Advisory Group Paper (00)16

CENSUS ADVISORY GROUPS

A Quality Assurance and Contingency Strategy for the One Number Census

1. This paper outlines the Quality Assurance (QA) and Contingency Strategy for the One Number Census (ONC) and the work that has been conducted to date on taking forward this strategy for 2001.

2. The QA and Contingency Strategy for the ONC was endorsed by the One Number Census Steering Committee in February 2000. The approach for taking the strategy forward was similarly endorsed by the Steering Committee in June 2000.

3. Although the paper mainly focuses on England & Wales, GROS and NISRA will be adopting the same general approach and discussions are ongoing to ensure that all three Census Offices use similar QA and Contingency procedures.

4. Members of the Advisory Groups are invited to:

a) note the paper; and

b) provide comments on the strategy and its implementation either at the forthcoming Advisory Group meetings or in writing to the address below within 2 weeks of the meetings.

Emma Wright
Census Division
Office for National Statistics
Room 4200W
Segensworth Road
Titchfield
Fareham
HANTS
PO15 5RR

Tel: +44 (0)1329 813507
Fax: +44 (0)1329 813407
E-mail: emma.wright@ons.gov.uk

September 2000

A QUALITY ASSURANCE AND CONTINGENCY STRATEGY FOR THE ONE NUMBER CENSUS

1. INTRODUCTION

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:

a) Maximising coverage in the Census; and

b) Developing a contingency should the Census not achieve 100% coverage.

In order to address the second point, the One Number Census (ONC) Project was initiated. This project is concerned explicitly with the measurement of underenumeration and the development of procedures for adjusting the census for this 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.

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 quality assurance (QA) procedures and a further contingency strategy to deal with 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.

The purpose of this paper is to describe:

a) the high level procedures that will be used to QA the ONC population estimates following the 2001 Census; and

b) the contingency strategy that will be used should these QA procedures find significant evidence to indicate that the ONC population estimates are implausible.

The paper describes the procedures and strategy proposed for England & Wales and Northern Ireland. In general, GROS propose to use similar QA procedures, but there are some differences between the contingency strategies proposed for Scotland and those for the rest of the UK[1].

2. OVERVIEW

The ONC Quality Assurance and Contingency Strategy was agreed by the ONC Steering Committee in February 2000. The key points of this strategy are:

- ONC population estimates should be quality assured to ensure they are plausible for all areas of the country. This QA process will involve demographic analyses and include broad comparisons with both demographic estimates and administrative sources;

[1] For further details see ONC Steering Committee Paper ONS(ONC(SC))00/04, available on request.

- ONC estimates for each Design Group (or 'Estimation Area') and its constituent unitary/local authorities will be systematically quality assured as they are produced. Once estimates for all Design Groups/local authorities have been generated and quality assured, the resulting national ONC population estimate will also be subject to quality assurance; and
- If the QA procedures should clearly indicate that it is necessary to adjust one or more ONC estimates at either a sub-national or national level, a Contingency Strategy will be invoked involving either borrowing strength (at a sub-national level) or using empirical Bayes estimation based upon ancillary demographic information (at a national level).

Proposals outlining how the ONC QA and Contingency Strategy should be implemented for 2001 were discussed and agreed by the Steering Committee in June 2000. That paper focussed on:

- The choice of comparators to quality assure the ONC estimates;
- The choice of comparisons and demographic analyses to be conducted, and the circumstances in which a decision would be made to adjust an ONC estimate; and
- The strategies to adjust an ONC estimate.

Both the QA & Contingency Strategy and the proposals for its implementation are summarised in the rest of this paper. Further details are contained in ONC Steering Committee papers ONS(ONS(SC))00/04 and ONS(ONC(SC))00/18, available on request from ONS.

3. QUALITY ASSURANCE PROCEDURES

3.1 Sub-national Quality Assurance Procedures

3.11 Quantitative Quality Assurance

It is proposed that ONC estimates for each individual Design Group (and its constituent unitary/local authorities) be compared against a range of comparators, including both demographic estimates (i.e. rolled-forward mid-year population estimates - MYEs) and data derived from administrative sources (health registers, benefit registers, etc.) This portfolio of different comparators will vary by age group as outlined in Table 1 overleaf, but for all age groups the different comparators will be combined to produce diagnostic ranges within which one would reasonably expect the ONC estimates to lie. Each of the proposed comparators are described in more detail in Appendix A.

The portfolio of comparators for each age group within each Design Group and unitary/local authority will be input into an automated QA program prior to ONC estimation. Using these comparators, pre-determined diagnostic ranges will be calculated by considering the variation between the comparators for each age-group[2]. The ONC estimates and their associated error bounds will then be fed into this QA program as they are generated and a series of pre-determined comparisons conducted; a variety of demographic analyses such as sex, age and dependency ratios will also be examined. If the ONC estimates for a Design Group and its constituent unitary/local authorities fall within the diagnostic ranges on all comparisons, these will automatically be accepted. However, if one or more of the ONC estimates should fail the diagnostics on any of the comparisons, they will be subject to further qualitative QA procedures as described in section 3.12.

[2] The fact that FHSA patient register data is prone to list inflation will mean that this comparator will tend to be higher than the corresponding demographic estimate for any given age-group. It is therefore proposed that demographically similar local authorities are grouped, and a mean percentage difference between the two comparators calculated for each age-group within each group of local authorities.

Table 1: Portfolio of Comparators by Age Group for Quality Assurance of ONC Estimates

Age Group	Comparators
0	1. Demographic Estimates (ONS) (including Birth registration data) 2. Child Benefit data (DSS) 3. Health Authority patient register data (FHSAs)
1 – 4	1. Demographic Estimates (ONS) 2. Child Benefit data (DSS) 3. Health Authority patient register data (FHSAs)
5 – 9 10 – 14	1. Demographic Estimates (ONS) 2. Child Benefit data (DSS) 3. Health Authority patient register data (FHSAs) and possibly: 4. Schools Census data (DfEE)
15 – 19 20 – 24 25 – 29 30 – 34 35 – 39 40 – 44 45 – 49 50 – 54 55 – 59 60 – 64	1. Demographic Estimates (ONS) 2. Health Authority patient register data (FHSAs) and for working age population sub-groups prone to underenumeration: 3. Student data (HESA) 4. HM Armed Forces data (DASA)
65 – 69 70 – 74 75 – 79 80 – 84 85 +	1. Demographic Estimates (ONS) 2. Retirement Pension data (DSS) 3. Health Authority patient register data (FHSAs)

3.12 Qualitative Quality Assurance

This second stage of quality assurance will involve a systematic assessment of the following information:

- Information from the quantitative QA procedures described in 3.11 above (graphs, tables of differences, extent to which diagnostic ranges are exceeded, etc.);
- Information from the ONC estimation procedures (model parameters, off-diagonal components of the Dual System Estimators, etc.);
- Comparators for specific sub-groups prone to underenumeration (e.g. DASA data on HM Armed Forces, HESA data on students in Higher Education – see Table 1 & Appendix A);
- Past feedback received by ONS from unitary/local authorities on the rolled-forward demographic estimates;
- Information from the field – both Census (Census Management Information System) and CCS (Team Reporting and Communication System – TRACS); and
- Details of the adjustments made to the 1991 Census estimates.

If this assessment of all the available evidence indicates that the ONC estimates for the Design Group and its constituent unitary/local authorities are indeed plausible (despite one or more estimates failing the pre-determined diagnostics), they would be accepted. However, if the evidence indicates that one or more ONC estimates require adjustment, a decision would be made to invoke the Contingency Strategy (see section 4).

3.2 National Quality Assurance & Contingency Procedures

Once the ONC estimates for every Design Group have been generated, quality assured and (if necessary) adjusted, they will be summed to produce the national ONC population estimate. This will be compared against the rolled-forward national demographic estimate for which plausibility bounds have already been estimated[3]. The national ONC population estimate will be accepted if it falls within the plausibility bounds of the rolled forward estimate. If it does not, however, the sub-national quality assurance results will firstly be re-examined for systematic patterns of the ONC estimates falling above or below the corresponding demographic estimates. (If the ONC estimates for some or all Design Groups fall within their diagnostic ranges but consistently fall below (or above) their corresponding demographic estimates, then it is possible that when summed the resulting national ONC estimate will fall outside the national plausibility range.) Secondly, a series of demographic and other sensitivity analyses will also be conducted to assess whether there is a need to make an adjustment at the national level.

3.3 Quality Assurance & Contingency Schedule

The QA of ONC estimates for each age-sex group will progress sequentially through the Design Groups and their constituent unitary/local authorities as follows:

1. If the ONC estimates for Design Group A are plausible, consider the constituent local authorities (a_1 to a_X):
    - 1.1. if the ONC estimates for local authorities (a_1 to a_X) are plausible, accept and move on to Design Group B;
    - 1.2. if the ONC estimates for local authorities (a_1 to a_X) are not plausible, make adjustments and repeat the QA of Design Group A.
2. If the ONC estimates for Design Group A are not plausible, consider the constituent local authorities (a_1 to a_X):
    - 2.1. if the ONC estimates for local authorities (a_1 to a_X) are not plausible, make adjustments and repeat the quality assurance of Design Group A;
    - 2.2. if the ONC estimates for local authorities (a_1 to a_X) are plausible, adjust the ONC estimates for Design Group A and then repeat the QA of Design Group A and its constituent local authorities.
3. Once ONC estimates for all Design Groups and their constituent local authorities have been quality assured and agreed, sum and consider the national ONC estimates:
    - 3.1. if the national ONC estimates are plausible, accept and end the QA process;
    - 3.2. if the national ONC estimates are not plausible, adjust at the national level and repeat the quality assurance of Design Groups and their constituent local authorities.

[3] For details see ONC Steering Committee Paper ONS(ONC(SC))99/05, available on request from ONS.

4. CONTINGENCY STRATEGY – ADJUSTMENT OF ONC ESTIMATES

4.1 Sub-national Adjustment: Borrowing Strength

If a decision is made to adjust the ONC estimates for a particular Design Group or unitary/local authority, the basic strategy to be used will be to impute the degree of underenumeration in a particular local authority or Design Group from the levels of underenumeration estimated in socio-economically and demographically similar areas. The regression methodology by which this strategy of "borrowing strength" will be operationalised is outlined in Appendix B.

For local authorities the most obvious, and in some cases, suitable areas to borrow strength from will be other local authorities within the same Design Group. However, this will not be the case for Design Groups that are single local authorities or those whose constituent local authorities are likely to have quite different patterns of underenumeration. It is therefore proposed that the choice of local authorities from which to borrow strength will be made using the *ONS classification of Local Authorities* developed by ONS' Methods and Quality Division. This classification combines 37 variables from the 1991 Census to calculate a "distance" between each and every other local authority in Great Britain.

4.2 National Adjustment: Bayes Estimation based upon Demographic Information

In the case where the evidence suggests that the national ONC estimates, formed by summing the agreed sub-national estimates, still require adjustment then this is likely to be because, across the nation, the degree of dependence between the Census and CCS collection procedures will have been too high (or low). A diagnostic would be that all estimates were a little low (or high). In this case it is proposed that adjustments be made to the sub-national dual system estimates and a new national estimate calculated. Elliott and Little (1999) show how, in the situation described above, empirical Bayes estimation can be used to make such adjustments as long as one has ancillary information such as expected sex ratios, dependency ratios or other demographic or administrative data. This information is used as a set of prior targets within the Bayes estimation process. It is proposed that this strategy will be used.

5. REFERENCES

Bailey, S. et al (1999), *'The ONS classification of local and health authorities of Great Britain: revised for authorities in 1999',* Office for National Statistics publication.

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

Charlton, J. & Chappell, R. (2000), *'Uncertainty Intervals for National Demographic Estimates – Update',* ONC(PB)00/06.

Cruddas, M. & Diamond, I (2000), *'A Quality Assurance and Contingency Strategy for the One Number Census',* ONS(ONC(SC))00/04.

Diamond, I & Wright, E (2000), *'Taking Forward the ONC Quality Assurance and Contingency Strategy',* ONS(ONC(SC))00/18.

Elliott, M.R. & Little, R.J.A. (1998). *'On Combining Information from a Census, a Coverage Measurement Survey and Demographic Analysis'* submitted to J.A.S.A.

Appendix A

COMPARATORS TO BE USED IN ONC QUALITY ASSURANCE PROCEDURES

Demographic Estimates

Rolled-forward mid-year population estimates (MYEs) are produced every year by the Population Estimates Unit (PEU) of Population & Vital Statistics Division (PVSD) within ONS. The MYEs are produced by age and sex for each unitary/local authority in England & Wales, and these demographic estimates will be used as one of the main comparators for all age groups in the ONC QA procedures both nationally and sub-nationally. However, because the estimates have been rolled-forward from the 1981 Census (using births, deaths, estimates of migration, and estimated underenumeration for the 1991 Census), they are subject to a degree of error. Plausibility bounds have therefore been estimated for the demographic estimates at a national level (see Chappell and Charlton 1999/ 2000), and it is proposed that plausibility at a sub-national level is assessed in two ways: (i) by assuming that for similar types of areas, errors will broadly be constant between 1991 and 2001, and (ii) by using the differences between the demographic estimate and the other administratively-based comparators for a particular age-sex group to produce a diagnostic range.

The rolled-forward MYEs for 2001 will not be available to feed into the ONC QA process (due to many of the constituent components that make up the estimates not being available in time). Alternative demographic estimates will therefore be required for use in ONC quality assurance procedures at both sub-national (Design Group and unitary/local authority) and national levels:

Sub-national level

ONC estimates for the first England and Wales Design Group should be quality assured around September 2001. It is currently proposed that published 2000 MYEs will be used in this QA process, although an alternative would be to further extrapolate these estimates forward to mid-2001 to allow for average annual population change between mid-1991 and mid-2000.

National level

The final rolled-forward estimate MYEs for England & Wales as a whole are not likely to be available until early July 2002. It is therefore proposed that the Government Actuary's Department (GAD) 2000-based population projections for 2001 are used as as the provisional national demographic estimate until the final 2001 MYEs are available.

Birth Registration Data

The civil registration system records all new births in the population, and this will therefore be a key source of data when quality assuring ONC estimates for children aged under one year (a group of the population which censuses worldwide seem to undercount to a greater extent). ONS's methods will be used for adjusting the registration data on births at both national and sub-national levels for infant deaths and migration to produce a high quality comparator for use in the ONC QA procedures.

DSS Child Benefit Data

The Benefits Agency administer the Child Benefit Claimant Register which holds information on all persons claiming child benefit in the UK and the children for whom the benefit is claimed. As Child Benefit is almost universally taken up for children in the UK, this administrative register is potentially a good source of information on those aged under 16 (the benefit can only be claimed

for those children aged 16 and over who remain in full-time 'school-type' education). In order to aggregate these records and use them as a comparator in ONC quality assurance procedures, certain problems need to be borne in mind and overcome:

i) Many of the postcodes on the records are either missing, contain errors, or are out of date. It is therefore necessary to validate the postcodes against address details using postcoding software before any data is aggregated geographically.

ii) Newborn children can take up to three months to appear on the Child Benefit register, due to delays in claims being made and the information subsequently being added to the register. It is therefore necessary to obtain an additional extract from the register approximately three months after the reference date required (to obtain information on children born on or before the reference date for whom benefit has subsequently been claimed) and to add these records to the extract relating to the reference date.

iii) Approximately 0.5% of children listed on the register are linked to claimants with overseas addresses (mainly BFPO addresses). It is not known how many of these children are educated (and therefore resident) in the UK, and how many are resident overseas; and

iv) Some claimant address information may be out of date due to claimants moving but not informing the Benefits Agency. This is only likely to be a problem for those customers who have their benefits paid directly into a bank account.

ONS are liaising with DSS to obtain “clean” child benefit data for both 1999 (for use in developing and assessing prototype QA programs and procedures) and 2000 (for use as a comparator in 2001).

DSS Retirement Pensions Data

The Benefits Agency administers the Retirement Pension Register which holds information on all persons claiming a state retirement pension in the UK. Almost all persons aged 65 or over are entitled to claim some form of state retirement pension, be it as a recipient of a full contributory pension, a partner of such a recipient, a recipient of a non-contributory pension, or a combination of these options. This administrative register is potentially a good source of information on those aged 65 and over, but once again there are problems that need to be overcome and issues that need to be borne in mind if the records are to be aggregated and used as a comparator in ONC quality assurance procedures:

i) In common with the Child Benefit Register, many of the postcodes are either missing, contain errors, or out of date and it is therefore necessary to validate these before any data are aggregated geographically.

ii) To ensure maximum coverage of the population aged 65 and over without duplication, records relating to each of the different types of pension (as well as benefits such as Severe Disability Allowance and Incapacity Benefit) need to be extracted and linked by National Insurance number.

iii) As well as men aged 65+ and women aged 60+ living in the UK, some expatriates resident overseas are also eligible to claim a state pension.

iv) Again, some address information may be out of date due to pension claimants moving but not informing the Benefits Agency. This is only likely to be a problem for those customers who have their pensions paid directly into a bank account.

ONS are again liaising with DSS to obtain “clean” retirement pension data for both 1999 (for use in developing and assessing prototype QA programs and procedures) and 2000 (for use as a comparator in 2001).

FHSA Patient Register Data

Patient registers administered by individual Health Authorities provide the most comprehensive administrative source in terms of coverage of the whole population. Previous work conducted by ONS, however, concluded that data derived from these registers were unsuitable for producing population estimates due to (i) coverage differences with the resident population (certain groups such as Armed Forces personnel are excluded from patient registers), and (ii) list inflation (caused by delays in people who have died or moved out of a Health Authority being removed from the register, duplicate entries, and people having more than one NHS number). It should be noted that the degree of list inflation is extremely variable geographically. Following on from work by individual Health Authorities to "clean" their registers, ONS are liaising with the NHS Executive to obtain suitable data to feed into the development of the QA system.

Schools Census Data

The Schools Census is a count of all children attending educational establishments, including schools which are privately funded. In England, information relating to the 1 January is collected from local authorities by DfEE each year, whilst in Wales, Scotland and Northern Ireland this role is carried out by the appropriate devolved Government authority, and has a different reference date. ONS are currently liaising with DfEE to establish how Schools Census data might be used in the quality assurance procedures for 5-14 year olds in 2001, and in particular how the problem of children living in one local authority/Design Group/country and attending a school in another might be overcome.

Data Sources relating to specific sub-groups

As well as comparators that cover one or more specific age groups within the population, other administrative data sources that relate to specific sub-groups known to be prone to underenumeration are also being considered for use as comparators in the ONC quality assurance procedures. The main sources of data that have been identified are the Higher Education Statistics Authority (HESA – these data on students are currently being assessed by ONS) and the Defence Analytical Services Agency (DASA – once they have been validated by local authorities, these data on the Armed Forces are already judged by ONS to be of acceptable quality to feed into the construction of MYEs).

It should be noted that quality assurance for specific sub-groups will only be possible once ONC imputation has been completed for a Design Group, and it is therefore proposed that this secondary analysis is conducted as part of the qualitative quality assurance procedures outlined in 3.12 above.

Appendix B

SUB-NATIONAL CONTINGENCY STRATEGY – BORROWING STRENGTH

Should the quality assurance procedures indicate that the ONC estimates for a Design Group or unitary/local authority need to be adjusted, the basic contingency strategy will be to use information on underenumeration from demographically similar areas. The strategy for both Design Groups and unitary/local authorities is outlined below:

Design Groups

In this situation the evidence suggests that the ONC estimates for a Design Group are too low and that 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). The strategy is as follows.

1. For Design Group (DG) '*A*' the ONC estimate has a ratio estimate:

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

   where *Census* is the Census count.

2. In three demographically similar DGs which have passed QA, *B-D*, the ratio estimates are:

   $$B \quad \text{ONC} = 1.03\,Census \qquad (2)$$
   $$C \quad \text{ONC} = 1.02\,Census \qquad (3)$$
   $$D \quad \text{ONC} = 1.04\,Census \qquad (4)$$

3. Taking the weighted average of these three comparator areas gives:

   $$\text{ONC} = 1.03\,Census \qquad (5)$$

4. DG *A* is then re-estimated using this new ratio estimate and the QA procedure repeated for DG *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/hard to count group.

Unitary/Local Authorities

In this situation sub Design Group analyses suggest that one local authority (LAD) in the Design Group has an implausibly low adjustment. The aim in this case would be to borrow strength from socio-economically and demographically similar LADs using the same strategy as above.