national STATISTICS

Census 2001

One Number Census methodology and Quality Assurance process report

Contents

	Page
1 Introduction	**3**
2 Overview and summary	**3**
2. 1 Overview of the ONC methodology and processes	**3**
2.1.1 Census Coverage Survey	5
2.1.2 Matching	5
2.1.3 ONC Estimation process	6
2.1.4 ONC Imputation process	8
2.1.5 Consultation and peer review of ONC methodology	9
3 ONC Quality Assurance and contingency strategy	**10**
3.1 Background information	**10**
3.2 Overview of the Quality Assurance stages	**11**
3.2.1 Stage 1 Quantitative Quality Assurance	11
3.2.2 Stage 2 Qualitative Quality Assurance	11
3.2.3 Contingency strategy	11
3.2.4 Stage 3 Population subgroup Quality Assurance	13
3.3 Formal process of quality assuring ONC estimates	**13**
3.4 Detailed sub-national quality assurance procedures	**14**
3.4.1 Stage 1 Quantitative Quality Assurance	14
3.4.2 Stage 2 Qualitative Quality Assurance	20
3.4.3 Stage 3 Population subgroup Quality Assurance	21
3.4.4 Population subgroup comparisons	23
3.5 Cumulative Quality Assurance	**28**
4 Summary	**28**
5 References	**28**
Annex A: Example illustrative estimation summary report	**29**
Annex B: Example imputation report	**31**
Annex C: Explanation of terms used in imputation report	**33**
Annex D: Example imputation Quality Assurance report	**34**
Annex E: The research team and steering group	**36**
Annex F: Comparators used in the ONC Quality Assurance process	**37**
Annex G: Sources of qualitative information	**40**
Annex H: Example Qualitative Quality Assurance report	**41**
Annex I: Example population profile	**43**
Annex J: Glossary of acronyms	**44**

One Number Census methodology and Quality Assurance process report

1 Introduction

1.1 The aim of this report is to provide users of Census 2001 with an understanding of the One Number Census (ONC) process and the methodology that underpinned this process. It has been written to accompany the "Key findings and Actions from the One Number Census Quality Assurance process" report, produced to outline the key findings and actions resulting from the ONC quality assurance process.

1.2 This report describes the stages of the ONC process as outlined in 'Guide to the One Number Censu' **www.statistics.gov.uk/census2001/pdfs/oncguide.pdf**. It further provides a comprehensive description of the steps taken to quality assure the ONC estimates in 2001. It draws on information from the paper 'A Quality Assurance and Contingency Strategy for the One Number Census' which was circulated to both the Census Advisory Groups and Liaison Group on Population Statistics and Local Government Statistical Liaison Officers in September 2001.

1.3 A range of detailed information on the ONC process has been published on the National Statistics website. This includes information on response rates and papers presented to the ONC Steering Committee outlining the ONC methodologies. In addition, a range of quality assurance material at the Local Authority level has been published. This includes the specific charts provided for each local authority that were used during the quality assurance process alongside the comparator administrative and demographic data used in the process.

1.4 Furthermore, an illustrative Quality Assurance pack has been produced and published on the website for a fictitious Design Group (DG) and its constituent Local Authority Districts (LADs). Although this does not reflect a real DG, data from a number of real areas was used to derive the information presented in the Quality Assurance pack. This report demonstrates the sorts of patterns shown throughout the quality assurance process as well as the material available to the quality assurance panel.

2 Overview and summary

2. 1 Overview of the ONC methodology and processes

2.1.1 The ONC project aimed to integrate the 2001 Census counts with the estimated level of underenumeration in the Census - that is the number of households and people not counted. Firstly, it provided a new base for the mid-year population estimates at the LAD level, and secondly it adjusted the Census database itself for the estimated undercount so that all statistics added to 'One Number' - the national estimate of the population.

2.1.2 There were a number of key stages involved in achieving a One Number Census. **Figure 1** on page 4 illustrates the key stages of the ONC process and can be summarised as follows:

a) A Census Coverage Survey (CCS), undertaken independently of the Census, was designed to establish the coverage of the 2001 Census. For the CCS, England and Wales was divided into one hundred and one areas, each with a population of about 500,000. These areas known as 'DGs' and were made up of whole LADs or groups of smaller LADs. The CCS took place in all of these DGs. The CCS is covered in more detail in **section 2.1.2**.

b) The CCS records were matched with those from the Census using a combination of automated and clerical matching. The matching is covered in more detail in **section 2.1.3**.

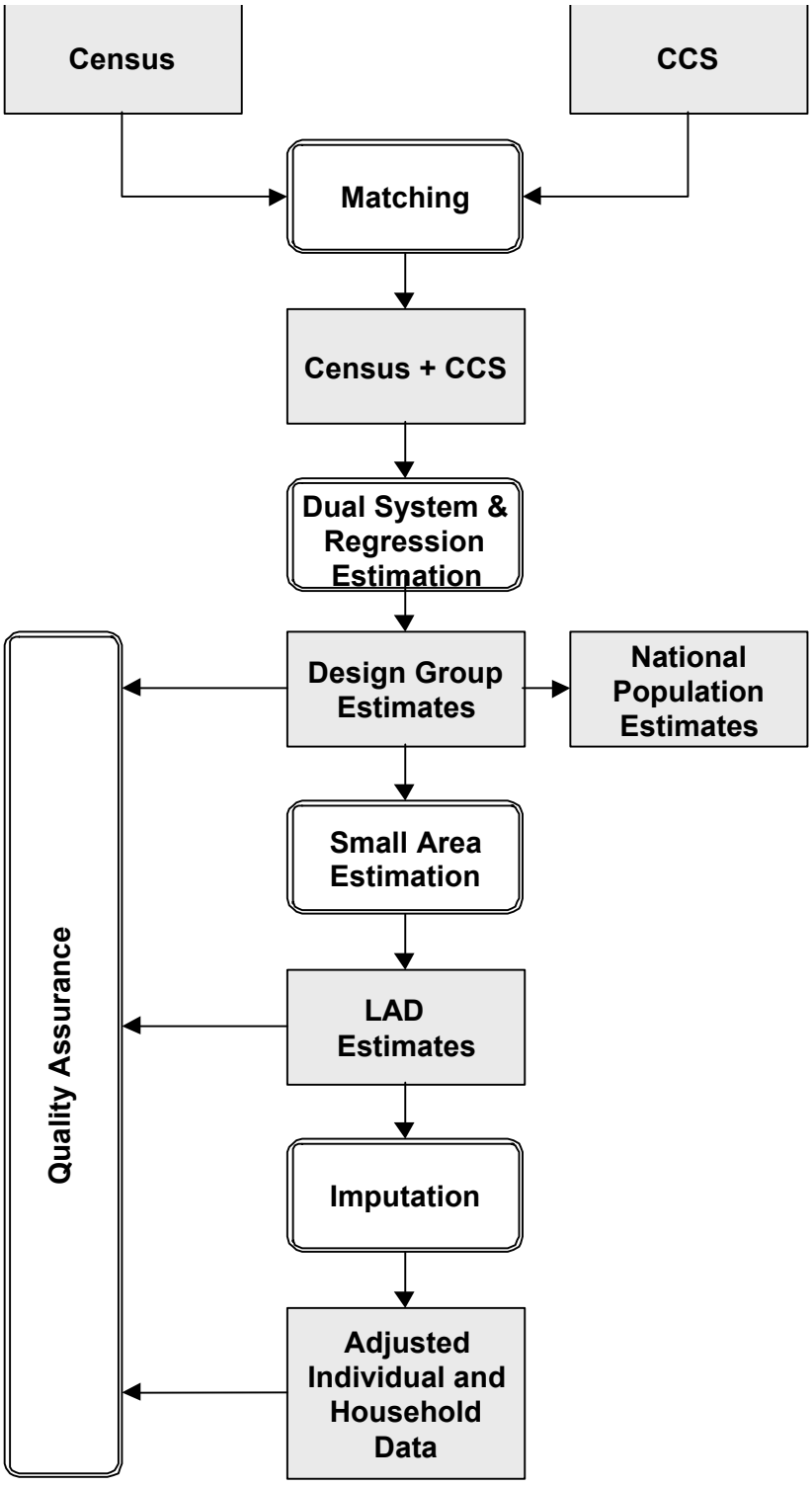
c) Populations for each DG, by age and sex, were estimated using a combination of standard estimation techniques. See **section 2.1.4** for more detail.

d) Small area estimation techniques were used to estimate LAD populations by age and sex. This estimation is also covered in **section 2.1.4**.

e) Households and individuals estimated to be missed by the Census were imputed to produce a fully adjusted Census database. The imputation process is covered in more detail in **section 2.1.5**.

f) All ONC population estimates were quality assured using demographic analysis and aggregate level administrative data. The quality assurance process is covered in **section 3**.

Figure 1
The ONC process



2.1.1 Census Coverage Survey

2.1.1.1 The CCS was a key element in the ONC methodology and it was designed to enable census population counts to be adjusted for underenumeration at the national, local and small area level.

2.1.1.2 The CCS consisted of a completely independent and intensive face to face survey of a sample of over 16,000 postcodes containing 320,000 households drawn from all local authorities in England and Wales. The CCS was operationally independent from the census enumeration exercise and it was undertaken during a four-week period starting three and a half weeks after Census Day.

2.1.1.3 The CCS was designed within a set of what ONS calls 'DGs', consisting of one or more whole Local Authorities. A sample of postcodes was selected across the whole DG. The objective of the sample of CCS postcodes was to provide the information necessary to be able to make robust estimates of the numbers of people missed by the census in the whole 'DG' and its constituent local authorities.

2.1.1.4 The sample of postcodes was not simply a random choice - information was used from the 1991 Census to derive a 'Hard to Count' index (HtC). This index had three levels - Easy, Medium and Hard - and each postcode was allocated to one of these groups based upon its levels of the following 1991 Census variables, which are believed to be a good indicator of where the census might miss people:

a) Multi-occupied households (i.e. bed-sits that are within the same building)

b) Privately rented households

c) 1991 Unemployment levels

d) Language difficulty - derived from country of birth

e) Imputed residents in the 1991 Census (i.e. where the 1991 Census had problems finding people)

2.1.1.5 The CCS was specifically designed and implemented to ensure that highly mobile difficult to count populations could be counted accurately. The sample of postcodes were selected within each of the three HtC categories, although a slightly higher proportion were selected in the hardest to count category (this was to ensure that the amount of information collected within the harder areas was boosted). A two stage clustered sample was used. Firstly Enumeration Districts (EDs) were clustered into groups with similar populations in the key age/sex groups associated with underenumeration in the 1991 Census. A sample of EDs was drawn, followed by a sample of postcodes from each ED. England and Wales was then divided into 101 DGs containing approximately 500,000 people.

2.1.1.6 The CCS sample postcodes were kept confidential and CCS interviewers did not have any sight of either the address lists produced in carrying out the census or the census forms returned in the area in which they were interviewing. The interviewers focused on making as many calls as necessary to achieve an interview, and the timing of these calls was varied to maximise the probability of making contact. Postback forms were left where no contact was made. The interviewers also sought proxy information from neighbours. The CCS contained a number of prompts to help identify all household members and reassure respondents of the confidentiality of the data.

2.1.2 Matching

2.1.2.1 Estimates of the total population were based on a methodology known as dual system estimation (DSE). This is further outlined in **section 2.1.4**. It was inevitable that some households and people would be missed by both the Census and CCS but DSE can be used to estimate this number by considering the relative numbers of the people observed by;

- both the Census and CCS;
- the Census but not the CCS; and
- the CCS but not the Census.

2.1.2.2 In order to identify the numbers in each of these groups it was necessary to match the records from the CCS with those from the Census. It was essential that this matching process was accurate, as the number of mismatches would have had a direct impact on the final population estimates.

The 2001 matching exercise involved a combination of automated and clerical matching. The matching process for a single CCS postcode is outlined below and illustrated in **Figure 2**. There are four key stages:

2.1.2.3 Stage1 Exact matching

Automatically link CCS and Census households and individuals where key details match exactly. The key details used to exact match households were postcode, address name/number, type of accommodation, number of people and surname of household representative. The

details used to exact match individuals were forename, surname, day and month of birth, marital status and relationship to the household representative. Households were only considered matched at this stage when all individuals within the household pair had been linked.

2.1.2.4 Stage 2 Probability matching
CCS and Census records that were not matched at Stage 1 of the process were then run through a probability matching process. Probability matching examined the same variables as used in exact matching. A probability weight was assigned to each pair of CCS and Census records based on the level of agreement between them. The higher the probability weight, the closer the agreement between the two records. For example, if a pair of records was identical with the exception of one detail, which may be due to recording error, then a high probability weight was assigned. Any household pairs with a high probability weight were linked and the individuals within them compared. Only very similar households and individuals were considered as matched at this stage.

2.1.2.5 The initial probability weights used were derived from the 1999 Census Rehearsal. When 2001 Census data became available, these weights were recalculated and thresholds reviewed so that they were more appropriate to the data being matched. 'Training' the weights in this manner helped to allow for any regional differences in the data.

2.1.2.6 Stage 3 Clerical resolution
Pairs of households and individuals with a reasonable level of agreement were presented for clerical resolution. At this stage operators were simply be asked to determine whether the pair of records shown constitute a matching pair or not. They were not expected to search for matching records.

The file of exact, high and medium (with clerical verification) quality matches were used to update the probability weights once again.

2.1.2.7 Stage 4 Clerical matching
The final stage of the ONC matching process involved a clerical search for any census records corresponding to unmatched CCS households and individuals. Operators had access to images of Census and CCS forms to assist in decision making.

2.1.2.8. Quality Assurance
As previously mentioned, the accuracy of the matching process was critical to the accuracy of the ONC population estimates. Quality Assurance procedures, similar to those used in the US, were built into the matching process to ensure that the necessary high levels of accuracy were met.

2.1.2.9 The output of the clerical matchers was checked by expert matchers to ensure that all matched pairs of records were legitimate matches. These experts also checked that all unmatched records did not have a possible match using extensive database searches.

2.1.2.10 A small number of supervisors checked the work of the expert matchers. These supervisors also assisted in marginal matching decisions. These processes ensured accuracy and a consistent approach.

2.1.3 ONC Estimation process
2.1.3.1 The next stage in the ONC process was to derive estimates of the population for all LADs using the combined Census and CCS data generated by the matching.

Stage 1 Design Group Estimation
2.1.3.2 The output from the matching process was used to estimate underenumeration by age and sex for each CCS postcode. This was achieved using DSE methodology.

The use of DSE requires a number of assumptions to be met to ensure the minimisation of error in the estimates:

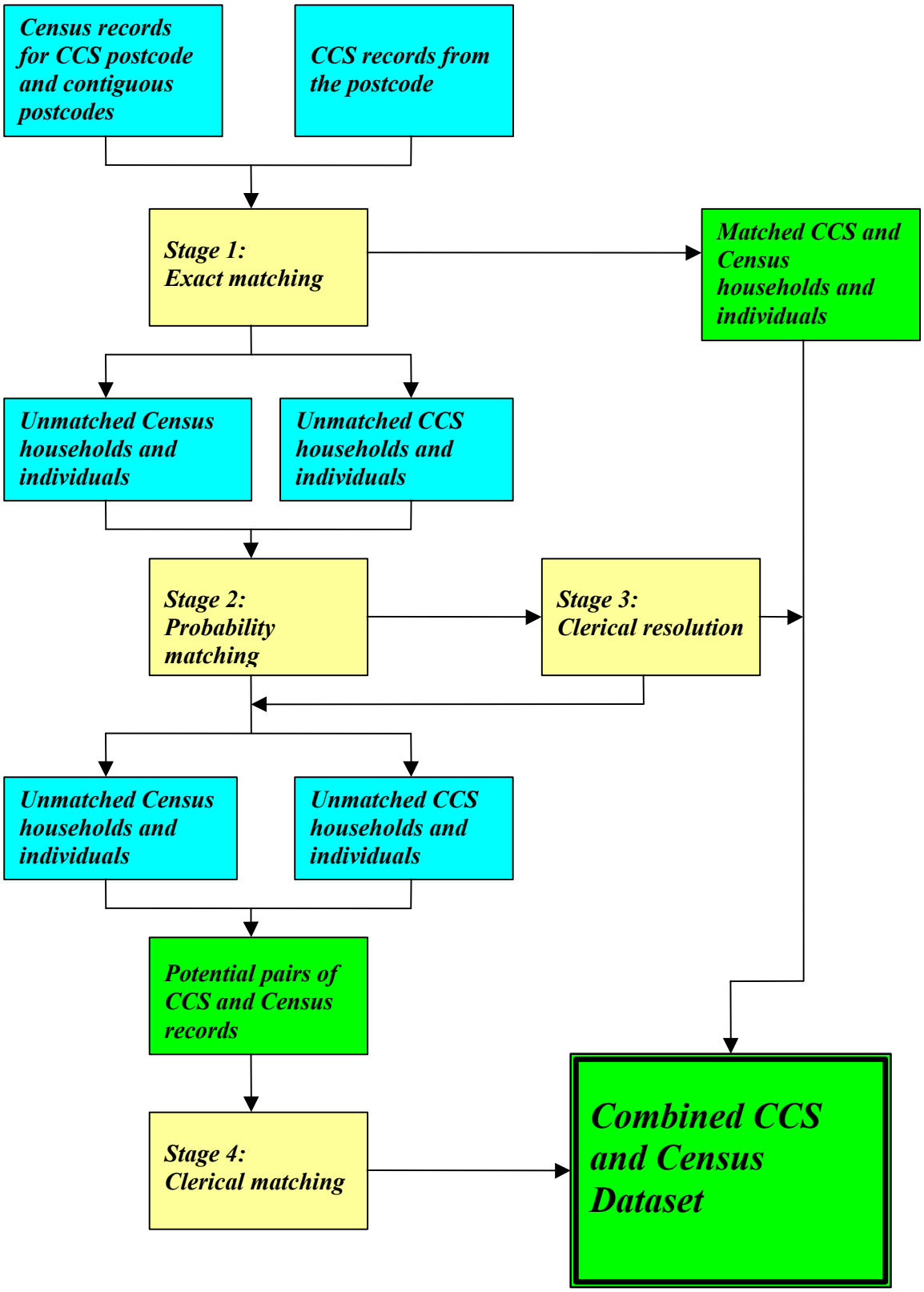
- Firstly, independence between the Census and CCS was assumed for an unbiased estimate. As a result the Census and CCS were operationally independent. Simulation work was undertaken during the development of the ONC to examine the impact of dependence between the Census and CCS on the ONC population estimates. These simulations showed the methodology to be reasonably robust under low levels of dependence.
- Secondly, it was assumed that the chance of a person being in the Census or CCS, for an age-sex group within a postcode, was the same across all people.

  Given these assumptions, DSE combined those people counted in the Census and/or CCS and estimated those people missed by both.

  Once postcode level DSEs were obtained, a modified regression equation was calculated for each age-sex group

Figure 2

An illustration of the ONC matching process



within each HtC group to estimate the relationship between the census counts and the dual system estimates. This equation was then used to estimate DG underenumeration for each age-sex group (of which there were 37) in each HtC group (of which there were three). Therefore there were 111 separate estimation strata in any given DG.

The output from this process were estimates of the population for each DG by age and sex, together with an indication of their accuracy. All of the subsequent processes described below were consistent with these population counts.

Stage 2 Local Authority District Estimation

2.1.3.3 Since many DGs consisted of more than one LAD, estimates of the age-sex population for each LAD were made. This formed the second stage of the estimation process.

Many LADs did not contain sufficient CCS postcodes to enable accurate direct estimates of underenumeration to be made. Small area estimation techniques were applied to produce accurate LAD level population estimates. The small area estimation technique used by the ONC used information from the whole DG to model the undercount within LADs. The resulting population estimates were calibrated to the DG estimates, and their accuracy was calculated.

2.1.3.4 However it was always planned to test the assumption of independence. When tested, it was found that the assumption was invalid. The direction of the dependence observed was such that a person missed by the Census was more likely to be missed by the CCS than one who was found by the Census. Therefore, the level of dependence was estimated and the population estimates adjusted accordingly. Further details on the dependency adjustment are outlined in the 'Key Findings and Actions from the One Number Census Quality Assurance Process' and more detailed information on the methodology to adjust for dependence can be found at **www.statistics.gov.uk/census2001/pdfs/dependency_paper.pdf**.

2.1.3.5 An estimation summary report was routinely produced for every DG. An example estimation report for the fictitious area used in the Quality Assurance Illustrative pack is outlined in **Annex A**. This information was available to the quality assurance panel on request though a summary of the information provided was supplied to the panel at the quality assurance meeting.

2.1.3.6 During estimation there were 111 separate estimation strata in any given DG, as mentioned above. In some instances, however, it was not possible to estimate each of the groups separately. Therefore a strategy was established known as 'collapsing strata' where these groups were combined with another group. Although this was applied as standard for DGs some judgement was required for individual cases. Further detail on the 'collapsing strata' strategy and the conditions that led to the implementation of the strategy are outlined in 'Key Findings and Actions from the One Number Census Quality Assurance Process".

2.1.4 ONC Imputation process

2.1.4.1 The main purpose of the ONC imputation system was to update the Census database with imputed households and people estimated to have been missed by the 2001 Census. Information on the characteristics of missed persons obtained in the CCS allowed the creation of a database that represented the best estimate of the entire population, whether counted by the census or not.

2.1.4.2 The ONC population estimates defined the number of people to be imputed along with some basic information about coverage patterns for other characteristics. However, it was important to identify the detailed characteristics of those households and individuals missed by the Census. For instance, it was anticipated that the characteristics of people within entirely missed households would differ from those missed from within otherwise counted households. Accurate matching of Census and CCS data allowed the identification of these key characteristics. Once these features were identified, prediction of both numbers and characteristics of missed individuals in the population not covered by the CCS was possible. The ONC imputation process can be summarised in three stages:

Stage 1 Imputation of missed households

2.1.4.3 The first stage of the process imputed missed households and the individuals within them. A weight was allocated to each Census household corresponding to the likelihood of households of that type being missed by the Census. These weights were derived from an analysis of households missed by the Census, but captured by the CCS. The Census households were ordered by these weights and cumulative actual and weighted counts calculated. The cumulative counts were compared and, if the weighted count exceeded the unweighted count

by more than 0.5, an imputed household was created with the characteristics of the current household.

Imputed households were placed into either a physical location identified by census enumerators where no response was received (e.g. absent household, refusal, non-contact), or into areas where similar households already existed.

Stage 2 Imputation of missed individuals

2.1.4.4 The second stage of the imputation process focused on individuals missed from households counted by the Census. A weight was created for each individual based on information obtained from analysis of the matched CCS/ Census data, which reflected the likelihood of people with their characteristics being omitted from the census return for their household. These weights were used to impute individuals into the types of households that were likely to have missed people from their Census return. It was this process that added people to real households.

Stage 3 Calibration to estimates of the population

2.1.4.5 A crucial requirement of the imputation process was that the overall distribution of imputed individuals and households should equal the ONC estimates of households and individuals missed by the 2001 Census. This calibration was accomplished by adjusting household and individual weights and by a final stage in the process, which either removed excess imputed individuals and households or topped up the Census database where necessary.

The result was an individual level database that represented the best estimate of what would have been collected had the 2001 Census not been subject to underenumeration. Tabulations derived from this database automatically included compensation for underenumeration for all variables and all levels of geography, and were consistent with the ONC population estimates.

2.1.4.6 As well as producing the files to allow for the update of the Census database with imputed households and people the imputation system produced diagnostic files which were used to obtain information on both the type of person and type of household imputed.

2.1.4.7 An imputation report was also produced, as standard, for each DG and was available on request from the quality assurance panel though a summary of the key findings was provided at each quality assurance meeting. An example Imputation report for the fictitious DG is available in **Annex B** with further explanation of the terms used within the Imputation Report available in **Annex C**.

2.1.4.8 Following the ONC imputation process, checks were made against the resulting underenumeration adjusted database. The purpose of this was to assess whether the imputation had created any inconsistencies or undesirable results. For all DGs, imputation was examined in order to assess the consistency of the data in depth. Checking the first few DGs in particular detail meant that any evidence of inconsistencies would have been noted at an early stage. However, it must be borne in mind that only very extreme inconsistencies were looked for - the point of the ONC was to correct for differential undercount and thus it was not expected that the imputed people would have the same general characteristics as the rest of the population. Some judgement had to be exercised when deciding whether a particular check had highlighted a problem.

2.1.4.9. An Imputation Quality Assurance summary report was also routinely produced as part of the imputation process. Similarly to the Imputation Report, this was available to the quality assurance panel on request and an example Imputation Quality Assurance Summary report can be found in **Annex D**.

2.1.5 Consultation and peer review of ONC methodology

2.1.5.1 It was important that users of census data had confidence in the estimates of the population produced by the ONC. Therefore, the ONC methodology was subjected to peer review and consultation throughout the research and developmental stage of the project. Acceptance was sought in a number of ways, including:
· the composition of the research team and the Steering Group (**Annex E**);

- the consultation process with census user groups;
- a Series A Royal Statistical Society paper (Brown *et al*, 1999);
- several RSS seminars;
- the Spring 1998 Census Consultation paper '2001 - A One Number Census';
- the Spring 1999 Consultation paper 'A Guide to the One Number Census'.

- a workshop devoted to the One Number Census project in May 1998; and
- special workshops held in conjunction with Census Output Consultation Roadshow meetings during April/May 1999 and September 2001.

3 ONC Quality Assurance and contingency strategy

3.1 Background information

3.1.1 The ONC Quality Assurance and Contingency strategy designed in collaboration with academics from Southampton University formed a fundamental part of the ONC methodology. It dealt with the possibility that the results of the ONC estimation may not have been plausible, either in some areas of the country or, indeed, in the nation itself.

3.1.2 The quality assurance process followed an agreed strategy, which had been the subject of wide consultation with census users. The ONC Quality Assurance and Contingency Strategy was agreed by the ONC Steering Committee in February 2000. The key points of this strategy identified that:

- ONC population estimates should be quality assured to ensure that they were plausible for all areas of the country. It was agreed that the quality assurance process would involve demographic analyses and include broad comparisons with both demographic estimates and administrative sources;
- ONC population estimates for each DG and its constituent unitary and local authorities should be systematically quality assured as they were produced; and
- If the quality assurance procedures clearly indicated it necessary to adjust one or more ONC estimates at either a sub-national or national level, that a contingency strategy would be invoked. This would involve either borrowing strength (at a sub-national level) or using an adjustment based upon ancillary demographic information (at a national level).

3.1.3 Proposals outlining how the Quality Assurance and Contingency Strategy, agreed in February, should be implemented were discussed and agreed by Steering Committee in June 2000. This focused on:

- comparator data used to quality assure the ONC estimates
- comparisons and demographic analyses to be conducted
- circumstances in which a decision would be made to adjust an ONC estimate
- strategies to adjust an ONC estimate

3.1.4 Both the Quality Assurance and Contingency Strategy and the proposals for its implementation were summarised in Advisory Group Paper 00 (16) which was circulated to members of the Census Advisory Groups and the Liaison Group on Population Statistics (LGPS) in October 2000. Further details are contained in ONC Steering Committee Papers ONS (ONC(SC)) 00/04 and ONS (ONC(SC)) 00/18. All papers are available from ONS on request

3.1.5 A Quality Assurance and Contingency Strategy for the One Number Census published in September 2001 follows on from Advisory Group paper (00)16 (LGPS paper (00)11). This paper, available at **www.statistics.gov.uk/census2001/pdfs/oncinfopaper.pdf**, circulated to both the Census Advisory Groups and LGPS and Local Government Statistical Liaison Officers described:

- the procedures to quality assure the ONC population estimates following the 2001 Census; and
- the contingency strategy to be used if the quality assurance procedures found significant evidence to indicate that the initial ONC population estimates were implausible.

3.1.6 This paper also highlighted the involvement of statistical liaison officers in local government in the consultation exercise undertaken in the winter of 2000/01 with regards the Contingency Strategy. If the quality assurance procedures indicated that the initial ONC population estimates at the sub-national level were implausible then information from similar areas that had already passed quality assurance was used to produce revised ONC estimates. This process was known as 'borrowing strength'.

3.1.7 The initial method for selecting LADs from which to borrow strength involved using the ONS classification of Local Authorities to identify the five most similar areas in terms of

socio-economic and demographic characteristics for each and every LAD in England and Wales. This classification, published by the Methods and Quality Division of ONS, combines 37 variables from the 1991 Census to calculate a 'distance' between each and every other LAD in Great Britain.

3.1.8 Statistical liaison officers were given the opportunity to review the list of the closest fifteen LADs to their own and were invited to comment on whether the closest five were acceptable to them for borrowing strength or whether alternatives were preferred.

3.1.9 Initial quality assurance took place at the LAD and DG level. This quality assurance, in broad summary, comprised a comparison of the ONC estimates with a range of both quantitative and qualitative information and involved three distinct stages as outlined below.

3.2 Overview of the Quality Assurance stages

3.2.1 Stage 1 Quantitative Quality Assurance

3.2.1.1 The process illustrated in Figure 3a below began with a comparison of the ONC estimates with the raw Census counts and with a number of comparator data sets. These comprised various administrative records together with the ONS mid-year population estimates for 2000.

3.2.1.2 The comparators were used to form a diagnostic range for each of the indicators (namely population counts by five-year age-sex group, sex ratios and dependency ratios). This range allowed an initial comparison to be made although ONC estimates were not signed off at this stage. The diagnostic ranges gave an indication as to whether the ONC estimates were broadly consistent with expectations from a range of demographic and administrative records. It was never intended to adjust the Census results on the strength of the diagnostic ranges alone. For a detailed description of the methodology to calculate the diagnostic range refer to 'A Quality Assurance and Contingency Strategy for the One Number Census' at **www.statistics.gov.uk/census2001/pdfs/oncinfopaper.pdf**.

3.2.2 Stage 2 Qualitative Quality Assurance

3.2.2.1 The second stage of the sub-national quality assurance process, illustrated in Figure 3a used a range of qualitative information to examine the ONC estimates further. This information was used in conjunction with the quantitative data to quality assure the ONC estimates. Further detail on the range of qualitative information is outlined in **section 3.4.3** and **Annex H**.

3.2.2.2 The qualitative indicators were used either to:

- affirm the ONC estimates because they were broadly consistent with expectations or because they identified population trends not picked up by demographic or administrative sources; or
- decide that the ONC estimates were not acceptable and that a contingency strategy needed to be invoked or that further actions should be investigated. Further detail on the implementation of the contingency 'borrowing strength' strategy and further actions investigated as a result of the discussions held at the quality assurance meetings is outlined in detail in the report on the key findings and actions from the One Number Quality Assurance process.

3.2.2.3 The overall objective was to look for a body of evidence on which to accept (or reject) an ONC estimate. It should be stressed that this stage (quantitative and qualitative) of quality assurance was not a 'black box' process. The broad range of evidence was always considered.

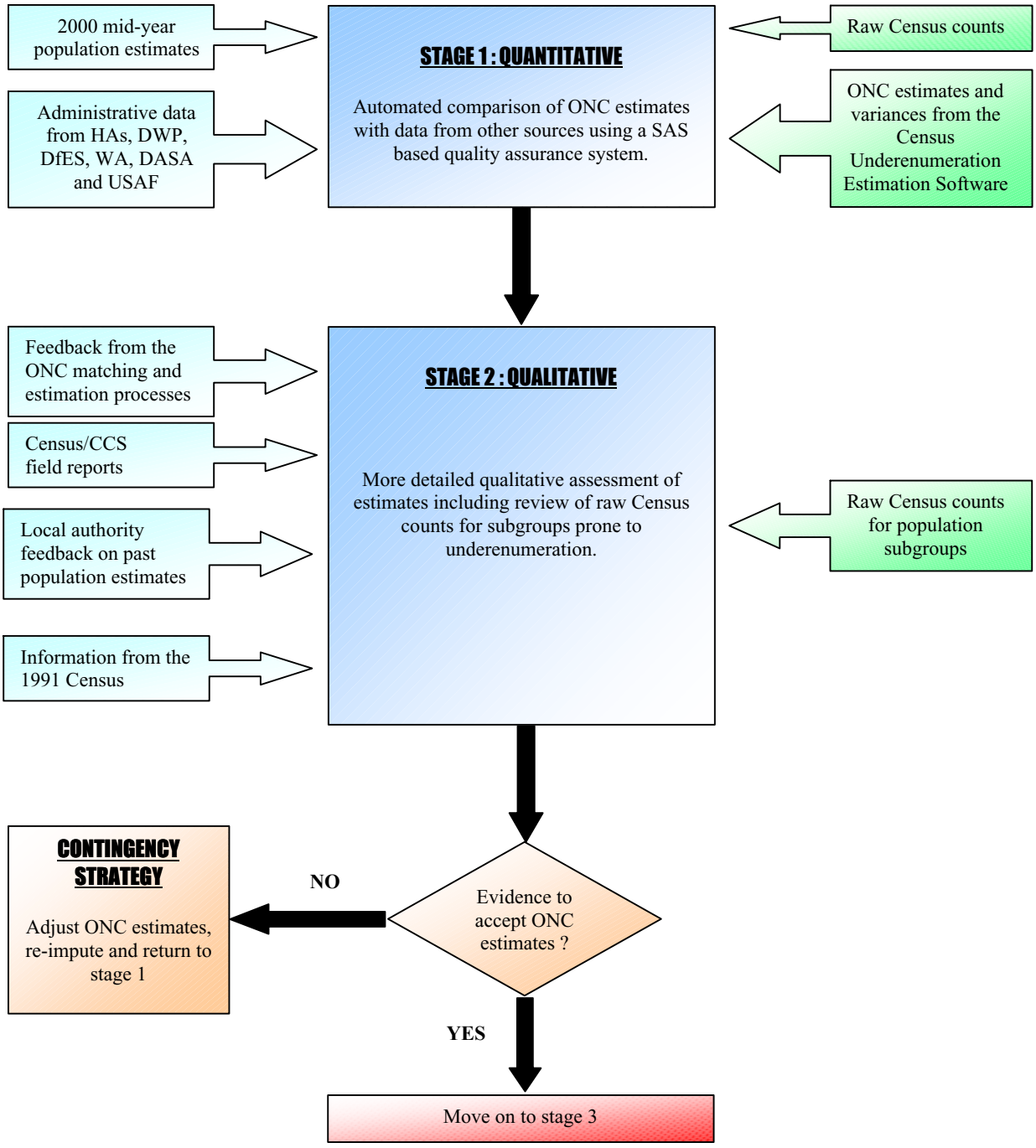
3.2.2.4 It should be noted that it was possible to accept the ONC estimates for some age-sex groups within an LAD while rejecting others. The contingency was invoked only for the age-sex groups that were rejected. The philosophy was that one or more ONC estimates could potentially fail the predetermined diagnostics but still be accepted

3.2.3 Contingency strategy

3.2.3.1 A contingency strategy was included as part of the ONC Quality Assurance strategy. This measure used the principle of 'borrowing strength' for situations when there was evidence, after having considered the qualitative and quantitative information that the CCS had failed within a DG or LAD. No other data sets were considered reliable and consistent enough to be substituted for the rejected ONC estimates. Therefore the contingency strategy used information about similar LADs that had already passed quality assurance to make adjustments to the rejected ONC estimates. The new estimates produced by this process were subject to the same quality assurance procedures as the initial estimates. Further detail on the contingency and the regression methodology underlying

Figure 3a

First and second stages of the sub-national quality assurance process



the strategy is outlined in 'A Quality Assurance and Contingency Strategy for the One Number Census' available from the National Statistics website at **www.statistics.gov.uk/census2001/pdfs/oncinfopaper.pdf**. Information on the occasions when the contingency strategy was invoked is available from the 'Key Findings and Actions from the One Number Census Quality Assurance Process'.

3.2.4 Stage 3 Population subgroup Quality Assurance

3.2.4.1 The quality assurance strategy included procedures to specifically monitor population subgroups known to be difficult to enumerate in censuses. This stage of the process, illustrated in **Figure 3b**, was undertaken once the ONC estimates had been quality assured and accepted at the DG and LAD level and once ONC imputation had been run. Separate comparisons of ONC estimates were undertaken for population counts with relevant comparative data: by single year of age and for specific population subgroups including prisoners, full-students, home armed forces personnel and foreign armed forces (FAF) personnel and their dependants.

3.2.4.2 If these estimates were accepted during this stage then the process moved on. However, if the ONC estimates were not accepted then this information was fed back to the imputation system, imputation was adjusted and the new estimates were quality assured

Further detail on the three distinct stages of the quality assurance at the sub-national level is described in **section 3.4**.

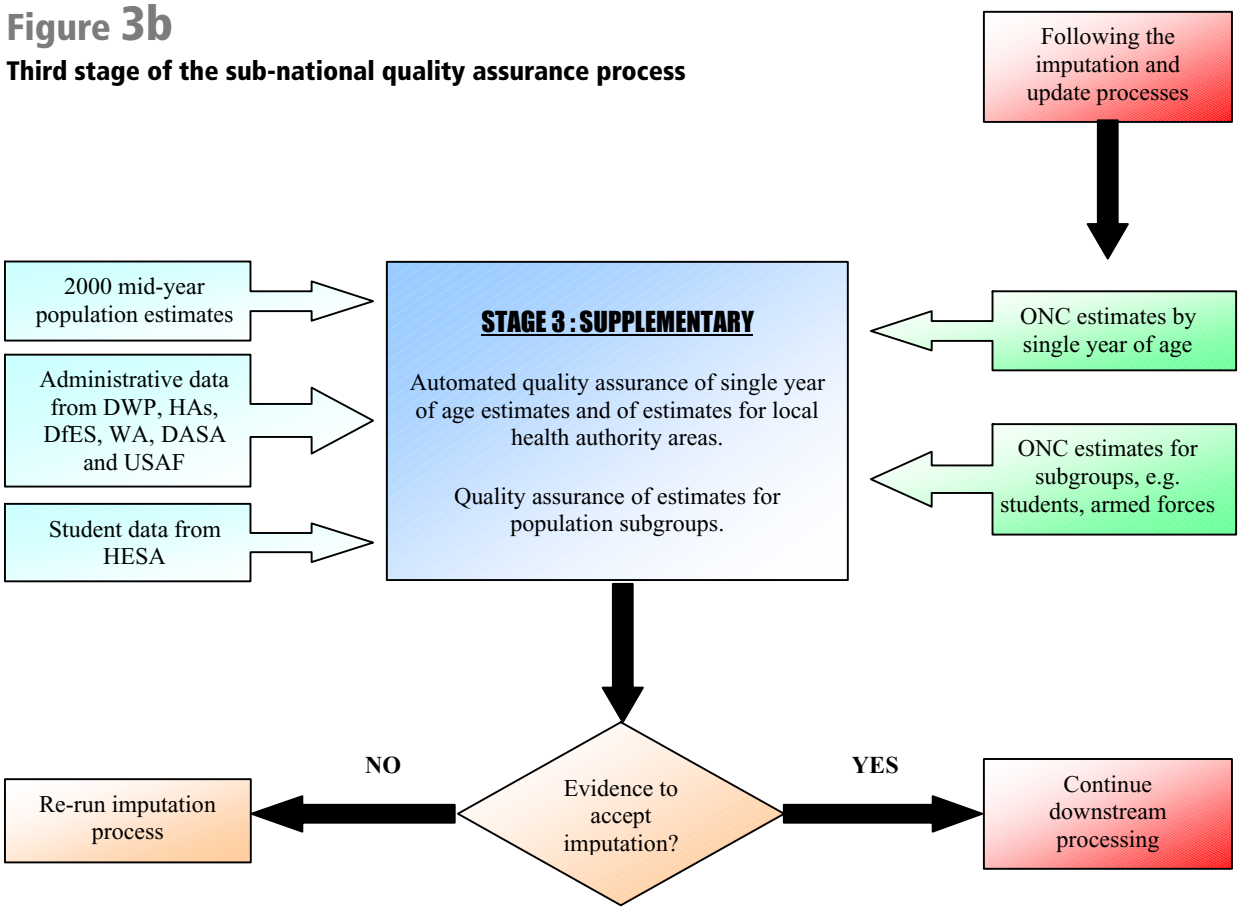
3.3 Formal process of quality assuring ONC estimates

3.3.1 The ONC estimates were subject to review and discussion by an expert panel at series of quality assurance meetings. The formal sub-national quality assurance process commenced November 2001 and involved weekly meetings. By the time the final ONC results were agreed, about 60 quality assurance meetings had been held and each of the 376 local authorities in England and Wales considered in turn.

3.3.2. Representatives from Northern Ireland Statistical and Research Agency (NISRA) and General Register Office for Scotland (GROS)

Figure 3b

Third stage of the sub-national quality assurance process



also attended a quality assurance meeting. The quality assurance process was also subject to external observation. Representatives from Tees Valley Joint Strategic Unit, Greater London Authority and the former Chief Statistician of the Scottish Office all attended a quality assurance meeting.

3.3.3 In addition to the weekly quality assurance meetings there were a series of high level quality assurance meetings involving members of senior management within ONS. At these meetings attendees discussed the cumulative ONC results presented and the differences between the mid-year population estimates and the ONC estimates. Discussions focused on both the ONC and MYE related issues attributing to the differences in the estimates. The effects of the 1991 base adjustments and international migration were covered, as were the issues of dependency and armed forces. A number of actions resulted from these meetings to further analyse the ONC and MYE related issues. The evidence presented at these high level meetings highlighted a large enough bias in the ONC estimates to warrant a dependency adjustment. Further information on the dependency adjustment is outlined in the 'Key Findings and Actions from the One Number Census Quality Assurance Process' and in the Dependency Paper available at **www.statistics.gov.uk/census2001/pdfs/dependency_paper.pdf**.

3.4 Detailed sub-national quality assurance procedures

The following section describes in more detail the stages in the sub-national quality assurance process summarised in **section 3.2**

3.4.1 Stage 1 Quantitative Quality Assurance

3.4.1.1 This first, quantitative stage involved quality assuring the ONC estimates for a number of indicators within each LAD and DG, namely: a count of all those aged under one (both sexes combined); a count for each sex for the 5 year age groups; a sex ratio for each of the age groups; and young and old dependency ratios.

A diagnostic range against which to compare each of these ONC estimates was constructed from appropriate comparative data.

3.4.1.2 The major criteria for selection of the comparators were that the data were available at both national and sub-national levels, were known to be reasonably accurate (or had been cleaned by an expert) and were reasonably timely.

3.4.1.3 **Table 1** outlines the different comparators that were used when quality assuring the ONC estimates at the sub-national level. It also details the sources and time references for these data and lists the age ranges

Table 1

Comparators for quality assurance of the sub-national ONC estimates by age

Comparator	Time reference	Source	Age groups
Demographic estimates		P&D	0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39,40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84 and 85+
Published rolled-forward MYEs for 2000	30 June 2000		
2000 MYEs extrapolated to mid-2001 based upon average annual changes	30 June 2001		
Estimates of under 1s	29 April 2001	P&D	0
Registered births during year prior to Census Day adjusted for infant mortality and migration			
HA patient register records	30 June 2000 & 29 April 2001	Individual health authorities via P&D	0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84 and 85+
Number of patients registered with NHS GPs			
Child benefit records	August 2000	Department for Work and Pensions (DWP), cleaned by Oxford University	0, 1-4, 5-9 and 10-14
Number of children aged under 16 for whom child benefit is claimed			
Pension records	May 2000	Department for Work and Pensions (DWP), cleaned by Oxford University	65-69, 70-74, 75-79, 80-84 and 85+
Number of people aged 65+ claiming state retirement pension and/or other benefits			
School census data	1 January 2000 & 1 January 2001	Department for Education and Skills (DfES) and the Welsh Assembly (WA)	5-9 and 10-14
Number of children aged 5-14 attending all educational establishments			

for which each of these comparators were used. A fuller description of the comparator data sources used in the quality assurance process is outlined in **Annex F**

3.4.1.4 The following points about the comparators were noted.

- Whilst patient registers provide migration information used in producing the rolled forward MYEs, they do not contribute to their absolute value. Therefore the population estimates from these two sources were considered to be independent.
- The 2000 MYES extrapolated to mid-2001 were not of sufficient quality to replace the 2000 MYEs and were used alongside them as a separate comparator. They were produced specifically for the ONC quality assurance and were not published for use for any other purpose.
- Health Authority (HA) patient register data was adjusted for coverage differences with the resident population (i.e. home armed forces and FAF and their dependents were added).
- The armed forces data used to make these adjustments based an individual's location on where they were stationed rather than their home address.
- Child benefit and pension records were cleaned to correct or add postcodes to enable all records to be allocated to a specific LAD.
- School census data bases an individual's location on where they study rather than their home address.

3.4.1.5 Individual charts for each indicator (population counts by five year age-sex group, sex ratios and dependency ratios) for each DG and LAD were available to the quality assurance panel as well as a figure sheet that contained population counts by five year age-sex groups for ONC estimates, administrative sources and mid year population estimates used in the quality assurance process.

3.4.1.6 **Table 2** is an example quantitative report for the fictitious area used in the Quality Assurance illustrative pack. Please note that only DG level information has been shown. The charts and figure sheets for the constituent LADs can be found by accessing **www.statistics.gov.uk/census2001/pdfs/onc_qa_pack.pdf**

Table 2

Example figure sheet

LP Elmbry and Wantown **Population estimates**

Age Group	Coverage	Census	ONC	2000 MYE	Adjusted Patient Records	Pension/ Child Benefit	Pop'tion Est.<1/ School Census	2001 Extrap'ns	Diag. Range Upper Bound	Diag. Range Lower Bound	ONC Upper 95% CI	ONC Lower 95% CI
MF0	80.1%	5,665	7,071	7,706	7,141	6,953	7,562	7,715	8,093	6,669	7,759	6,383
M1-4	82.5%	11,984	14,532	15,238	15,391	14,549		15,512	16,043	14,114	15,546	13,519
M5-9	83.9%	15,194	18,117	18,885	19,239	18,023	18,130	19,156	19,936	17,353	19,507	16,728
M10-14	87.6%	15,503	17,691	17,324	18,680	17,577	17,237	18,115	19,369	16,074	18,896	16,487
M15-19	85.4%	14,177	16,600	16,608	17,903			16,744	18,461	15,981	17,660	15,541
M20-24	77.0%	14,174	18,418	21,590	20,533			21,498	23,705	18,574	19,819	17,018
M25-29	78.0%	17,369	22,260	24,029	27,078			22,944	27,630	22,403	23,919	20,601
M30-34	80.3%	18,579	23,138	26,247	29,520			24,907	29,988	24,713	25,069	21,208
M35-39	82.0%	18,690	22,785	26,648	29,138			27,940	30,247	25,539	24,315	21,255
M40-44	86.1%	16,259	18,894	20,613	23,927			22,013	24,471	20,069	20,142	17,647
M45-49	90.0%	13,970	15,514	17,098	19,729			17,500	20,046	16,782	16,342	14,685
M50-54	93.0%	13,912	14,965	16,248	18,510			16,437	18,939	15,806	15,598	14,333
M55-59	90.7%	11,069	12,209	12,265	14,311			12,829	14,617	11,959	12,974	11,444
M60-64	93.1%	9,806	10,531	11,040	12,586			10,834	12,865	10,555	11,029	10,034
M65-69	91.9%	8,630	9,390	9,288	10,622	9,292		9,334	10,835	8,945	9,921	8,859
M70-74	91.6%	6,940	7,578	7,380	8,430	7,434		7,435	8,635	7,118	8,055	7,102
M75-79	94.6%	5,322	5,625	5,711	6,261	5,616		5,633	6,458	5,385	5,872	5,378
M80-84	93.9%	3,307	3,521	3,317	3,820	3,367		3,322	4,010	3,091	3,784	3,258
M85+	94.5%	2,377	2,515	2,551	2,951	2,469		2,540	3,193	2,198	2,706	2,324
MF0	80.1%	5,665	7,071	7,706	7,141	6,953	7,562	7,715	8,093	6,669	7,759	6,383
F1-4	82.5%	11,632	14,105	14,607	14,821	13,978		14,712	15,291	13,593	15,245	13,239
F5-9	83.8%	14,493	17,287	18,114	18,572	17,301	17,272	18,056	19,024	16,655	18,571	16,003
F10-14	87.8%	14,852	16,909	16,576	18,114	16,754	15,980	17,258	18,549	15,240	17,937	15,882
F15-19	85.3%	13,838	16,217	15,201	16,990			15,838	17,518	14,673	17,393	15,041
F20-24	78.9%	15,616	19,804	20,396	20,660			20,499	23,072	18,203	21,435	18,173
F25-29	79.2%	19,425	24,531	20,985	26,677			20,322	26,847	19,989	26,445	22,616
F30-34	82.2%	21,101	25,684	23,405	27,565			22,692	27,629	22,605	27,482	23,886
F35-39	86.0%	20,849	24,247	24,221	25,642			24,793	26,441	23,356	25,677	22,817
F40-44	87.9%	17,886	20,352	20,133	21,401			20,852	22,160	19,463	21,450	19,253
F45-49	91.6%	15,179	16,570	16,810	17,341			17,073	18,012	16,164	17,336	15,804
F50-54	92.5%	15,028	16,241	16,641	17,079			16,674	17,399	16,225	16,905	15,577
F55-59	93.9%	12,115	12,904	13,143	13,928			13,678	14,524	12,677	13,330	12,478
F60-64	91.8%	10,541	11,476	11,453	11,891			11,210	12,250	10,807	12,069	10,883
F65-69	93.3%	9,334	10,003	9,814	10,486	9,857		9,615	10,861	9,216	10,417	9,590
F70-74	91.8%	8,420	9,168	8,924	9,587	8,907		8,699	9,926	8,281	9,677	8,658
F75-79	95.8%	7,698	8,037	8,160	8,636	8,020		7,807	8,955	7,436	8,297	7,776
F80-84	95.6%	5,663	5,926	5,796	6,320	5,843		5,738	6,621	5,395	6,152	5,700
F85+	92.2%	6,049	6,564	6,807	7,098	6,376		6,629	7,787	5,582	7,037	6,091
Male	85.4%	220,162	257,906	275,995	302,264			278,648				
Female	86.8%	242,483	279,475	274,978	296,317			275,905				
TOTAL	**86.1%**	**462,646**	**537,381**	**550,972**	**598,581**			**554,553**				

LP Elmbry and Wantown **Sex ratios**

Age Group	Census	ONC	2000 MYE	Patient Records	Adjusted Patient Records	Pension/ Child Benefit	Pop'tion Est.<1/ School Census	01 Extrap'ns	Diag. Range Upper Bound	Diag. Range Lower Bound	ONC Upper 95% CI	ONC Lower 95% CI
0		104.9	103.2	103.6	103.6	103.2	104.8	105.2	106.2	102.2		
1-4	103.0	102.0	104.3	103.1	103.8	104.1		105.4	106.6	102.0	107.8	97.1
5-9	104.8	104.8	104.3	103.7	103.6	104.2	105.0	106.1	107.3	102.3	111.8	95.7
10-14	104.4	104.6	104.5	104.8	103.1	104.9	107.9	105.0	110.2	100.8	110.4	98.0
15-19	102.4	102.4	109.3	103.1	105.4			105.7	112.4	100.0	107.6	94.3
20-24	90.8	93.0	105.9	92.5	99.4			104.9	112.5	85.8	95.6	83.9
25-29	89.4	90.7	114.5	98.6	101.5			112.9	122.5	90.7	98.0	83.8
30-34	88.0	90.1	112.1	110.4	107.1			109.8	114.7	104.6	99.0	85.2
35-39	89.6	94.0	110.0	116.0	113.6			112.7	118.9	107.0	99.1	88.3
40-44	90.9	92.8	102.4	116.8	111.8			105.6	124.0	95.2	99.5	88.6
45-49	92.0	93.6	101.7	116.5	113.8			102.5	123.9	94.3	97.2	90.2
50-54	92.6	92.1	97.6	112.8	108.4			98.6	120.4	90.1	94.5	86.1
55-59	91.4	94.6	93.3	108.9	102.8			93.8	116.8	85.5	97.6	89.3
60-64	93.0	91.8	96.4	109.0	105.8			96.6	115.3	90.1	97.9	83.2
65-69	92.5	93.9	94.6	103.6	101.3	94.3		97.1	108.3	89.6	99.2	86.7
70-74	82.4	82.7	82.7	90.2	87.9	83.5		85.5	94.0	78.9	82.4	82.3
75-79	69.1	70.0	70.0	75.0	72.5	70.0		72.2	77.4	67.5	72.0	67.7
80-84	58.4	59.4	57.2	61.5	60.4	57.6		57.9	63.6	55.1	60.2	54.8
85+	39.3	38.3	37.5	45.5	41.6	38.7		38.3	49.5	33.5	40.2	34.9

Table 2

Example figure sheet - *continued*

LP Elmbry and Wantown Dependency ratios

	Census	ONC	2000 MYE	Patient Records	Adjusted Patient Records	2001 Extrap'ns	Diag. Range Upper Bound	Diag. Range Lower Bound	ONC Upper 95% CI	ONC Lower 95% CI
Young	28.9	29.1	28.9	26.4	27.1	29.3	30.7	25.0	29.1	29.1
Old	20.6	18.8	18.1	18.1	18.0	17.7	18.4	17.5	18.8	18.8

$$\text{Sex Ratio} = \frac{\text{Male population estimate in age group x}}{\text{Female population in age group x}} \times 100$$

$$\text{Young Dependency Ratio} = \frac{\text{Population aged 0 – 14}}{\text{Population aged 15 – 64}} \times 100$$

(Those economically inactive aged 0 – 14 as a percentage of those who are economically active aged 15 – 64)

$$\text{Old Dependency Ratio} = \frac{\text{Population aged 65+}}{\text{Population aged 15 – 64}} \times 100$$

(Those economically inactive of pensionable age as a percentage of those who are economically active aged 15 – 64).

3.4.1.7 The figure sheet contained data for each of the comparator data sources used to derive the diagnostic range for each DG and LAD. Comparisons could be made between the ONC estimates and the comparator data sources as well as between the Census counts for each age-sex group. Comparisons were generally conducted between the ONC estimates and the 2000 MYEs though it was noted that the ONC estimates would not be adjusted on the basis of the MYEs alone. Moreover, the figure sheet gave the quality assurance panel information on coverage rates (census count as a proportion of the ONC estimate). This information was used when quality assuring the ONC estimates and was of interest to the quality assurance panel to identify which age-sex groups had been subjected to the largest adjustments.

3.4.1.8 Below is a set of example quantitative charts (**Figure 4a**, **4b**, **4c** and **4d**) for the fictitious DG presented in the illustrative Quality Assurance pack. The information conveyed in the charts was used in conjunction with the figure sheet in the quantitative quality assurance of the ONC estimates.

Figure 4a

ONC Estimates of the male population by age group

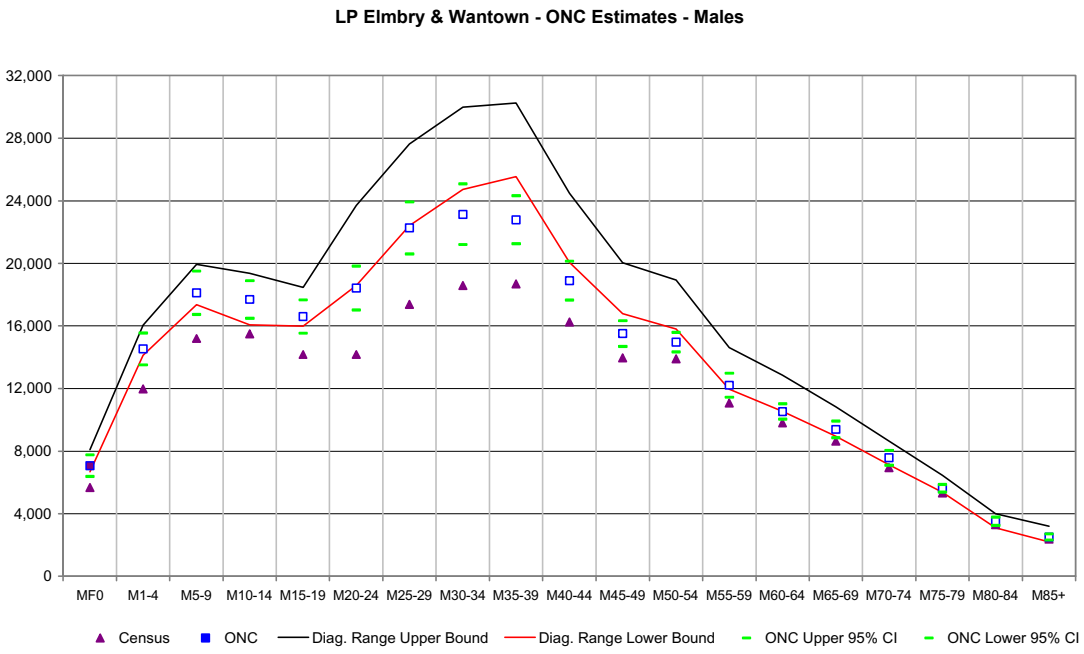


Figure 4b

ONC Estimates of the female population by age group

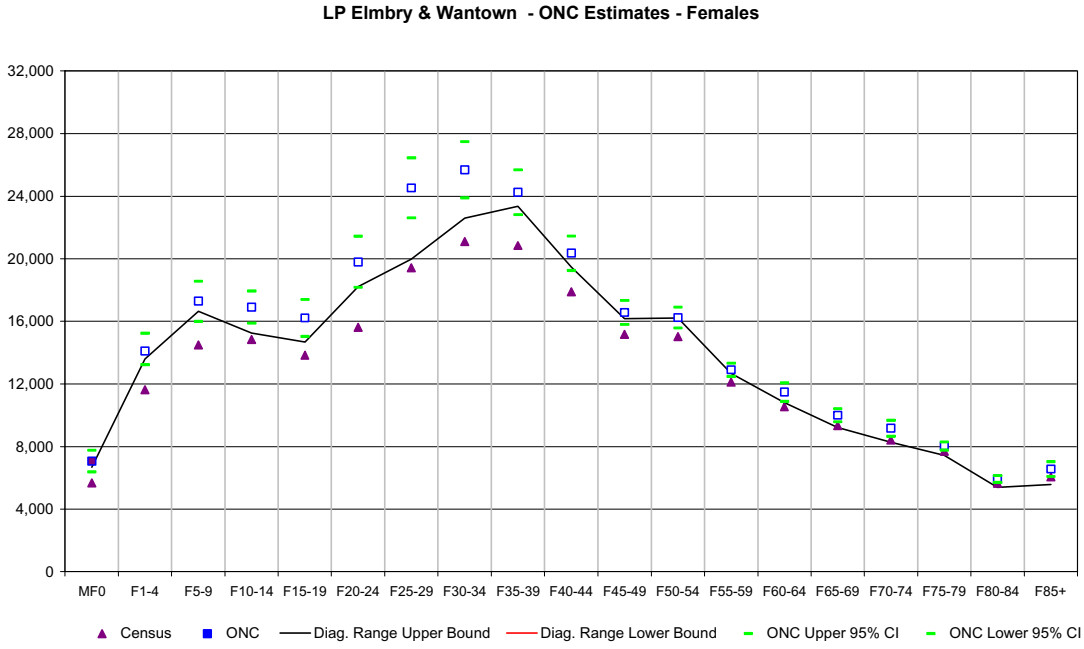


Figure 4c

ONC estimates of the sex ratio by age group

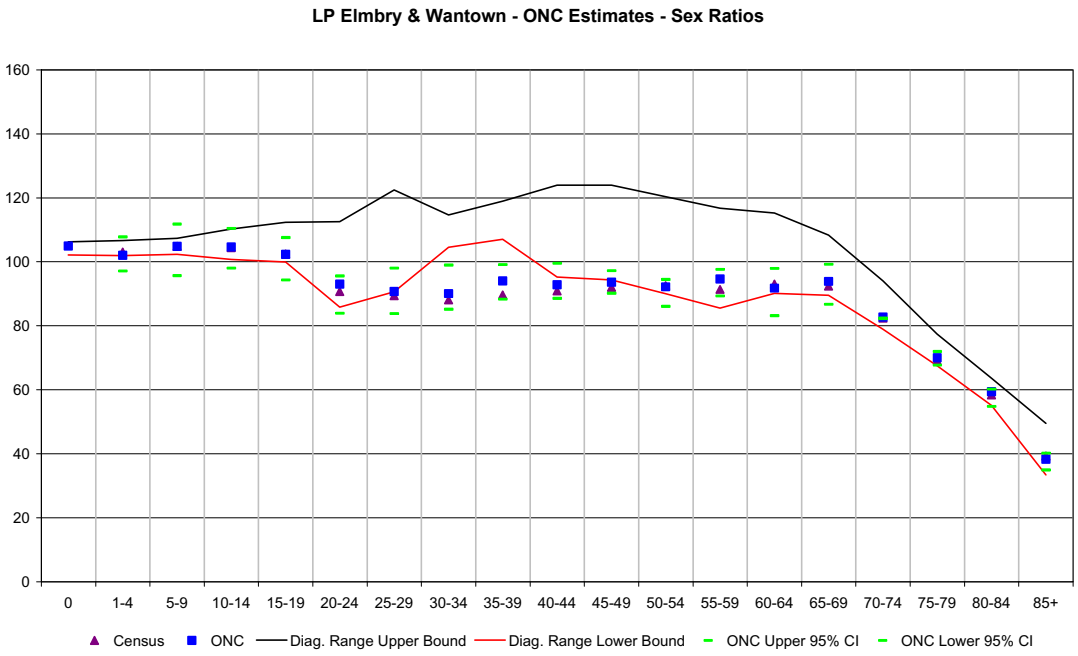
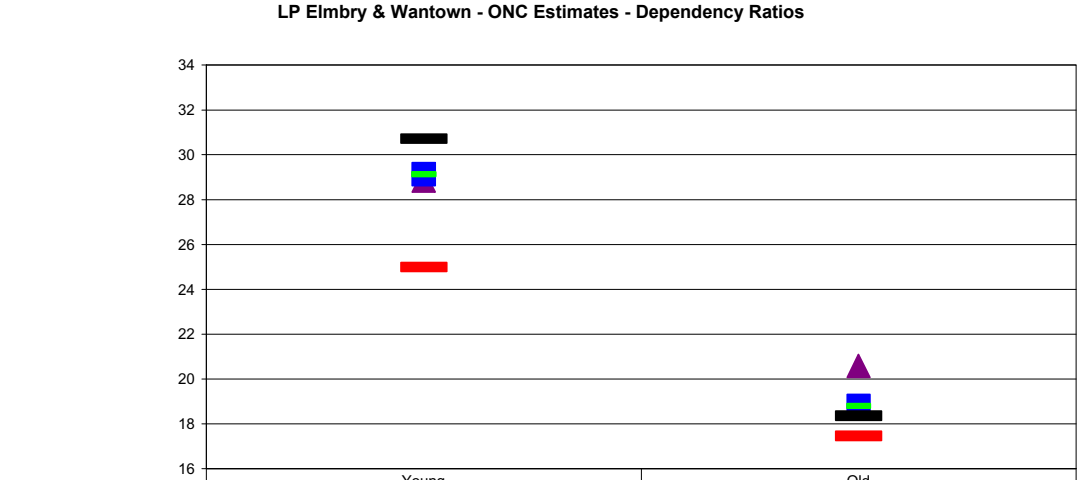


Figure 4d

ONC estimates of the young and old dependency ratio



	Young	Old
Census	28.9	20.6
ONC	29.1	18.8
Diag. Range Upper Bound	30.7	18.4
Diag. Range Lower Bound	25.0	17.5
ONC Upper 95% CI	29.1	18.8
ONC Lower 95% CI	29.1	18.8

3.4.1.9 If the estimate and the entire confidence interval fell within the diagnostic range for all charts presented to the quality assurance panel, there was strong evidence to support accepting the estimate. However it should be emphasised that no decision to accept or reject was made without considering the full range of information available, including the qualitative information detailed in **section 3.2.2**

3.4.1.10 Similarly, if the ONC estimates tended to be above or below the diagnostic range then this was not taken as final evidence to reject the ONC estimates. Instead it was seen as an indication of cause for concern, again to be investigated fully using the qualitative information.

3.4.1.11 From the figure sheet and charts above it can be seen that the adjusted Census count falls within the diagnostic range for all female and most male age groups, showing that the Census is generally in line with other sources for these groups. However, the Census count in some of the young male groups is below the diagnostic range The quality assurance panel would have noted that ONC estimates for males aged between 20 and 54 all fell below the lower bound of the diagnostic range. In particular, that the estimates for males aged 30 – 54 fell noticeably below the diagnostic ranges despite having sizeable adjustments. The quality assurance panel would have noted that the apparent shortfall in the ONC estimates of males aged 30 – 54 would cause low sex ratios particularly as the estimates for females in the same age groups fell within the diagnostic range and were higher than the corresponding estimates for males. Under such circumstances the quality assurance panel would have agreed further actions for the ONC team or Population estimates team to investigate. This may have included checking the proportion of single young males in the Census and CCS data as well as looking at the imputation weights and comparing with the Labour Force Survey data and other external expectations.

3.4.1.12 The quality assurance panel would have noted the difference between the total ONC estimate and the Health Authority adjusted patient record figure. (For further information on the adjusted patient record figure used in the ONC quality assurance process see 'A Quality Assurance and Contingency Strategy for the One Number Census' at **www.statistics.gov.uk/census2001/pdfs/oncinfopaper.pdf** for further information). In this instance the ONC estimates would not have been rejected by the quality assurance panel but would have been investigated further using the qualitative information (outlined in **section 3.4.3** and **Annex G** below).

3.4.2 Stage 2 Qualitative Quality Assurance

3.4.2.1 The second stage of the quality assurance procedure involved a systematic assessment of the following sources of information:

- information from the quantitative quality assurance procedures (e.g. graphs, tables of differences, extent to which diagnostic ranges are exceeded);
- information from the ONC estimation procedures (e.g. model parameters, off-diagonal components of the dual system estimators);
- past feedback received by ONS from LADs on the rolled-forward demographic estimates;
- information from the field, both Census (Census Field Management Information System - FMIS) and CCS (Team Reporting and Communication System - TRACS); and
- details of the adjustments made to the 1991 Census estimates.

Further detail on the sources of qualitative information available is outlined in **Annex G**.

3.4.2.2 A qualitative report was produced for each LAD drawing together the field information from the Census and CCS from various sources at various geographical levels (DG, LAD or CCS Team Manager (TM) area) as well as information provided by the ONC estimation process. An example qualitative report for one of the fictitious LADs used in the Quality Assurance illustrative pack is outlined in **Annex H**.

3.4.2.3 These reports were used to assess the success of the Census and CCS field operation in each LAD. Any issues identified within the field during the Census and CCS operation, or information from the estimation process was used in conjunction with the quantitative data to assess the plausibility of the ONC estimates.

3.4.2.4 Further qualitative information was provided by population profiles provided by the PEU in ONS. These were used alongside the qualitative and quantitative reports to quality assure the ONC estimates and highlighted

any past feedback received by ONS from LADs on the rolled-forward demographic estimates. A population profile was available for each constituent LAD within a DG and also provided mid year population estimates from 1991 onwards and information on annual population change and migration. An example population profile for one of the LADs used in the Quality Assurance illustrative pack is outlined in **Annex I**. The population profile for the other LAD can be found by accessing the following link: **www.statistics.gov.uk/census2001/pdfs/onc_qa_pack.pdf**.

3.4.2.5 The population profile was used to look at population change over the intercensal period and to note whether the area was one of high migration. This was used to feed into the quantitative review of the ONC population estimates.

3.4.2.6 Despite the wide range of detailed quantitative and qualitative information available, the quality assurance panel frequently requested additional information or analysis to be carried out. These additional analyses are outlined in detail in 'Key Findings and Actions from the One Number Census Quality Assurance Process"'

3.4.3 Stage 3 Population subgroup Quality Assurance

3.4.3.1 The third stage of the sub-national quality assurance process followed the ONC imputation process. Separate comparisons of ONC estimates were undertaken for population counts with relevant comparative data: by single year of age and for specific population subgroups including prisoners, full-students, home armed forces personnel and FAF personnel and their dependants.

Single years of age comparison

3.4.3.2 The single year of age comparisons were conducted at the LAD and DG level but only presented to the quality assurance panel at the DG level. LAD comparisons were conducted but only presented to the panel if there were concerns. The comparator data used in the comparisons included the HA patient register records, the 2000 mid-year population estimates and the 2000 mid-year estimates (MYEs) extrapolated to mid-2001 for all age groups and the pension records, school census data and child benefit records for specific ages.
The following, **Figure 7a** and **7b**, are example single years of age charts for males for the fictitious area used in the Quality Assurance illustrative pack:

Figure 7a

Example single year of chart for males aged 1 – 44

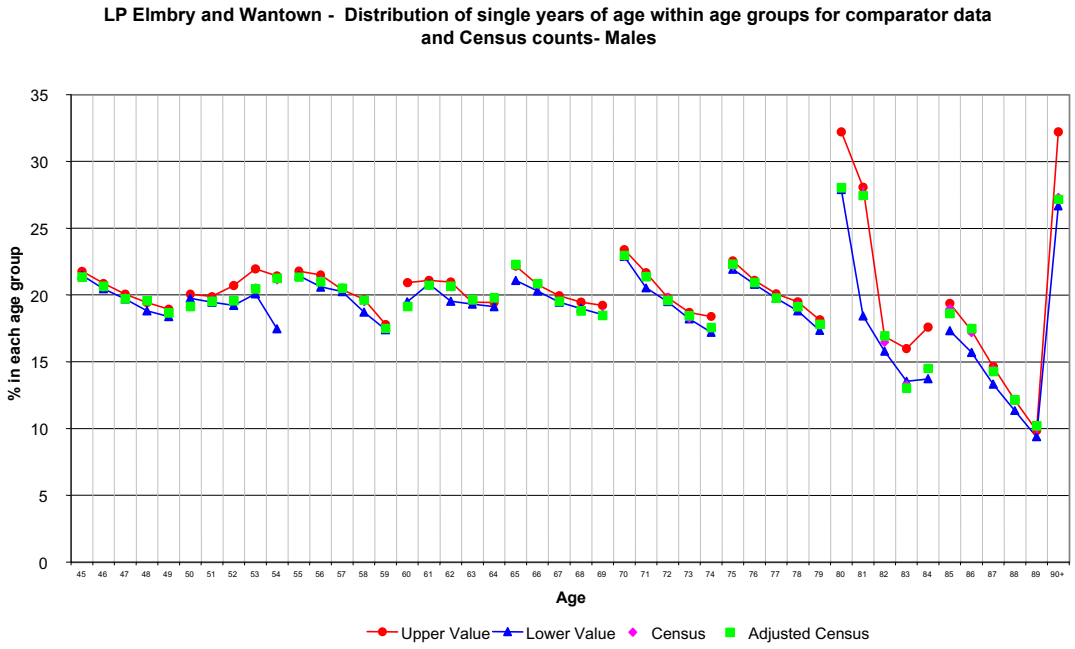
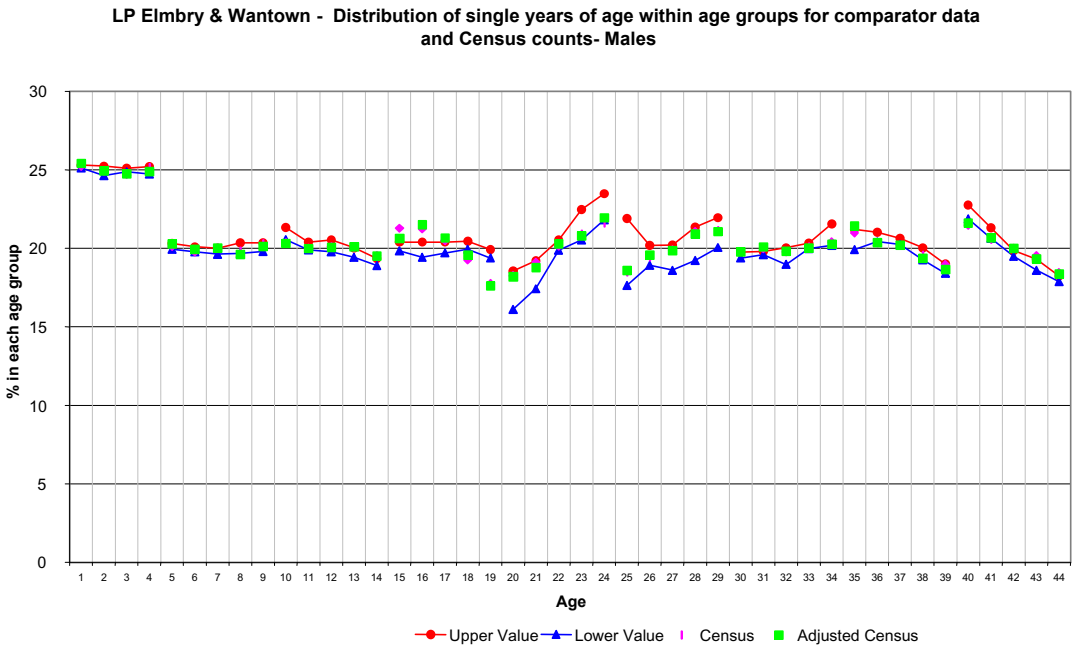


Figure 7b

Example single year of chart for males aged 45 – 90+



3.4.3.3 Concerns were raised if the adjusted Census figure deviated from the Census figure and fell outside of either the upper or lower value of the comparator data. This process ensured that the distribution of ages within age-sex groups were consistent with the same sources of demographic and administrative comparator data sources and ensured that the ONC imputation system had not imputed more people into one age at the expense of other ages.

3.4.4 Population subgroup comparisons

Home armed forces personnel

3.4.4.1 Home armed forces personnel were one of the subgroups known to be difficult to enumerate. They are subject to frequent changes in location, often at short notice. Many also live in large communal accommodation blocks, which can lead to problems, which can lead to problems in common with other establishments such as student halls of residence. There may also have been confusion amongst younger members of the services as to whether they should have been counted where they were stationed or at their parents' home address.

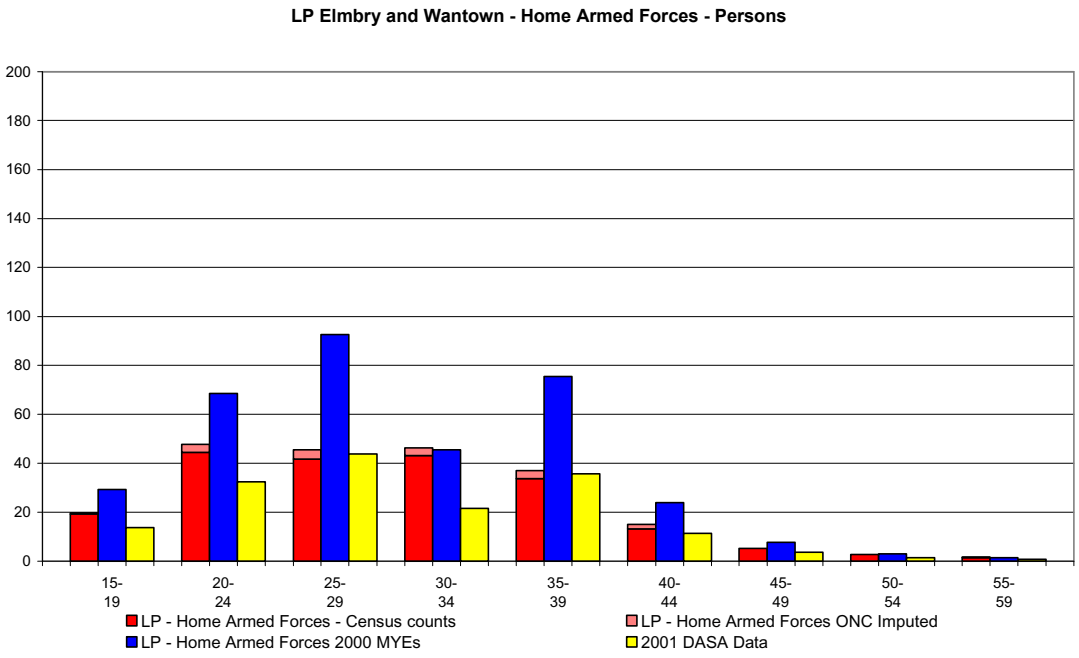
3.4.4.2 As part of the quality assurance process, the numbers of home armed forces by age were compared with comparative data for each DG. The comparative data were 2000 MYEs of the home armed forces population supplied by the PEU. They consisted of estimates of the number, by age, in each LAD. These estimates were based on data supplied by the Defence Analytical Services Agency (DASA) which detailed the number stationed in each LAD. The PEU applied a base to residence matrix, based on 1991 Census data, to estimate the number of those serving at each base who were resident in each of the local LADs. 2001 DASA data was available to the ONC team at LAD level but only for total population. This data did not have the base to residence matrix applied to it and reflected the armed forces personnel stationed in each LAD. This comparator was used in the quality assurance process to look at home armed forces personnel but involved applying the age-sex distribution of the 2000 MYEs of home armed forces to the total for each area.

3.4.4.3 Initial comparisons were drawn at the DG level and by five-year age groups. However both the census and comparative data were available at LAD level should a lower level of detail need to be investigated.

Figure 8 is an example home armed forces chart for the fictitious area used in the QA illustrative pack.:

Figure 8

Example home Armed Forces chart



Differences between the ONC estimates and 2000 MYEs of home armed forces personnel were noted early on in the quality assurance process. The resulting work that was instigated as result of this finding is outlined in further detail in 'Key findings and Actions from the One Number Census Quality Assurance Process' that accompanies this paper.

Foreign Armed forces personnel and dependants

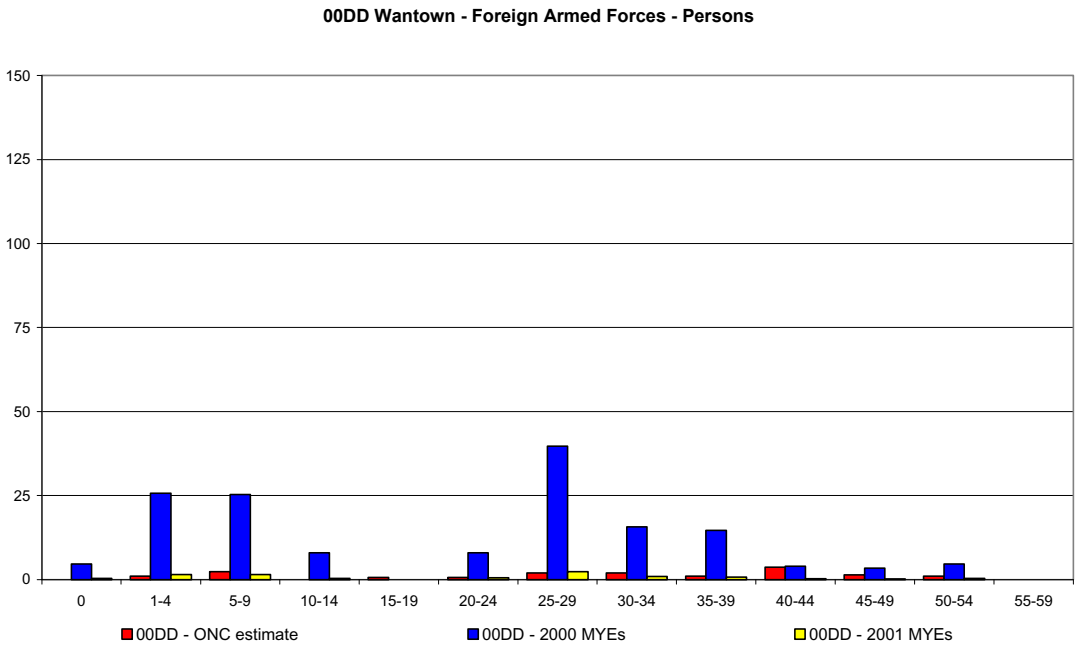
3.4.4.4 The numbers of FAF personnel and dependants by age were compared with comparative data for each DG. The comparative data were 2000 MYEs of the FAF population supplied by the PEU. They consisted of estimates of the number, by age and sex (though comparisons were only made by age), in each LAD. These estimates were based on data supplied by the United States Air Forces (USAF) which detailed the number of personnel and dependants at their home address who had resided or who had intended to reside in the UK for 6 months or more beginning on or before 29th April 2001.

3.4.4.5 It was originally proposed that the comparisons of FAF personnel and dependants would be undertaken for each DG as part of the population subgroup quality assurance process. It transpired, however, that this comparison would have to be done at a further stage in processing given the restriction of processing tools to derive the exact counts required.

3.4.4.6
Initial comparisons were drawn at the DG level and by five-year age groups. However both the census and comparative data were available at LAD level should a lower level of detail need to be investigated. **Figure 9** is an example FAF Chart

Figure 9

Example foreign Armed Forces chart



3.4.4.7 Differences between the ONC estimates and 2000 MYEs had been noted early in the stage two process for home armed forces personnel and it was therefore anticipated that there might be a similar issue with the FAF personnel and dependants. On inspection of the ONC estimates for FAF the quality assurance panel agreed that further investigation was needed to assess the impact of the presence of FAF on the ONC estimate. Further analysis was undertaken that focused on this population subgroup and this is explained in further detail in "Key findings and Actions from the One Number Census Quality Assurance Process"

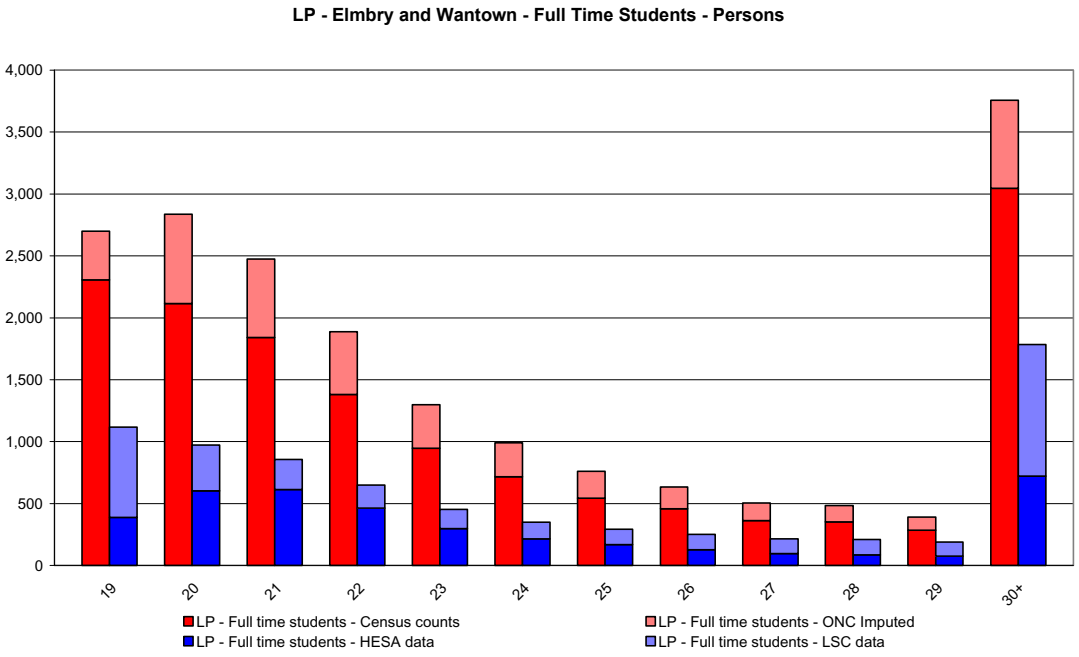
Full-time students

3.4.4.8 As part of the ONC quality assurance process, comparisons were also made between overall ONC counts of full-time students with comparator data. These comparator data included counts of full-time students by single year of age 19 to 29 and 30+ from the Higher Education Statistics Agency (HESA), the Learning and Skills Council (LSC) and the Welsh Funding Council (WFC). Data received from HESA related to ages as at August 2000 and data received from LSC related to December 2000. Data received from WFC related to the beginning of December 2000. These data sets were then adjusted to reflect ages as at Census day.

Figure 10a is an example full-time students chart for the fictitious area used in the Quality Assurance illustrative pack:

Figure 10a

DG level full-time student chart



3.4.4.9 Some discrepancies between ONC student counts and comparator data were expected as HESA and LSC data record students at their place of study rather than their home address, and the 2001 Census enumerated students at their term-time address. It was therefore expected that some students would reside in a different DG to the one in which they were studying.

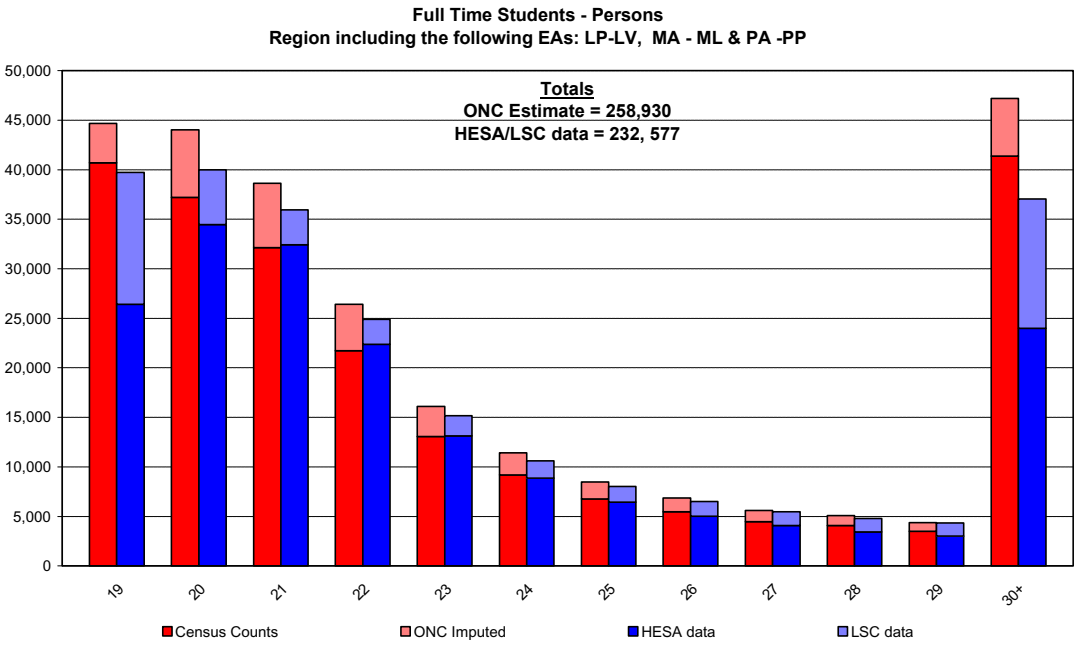
3.4.4.10 HESA data also assigns students to the administrative centre of the university rather than where the students are actually studying. In some cases a university campus is located in a different LAD to the administrative centre but the HESA data will include the students in the LAD where the administrative centre is.

3.4.4.11 Initial comparisons were drawn at the DG level by single year of age but this was extended to include a regional comparison to try and capture cross boarder flows of students.

Figure 10b is an example regional full-time students chart for the fictitious region:

Figure 10b

Regional level full-time student chart



3.4.4.12 If the quality assurance panel believed that the ONC estimates for students looked inconsistent with expectations then there were several actions undertaken to look at full time students. This involved contacting the Department of Health (DoH) to enquire over the possibility that students (both home and international) do not de-register from the GP patient records when they leave university. This would inflate the patient record figures (one of the administrative comparator data sources used in the quality assurance process) which may potentially inflate the diagnostic ranges. This and additional work conducted to look at the enumeration of halls of residences is outlined in more detail in 'Key Findings and Actions from the One Number Census Quality Assurance Process'.

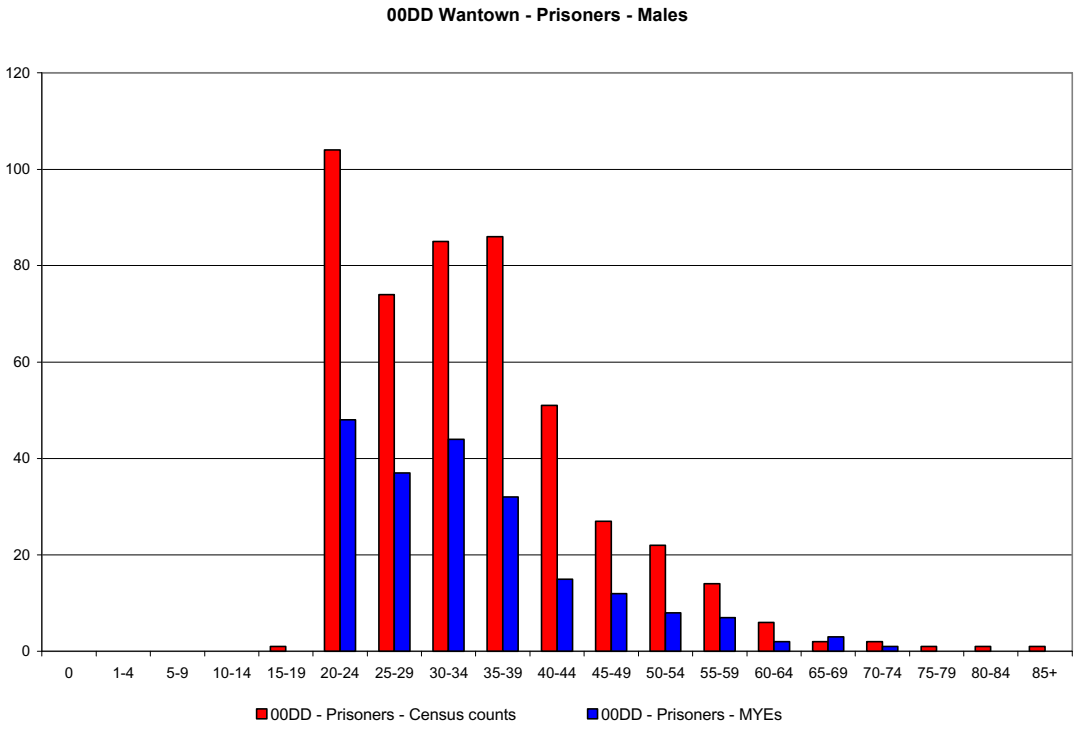
Prisoners

3.4.4.13 The numbers of prisoners by age and sex were compared with comparative data for each LAD. The comparative data were 2000 MYEs of the prison population supplied by the PEU. They consisted of estimates of the number, by age and sex, in each LAD. These estimates were based on data supplied by the Home Office (HO) which detailed the number of prisoners in each individual establishment.

Figure 11 is an example prisoners chart for one of the constituent LADs that comprise the fictitious area used in the Quality Assurance illustrative pack:

Figure 11

Example prisoners chart



3.4.4.14 Due to definitional differences, the 2001 Census generally enumerated more people in prisons than expected by the 2000 MYEs as is reflected in the example chart above. Further investigative work was undertaken to look at individual prison establishments for evidence of significant underenumeration where the ONC estimates of prisoners was lower than the comparator 2000 MYEs (this is outlined in 'Key Findings and Actions from the One Number Census Quality Assurance Process').

3.5 Cumulative Quality Assurance

3.4.4.15 Initial quality assurance took place at the LAD and DG level. This involved comparing the ONC estimates for the entire population of an area as well as for specific subgroups of the population with a range of quantitative and qualitative information. Another stage of the quality assurance process was at the national level. A cumulative approach was taken at this level, with the estimates from each DG being added to the cumulative total as they were accepted through the sub-national process. The cumulative charts for the population of an area and specific subgroups were available to the quality assurance panel and allowed the panel to monitor the relationship between the ONC estimates, the mid-year population estimates and the diagnostic ranges.

4 Summary

4.1 This report has outlined the ONC methodology and Quality Assurance procedure and further highlighted the range of information that was routinely available to the Quality Assurance panel. The ONC methodology and quality assurance process have ensured the ONC population estimates are sensible and of the right overall magnitude. It accompanies the 'Key Findings and Actions from the One Number Census Quality Assurance Process' report that has been produced to describe and explain the key findings resulting from the quality assurance process.

5 References

Abbott, O, Brown, J. J., Diamond, I. D. (2002) Dependence within the One Number Census. Draft technical paper available at **www.statistics.gov.uk/census2001/pdfs/dependency_paper.pd**f.

Brown, J. J., Diamond, I. D., Chambers, R. L., Buckner, L. J. and Teague, A. D. (1999) A methodological strategy for a one-number census in the UK. *J. R. Statist. Soc* **A, 162**, 247-267.

ONS (1998) Evaluation of the main objectives of the 1997 GB Census Test. AG(98)01

ONS (1999) A Guide to the One Number Census, Census Consultation Paper.

ONS (2000) A Quality Assurance and Contingency Strategy for the One Number Census. Advisory Group Paper AG(00)16. Available at **www.statistics.gov.uk/census2001/pdfs/oncinfopaper.pdf**

ONS (2001) A Quality Assurance and Contingency Strategy for the One Number Census.

ONS (2002) One Number Census Illustrative Quality Assurance Pack: **www.statistics.gov.uk/census2001/pdfs/onc_qa_pack.pdf**

ONS (2003) Key Findings and Actions from the One Number Census Quality Assurance Process

OPCS (1993) Rebasing the annual population estimates. Population Trends, **73**, 27-31.

OPCS (1994) Undercoverage in Great Britain. 1991 Census User Guide **58**. HMSO, London.

Simpson, S., Cossey, R. and Diamond, I. (1997) 1991 population estimates for areas smaller than districts. *Population Trends*, **90**, 31-39.

Vickers, P. and Yar, M. (1998) The development and evaluation of the donor imputation system (DIS) for the 2001 UK Census of population and housing. *Proceedings of the Joint IASS/IAOS Conference*, Mexico, September 1998.

Annex A: Example illustrative estimation summary report

LP Estimation summary report

Executive summary

The two LADs included in this DG are Wantown (00DD) and Elmbury (00DE). This area contains some of the residential buildings for Wansbury University.

Also, females aged 75 – 79 have been collapsed with females aged 80 – 84 as that age group had small sample sizes in the medium to count areas. Males aged 20 – 24 have been collapsed with males aged 25 – 29 as that age group had small sample sizes in the medium to count areas. Also, males aged 1 – 4 have been collapsed with females aged 1 – 4, and males aged 5 – 9 have been collapsed with females aged 5 – 9, due to differing coverage rates between the sexes.

The overall estimated coverage for this area is 86.1 per cent - which is as expected for an area of this type. The HtC areas have large estimated undercounts, with the hardest to count areas having a lower coverage as expected.

The quality assurance charts show that the estimates seem low for males aged between 30 and 54.

Design group level

Overall coverages

The overall coverage for this area is estimated to be 86.1 per cent. The coverages for the HtC index levels are 88.3 per cent, 85.9 per cent and 84.4 per cent for the easier and harder of the remaining HtC levels respectively, which is as we would expect.

Investigations of some influential postcodes have highlighted fieldwork problems in one postcode where very few households were found in the CCS. Postcode AA1 1AB has therefore been excluded from the sample in an attempt to remove the area where we have evidence that the CCS was poor.

Age group coverage

The females aged 75 – 79 strata has been collapsed with the females aged 80 – 84 strata, due to the fact that the medium to count level had a very small sample size. Also, the males aged 20 – 24 strata has been collapsed with the males aged 25 – 29 strata due to the fact that the medium to count level had a very small sample size. Also, the following age groups have been collapsed to smooth out some unusual looking estimates and sex ratios: males aged 1 – 4 with females aged 1 – 4, and males aged 5 – 9 with females aged 5 – 9.

The age groups with the highest estimated undercount were males aged 20 – 24 (77.0 per cent), males aged 25 – 29 (78.0 per cent), females aged 20 – 24 (78.9 per cent), and females aged 25 – 29 (79.2 per cent). These are the age groups we would expect to have relatively low coverages. However, the low coverages are all in the hardest to count levels and the males aged 25 – 29 are particularly low (76.1 per cent). For this stratum, there are many postcodes where a lot of extra people have been found by the CCS and as such they do not really class as outliers. As this is a London area so we would expect to have a large undercount in these age groups, and the quality assurance graphs show that the estimates are already a lot higher than the census count and fall inside the diagnostic ranges so it may be correct to leave it.

Babies have a low coverage at 80.2 per cent, but this is spread across all the HtC levels. The males and females aged 85+ have relatively reasonable coverages of 94.5 per cent and 92.2 per cent respectively.

At the age group by HtC level, the lowest coverages are all in the hardest to count areas, as we might expect. The lowest coverage is for males aged 25 – 29 as discussed above. The next two lowest coverages are for males and females aged 20 – 24 in the hardest to count areas (74.4 per cent and 75.6 per cent respectively). Again, there are lots of postcodes that have extra people in the CCS, but none that stand out as outliers.

Looking at the quality assurance charts, males aged between 30 and 54 are all low and this is causing low sex ratios for many of these age groups. However, we have estimated a lot of extra people in these age groups and the coverages are not largely dissimilar to those of the surrounding age groups. It may be worth checking that the diagnostic ranges are realistic.

Variances

The age group variances are reasonably small at DG level. The largest variances relative to the estimates are those for males aged between 20 and 29. This is consistent with the low coverages. Babies and females aged between 20 and 29 also have high variances (see below).

At age group by HtC level, males and females aged between 20 and 29 in the hardest to count areas have the highest variances relative to their estimates, and this is consistent with the investigations already carried out.

For babies, both HtC levels have high variances relative to the estimates, and this is probably due to the sample sizes being quite small: for the easier to count areas, there were 22 persons in the CCS postcodes, of which 3 were in the CCS only; for the medium to count areas, there were 14 persons in the CCS postcodes, of which 1 was in the CCS only; for the harder to count areas, there were 17 persons in the CCS postcodes, of which 2 were in the CCS only.

Local Authority District level

Overall coverage

The overall coverages for the LADs are as follows:

Wantown (00DD) 87.6 per cent
Elmbury (00DE) 84.6 per cent

These are fairly similar and are consistent with the overall coverage. Wantown has more of the population in the easiest to count areas (65.2 per cent), whereas Elmbury has most of it's population in the hardest to count areas (67.5 per cent).

To examine the amount of undercount in each LAD in relation to the amount of undercount in the whole DG, the following table was produced.

LAD	Census only	CCS only	Both	Total Census	% of total Census	% of CCS
Wantown (00DD)	237	192	2641	2878	51	46
Elmbury (00DE)	328	223	2397	2725	49	54

The LADs split the DG up fairly evenly, although slightly less people were found in the census in Elmbury and more were found in the CCS so this LAD will probably take slightly more of the undercount. However, the LAD effects are fairly small.

Age group coverage

The lowest four coverages for age group by LAD are for the groups we have already investigated: males aged between 20 and 29, and females aged between 20 and 29. Elmbury has the lowest seven coverages, but we would probably expect this given that most of the hardest to count areas fall into this LAD.

Looking at the quality assurance charts, males aged between 30 and 54 look low in both LADs and this is causing some low sex ratios. However, we have estimated a lot of extra people in these age groups so it may be worth checking that the diagnostic ranges are realistic.

Variances

Again, the results with the highest variances relative to their estimates are all for the age groups mentioned already. There are no other variances that are particularly high in relation to their estimates.

Annex B: Example imputation report

Imputation report for LP

Imputation summary

The imputation system imputed 30,885 households and 74, 261 individuals in total.

Household level summary

Analysis of the household weights showed the following:

a) The highest weights were for households containing two adults aged 15 – 34 or over 80 with no children, of Black/Black British ethnic origin, living in households that were part rented and part mortgaged, in the HtC 1 areas of both LADs;

b) Households of the following types were most likely to be missed: part rented and part mortgaged (1.39) or rented from a private landlord (1.31), containing people of Black/Black British ethnic origin (1.37) or Chinese ethnic origin (1.29), in the hardest to count areas (1.14), who are single men and women aged 15 – 34 (1.19 for men, 1.16 for women); and;

c) There was little difference in household weights across LADs (as this is driven by the estimation totals).

For the imputed households, the following table shows how the selection of donors performed (Search 0 is the best selection possible, searches 8 – 10 search the entire DG):

Donor type	Frequency	Cumulative percent	Cumulative frequency	Percent
0	23,579	76.36	2,357	76.34
1	2,041	6.62	25,620	82.95
2	1,070	3.48	26,690	86.42
3	29	0.11	26,719	86.51
4	57	0.04	26,776	86.70
5	111	0.37	26,887	87.06
6	2	0.02	26,889	87.06
7	29	0.11	26,918	87.16
8	1,130	3.67	280,487	90.81
9	7	0.04	28,055	90.84
10	2,830	9.18	30,885	100.00

The table shows that over 76 per cent of the donors were selected using the best possible search - and under 13 per cent required a search across the DG. This is an indication that the majority of donors were of good quality.
Of the imputed households, 27,276 of them were placed into dummy forms and 3,609 were placed into random postcodes.

Of the dummy form placements, almost 35 per cent were placed using the best possible match, as shown in the following table:

Score	Frequency	Cumulative percent	Cumulative frequency	Percent
3	1,088	3.99	1,088	3.99
4	3,832	14.05	4,920	18.04
5	5,179	18.99	10,099	37.03
6	7,686	28.18	17,785	65.20
7	9,491	34.79	27,276	100.00

In terms of the location of the dummy forms used, 4 postcodes have over 20 dummy forms used whilst 9 EDs had over 40 households placed into dummy forms.

Person level summary

11,911 persons were imputed into counted households. This proportion (16 per cent) of the total number of imputed persons shows that the majority of person undercount was still from within missed households. Therefore, the person weights will be quite small.

Analysis of the person weights showed the following:

a) The highest weights were for females aged 20 – 24 in households with an Other tenure, who are single, and live in the hardest to count areas of 00DE (Elmbry);

b) Persons of the following types were most likely to be missed in counted households: people who are in part rented and part mortgaged properties (1.04), who are looking for work (1.10) or in full time education (1.06), who are single (1.04), who are males aged 20 – 24 (1.09) or males aged 25 – 29 (1.07) or females aged 20 – 24 (1.08) or females aged 25 – 29 (1.08); and

c) There was little difference in person weights across, size of the household, HtC group or LAD.

Of the 11,911 persons imputed into counted households, 46.7 per cent were placed using the best possible search and no people were placed using the final searching algorithm as shown in the following table. This shows that the placing of the imputed persons was generally good.

Search	Number of persons placed	Percentage of persons placed
1 (Best)	5,564	46.7%
2	605	5.1%
3	3,266	27.4%
4	2,476	20.8%
5 (Worst)	0	0.0%

Pruning and Grafting summary

In total, there were 1,756 grafts and 2,031 Prunes. The initial prune and graft choice matrix is shown below, and indicates that in general we imputed too many households of size 5 since we have to do a lot of pruning and grafting to reduce the number of households in this group.

Household size	Current total	Target total	Grafts to do	Prunes to do	Prunes in size 7+
1	58,888	58,548	372	0	-
2	37,624	38,135	0	0	-
3	14,275	14,402	0	837	-
4	9,367	9,076	0	956	-
5	5,423	4,386	1,376	432	-
6	2,856	3,014	0	0	273

Of these initial grafts, the following table shows the breakdown of the quality of placement. This shows that the quality of placement was not too bad, with only 38.9 per cent placed into the correct household size using the best possible placement.

Search	Number of persons placed	Percentage of persons placed
1 (Best)	684	38.9%
2	128	7.3%
3	270	15.4%
4	670	38.2%
5 (Worst)	4	0.2%

Of the prunes, there were 1,768 people removed from Imputed households and 263 persons imputed in counted households were removed. The final pruning and grafting stage required 313 prunes and grafts to be able to achieve the correct age group by LAD targets. This is quite a large number of prunes. Some imputed households were replaced hence the difference in the dummy form numbers above.

Annex C: Explanation of terms used in imputation report

Imputation weights - The first stage of the imputation process imputes individuals in missed households onto the Census database. This was done by giving a weight to every household counted by the Census corresponding to its propensity to have been missed by the Census - these weights were derived from an analysis of missed households in the matched CCS and Census data. Households with high weights (and hence more likely to have been missed) were duplicated on the Census database using a systematic procedure which spread these duplications over areas where missed households were most likely. The second stage of the imputation process focussed on individuals that were missed in households actually counted by the Census. Weights were calculated for each individual in a similar way to that for households.

Each imputation report carries a commentary on the types of households and individuals most likely to be missed in the Census.

Donor types - In order to assign characteristics to imputed households, a donor imputation method was used. For each imputed household, a donor household with the same weight was selected from the counted households, and the characteristics of the household and its occupants were copied to the imputed household. The imputation system used a number of different levels of search (from 0 to 10) in order to find donor households. A high percentage of donors found using the better searches indicates a better quality of imputation.

Dummy forms - enumerators completed Dummy forms when they could not gain a response from a household. Certain types of dummy form were used to place imputed households (e.g those completed for non-response) where possible, otherwise they were assigned to random postcodes within the same ED as the donor household. The more dummy forms that were used in comparison to random postcodes, the better the quality of placement. The imputation system tried to imputed households into close matches using information on the dummy form (e.g. accommodation type) and information on this placement of dummy forms is also given in the imputation reports.

Where imputation of dummy forms was clustered into certain postcodes or EDs, this indicates that there was a high level of underenumeration in that area.

Person level imputation searches - Each imputed person must be placed into a suitable 'recipient' household. The criteria used by the imputation program to place individuals in households were (in order) the size and age/sex structure of the household, the tenure, HtC index and ethnicity of the household, and the distance of the recipient household from the donor household. The aim was to find a recipient household that, once the imputee had been added, looked as much like the donor household as possible. The higher the search number, the more of these constraints had to be relaxed, therefore placement of imputed people was better the more that are placed in the first searches.

Pruning and grafting - A crucial requirement of the imputation process was that the overall distribution of imputed households and people should be equal to the ONC estimates of the population. This calibration was accomplished by adjusting household and individual weights appropriately in the imputation process, and by a final stage in the process, which either removed excess imputed individuals and imputed households (pruning) or topped up the database (grafting) where necessary to ensure consistency with the ONC estimates. The Pruning and Grafting summary section of the imputation reports contain information on the numbers of prunes and grafts that the system needed to carry out. The quality of the placement of grafted individuals also varied, and this information was also included in each report.

Annex D: Example imputation Quality Assurance report

Report on imputation Quality Assurance for DG LP

Summary

The ONC Imputation process appears to have worked well for LP. Imputation has generally maintained the consistency of both LADs and postcodes containing imputed households and people.

Distribution of synthetic households

Imputation was fairly evenly spread throughout the two LADs: Wantown (00DD) has 7 per cent, and Elmbry (00DE) has 7 per cent.

At the postcode level, there were a number of postcodes that contained more imputed households than counted households, but only one had a significant number of people in it. Postcode AB1 2CD had ten counted people and twelve imputed people. This is investigated further below.

Uncontrolled characteristics of households at LAD level

Queries for accommodation type showed that the proportions of imputed detached and semi-detached households were lower than the proportions of counted detached and semi-detached households, with the proportions of imputed purpose built flats and converted or shared households higher than counted. This pattern was the same for both LADs. The data for Wantown (00DD) is below as an example:

00DD Wantown

Accommodation type	% Imputed households	% Counted households
Detached	7	14
Semi-detached	29	38
Terraced	21	19
Purpose built flat	30	19
Part of converted or shared house	7	5
Commercial building	6	4
Caravan, mobile or temporary	0	0

Queries for number of rooms showed that the proportion of imputed households with 3 rooms was generally higher than for counted data, and the proportion of imputed households with 6 rooms was generally lower than for counted data. This was particularly pronounced for Elmbry (00DE):

00DE Elmbry

Number of rooms	% Imputed households	% Counted households
01	2	2
02	5	4
03	21	11
04	23	19
05	23	22
06	16	25
07	5	10
08	2	2
09	1	2
10	1	0
11	0	0
12	0	0
13	0	0
14	0	0
15	0	0
16	0	0
17	0	0
18	0	0
19	0	0
20	0	0
21	0	0
25	0	0
27	0	0
41	0	0
53	0	0
70	0	0
86	0	0
87	0	0
XX	0	2

Queries for the number of cars per household showed that the proportion of imputed households with no car or one car was generally higher than counted, and the proportion of imputed households with two cars was lower than counted. This was most pronounced for Elmbry (00DE):

00DE Elmbry

Number of Cars	% Imputed households	% Counted households
No Cars	28	21
One Car	47	40
Two Cars	19	28
Three Cars	5	4
Four Cars	1	1
Five Cars	0	0
Six Cars	0	0
Seven Cars	0	0
Eight Cars	0	0
Nine Cars	0	0
Ten to Twenty Cars	0	0
No Code Required	0	3

The proportions for visitors in imputed households were similar to the counted data. The query for number of rooms by household size indicated that we had not caused obvious overcrowding by adding synthetic people into counted households, although we have imputed one household of size 5 into a one bedroom house in Wantown (00DD), and two households of size 5 and one of size 6 into one bedroom houses in Elmbry (00DE).

Uncontrolled characteristics of individuals at LAD level

Imputed individuals have similar characteristics to the counted population in terms of single year of age, long term illness, occupation and qualifications. In addition, imputation has not added in too many economically active students to the area.

Query output indicates that we have imputed less students than we have found in the counted data in both LADs, as we would expect.

Imputation has added in more people that have 'elsewhere' as their usual address one year ago than have been counted and less people with 'address on the front of the form', particularly for Wantown (00DD):

00DD Wantown

Address one year ago	% Imputed persons	% Counted persons
Address shown on front of form	71	85
No usual address one year ago	3	1
Same as person one	10	7
Elsewhere	16	6
No Code required	0	1

Imputation has generally added in more people with a religion of Hindu, Muslim or Sikh than were found in the counted data, and less Christians. This was particularly pronounced in Elmbry (00DE):

00DE Elmbry

Religion	% Imputed persons	% Counted persons
No religion	10	9
ES: Christian, NI: Roman Catholic	34	46
ES: Buddhist, NI: Protestant and Other Christian	1	3
ES: Hindu, NI: Other Religions and Philosophies	24	17
ES: Muslim, NI: Invalid	14	7
ES: Sikh, NI: Invalid	7	1
ES: Jewish, NI: Invalid	3	7
ES: Any Other Religion, NI: Invalid	1	3
No Code Required	0	2
Missing	7	7

Imputation has generally added in less White people than have been found in the counted data, and more Asian and Black people. This was especially noticeable in Wantown (00DD):

00DD Wantown

Ethnic group	% Imputed persons	% Counted persons
White	38	57
Mixed	4	3
Asian	36	28
Black	19	7
Chinese and other	3	3
No code required	0	2

Inconsistency of households at the postcode level

Imputation has maintained the characteristics of households for the postcode that has been investigated.

Inconsistency of Individuals at the postcode level

Imputation has maintained the characteristics of individuals for the postcode that has been investigated.

Annex E: The research team and steering group

Development of the ONC methodology was undertaken by a joint team of government and academic researchers, under the direction of **Professor Ian Diamond** and **Professor Ray Chambers** at the University of Southampton and including **Dr Ludi Simpson** of Bradford City Council. The Project Board, chaired by **Mr Tim Jones**, then Director of Methods and Quality Division at ONS, reported to a Steering Committee which oversaw the methodological development. The Steering Committee included representatives of the academic and local authority communities and a senior representative of the Australian Bureau of Statistics as well as officials from the UK Statistical Service. The members of the Steering Committee were:

Chair:

Dr John Fox (Chair), Group Director, Census, Population and Surveys, ONS (succeeded by John Pullinger from February 2000).

External members:

Dr Jim Cuthbert, Consultant, formerly Government Statistical Service
Professor Denise Lievesley, Director of the UNESCO Institute for Statistics
Professor Mike Murphy, London School of Economics
Mr Tim Skinner, Australian Bureau of Statistics
Professor Mike Titterington, Glasgow University
Mr Steve Turner, Tees Valley Joint Strategy Unit

Official members:

Dr Norman Caven, Registrar General for Northern Ireland
Mr James Meldrum, Registrar General for Scotland (succeeded by Mr John Randall from February 1999)
Mr Julian Calder, Group Director, Survey and Statistical Services
Mr Graham Jones, Director of Census
Mr Tim Jones, Director of Methods and Quality and Chair of the ONC Project Board
Ms Judith Walton, Director of Population and Demography
Dr Marie Cruddas (Secretary), Statistician, Census Division

Annex F: Comparators used in the ONC Quality Assurance process

1 Demographic estimates

Rolled-forward mid-year population estimates (MYEs) are produced every year by the PEU of P&D Division within ONS. The MYEs are produced by age and sex for each LAD in England and Wales, and these demographic estimates were used as one of the main comparators for all age groups in the ONC quality assurance procedures both nationally and sub-nationally. However, because the estimates were rolled forward from the 1981 Census (using births, deaths, estimates of migration, and estimated underenumeration for the 1991 Census), they were subject to a degree of error. This was further highlighted throughout the QA process.

The rolled-forward MYEs for 2001 were not available to feed into the sub-national ONC quality assurance process (due to many of the constituent components that make up the estimates not being available in time). Alternative demographic estimates were therefore used in ONC quality assurance procedures at the sub-national (DG, LAD and regional) level. ONC estimates for the first England and Wales DG were quality assured in November 2001. Published 2000 MYEs were used in this process. 2000 MYEs extrapolated forward to mid-2001 to allow for average annual population change between mid-1991 and mid-2000 were used as an additional comparator data set.

2 Birth registration data

The civil registration system records all new births in the population, and this was therefore a key source of data when quality assuring ONC estimates for children aged under one year. The numbers will differ where children have died, or moved in or out of the area, before their first birthday.

The Census figure for under 1s in England & Wales is 1.8 per cent lower than the birth registration data. By local authority, the difference as a percentage of the birth registration data shows the Census figure ranging from 20 per cent below the birth registration figure to 22 per cent above. Many of the areas where the Census was furthest below the registration data were areas with high concentrations of ethnic minorities or asylum seekers, suggesting that the 'missing' babies were no longer in the country.

3 DWP child benefit data

The Benefits Agency administers 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. For the purposes of comparator data, ONS liased with DWP and Oxford University to obtain 'clean' child benefit data relating to August 2000 (some postcodes were originally missing or needed correcting, and some new born children did not appear for up to three months after they were born). The Census counts those children normally resident in the UK, some of whom are not eligible for child benefit (for example children of FAF). Differences between the number of children claiming benefit and the Census figure could be explained by poor address data or this difference in definitional base. Also, there can often be a lag in updating records when a claimant moves, particularly now that payments are made directly into bank accounts.

For England and Wales as a whole, the Census figure is 0.9 per cent higher than the number of children claiming child benefit. However the differences as a percentage of the child benefit data vary widely by local authority, from 7 per cent lower to 51 per cent higher.

4 DWP retirement pensions data

The Benefits Agency administers the Retirement Pension Register that holds information on all persons claiming a state pension in the UK. Almost all persons aged 65 or over are entitled to claim some form of state retirement pension. For the purposes of comparator data, ONS liased with DWP and Oxford University to obtain 'clean' retirement pension data for May 2000 (some postcodes were originally missing or incorrect and some records were duplicated). Differences between those claiming pensions at a national level occur because not everyone is eligible. In addition, some expatriates resident overseas are eligible to claim a state pension. At a local level, differences could be explained

by poor address data. Also, there can often be a lag in updating records when a claimant moves, particularly now that payments are made directly into bank accounts.

The Census figure for pensioners in England and Wales was 1.3 per cent higher than the corresponding pensions data. The difference between Census figures and pension data as a percentage of the pensions data for local authorities ranged 3 per cent lower to 30 per cent higher.

5 HA 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 to be used alone for producing population estimates due to:

- coverage differences with the resident population, as certain groups such as armed forces personnel are excluded from patient registers; and
- errors caused by delays in people who have died or moved out of a HA being removed from the register, duplicate entries and people having more than one NHS number.

It is believed, however, that these data provide a useful additional source of population estimate as they are based on more recent data than the MYEs and do include a high proportion of the population resident in an area. An adjustment will be made to address the issues above (see stages 1 and 2 of **section 4.1.1**).

For England & Wales as a whole, total GP patient registrations in 2001 were 5 per cent higher than the 2001 Census figure for the total population. However the differences as a percentage of the patient registration data vary widely by local authority, from 23 per cent lower to 19 per cent higher.

6 School Census data

The School Census is an annual count of all children attending educational establishments, including schools which are privately funded. In England, information relating to January 1st is collected from local authorities by DfES each year, while in Wales, Scotland and Northern Ireland this role is carried out by the appropriate devolved government authority and has a different reference date. The main problem with this data is that it relates to place of study, which may be in a difference local authority to where the child lives. For this reason, less weight was given to these figures in the quality assurance procedure.

For England and Wales, the Census figure was 0.5 per cent lower than the Schools Census. Differences at local authority level as a percentage of the school census data ranged from the Census being 27 per cent lower to 43 per cent higher. In one exceptional case, the Census figure was 72 per cent below the Schools Census due to large numbers of pupils living outside the area and attending schools inside.

Further administrative data sources were researched for inclusion in the quality assurance process. These included the Council tax register and the electoral roll. However, after investigation into definitional differences it was agreed that these were not suitable comparator data sets for use in the quality assurance process. The definitional differences identified between the Census and the Council tax register and electoral roll are outlined below:

7 Council tax register

There are a number of reasons why a Census could identify a different number of household spaces from those identified on a different source collected for other purposes. For example:

Timeliness of the data: One measure of the quality of any data source is how up-to-date it is. In particular, how quickly are demolished / derelict properties removed from the Council Tax database and how quickly are new or converted properties added? Many data sources are prone to inflation because 'old' records are not removed.

Definitional differences: Any two data sources will almost always provide different numbers because of definitional differences. In this case the major sources of definitional difference are likely to be multiply-occupied properties and communal establishments. These could appear on one list as a number of single person households and on another as a single dwelling containing multiple people. This would lead to differences in the household count.

It should be noted that the Census only counts people at the address which they consider their usual residence. It is quite possible that some Census households were counted as vacant, even

though the household was paying Council Tax. In an area where Council Tax is low, and where it is feasible that a number of people have family homes elsewhere, it could be financially beneficial for some people to declare their [name of LAD] accommodation as their main home, and their family home elsewhere as a second home. However, for Census purposes at least some of these people would be recorded at their family home.

Duplication: Another common difficulty with data sources is duplication. Special procedures were used on the Census to search for and remove duplicate households. Council Tax records have been used by a number of areas within the ONS and, in general, problems have been found with double counting households. To examine this for a specific local authority, and quantify the impact, we would need to ask those area to do further work.

8 Electoral roll

The definitional differences between the electoral roll and the Census include the following:

Students may be registered at both their term-time and vacation addresses, although they may only vote at one of these in any election. The published Census figures only count students at their term-time address.

People with two homes may be registered at both addresses, although again they may only vote at one of them. The Census only counts people at the address which they consider their usual residence.

British citizens who have left the country within the last 15 years may still register to vote in the UK. As they are no longer resident in the UK, they will not be included in the Census figures. Members of the British armed forces and their families are registered to vote in the UK, even when stationed overseas.

Citizens of countries outside the EU and Commonwealth are not eligible to vote in the UK. These persons are included in the published Census figures if they are resident in the UK. A number of other groups are not eligible to vote, in particular most prisoners. Such groups are included in the Census figures.

In addition, residents who have died or moved out of the area may not immediately be removed from the register. How quickly this happens will depend on the electoral registration officer in

Annex G: Sources of qualitative information

1 Parameters from ONC estimation process

Data was collected throughout the ONC estimation process.

2 Census/CCS information sources

Information	Source
CENSUS	
Census Field Information, e.g. forms delivered, forms collected, refusals.	FMIS feeding into: dedicated field Lotus Notes database
Detailed Census field information	Enumerator Record Books (ERBs)
Checks done by Lockheed Martin **Checks done by ONS** **Information about problems encountered**	Data Quality Management Programme (DQMP)
Summary of results for each LAD fed into Lotus Notes Database	Data Quality Monitoring System (DQMS)
Issues that could lead to a request for change (RFC) or a datafile amendment (DFA) recorded on Lotus Notes Database	Data Quality Reviewing Procedure (DQRP)
Information on where enumerators have 'doubled up' on workloads	Payroll
Cumulative Postback Response rates by area	CenIntel
Information on number and type of calls to Census Helpline	CenIntel
Information on where Census forms were posted out (rather than delivered by enumerators): both **a) planned (e.g. foot & mouth areas)** **b) unplanned (e.g. calls to helpline)**	
Information on Direct Returns	Direct Returns database
Some enumerators staying on in field up until 28 May 2001	Data Collection Development
Creation of Emergency EDs where the workload for an enumerator was greater than originally planned	Geography Database
General geography problems such as out of date AP, results of ERBs or Maps, old map backgrounds etc.	ERBs, (Census Geography Notes database)
Census Query Resolution - give us an idea of whether an area was throwing up lots of problems	Database of queries
Enumerator questionnaires/debriefing?	
- CAMs debriefed only	Data Collection Development
Issues arising in processing (including TOAST)	Operational Processing on CENEXT1
CCS	
CCS Field Information - e.g. households identified interviews achieved, refusals, etc.	TRACS
Detailed CCS Field Information	Property Listing Sheets
Information on occurrence of foot & mouth restrictions in sample areas	Field checks
Problems identified early by TMs	Geography Field checks (on TRACS)
Information on where interviewers have "doubled up" on workloads & information on problems encountered in the field	Debriefing questionnaire
Problems with processing CCS data	Database maintained by ONC
Information on location of SSD interviewers	

3 Local authority feedback on past population estimates

P&D to produce profiles of each LAD. Profiles to include mid year population estimates from 1991 onwards; average annual change since mid 1991; natural change since mid 1991; change in migration since mid 1991; annual estimates of net migration; post 1991 estimated error of rolled forward population estimate; significant presence of difficult to estimate groups; boundary changes involving a net movement of at least 100 people; details of any communication between each LAD and P&D.

4 Information on the 1991 Census

Information gathered on the adjustments made in the 1991 Census and the rebasing of population estimates.

Annex H: Example Qualitative Quality Assurance report

00DD Wantown - Qualitative Report

Key Points to Note

- Census Coverage Figure = 87.6 per cent
- Population Profile states that this is an area of average change.
- The percentage of Direct Returns & helpline calls are outside of the acceptable range. The percentage of non-returns is fairly high at 8.87 per cent.
- Operational issues highlights a problem with one postcode with sparse CCS data. This postcode was subsequently removed from the sample.

Census

Census field information:

% Forms delivered:	97.0%	(Acceptable range 96% - 106%)
% Forms received (Royal Mail weights):	78.9%	(Acceptable range 70% - 104%)
% Refusals:	0.07%	(Acceptable range 0% - 0.12%)

Information on Direct and Late returns:

Percentage Direct Returns:	3.06%	(Acceptable range 0% - 2.5%)
Percentage Late Returns:	0.24%	(Acceptable range 0% - 0.3%)

Detailed Census Field Information (ERBs) and Field Incidents:

ERB's not examined.
No reported field incidents.

Data Quality Management Programme:

DCR report for date of birth shows that all CD's met service level accuracy of > 99.5%.

Data Quality Monitoring System:

Data Quality Executive Summary states there are no obvious problems with the data.

Enumerators doubling up on workloads:

Yes Number: 7 (out of approx 266 Enumerator jobs - 2.63%)

Information on number of calls to Census helpline:

% of calls made to helpline:2.51% (Acceptable range 0% - 2%)

Information on where Census forms were posted out:

a) planned (ERBs)
% forms sent out: refer to ERBs if considered necessary

b) unplanned (helpline)
% forms sent out: 0.34% (Acceptable range 0% - 4%)

Foot and Mouth areas:

Number of CDs affected: 0
Total number of CDs in LAD: 7

Dummy Form Information:

% of absent households:	0.52%
% of Refusals:	.27%
% of non-returns:	8.87%
% of second residences:	0.08%
% of vacant household spaces:	1.73%

Some enumerators staying on in field until 28th May:

No

Creation of emergency EDs/SEDs:

None.

Authority Query Resolution:

% of estimated addresses needing an AQR: 1.52%

Operational processing issues:

Estimation Summary Report mentions that one postcode had very sparse CCS data and was therefore excluded from the sample.

CAM Debriefing:

Region 3

Address lists and Maps

- Address lists could have been more accurate and maps more up to date.

Enumeration

- Using Local Authority gypsy liaison officers as special enumerators was very successful.

- There were a number of households who reported that they had not received a census form.
- Despite a disparate number of languages to contend with, the enumeration of asylum seekers went well.
- Wantown Council initiated a very effective publicity campaign telling people about the local benefits of completing their census form. Consequently enumerators in that borough reported a very positive response on the doorstep.
- Some enumerators had high or double workloads principally as a result of recruitment problems.

Census Coverage Survey

CCS Field Information (households identified, interviews achieved, refusals):

Workloads covering this LAD: NC04, NC05, ND02, ND03, ND04, ND05, NE01, NE02, NE03, NE04, NE05.

No households listed :	3,492
% interviews done :	78.24%
% refusals :	8.21%

Detailed CCS Field Information and Field Incidents: Property Listing Analysis

Team Manager NC

The housing type is mainly self contained houses. However, about 20 per cent of the postcodes include purpose built flats, houses converted into flats/bed-sits or flats above shops. Apart from comments about language problems amongst asylum seekers, few difficulties appear to have been encountered and the Property Listing Sheets have been completed to a good standard of neatness, quality and completeness. There are also plenty of comments about checks to ensure correct assignment of addresses to postcodes.

Team Manager ND

The housing type is mainly self contained houses. There are only four postcodes showing houses converted into flats and purpose built flats, but there are the usual difficulties with access and getting information about occupancy reported. Otherwise few problems seem to have been encountered and overall the Property Listing Sheets have been completed to a good standard of neatness, quality and completeness.
One postcode is dedicated to 23 warden controlled flats housing elderly people.

Team Manager NE

The housing type is mainly self contained houses with only a small number of postcodes including houses converted into flats. Few problems seem to have been encountered and overall the Property Listing Sheets have been completed to a good standard of neatness, quality and completeness.

CCS Field Incidents

none recorded.

Occurrence of Foot and Mouth:

No

Problems identified early by Team Managers during Geography check:

NC,ND and NE - generally all as expected. Some new construction in NC04 and some difficulty in locating postcodes in NE but starting addresses given.

Problems with processing CCS data (DDS AQR's):

% of listed households in NC needing an AQR: 7.14%
% of listed households in ND needing an AQR: 2.29%
% of listed households in NE needing an AQR: 5.97%

Information of location of SSD interviewers:

Number of SSD Interviewers in TM Area: 1 in NE

Double workloads (from debriefing questionnaires):

NC: Number of double workloads: 1 Number of single workloads: 3
ND: Number of double workloads: 1 Number of single workloads: 4
NE: Number of double workloads: 1 Number of single workloads: 3

Debriefing questionnaire comments:

Field Manager - FM35

NC - there were vacant properties in one workload
ND - no relevant comments
NE - there were difficulties with recruitment within this workload and postcodes were spread out across a large geographical area.

Team Manager

NC04 - there were a few vacant properties as well as new constructions within this workload.
NC05 - no relevant comments
NE01 - 03 - contacting people and finding addresses was found to be fairly difficult due to the large geographical area.
NE04 -05 - no relevant comments.
ND - no relevant comments.

Interviewer

NC04 - vacancies in block of flats.
NC05 - no relevant comments
ND02-05 - no relevant comments
NE01-03 - people were fairly difficult to contact and finding addresses was also found to be fairly difficult due to a large geographical area.
NE04 -05 - no relevant comments

Annex I: Example population profile

Population Profile:

Wantown **00DDQA9/F85**

Mid Year	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Population Estimates	269.5	268.9	268.1	266.1	267.6	268.9	271.4	271.2	274.1	275.8

1 Is it an area of large change since 1991? How reliable is our estimate of change?

Thousands

Average annual change since mid-91 (thousands)	E & W	Wantown
Average annual change	0.5	0.7
Average annual natural change	0.3	1.4
Average annual change in migration	0.1	-2.1

Thousands

Migration	Mid-92	Mid-93	Mid-94	Mid-95	Mid-96	Mid-97	Mid-98	Mid-99	Mid-00
Net Internal							-2.8	-1.4	-0.7
Net International							-0.2	-0.3	-1.1
Total Net	-3.4	-3.4	-3.7	-0.5	-1.2	0.2	-3.0	-1.7	-1.8

2 Has there been any communication with the Local Authority District (LAD)?

No

3 Other information

Post 1991 Census estimated error of rolled-forward population estimate (Average absolute error at district level for the whole of England and Wales was 2.5%)	Overestimate of 0.9%
Significant presence of difficult to estimate groups.	No
Boundary Changes since 1991 involving a net movement of at least 100 people.	None

Annex J: Glossary of acronyms

CCS
Census Coverage Survey

DASA
Defence Analytical Services Agency

DfES
Department for Education and Skills

DG
Design Group

DoH
Department of Health

DQMP
Data Quality Management Programme

DQMS
Data Quality Monitoring System

DQRP
Data Quality Reviewing Procedure

DSE
Dual System Estimation

DWP
Department for Work and Pensions

ED
Enumeration District

ERB
Enumerator Record Book

FAF
Foreign Armed Forces

FMIS
Field Management Information System

GROS
General Register Office for Scotland

HA
Health Authority

HESA
Higher Education Statistics Agency

HO
Home Office

HtC
Hard to Count

LAD
Local Authority District

LSC
Learning and Skills Council

MYEs
Mid-year estimates

NISRA
Northern Ireland Statistical and Research Agency

ONC
One Number Census

P&D
Population and Demography Division

PEU
Population Estimates Unit

USAF
United States Air Force

TM
Team Manager

TRACS
Team Reporting and Communication System.

WA
Welsh Assembly

WFC
Welsh Funding Council