

ONE NUMBER CENSUS STEERING COMMITTEE**A Quality Assurance and Contingency Strategy for the One Number Census**

1. This paper describes the QA and Contingency strategy for England & Wales and Northern Ireland agreed at a meeting of senior managers held on 13th October 1999.
2. GROS will adopt the same general approach but they plan to implement slightly different contingency actions – these are outlined in Appendix A of the paper.
3. **The Steering Committee are asked to:**
 - a) **note the paper;**
 - b) **endorse the proposed Quality Assurance and Contingency Strategy for England & Wales and Northern Ireland; and**
 - c) **comment on the differences to the strategy outlined for Scotland.**

**Marie Cruddas
Census Division
Office for National Statistics
Room 4200 W
Segensworth Road
Titchfield
Fareham
HANTS
PO15 5RR**

February 2000

A QUALITY ASSURANCE AND CONTINGENCY STRATEGY FOR THE ONE NUMBER CENSUS

1. INTRODUCTION

The purpose of this paper is to outline procedures to Quality Assure (QA) the One Number Census (ONC) population estimates following the 2001 Census.

The paper also describes, at a high level, the Contingency Strategy to address the situation that the QA process finds evidence to suggest that the ONC based population estimates are not consistent with the rolled forward Demographic Estimates.

The paper considers these issues by describing the following:

- * The ONC estimates, Demographic Estimates and Administrative Registers involved in the comparisons;
- * How the estimates will be compared;
- * The planned resolution in the event of a difference (The Contingency Strategy)¹;
- * Consultation that will result in a set of quality assurance procedures widely accepted by users.

The paper describes the procedures proposed for England & Wales and Northern Ireland. (For ease of reading, the paper generally refers to data sources and administrative units in England and Wales; these references should be interpreted as also referring to the equivalent data sources and administrative units in Scotland and Northern Ireland.) In general, GROS propose to use similar quality assurance procedures, but there are some differences between the contingency strategies proposed for Scotland and those for the rest of the UK (see Appendix A for a description of the GROS QA procedures and Contingency Strategy).

1.1. Background

Following the 1991 Census it was acknowledged that the Census suffered from a degree of undercount the extent and nature of which was not identified by the Census Validation Survey (CVS). To avoid a similar situation following the 2001 Census, development and planning has focused on two main areas:

1. Maximising coverage in the Census;
2. Developing a contingency should the Census not achieve 100% coverage.

In order to address the second point, the One Number Census Project was initiated. This project is concerned explicitly with the measurement of underenumeration and development of procedures for adjusting the census for the undercount. It ensures that the most appropriate estimation procedures are used. The work is conducted by the ONS in collaboration with academics from Southampton University and additional consultation with experts in the fields of matching and dual system estimation. The work has been subject to expert review at all stages resulting in the design of the Census Coverage Survey (CCS) together with the associated estimation methodology. Thus the CCS and the ONC methodology combine to constitute the 'contingency' for the 2001 Census.

¹ The description of the strategy assumes that the ONC only gives underestimates of the population. The strategy works equally well with overestimates.

1.2. The need for a further Contingency Strategy

Throughout the ONC a number of features have been developed to assure the quality of the process. One of these is the need to have a strategy to account for the possibility that the results of the ONC estimation may not be plausible either in some areas of the country or, indeed in the nation itself. This could be the result of dependence between the observations in the Census and the CCS or because of fieldwork problems in the CCS. This paper discusses this quality assurance process

2. ONC ESTIMATES AND DATA SOURCES

This section simply lists the data to be used as part of the quality assurance process. As well as the ONC estimates, use will be made of demographic estimates and administrative records.

2.1 ONC Estimates

Those parts of the ONC Project relevant to this paper are briefly summarised as follows:

- * Design Groups are clusters of Local Authority Districts (LADs) which contain about 500,000 population. A Design Group may comprise a single large LAD.
- * Within each Design Group ONC estimates are derived for age-sex groups with associated variances – these are calculated from combined Census and CCS data using a dual system / regression estimation approach.
- * From the Design Group estimates subsequent LAD estimates are calculated using synthetic estimation techniques. These are constrained to the Design Group totals.
- * The estimates for all Design Groups are summed to give the 2001 Census based National population estimate.
- * The Census based National and Design Group population estimates, by the age-sex groups, will have associated error bounds.

The ONC estimation methodology is not described here. For an overview see Brown et al (1999).

2.2 Demographic Estimates

Demographic Estimates produced by the Population and Vital Statistics (P&VS) Division of the ONS will be made for 2001. All Offices (ONS(PVSD), NISRA and GROS) are developing population estimates to use as comparators in the quality assurance process.

It is intended that there will be a plausibility range around all population estimates and current work is investigating this. Two strategies are being investigated: Firstly, using advice from an independent panel of experts, upper and lower variants of the national population are being estimated, including variants on levels of fertility, mortality and migration as well as on various hard to count groups such as refugees; these intervals involve a degree of subjectivity and the methodology used to produce them is not suitable to produce upper and lower bounds at sub-national levels for which a different strategy is required. At a sub-national level, the assumption that for similar types of areas, errors will broadly be constant between 1991 and 2001 is being examined.

The Demographic Estimates make some use of the higher quality Administrative Registers and provide the best plausible single comparators for QA purposes. However, research is continuing into the further potential for use of administrative records. These will provide important comparators for specific age groups or areas and more detail is given in Appendix B. Although

alternative, independent, population estimates are being used in a quality assurance role, they are only a guide to detecting possible implausibilities in ONC estimates. This is because there is currently no alternative to the Census that can provide a 'Gold Standard' population estimate.

2.3 Administrative Registers

Aggregate level data from Administrative Registers will also be available for comparisons. Registers such as the FHSAs can be aggregated to give National, Design Group and LAD level counts of the population by specific age-sex groups. However, it should be emphasised that these national registers are administrative data that are not maintained for the purpose of making population estimates and that there is evidence of list inflation at the national level. Other Registers focus on population sub-groups that are believed to be at particular risk of underenumeration in the Census: students, the elderly and young children. The availability, reliability and quality of these data sources are currently being investigated within the ONS.

Table 2.1 shows the Administrative Registers under consideration with a brief description of the population covered.

Table 2.1 Administrative Registers

<i>Register</i>	<i>Population Covered by Register</i>	<i>Used in Demographic Estimates?</i>
<u>Health Service</u> 1. Health Authority	Population registered with NHS GP Practices – this provides postcoded information	No, but might be used for some migration in the future
2. NHSCR	Aggregated population registered with NHS GP Practices at Health Authority level - does not currently include postcodes	Not for population stocks, but is used to estimate migration within the UK
<u>Benefits Agency</u> 1. Child Benefits	Children aged under 16 for whom Child Benefit is paid	No
2. Retirement Pensions	Females aged 60 and over in receipt of a pension Males aged 65 and over in receipt of a pension	No (but information on those aged 85+ was used in the 1991 Census base)
Higher Education Students Authority (HESA)	All students in publicly funded higher education institutions	No
<u>Registration System</u> 1. Birth data	All children under 1 born in the UK	Yes
2. Deaths data	Deaths for the whole population that occur in the UK	Yes
<u>Armed Forces Information</u> 1. Defence Analytical Services Agency (DASA)	UK Stationed Armed Forces	Yes
2. US Embassy and RAF Mildenhall	US Armed Forces and their dependants resident in the UK	Yes
Home Office Prisons Information	All prisoners who have been sentenced and have served 6 months or more	Yes
<u>Schools Information - DfEE/WA</u> 1. School Boarders (DfEE/WA)	All pupils who board at schools in England & Wales	Yes
2. Schools Census (DfEE/WA)	All pupils in schools in England & Wales	No

Appendix B contains further information on the main Administrative Registers in Table 2.1 together with an indication of their potential use.

3. COMPARISON OF THE ONC ESTIMATES

3.1 Background

Quality Assurance will occur at three levels: National, Design Group and LAD. Where the quality assurance procedures outlined below identify a 'difference' at any of these three levels, the Contingency Strategy will be invoked, as described in Section 4.

3.2 Timetable Implications

The QA/Contingency Strategy has crucial implications for the ONC timetable. Sub-national population estimates must be supplied to P&VS by August 2002 to comply with resource allocation requirements.

To date it is unknown when the data will be supplied to ONC for the matching process to begin. It is imperative that sufficient time is allowed for the QA/Contingency Strategy to take place to allow production of high quality output. Therefore, it is important that an agreed plan is put into place well in advance of the 2001 Census date.

Design Groups form the processing block for both the Census and the CCS. Thus the ONC processing will form a sequential process i.e. Match, Estimate, Quality Assure. In an ideal world we would QA each Design Group, identify any deficiencies and delay resolution until all Design Groups had been Quality Assured and the ONC National estimate of the population was available.

However, this is not feasible as time scales will be tight in 2001. In order to maximise efficiency, it is proposed that the Contingency Strategy be invoked as soon as a decision has been made and there is a need for adjustment of the ONC estimates at a sub-national level.

3.3 Quality Assurance Process

The QA process will comprise demographic analyses. These analyses will be initially developed by the ONC team and then be the subject of a consultation process with the user community, planned to take place in the summer/autumn of this year.

The process will be based on the following:

1. Demographic analyses will be undertaken at all three levels of aggregation. These will include analyses of sex ratios, dependency ratios and the components of any adjusted figures.
2. At specific age groups certain administration records will prove robust. For example birth registrations are excellent comparators for the very young; similarly at a local level, school rolls may prove useful and at a national level pensions data are likely to provide useful comparators.
3. The distributions of absent households and estimated CCS response/nonresponse rates will be examined.
4. Broad comparisons will be drawn with (a) the demographic estimates from P&VS together with the plausibility ranges on which work is progressing (see section 2); (b) estimates of special groups such as Armed Forces Personnel. Each comparator must be such a good measure of the

population that *if the census estimate is significantly wrong*, then the comparison is very likely to show this up.

The initial unit of comparison will be the Design Group but acceptance of a Design Group will be contingent on an analysis of the constituent LADs.

The QA process for each level of data is similar and hence that for a single Design Group is described. When the QA process moves to the smaller Local Authority District level we must be aware that the other comparators are only broad indicators and be cautious about their reliability. It is at the National level that the comparators are most reliable.

The QA process is sequential, and best described by the following algorithm. We start with a single Design Group:

1. If the Design Group estimates are consistent then consider the constituent LADs:
 - 1.1. if the LADs are consistent then move to next Design Group;
 - 1.2. if the LADs are not consistent make adjustments and re QA the Design Group;
2. If the Design Group estimates are inconsistent then consider the constituent LADs:
 - 2.1. if LAD/s inconsistent adjust LAD/s estimates and recommence QA;
 - 2.2. if LAD/s consistent adjust Design Group then recommence QA.
3. If all Design Groups have been Quality Assured and agreed then sum to form ONC National estimate
 - 3.1. if National estimates consistent then finish
 - 3.2. if National estimates not consistent then adjust and QA

This will be further discussed in Section 4. Thus whatever level of data is under consideration, the QA will comprise a series of standard demographic techniques.

It is important to note that decisions on whether or not to adjust a specific LAD or Design Group will be made on a holistic evaluation of the evidence from each of the four steps of the process described at the beginning of this section.

4. CONTINGENCY STRATEGY

This section sets out some possible adjustment strategies, should adjustment prove necessary for one or other sub-national area. It is preliminary and gives ideas of what may be possible. There are two potential strategies being considered: (i) borrowing strength at the Design Group and Local Authority District level, and (ii) merging ONC estimates with demographic information at the national level.

4.1 Borrowing Strength

The basic strategy here is to use information on underenumeration from surrounding areas in order to readjust the ONC estimates for the Design Group or LAD in question. The strategies for Design Groups and LADS are similar but the differences are such that two examples will be given.

Design Groups

In this situation the evidence will suggest that the ONC estimates for a Design Group are too low and that it appears the reason is that the dependence between the census and CCS has been high across the design group. It should be noted that this might occur for one or more Hard to Count

groups (in which case adjustment could be restricted to only those Hard to Count groups affected). In this case the strategy is as follows.

1. For Design Group (DG) 'A' the ONC estimate is 479, 000 people with a ratio estimate

2.

$$A \quad \text{ONC} = 1.01 \text{Census} \quad (1)$$

where *Census* is the Census count.

2. In three contiguous DGs, *B-D*, the ratio estimates are

$$B \quad \text{ONC} = 1.10 \text{Census} \quad (2)$$

$$C \quad \text{ONC} = 1.12 \text{Census} \quad (3)$$

$$D \quad \text{ONC} = 1.11 \text{Census} \quad (4)$$

3. Consider DGs *B, C, D* as one Design Group and re-estimate the ratio model gains

$$\text{ONC} = 1.11 \text{Census} \quad (5)$$

4. Re-estimating DG *A* using the ratio model in Equation (5) gives a revised ONC estimate of 526, 426.

5. Repeat QA procedure for Design Group 'A'.

It should of course be noted that one would expect that there would be different levels of dependence, say, for men and women. Hence it would not be sensible to make a uniform adjustment or, for example, one would end up with erroneous sex ratio. Therefore, adjustments will be made separately for each age/sex group.

Local Authority Districts

In this situation sub Design Group analyses would suggest that, say, one LAD in the Design Group would have an implausibly low adjustment. The aim, then, would be either to borrow strength from other LADs in the district or possibly (but less likely) to borrow strength from socio-economically and demographically similar LADs in adjacent or close Design Groups. In this case the strategy would entail the following steps:

1. For DG *A* comprising LADs *b, c, d, e* the ONC estimate is 490,000 with a ratio estimate

$$\text{ONC} = 1.00 \text{Census} \quad (6)$$

2. Evidence from a variety of sources means it is agreed that the estimates for LAD *e* are implausible. The ratio estimates for the four LADs are

$$b \quad \text{ONC} = 1.10 \text{Census} \quad (7)$$

$$c \quad \text{ONC} = 1.11 \text{Census} \quad (8)$$

$$d \quad \text{ONC} = 1.09 \text{Census} \quad (9)$$

$$e \quad \text{ONC} = 1.01 \text{Census} \quad (10)$$

The estimated populations are *b*: 150,000; *c*: 120,000; *d*: 100,000; *e*: 120,000. These estimates have been made by allocating the DG population to the LADs using small area estimation techniques.

3. Remove LAD e from the design group and re-estimate the DG ratio model for a DG comprising b , c and d . This becomes

$$\text{ONC} = 1.10\text{Census} \quad (11)$$

4. Use this model to re-estimate the underenumeration for e . This leads to an LAD estimate of 130,693.
5. Re-adjust the DG estimate to a population of 500,693.
6. Repeat QA procedure for Design Group 'A'.

4.2 Merging ONC estimates with Demographic Information

There are some situations which can mean that estimates at the National level are not plausible. This could happen for example if the CCS fieldwork results in high levels of dependency between the Census and CCS and hence, nationally, there would be an underestimate of underenumeration. If this were the case then there will be early indications from the QA Process.

It is necessary to have a strategy for calibration at the National Level. Two possibilities are proposed:

1. Accept the Demographic Estimates based on the 1981 Census as supplied by the P&VS; or
2. Up-rate the partially adjusted 2001 Census figures using an estimate of the dependence implied in the Dual System Estimator by the lack of success nationally

It is accepted that (1) is not an option. The working strategy starts from the position used in 1991 when the Census estimates were uprated using a deterministic model based on achieving a target sex ratio. At certain ages the Census estimates had largely implausible sex ratios. In 2001 it is proposed to use a statistical approach also based on sex ratios. The method will be an adaptation of that proposed by Elliott and Little (1999). It is based on the premise that the 'underestimate' occurs because of dependence between the results of the Census and those of the CCS in the DSE. To address this, ancillary information, in the form of a target sex ratio, is added to the dual system estimation process and an empirical Bayes estimator used to adjust the DSE.

For some specific sections of the population it may be possible to adjust the ONC estimates directly from the Demographic Estimates (e.g. babies under one year) where one is confident of reliability. This adjustment could be made at the National level but consideration must be given as to how this can be cascaded down to the sub-national levels.

7. REFERENCES

Brown J.J. et al (1999), *'Methodological Research leading to a One Number Census in England and Wales'*. JRSS Series A, 161.

Charlton, J., Chappell, R., Diamond I., *'Demographic Analyses in Support of a One Number Census'*, Proceedings of the Statistics Canada Symposium 1997.

Charlton, J. & Chappell, R. (1999), *'Uncertainty Intervals for the 2001 National Demographic Estimates'*, ONS(ONC(SC))99/05.

Diamond, I.D. (1993), *'Where and who are the "missing million"? Measuring census of population undercount'*, In Statistics Users' Council Conference Proceedings on Regional and Local Statistics, 16 November 1993. Published by IMAC Research, Esher.

Elliott, M.R. & Little, R.J.A.(1999), *'On Combining Information from a Census, a Coverage Measurement Survey and Demographic Analysis'*, submitted to J.A.S.A.

Lunn D.J., et al (1998) *'The accuracy of age-specific population estimates for small areas in Britain'*, Population Studies **52** (1998), 327-344.

OPCS (1991), *'Making a Population Estimate in England and Wales'*, Occasional Paper 37.

OPCS (1993), *'Rebasing the annual population estimates'*, Population Trends 1993.

Simpson, S., Cossey, R. & Diamond, I.D. (1997), *'1991 population estimates for areas smaller than districts'*, Population Trends 90.