



ONS(ONC(SC))98/17

ONE NUMBER CENSUS STEERING COMMITTEE

Dress Rehearsal of One Number Census Methodology

1. This report describes the 'dress rehearsal' of the One Number Census (ONC) methodology as part of the Census Dress Rehearsal (CDR) in 1999. A diagram illustrating the ONC process in the CDR is also presented.
2. **The Steering Committee are asked to:**
 - a) **note the paper;**
 - b) **endorse the approach to implementation and evaluation of the ONC in the CDR;**
 - c) **provide any comments at the meeting on the 13 November 1998, or in writing by 27 November 1998.**

**Lisa Buckner
Census Division
Office for National Statistics**

**Room 4200W
Segensworth Road
Titchfield
Fareham
HANTS
PO15 5RR**

October 1998

Dress Rehearsal of One Number Census Methodology in England and Wales

1. Introduction

1.1 The Census Dress Rehearsal (CDR) in April 1999 offers the opportunity to carry out the entire One Number Census (ONC) process. This will allow the operational aspects of the ONC methodology to be tested for the individual processes within the ONC, as well as the flow of data from one process to the next. The integration of the ONC with the census process will be explored and the interfaces tested.

1.2 In order to get maximum benefit from this exercise it is essential that the ONC methodology applied in the CDR is as near as possible to that planned for 2001 (as described in ONS(ONC(SC))98/16). As far as possible it is planned to run the processes in real time which will give a indication of how long each process will take and the amount of resources required for each part of the process. It will also be the only opportunity to test the interfaces in a 'live' situation. Evaluation of this ONC dress rehearsal will feed into the decisions on the methodology for the 2001 Census.

1.3 Although the operational aspects of estimating population totals will be tested it is obviously very unlikely that these population estimates will be of the correct magnitude due to:

- the voluntary nature of the CDR; and
- the fact that the CCS is designed to produce estimates of underenumeration of pseudo counties of a much larger size than the CDR.

The method for comparing these with demographic estimates and aggregate level administrative sources will be tested to ensure that the data are available when required and are in the correct form.

1.4 This paper describes the application of the ONC methodology to the CDR in 1999 in which different aspects of the methodology will be evaluated, including:

- CCS design and sample selection
- CCS practicalities
- Matching census and CCS data
- Estimation of pseudo county and Local Authority District (LAD) population totals
- Coverage modelling and adjustment of the database.

1.5 Section 2 and 3 of this paper give a brief overview of the CDR and ONC process in 1999 respectively. Evaluation of the ONC process as part of the CDR is addressed in Section 4 although the criteria for evaluation are still under discussion. Section 5 highlights some issues regarding the dress rehearsal of the ONC.

2. Census Dress Rehearsal

2.1 The CDR is a large-scale test of the 2001 Census methodology. It will take place on 25 April 1999 in Leeds, Lincoln, Bournemouth, Ceredigion and Gwynedd, covering approximately 102,900 households in England and Wales (22,700 in Scotland and 8,000 in Northern Ireland). These areas have been chosen to include areas of:

- students and student halls of residence

- high multi-occupancy
- high ethnicity
- welsh language speakers in all age groups
- small communal establishments
- a wide range of occupations.

2.2 Enumerators will deliver CDR forms together with a freepost envelope to allow the public to post their CDR forms back. After a short length of time to allow forms to be posted back, enumerators will call back to collect any forms which have not been returned by post. Completion of Census forms by householders in the CDR will be voluntary

2.3 The CDR will involve a test of the fieldwork and also the processing methodology for 2001. Data will be captured, coded, edited and missing data imputed.

3. ONC Dress Rehearsal

3.1 Dress rehearsal of the ONC process will take place as part of the CDR. Following the CDR, a CCS will be undertaken in all the areas of the CDR. CDR and CCS data will be captured, coded, edited and missing data imputed. These data will be matched to identify those extra people found by the CCS. Population estimates will be made, coverage weights calculated and the database adjusted.

3.2 Figure 1 is a flow diagram for the ONC process in the Census Dress Rehearsal in 1999. For each process, the diagram illustrates the data input required and data produced. Individual aspects of the methodology are discussed below.

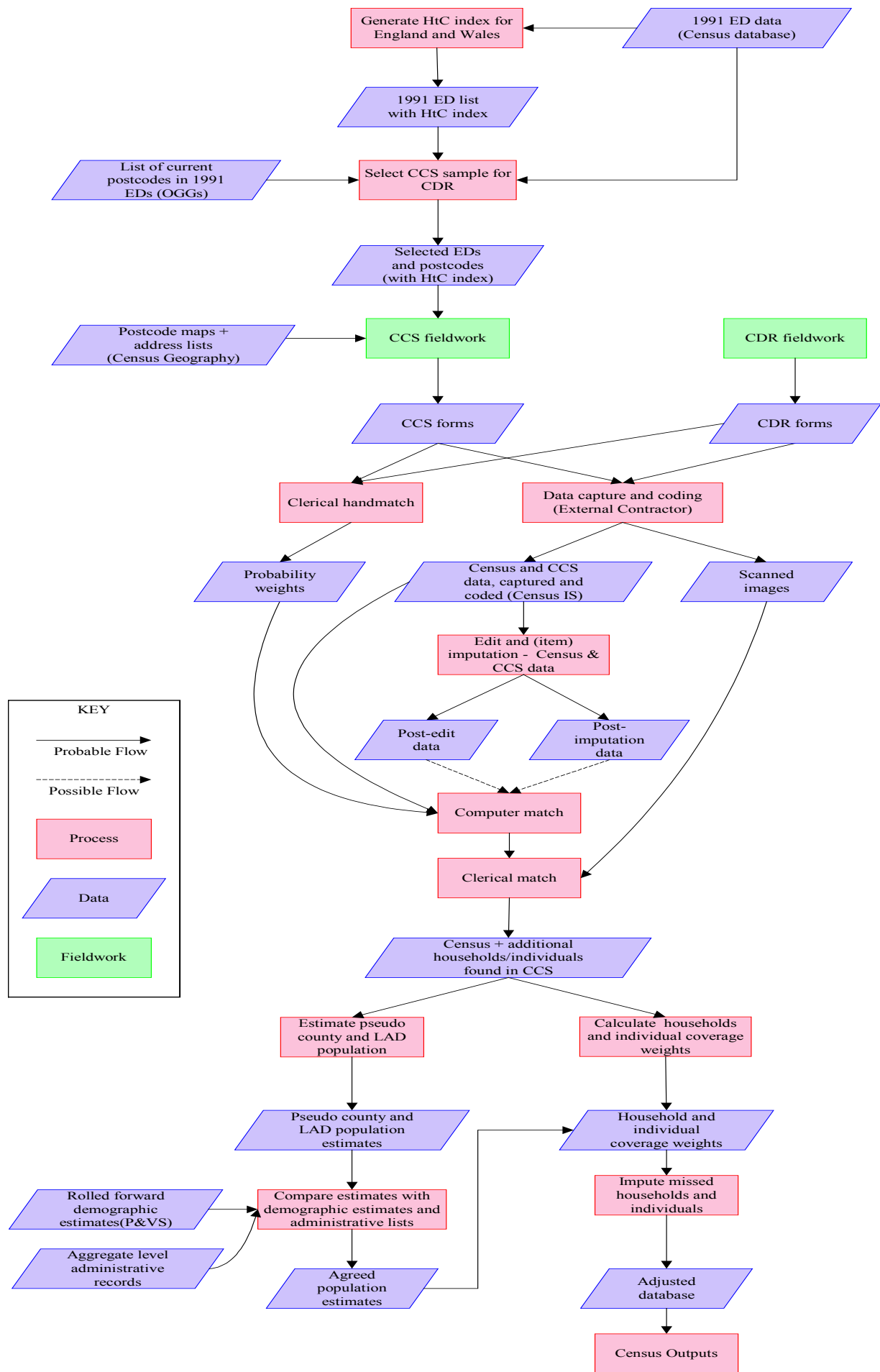
Selection of CCS sample

3.3 1991 Enumeration districts (EDs) will be stratified into groups by a 'Hard to Count' (HtC) index since it is expected that underenumeration will, at a local level, be higher in certain areas characterised by particular social, economic and demographic characteristics. This will be a national index the components of which will be chosen to represent characteristics found to be important after the 1991 Census by the Office for National Statistics (ONS) and the Estimating With Confidence Project (Simpson *et al.*, 1997).

3.4 After the composition of the Hard to Count index is finalised (late November 1998), 1991 Census data for EDs in England and Wales will be used to derive the HtC index. All 1991 EDs in the CDR will then be assigned their HtC index. A list of current postcodes for these EDs will be obtained. Postcodes which cross ED boundaries will be assigned to the ED which contains the majority of the households of the postcode so all postcodes will be assigned to one ED only. EDs and postcodes for the CCS Dress Rehearsal will then be selected.

3.5 This will enable the interfaces with Census Geography and Customer Services to be tested to ensure that the most recent postcodes are included in the population from which the sample will be drawn. It will also mean that the procedure for excluding certain postcodes (e.g. business postcodes) can be tested. For the CDR it will be important that no postcodes are selected that cross CDR boundaries as this would result in part of the postcode being outwith the CDR area.

Figure 1. Flow diagram for ONC process in the Census Dress Rehearsal in 1999



3.6 The number of households in the whole CDR area is much smaller than for any of the proposed pseudo counties for 2001 (ONC(ONS(SC))98/12). However, in order to test the routines for making population estimates for pseudo counties and their constituent LADs it is proposed that in the CDR the five areas (Leeds, Lincoln, Bournemouth, Ceredigion and Gwynedd) are taken together to represent a pseudo county of approximately 100,000 households. The individual areas can be thought of as LADs of different sizes, ranging from approximately 6,900 – 34,200 households, within this pseudo county.

3.7 The problems surrounding the size of the CDR in comparison to the pseudo county is more acute for Scotland and Northern Ireland. It is also not possible to include these areas in the CDR pseudo county due to the use of different Hard to Count indices by Scotland and Northern Ireland.

CCS Practicalities and Fieldwork

3.8 It is essential that the CCS methodology is dress-rehearsed as part of the CDR in 1999. Approximately 3-4 weeks after Census day in 2001 the CCS will re-enumerate between 1- 2.4 per cent of the households in England and Wales. Applying this proportion to the CDR would mean that only 1,000 – 2,500 households would be included in the survey. It is considered that this would be insufficient to test properly the practicalities of undertaking a large CCS.

3.9 For the CDR it is therefore proposed that the CCS re-enumerate approximately 18,000 households which comprises approximately 17 per cent of the households in the CDR areas. The reasons for taking a proportionally larger CCS are:

- a) because the CDR is not compulsory a smaller proportion of CDR forms will be returned than in the census;
- b) to dress rehearse not only the CCS but also the ONC process; and
- c) to test the CCS methodology on a large scale¹ before 2001, including:
 - selection of enumeration districts and postcodes;
 - production of a large quantity of postcode maps;
 - retaining a significant proportion of enumerators;
 - re-training enumerators to undertake the CCS interviewing;
 - field management of a large survey;

Interfaces include those with the external contractor, Census Field, Logistics, Geography.

Matching

3.10 Following the CDR and CCS in April 1999 the CCS and CDR data will be captured and coded, edited and missing data imputed. Extracts of data will be available pre- and post edit and post imputation. Data from the CDR and CCS will be matched in order to identify those people who were missed in the CDR. Different matching strategies will be employed in

¹ It should be noted that the development of the CCS process has, to date, been based on two relatively small pilots (20 postcodes in Brent in 1997 and 100 postcodes in Southampton in 1998).

order to access the improvement obtained by including name and/or address and to identify the optimum matching strategy.

3.11 The matching strategy will follow that developed from ONS(ONC(SC))98/14. It will first require a certain amount of clerical matching will be required in order to calculate the probability weights prior to undertaking the automatic computer match. This clerical matching will be undertaken at the processing centre and will involve using census and CCS forms to identify a number of 'true matches'. These will then be used to calculate the probability weights. It may be possible in 2001 to use this information from the CDR and the Estimation and Maximisation (EM) algorithm to train the probability weights instead of undertaking another clerical match prior to computer matching. This method will be investigated as part of the CDR.

3.12 A further clerical match will be required following the computer match in order to resolve outstanding unmatched CCS records. This will involve using scanned images of forms to ensure that unmatched CCS records are 'true' non-matches and will take place at ONS Titchfield.

3.13 Important interfaces for this part of the process are those with the contractor who is capturing the data and with Census IS to ensure that the data required by the ONC process can be obtained when they are needed and in the form required.

Production of pseudo county and LAD estimates

3.14 These data will then feed into the estimation procedure to produce pseudo county and LAD estimates as described in ONS(ONC(SC))98/16. This will be largely automatic as both research and development of this process will have taken place although the practical aspects will need to be investigated.

Comparison of estimate with demographic estimates and administrative records

3.15 The pseudo county estimate for the CDR area will be compared with the demographic estimates and administrative records. However, demographic estimates do not necessarily exist for these areas since pseudo counties will be composed of groups of contiguous whole LADs in 2001 whereas for the CDR the only part LADs are included. Therefore, it may be necessary to synthetically aggregate available data so it will not be possible to undertake a true comparison. However, it is important to test the procedures for comparing these estimates to ensure that is undertaken systematically and efficiently after the census in 2001.

Coverage modelling and adjustment of database

3.16 Coverage weights for households and individuals will be calculated and will be used to impute missed households and individuals (ONS(ONC(SC))98/15). This process will be mainly automatic but will require some intervention to ensure that the statistical models used are appropriate.

3.17 After the census database has been adjusted it will then be passed to Census Output. Interfaces for this process include Census IS and Output.

4. Evaluation of the ONC process as part of the CDR

4.1 The CDR will provide the opportunity to run the ONC process in real time. This will allow the operationalisation of the methodology to be tested and also provide estimates of the time taken for each of the processes and the resources required.

4.2 As well as testing the interfaces with other Census processes such as the external contractor, Census Field, Logistics, Geography, IS and Outputs, specific areas tested in the CDR of the ONC include:

CCS design and sample selection

- Hard to Count index
- Postcode selection
- Questionnaire design

CCS practicalities

- Fieldstaff pay and recruitment
- Timing of recruitment exercise
- Training methods and materials
- Timing of training with respect to CDR and CCS fieldwork
- Use of the Hard to Count index and information from Geography database to plan Interviewer workloads
- Most cost-effective type of postcode map
- Interviewer workloads and length of fieldwork versus coverage
- Field management including: ratio of team supervisors to interviewers; staff resignations; illness; etc.; management information required
- Number and nature of public queries
- Requirement for second level of field manager
- Best strategies for persuading wavering respondents
- The best design for the data collection materials (property listing, questionnaire, showcards, etc.)
- Public understanding of purpose of CCS, definitions, questions
- The information required from the debriefing
- The time required and best format for the debriefing

Matching, modelling and adjusting the database

- Size and feasibility of initial handmatch to calculate probability weights
- Length of time and number of staff required to undertake this handmatch
- Possible use of EM algorithm to train weights
- Improvement achieved by utilising names and/or addresses in matching procedure
- Time and resources required for computer matching process
- Time and resources required for final clerical matching stage
- Ease of use of scanned images for final clerical matching
- Time and resources required for estimation of pseudo county and LADs
- Amount of intervention required in estimation procedure
- Time and resources required for coverage modelling and imputation
- Amount of intervention required for coverage modelling and imputation

5. Issues

Edit and Imputation

5.1 It is currently planned that Edit and Imputation of CDR (and CCS) data will not take place until some time in 2000. This may lead to the decision on the capture of names and/or addresses being too late to feed back to the external contractor in time for the 2001 Census. It may therefore be necessary to use pre-edit data to assess the matching strategy in the CDR regarding the use of names and/or addresses

5.2 The dress rehearsal of ONC process after the matching procedure will require edited and imputed CDR and CCS data. The evaluation of post matching processes is therefore heavily dependent on the current timetable for the edit and imputation process.

Key Decision Dates

5.3 A number of key dates will influence the timing of the evaluation of the ONC dress rehearsal. In particular:

- Timetable for feeding decisions for 2001 back to the external contractor
- Agreement of the Census White Paper which will determine the topic content of the census
- Timetable for printing of Census and CCS forms

6. Conclusions

6.1 The CDR will provide an invaluable opportunity to test the operational aspects of the ONC methodology. Although individual aspects of the process will be tested and simulated data will be used to test the whole process from start to finish prior to the CDR, this will be the only opportunity to test the ONC in a 'live' situation.

6.2 The evaluation of the dress rehearsal of the ONC process will feed into the final decisions on the methodology for the 2001 Census.