2011 Census Advisory Group papers

16 2011 Census Advisory Group papers on data quality

17 Minutes and actions from the 2011 Census Advisory Groups

18 2011 Census roadshows (September 2011) – questions and answers

19 2011 Census Methodology tutorials – feedback (questions and answers)

20 2011 Census Methodology workshops – June/July 2012 - information
21 Data quality assurance - Further information on the local authority QA studies

22 Evaluation of the local authority QA studies

23 An Independent Review of Coverage Assessment, Adjustment and Quality Assurance

24 Quality assurance of 2011 Census population estimates

25 Methodology for Quality Assuring the 2011 Census Population Estimates

26 Quality assuring the 2011 Census population estimates

27 ONS (2012d) Household bias adjustment

28 ONS (2012e) Within-household bias adjustment.

29 ONS (2012f) CCS sample balance adjustment

30 Making a National Adjustment to the 2011 Census (July 2012)
31 Estimation and Adjustment for Communal Establishments

32 Independent Review of Coverage Assessment, Adjustment and Quality Assurance

33 UK Statistics Authority website
http://www.statisticsauthority.gov.uk/

34 2011 Census interactive

35 What the census tells us – 2011 Census analysis

36 ONS 2013/14 customer satisfaction survey

37 Census Regional Champions – Evaluation of local authority engagement

38 Supporting material published with the first release of population estimates

39 Results from using routinely-collected government information for 2011 Census quality assurance
10. Annexes

10.1 Annex A: Detailed checking and data sources

This Annex A provides a more detailed specification of those checks identified in the methodology section of the document (Chapter 6).

A series of tolerances was derived for each of the core checks to help distinguish between potential errors in the census estimates and definitional/coverage issues with the comparators.

A pre-defined tolerance range for each of the core comparator checks was used to identify potential discrepancies between census estimates and the comparator data. Tolerances were used to identify where more detailed ‘supplementary’ quality assurance was needed.

Using tolerances recognised that although comparator data provide a good estimate of the indicator in question, the comparators themselves might contain errors 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 improved during the quality assurance process. During the quality assurance process tolerances (and the associated guidance) were reviewed to assess how effectively they identified census estimates requiring further investigation.

**Tolerance methodology**

Four types of method were used to set tolerances:

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

Further details of tolerance methods can be found in: 2011 Census Methodology for Quality Assuring the Census Population Estimates.
10.1.1 Core checks and data sources at LA level

Checks against comparator sources

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

Comparator sources:
• Birth registration
• ONS Mid-year population estimates
• NHS Patient register
• Department for Education (DfE) and Welsh Government (WG) School census
• HM Revenue and Customs (HMRC) Child benefit data
• Department for Work and Pensions (DWP) Pension claimants data
• Customer Information Service data from DWP/HMRC on interactions with benefits and taxations systems

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 was assessed across each LA. Differences between the census and comparator sources across LSOAs and MSOAs were reviewed to identify any systematic differences (for example where the census estimates were 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 was checked based on census household returns. This could only be done where the comparator sources indicated whether an address was occupied (for example on the patient register).

Included in the check was a household count made up of the number of occupied census household returns and ‘dummy forms’ (which were completed by census field staff for non-responding addresses) which indicated that an address was a main residence. It also included addresses where no information was collected - with an assumption made about which of these addresses was a main residence.

Other sources which do not identify occupancy were also be included on the check for the sake of completeness. This included both the 2011 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
• Department for Communities and Local Government (DCLG) Household projections
• Alternative household count
d. Household Size
Census estimates of household size were checked by assessing average household size and the frequency distribution of household size. The frequency distribution was particularly useful to identify whether census household continuation forms were used (or were not used as intended). The main household census form had only space for six individuals, a sudden drop between the number of six and seven households might have indicated that large households often did not request a continuation questionnaire. This was validated against comparator data.

Comparator sources:
- NHS patient register
- Integrated Household Survey
- ONS population projections and Department for Communities and Local Government (DCLG) on interactions with benefits and taxations

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

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

An assessment of the number of census forms returned was 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 were checked.

Comparator sources:
- ONS Population estimates by ethnic group
- Department for Education and WAG School census
- Integrated Household Survey
g. Students
Census estimates of students in full-time education aged 18 or over were checked by single year of age and sex. This included students in further and higher education. The check covered all students (regardless of whether they lived 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 were carried out within an LA at MSOA level.

Comparator sources:
• Higher Education Statistics Agency data
• Department for Business Innovation & Skills (BIS) and Welsh Government (WG) 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 assessed 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 included both personnel resident on the base and in private accommodation. The comparator data for both UK and US armed forces was by base only. There was no information about where individuals are resident. It was necessary to apportion armed forces personnel by base to their LA of residence using the 2001 Census. The process of apportioning was used in the production of mid-year population estimates but was 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 was 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 and Wales and from Northern Ireland and Scotland are based on the address one year ago question. Internal migration to each local authority was checked by age and sex. It was not possible to check internal migration from each local authority until all areas were processed.

Comparator sources:
• NHS Patient register

j. Migration (International)
The address one year ago question was used to identify international migrants who previously lived outside the UK. The number of short- and long-term migrants were both checked by age and sex. Short-term migrants were 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 were calculated based on census population estimates of women aged 15-44. Calculating and analysing rates of fertility and mortality was a useful alternative way of assessing census population estimates given the accuracy of compulsory birth and death registration. These rates were 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) were 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 were checked against existing MYEs and other comparator sources. Analysis of fertility also assessed variation in fertility rates within each LA.

Comparator sources:
• Births registrations data
• ONS Mid-year population estimates

b. Mortality Rates
Similar comparisons were made for mortality rates as set out for fertility. Age-standardised mortality rates were calculated using census population estimates and compared to the equivalent rates calculated using mid-2011 population estimates. Analysis was 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 was checked by single year of age, to see whether the census population estimates resulted in implausible values.

Operational checks
Operational intelligence also included core checks on the census return rates including:

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

Where potential problems were identified a range of additional intelligence was considered. Additional intelligence included the findings from debriefings with area managers and local authorities on any problems with the field operation. It also drew on the detailed information provided by the questionnaire tracking system.
10.1.2 Core checks at region and national level

a. Aggregate checks against comparator sources
All of the core checks listed in the section above (at LA level) could be aggregated to regional and national levels. The age/sex and household checks were routinely repeated at these levels as these were central to the quality assurance process.

At regional level the number of armed forces personnel by base was routinely checked. It was assumed that region of residence (from the census) was directly comparable to workplace (from the comparator data). No comparator data were available which directly identify where armed forces personnel were resident, only on where they were 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 were repeated at both regional and national levels. Experience from 2001 suggested that sex ratio analysis would be particularly valuable. For 2011, the ONS Centre for Demography (ONSCD) undertook research to determine plausible national age/sex ratios before the census QA process. Work focused on sex ratios based on an analysis of the Department for Work and Pensions’ (DWP) Lifetime Labour Market Database (known as L2). The L2 was a one per cent sample of National Insurance numbers. For these individuals, interactions with both Her Majesty’s Revenue and Customs (HMRC) and DWP were linked to provide a longitudinal picture of economic activity (including registration for benefits).

c. Analysis of the ONS Longitudinal Study
The Longitudinal Study (LS) contains linked census and vital events for 1 per cent of the population of England and 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 and Wales level. Recognising the potential value in analysis of the LS for quality assurance, ONS brought forward the process of LS matching to the 2011 Census. As a result of bringing forward this matching the evidence was checked during the quality assurance process.

d. Census Non-Response Link Study (CNRLS)
The CNRLS matched census respondents to relevant ONS social surveys, including the Integrated Household Survey. Matching identified those people who responded to a survey but not to the census. This included 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 were useful to assess whether there was systematic under-coverage of individuals from households where a census form was returned.

e. Modelling non-response
The coverage adjustment process was 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 were monitored as they emerged in the accumulating data. This modelling assessed 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 and Wales level the census population estimates were compared to mid-year population estimates rolled forward from the 2001 census. Any difference between these estimates (by sex) was 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 was undertaken as part of the preparations for the quality assurance.

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

Additional operational intelligence
Information from the census and census coverage survey field operation was reviewed to identify systematic coverage issues which may have been reflected in the census estimates. In addition to information on return rates discussed above this included:
- Notes from field staff debriefs
- Information about field incidents
- Information from the questionnaire tracking (QT) system
Diagnostics from the data processing stages was also available (such as the amount of data editing and imputation which was carried out).

Area profiles
Background information was been compiled for each LA to help understand recent demographic, social and economic change. This included how official estimates of population changed since the 2001 Census with detail on births, deaths and migration over this period. This was set alongside change in other sources such as patient registers and housing stock. To provide context, the information provided for the LA was set against the wider region and against estimates for England and Wales as a whole. A summary of correspondence between each LA and the ONS was made available for reference. This included 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 were directly involved in the preparation for census quality assurance. This 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 was 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 local authority 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 subsequently requested this information from all 348 LAs, referring to March 27th 2011. All LAs compiled a Census Local Partnership Plan (CLPP) in conjunction with the census area managers which included data and intelligence of potential use in quality assurance.
10.1.4 Supplementary checking

Where the core checks at LA level identified that further analysis was required, more detailed supplementary quality assurance was carried out. The further analysis to be undertaken depended on which of the core checks highlighted potential problems. It included analysis of specific geographic areas within an LA and analysis of specific population subgroups. This section provides an overview of supplementary analysis carried out.

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 carried out are provided below.

Households and 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 stated that they had a second residence within the LA, based on census questionnaires completed at a different usual residence. (Not all second residences identified in this way were available until all questionnaires were processed.)
- The above information collected from census forms was matched to assess over-count, the accuracy of dummy form information, and the accuracy of questionnaires returned from second residences. Accuracy was also 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 and 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
Analysis of Specific Geographic Areas
The accuracy of census population estimates at local authority level was assessed as part of the core check work carried out for all areas. Estimates were also checked at LSOA and MSOA as part of this work. To fully understand differences between census estimates and comparator data it was in some cases necessary to undertake supplementary analysis at even lower geographic levels. In some cases this involved 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 incorporate detailed studies ahead of the planned first release in July 2012. Analysis of specific geographic areas involved two stages, 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 were subject to rigorous quality assurance. It was in these postcode clusters that Dual System Estimation (DSE) was the most robust. The estimates were particularly important as they were 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, were compared with aggregate statistics from each administrative dataset for the same postcode cluster. This analysis, carried out for individuals and households, made use of confidence intervals around the census estimates.

b. Unit Record-level Matching
Record-level matching between census and administrative sources helped to validate census population estimates. Initially this work focused on CCS postcode clusters, for which ONS had the most comprehensive snapshot of the 2011 Census-day population and which were used to adjust for undercount across estimation areas. The same approach was applied to non-CCS areas where census counts posed QA challenges.

Address matching between Valuation Office Agency data, electoral rolls, the GP NHS patient register, School census, CCS address lists and the Address Register helped to identify addresses that are held in common or are unique. This involved both automatic matching and the clerical resolution of potential matches and unmatched residuals. A set of rules was devised to identify addresses that administrative sources suggested were occupied but which had been missed by census processes.

Individual person matching between Census, CCS, GP NHS patient register, School census and Higher Education Statistics Authority data created additional evidence based on administrative sources for validating Census/CCS matches and dual system estimates. This analysis drew 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 had no corresponding census record at the same address. Adjustments were made for short-term migrants. Analysis of the linked data provided some evidence of within-household under-enumeration.
10.2 Annex B: Glossary of sources used in the QA process

Compulsory Registration

- Birth registration
  Compulsory registration of births in England and Wales available at record level. QA also uses 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 and Wales at record level. Age specific mortality rates use age at death.

Administrative Sources

- GP NHS Patient register
  Registration with a GP in England and 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.

- GP 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 and PAYE, DWP Benefit Systems and 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.