



## **Economic and Social Council**

Distr.: General  
5 May 2010

Original: English

---

### **Economic Commission for Europe**

#### **Conference of European Statisticians**

#### **Group of Experts on Population and Housing Censuses**

##### **Thirteenth Meeting**

Geneva, 7-9 July 2010

Item 3 of the provisional agenda

##### **Plans for census quality evaluation**

### **A Quality Assurance Strategy for the 2011 United Kingdom Census**

**Note by the Office for National Statistics, United Kingdom**

#### *Summary*

Quality is at the heart of planning the United Kingdom 2011 Census, and the focus of many of its innovations is about improving the quality of the census processes and, thus, the data. This paper summarises two central strands of the ONS Data Quality Assurance (QA) Strategy which will contribute to the improvement of population counts and structures and the information on topic characteristics collected on the Census.

## I. Introduction

1. A census, by its nature, is designed to cover the whole population, but errors inevitably arise. A good census design will recognise this and take account of them. For example, the Address Register – no matter how up to date – may fail to record all residential accommodation in an area at the time of the Census; field staff may fail to identify all households within a multi-occupied dwelling; not all members of a recorded household may be included; non-response bias occurs when people do not complete a census form or do not answer all the questions; and there may also be system or processing errors.

2. The improvements which are being introduced for the 2011 Census have been designed to minimise these errors; for example:

(a) the creation of an up-to-date Address Register agreed with stakeholders, together with a form-tracking system functioning in real time, will provide a reference to ensure that questionnaires are delivered to, and returns collected from, all residential addresses and that monitoring of return rates can be done in real time;

(b) the re-design of the Census questionnaire and the facility to return it online will make it easier for respondents to complete the form and for ONS to process the data; and

(c) the focusing of field resources in areas of lower response is aimed at improving the overall coverage.

3. Non-response or under-enumeration is the most significant error. While the estimated overall under-coverage in the 2001 Census (6 per cent) was small compared with national government surveys, it was greater than in 1991 (4 per cent), reflecting an international pattern of increasing non-response in the Census. But of more significance was its variation across population subgroups and different parts of the country.

4. The aims for 2011 are to maximise the overall level of quality of data and to minimise the differences in quality between areas. ONS aims to develop a strategy for assuring the quality of the 2011 Census database before any release of output.

5. To achieve those aims, a series of critical success factors (CSFs) are being developed which will give tangible evidence of whether those aims have been met. Quality throughout the census operation will be managed using a quality model (defined in the Office for National Statistics 2011 Census Quality Strategy<sup>1</sup>) that involves:

(a) design quality;

(b) operational quality management;

(c) quality assurance;

(d) quality measurement and reporting; and

(e) the production of high quality population statistics.

6. This paper summarises the two central strands which Data Quality Assurance (QA) Strategy which contributes primarily to the third and fourth strands. Further detail can be found in the full strategy document<sup>1</sup>. Reference is made to two specific aspects of quality assurance, demographic (population counts and structures) and topic (characteristic information collected on the Census).

7. Quality assurance will start in the two to three weeks before census day, when management information and early census returns will provide early evidence of response patterns and characteristics, and will continue through to the publication of outputs.

## **II. Quality assurance of population counts and structures (Demographic QA)**

9. Demographic QA is concerned with ensuring that where census population estimates and structures diverge from non-census sources, including mid-year population estimates, the differences are understood and can either be explained or prompt a review of the census estimates.

10. The 2011 Census will build on 2001 methodology to adjust for under- and over-enumeration<sup>2</sup>.

11. Individuals and household records from a large survey, the census coverage survey (CCS), of approximately 500,000 individuals, will be matched against census records (see Box 1). Dual system estimation will estimate census undercount and over-count and further modelling will create estimated populations. Population totals for five-year age groups by sex will be estimated for each local authority district (LAD). Record imputation will add the population estimated to have been missed to the census database and item imputation will follow, to correct data anomalies arising from the coverage imputation.

### **Box 1. The 2011 Census Coverage Survey**

The Census Coverage Survey (CCS) will be the key source of information on the extent and distribution of the census undercount. The CCS will be a separate sample survey currently planned to be carried out over a three to four-week period starting six weeks after the Census itself, and will take the form of a short interview to check on the coverage of households and people within households, and to collect basic demographic characteristics (such as age, sex, marital status, ethnic group and economic activity).

The results of the CCS will then be matched, at the individual level, to the corresponding 2011 Census data, identifying the number and characteristics of those missed in the Census. The combined Census and CCS information, along with statistical models, will be used to produce an estimate of the numbers of people missed by the Census. The people and households estimated to have been missed will then be added to the database using similar techniques and processes to the edit and imputation stage.

A similar approach was taken in the previous census, but for 2011 there are added dimensions that were not present in 2001. In particular, the intention to collect information on visitors and second residences, in order to improve understanding of changing patterns of usual residency, will tend to complicate coverage assessment. For example, through matching visitors to Census and CCS records, additional information may be available to help identify residents who were recorded neither in the Census nor in the CCS. Furthermore, matching information from administrative sources to the Census so as to better inform the assessment of coverage (and quality) is being considered. Matching the Census database to itself, particularly where people give details of second addresses, is also being considered to assess over-count.

The information obtained from the survey will be used, in conjunction with the Census data itself, to help produce a consistent set of census-based counts, which will form the new base for the series of annual mid-year population estimates for local and health authorities.

This approach will develop and improve the methodology adopted for the 2001 Census that was widely welcomed by users. ONS will consult and inform users on the methodological

approaches to be adopted, and will assess and report on the quality of the data produced.

A small, separate Quality Survey will be undertaken after the Census to measure the accuracy of responses to individual questions.

12. A number of comparator data sets, including various administrative sources and mid-year population estimates, will be used to create expected values for the population distributions and a number of key indicators. The pre-adjusted, raw census distributions will also be compared against the estimated populations. Key indicators will include, but not be limited to:

- (a) population aged under one;
- (b) population by five-year age group up to the age 85+, by sex;
- (c) sex ratios by age group;
- (d) number of households;
- (e) household size;
- (f) fertility rates; and
- (g) mortality rates.

13. This LAD-based analysis will be supplemented with a comparison of population distributions and key demographic indicators for each LAD against a growing cumulative total, at regional or national level.

### **III. Quality assurance of census topics (Topic QA)**

14. The principal focus will be on variables that create population and housing counts and those used to support key uses of census demographic information, for example mid-year population estimation and the distribution of resources to local authorities. This prioritisation follows the practice used by Statistics New Zealand<sup>3</sup> and has involved categorising census variables as high, medium or low priority. A lesson from 2001 census was that data quality monitoring needs to be targeted to identify errors that might have a substantial impact on outputs rather than getting bogged down on quality issues relating to few records or lower priority variables.

15. Pre-programmed checks will be made on census data at different geographic levels and at different processing stages. Checks that are actionable will be the highest priority. Automating the checking means that data validation will be faster, allowing more scope for correcting problematic processes (the most likely solution being through edit and imputation) and freeing up resource in the quality team to focus on unexpected or anomalous results. The full list of checks and processes being checked will be agreed in advance.

16. In 2001, a number of 'soft checks' were made on the data, for example numbers of step-children. It is proposed to check the plausibility of these distributions as part of the data QA.

17. At the stages of data processing identified by the checking framework, 2011 variables will be compared with external sources. Comparator datasets will be identified through consultation with ONS specialists and topic experts. Variables known to vary geographically will have additional checks for implausible distributions. An example would be farmers in occupational data for Greater London.

18. The quality team will identify, through consultation with colleagues in the ONS Methodology Division, tolerance levels within which 2011 Census data will be permitted to fall. Those outside of the specified tolerances will be referred for further investigation. A Red-Amber-Green system built into the automated checks will highlight distributions that can be accepted with caution and those that need further checks. It is likely that tolerance levels will vary by geographic level and area, with greater variation and therefore higher tolerance anticipated where areas have experienced rapid recent population change, as signalled by the latest information on population flux from the ONS Centre for Demography. Consideration will be given to adjusting comparator data in areas that have experienced high known levels of population change since the data were originally captured. The QA team will draw on expertise in ONS Methodology to draw up a unified approach to setting tolerance levels within the various census topic areas. Topic experts will approve expected values for variables relating to their subject specialism.

19. The topic QA will monitor and collaborate with the demographic-QA processes on population subgroups including but not limited to:

- (a) babies aged under one;
- (b) young men, who had high rates of missingness in 2001;
- (c) over-85s;
- (d) those serving with UK armed forces;
- (e) those serving with foreign armed forces;
- (f) international migrants, both long and short-term;
- (g) prisoners;
- (h) residents and staff in communal establishments;
- (i) minority ethnic groups;
- (j) same-sex couples;
- (k) people in civil partnerships;
- (l) private renters;
- (m) students;
- (n) those with a second address;
- (o) non English speakers; and
- (p) internal migrants.

20. Special attention will be paid to areas where problems of enumeration are anticipated, or where examination of management information from field operations, including the results of address checking, has highlighted particular difficulty in enumeration, for example:

- (a) areas with lots of holiday homes;
- (b) areas containing caravan sites;
- (c) areas with large numbers of second residences;
- (d) areas with high multi-occupancy;
- (e) regeneration areas; and
- (f) areas with low enumeration or high variability.

#### **IV. Improvements in data QA since the 2001 Census**

21. This paper has set out the two central elements of the QA strategy, demographic and topic QA. These elements should be seen in the context of the wider QA strategy previously published by ONS and enhancements to data quality introduced since 2001.

22. The methodology for 2011 QA has explicitly drawn on 2001 experience, through the examination of 2001 policies, systems and publications, a review of the literature following the 2001 census and interviews with those responsible for 2001 QA. The lessons learned have shaped the proposed strategy so it should cope with data quality challenges that can be anticipated because they were present in 2001. Unanticipated challenges, such as a 2011 variant of the foot-and-mouth epidemic, will be the ultimate test of the strategy's effectiveness.

23. Specific enhancements in 2011 include:

- (a) an enhanced address register against which returns will be monitored;
  - (b) availability of management information, including 'soft' data from field operations, as qualitative and quantitative evidence to inform the quality of LAD estimates;
  - (c) explicit and planned prioritisation of QA processes;
  - (d) planned unit record linkage to address quality concerns prior to data release;
- and
- (e) use of newly available administrative datasets for validation purposes
  - (f) integration of topic and demographic QA within a single strategy, carried out by a single team.

#### **V. References**

2011 Census Data Quality Assurance Strategy (2009) <http://www.ons.gov.uk/census/2011-census/process-info/data-quality-assurance/2011-census---data-quality-assurance-strategy.pdf>

Abbott, O. and Brown, J. (2006), "A review of the 2001 One Number Census methodology and lessons learnt." Paper presented at GSS Methodology Conference, London, June 2006 [www.statistics.gov.uk/events/gss2006/downloads/D1Abbott.doc](http://www.statistics.gov.uk/events/gss2006/downloads/D1Abbott.doc)

Statistics New Zealand (2008) 2006 Quality Management Strategy (QMS) Summary Report. Available at [www.stats.govt.nz/census/about-2006-census/methodology-papers/quality](http://www.stats.govt.nz/census/about-2006-census/methodology-papers/quality)