



ONS(ONC(SC))00/06

ONE NUMBER CENSUS STEERING COMMITTEE

1999 Census Rehearsal: Evaluation Plan

1. This paper outlines the proposed One Number Census evaluation of the 1999 Census Rehearsal.
2. **Members of the Steering Committee are asked to endorse the proposed evaluation.**

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

January 2000

1999 CENSUS REHEARSAL: ONC EVALUATION PLAN

TABLE OF CONTENTS

1. INTRODUCTION.....	4
2. CCS SAMPLE SELECTION.....	5
2.1 BACKGROUND	5
2.2 CHANGES IN METHODOLOGY	5
2.3 DIFFERENCES IN METHODOLOGY SPECIFIC TO THE REHEARSAL.....	5
2.4 EVALUATION TOPICS.....	5
2.4.1 <i>Sampling Frame</i>	5
2.4.2 <i>Hard to Count Index</i>	6
2.4.3 <i>Size Stratification</i>	6
2.4.4 <i>Prototype Sampling System</i>	7
2.4.5 <i>Timing</i>	7
3. MATCHING THE CCS AND CENSUS REHEARSAL DATA.....	8
3.1 BACKGROUND	8
3.2 CHANGES IN METHODOLOGY	8
3.3 DIFFERENCES IN METHODOLOGY SPECIFIC TO THE REHEARSAL.....	8
3.4 EVALUATION TOPICS.....	8
3.4.1 <i>Matching Feasibility</i>	9
3.4.2 <i>Derived Variables</i>	9
3.4.3 <i>Matching Variables</i>	9
3.4.4 <i>Input Data</i>	10
3.4.5 <i>Matching Techniques</i>	10
3.4.6 <i>Matching Software</i>	10
3.4.7 <i>Clerical Matching Training</i>	11
3.4.8 <i>Calculating / Updating Matching Weights</i>	11
3.4.9 <i>Timing</i>	11
4. ESTIMATION	12
4.1 BACKGROUND	12
4.2 CHANGES IN METHODOLOGY	12
4.3 DIFFERENCES IN METHODOLOGY SPECIFIC TO THE REHEARSAL.....	12
4.4 EVALUATION TOPICS.....	12
4.4.1 <i>Design Group Estimation Software</i>	12
4.4.2 <i>Timing of Design Group Estimation</i>	13
4.4.3 <i>LAD Estimation Software</i>	13
4.4.4 <i>Timing of LAD Estimation</i>	13
5. DEMOGRAPHIC ESTIMATES AND ADMINISTRATIVE RECORDS	14
5.1 BACKGROUND	14
5.2 REHEARSAL METHODOLOGY	14
5.3 EVALUATION TOPICS.....	14
5.3.1 <i>Design Group Estimates</i>	14
5.3.2 <i>Quality of Data from Administrative Records</i>	14
6. IMPUTATION.....	15
6.1 BACKGROUND	15
6.2 CHANGES IN METHODOLOGY	15
6.3 DIFFERENCES IN METHODOLOGY SPECIFIC TO THE REHEARSAL.....	15
6.4 EVALUATION TOPICS.....	15
6.4.1 <i>Household and Individual Coverage Weights</i>	15
6.4.2 <i>Selection of Donor Households</i>	16
6.4.3 <i>Imputation of Households</i>	16
6.4.4 <i>Selection of Donor Individuals</i>	16

6.4.5 Imputation of Individuals.....	16
6.4.6 Constrain Marginal Totals.....	17
6.4.7 Imputation Software.....	17
6.4.8 Timing.....	17
7. ONC OVERVIEW.....	18
7.1 BACKGROUND.....	18
7.2 EVALUATION TOPICS.....	18
7.2.1 Measurement of Change of Data through ONC Processes.....	18
7.2.2 Communal Establishments.....	18
7.2.3 Project Interfaces.....	19
7.2.4 Data Quality.....	19
7.2.5 Timetable and Staffing.....	19
APPENDIX A: TIMETABLE.....	20
APPENDIX B: DATA SOURCES.....	22
APPENDIX C: EVALUATION AND QUALITY ASSURANCE TEAMS.....	23
BIBLIOGRAPHY.....	24

1. INTRODUCTION

The recent Census Rehearsal and subsequent Census Coverage Survey (CCS) Rehearsal allow the evaluation of many aspects of the 2001 Census that has not been possible in the past.

This document outlines the planned evaluation of the One Number Census (ONC) processes. The ONC project aims to adjust the Census database for estimated undercount so that all statistics add to 'One Number' – the national estimate of the population. Much of the methodology required to achieve this aim has been developed specifically for this purpose. Therefore, careful and extensive evaluation is required. The ONC processes are outlined in the consultation document 'A Guide to the One Number Census' (ONS(1999)).

The purpose of the evaluation is to assess and develop the ONC methodology and test its implementation in order to ensure that the requirements of Census Users are met.

The key stages of the ONC evaluation are as follows:

- CCS sample selection;
- Matching the Rehearsal CCS and Census data;
- Estimation;
- Demographic estimates and administrative records;
- Imputation;
- ONC Overview.

For each area of evaluation, the following information is supplied:

- Evaluation objective;
- Areas of evaluation;
- Possible areas of further evaluation.

A timetable for the evaluation, also showing milestones and key dates, is given in Appendix A. This timetable is dependent on the timely delivery of accurate rehearsal data. Appendix B shows the data sources required for the ONC evaluations. Appendix C gives the members of the evaluation and QA teams for each of the above evaluation topics.

The evaluation of the practicalities of conducting the CCS is covered in the CCS Project Board paper CCS(PB)99/13.

2. CCS SAMPLE SELECTION

2.1 Background

The CCS involves an intensive enumeration of a large, nationally representative sample of postcodes. The sample of postcodes will be drawn from all Local Authority Districts (LADs) to enable population estimates to be made for all districts. LADs will be grouped together to form 'design groups' which have a population of approximately half a million people. The CCS sample design is applied to each design group independently. The strategy for the selection of postcodes within each design group is given in papers ONS(ONC(SC))00/01 and ONS(ONC(SC))98/12. The design groups for 2001 are presented in ONS(ONC(SC))00/10.

2.2 Changes in Methodology

For the 1999 Rehearsal the Hard to Count (HtC) index was calculated using 1991 Census data for multi-occupancy, private rented accommodation and young migrants. These variables were chosen as they are regarded as being strongly related to Census undercount and they form the minimum set of variables required for the purposes of the HtC index.

2.3 Differences in Methodology Specific to the Rehearsal

A small number of differences in the methodology outlined in the above papers were necessary for the successful implementation of the 1999 Rehearsal sample selection. These differences were as follows:

- To allow the CCS Team to evaluate the practicalities of co-ordinating a large-scale survey, the CCS had to involve a large sample of postcodes. Therefore, a significantly higher sampling fraction than proposed for 2001 was used. The rehearsal CCS was budgeted for approximately 18,000 households.
- The 1999 Rehearsal areas aggregate to form a Design Group of approximately 200,000 persons. Each area is treated as an LAD. This difference is not critical to the sample selection strategy due to the large sampling fraction.
- A constraint to sample at least one Enumeration District from each Hard to Count level within each Rehearsal area was imposed. The 1999 Rehearsal areas are heavily weighted toward the hardest to count category of the index. The constraint ensures that some data are available for all HtC categories within all Rehearsal areas.

2.4 Evaluation Topics

The evaluation of the sample selection will focus on five areas:

- The sampling frame;
- The Hard to Count index;
- Size stratification;
- The prototype sampling system;
- Timing.

2.4.1 Sampling Frame

Objective: To evaluate the quality of the CCS sampling frame and determine an appropriate sampling frame for 2001.

Areas of Evaluation:

- Were the data sources used in the construction of the sampling frame sufficiently up to date and of a suitable quality?
- Did the methodology for constructing the sample frame work in practice?
- What issues have been identified due to the methods used?
- How can the quality of the sampling frame be maximised for 2001?

Possible Areas of Further Evaluation:

- The frequency and type of changes to the CCS postcodes from the time they were sampled to the commencement of CCS fieldwork.

2.4.2 Hard to Count Index

Objective: *To evaluate the Hard to Count index and determine whether the variables used are appropriate for inclusion in the 2001 index.*

Areas of Evaluation:

- Are the data used for the proposed variables contributing to the index sufficiently up to date at the 1991 ED level?
- Did the index categories reflect the levels of difficulty the Census Rehearsal enumerator or CCS interviewers perceived at both the ED and postcode levels?
- How correlated was the index with the estimated levels of underenumeration?

Possible Areas of Further Evaluation:

- How valid is the assumption that postcodes have the same hard to count category as the 1991 ED into which the majority of the postcodes addresses fall?
- Are the data for other variables proposed for inclusion in the index (unemployment, single elderly households and language difficulty) sufficiently up to date for the 1991 Census EDs?
- The correlation of any other index with an estimate of the underenumeration levels.

2.4.3 Size Stratification

Objective: *To evaluate the data used to implement the CCS design size stratification that occurs within each Hard to Count category.*

Areas of Evaluation:

- Are the data used for the proposed variables used in the size stratification (which are 1991 Census counts of males aged 0-4, females aged 0-4, males aged 20-24, males aged 25-29, males ages 30-34 and females aged 85 or more) sufficiently up to date at the 1991 ED level.

2.4.4 Prototype Sampling System

Objective: *To evaluate the prototype sampling system and determine the most appropriate language for implementing the 2001 system.*

Areas of Evaluation:

- Ease of use of the system.
- Adaptability of the system for 2001.
- Whether SAS is appropriate for the development of the 2001 System.

2.4.5 Timing

Objective: *To determine the time required to perform the sample selection for 2001.*

Areas of Evaluation:

- Speed of the system.
- Time required to check the sampling.
- Time estimates of any additional system functionality identified for 2001.

3. MATCHING THE CCS AND CENSUS REHEARSAL DATA

3.1 Background

The 2001 matching exercise will involve a combination of automated and clerical matching. There are four key stages:

- Automatically link households and individuals where key details match exactly.
- Automatically link households and individuals where key details are very similar. Similarity is determined by probability weights. The higher the probability weight, the closer the agreement between two records.
- Automatically select similar pairs of records for a clerical decision on their matching status.
- Clerically search for Census records corresponding to any unmatched CCS records.

The matching process is described in detail in the paper ONS(ONC(SC))98/14.

3.2 Changes in Methodology

A small number of changes to the methodology outlined in the above paper have resulted from the ongoing research. These changes are as follows:

- The matching weights derived from the 1999 Census Rehearsal will be used as starting weights for the probability matching in 2001. These weights will be tuned to the data being matched.
- The algorithm for deriving the household structure has now been developed and will be tested as part of the 1999 Rehearsal Evaluation.
- Tenure will not be used as a matching variable since it is a key household analysis variable. Type of accommodation may be used instead. This evaluation will determine the precise matching variables to be used.

3.3 Differences in Methodology Specific to the Rehearsal

Prior to the analysis described here, it will be necessary to match the rehearsal data clerically. The weights calculated from this clerically matched data will be used as starting weights for the 2001 matching process. The rehearsal weights can be iteratively tuned to the data being matched. This pre-automated matching clerical matching process will not be repeated in 2001 and therefore is not included in this evaluation. For the purposes of the following evaluation, the clerically matched records are deemed to be 'true' matches.

3.4 Evaluation Topics

The evaluation of the matching process will focus on nine areas:

- Matching feasibility;
- Derived variables;
- Matching variables;
- Input data;
- Matching techniques;
- Matching software;
- Clerical matching training;
- Calculation / updating of matching weights;
- Timing.

3.4.1 Matching Feasibility

Objective: To determine whether automated matching can be performed to requirements in 2001.

Areas of Evaluation:

- Can the necessary accuracy be achieved?
- Can the matching be performed to an acceptable timetable?

Possible Areas of Further Evaluation:

- Develop a methodology for estimating the accuracy of the matching in 2001.

3.4.2 Derived Variables

Objective: To evaluate the derived variables used in the Rehearsal matching and define the most appropriate derived variables for 2001 matching purposes.

Areas of Evaluation:

- Did the household representative algorithm consistently identify the same member of the household in the Census and CCS?
- How can the household representative algorithm be improved?
- Did the household type algorithm successfully partition the data as required?
- How can the household type algorithm be improved?

3.4.3 Matching Variables

Objective: To select appropriate matching variables for the 2001 automated matching process.

Areas of Evaluation:

- Which variables provide the most matching information?
- Can a set of mutually independent matching variables be selected?
- Should some variables be combined?
- Which variables should be used for blocking and which for matching?
- Which variables should be used for exact matching?
- What impact does the availability of house name / number and individual's names have on the accuracy of the matching process?
- How should names be compared? (Soundex, NYSYS?)

Possible Areas of Further Evaluation

- Consider lists of frequency of Soundex codes of names in population
- What impact have Asian dates of birth on the matching?
- Can the household type algorithm be improved?
- Can the household representative algorithm be improved?

3.4.4 Input Data

Objective: To determine the set of data that facilitates the most accurate matching.

Census and CCS data will be available in two forms:

- Load data
- Post DEIS data

Areas of Evaluation:

- Which data facilitates the most accurate match?
- Investigate timing implications of each data set.

3.4.5 Matching Techniques

Objective: To determine the most appropriate matching strategy for 2001.

Once the data have been analysed and practical experience gained of matching census data, it is sensible to review the matching strategy to determine whether it is appropriate.

Areas of Evaluation:

- Consider structure of match; is matching households and then individuals within households appropriate?
- Consider emphasis on address information (i.e. blocking by postcode and building number); is the address information accurate enough to justify this?
- Are contiguous postcode lists necessary?

Possible Areas of Further Evaluation

- Are house-movers dealt with effectively?
- How should communal establishments be matched?
- Is the 'fuzziness' at postcode edges accounted for effectively?
- Are vacant properties dealt with effectively?
- How should we deal with single households on one dataset that are enumerated as multiple households on the other dataset?

3.4.6 Matching Software

Objective: To specify an automated matching system appropriate for linking the 2001 Census and CCS data.

The 1999 Rehearsal data will be matched using a prototype matching system.

Areas of Evaluation:

- How could the usability of the prototype system be improved?
- Did the prototype system operate at an acceptable speed?
- Did the system output satisfy requirements?
- Could the system be easily written in any other language?

Possible Areas of Further Evaluation

- Consider ways of improving the efficiency and error trapping in the prototype programme.

3.4.7 Clerical Matching Training

Objective: *To produce comprehensive training materials for 2001 clerical matching.*

Areas of Evaluation:

- How can the training for the 1999 Rehearsal be improved?
- Were the aims of the matching process clearly explained?
- Was the use of the matching system clearly explained?
- Was the definition of what constitutes a match clear?

3.4.8 Calculating / Updating Matching Weights

Objective: *To produce starting weights for the 2001 automated matching process and determine appropriate methodology for updating them as the matching process progresses.*

The current matching proposal involves using weights calculated from the 1999 Rehearsal data as starting weights for the 2001 automated matching process. These weights will be updated as matched 2001 data becomes available.

Areas of Evaluation:

- Is it appropriate to use 1999 weights as starting weights?
- Which methods of calculating matching weights are appropriate for each variable?
- At what level should the weights be updated? (Postcode, ED, HtC?)
- How can weights be updated?

3.4.9 Timing

Objective: *To determine the time and number of people required to perform the matching in 2001.*

Areas of Evaluation:

- How many records will require clerical validation in 2001?
- How many records will require clerical matching in 2001?
- How many people will be able to match simultaneously in 2001?
- How long will the matching take in 2001?

4. ESTIMATION

4.1 Background

The CCS is designed to produce direct estimates for age-sex groups at the design group level. First, a Dual System Estimation (DSE) method is used to estimate the number of people in different age-sex groups missed by both the Census and CCS within each postcode in the CCS sample. Second, the postcode level population counts obtained from these DSEs are used with a modified ratio estimator to obtain final counts for the design group as a whole. This process is carried out separately in each Hard to Count category. The methodology for estimating the design group populations is given in more detail in ONS(ONC(SC))00/03A.

Many LADs will not have enough CCS postcodes to allow accurate direct population estimates to be made. Therefore, it is proposed to use synthetic (or small area) estimation to produce accurate LAD level population estimates. Synthetic estimation uses information from the whole design group to apportion the estimated undercount to the LADs. An LAD adjusted synthetic estimator will be used to derive the LAD estimates as detailed in ONS(ONC(SC))00/03B.

4.2 Changes in Methodology

The above papers represent the most recent ONC estimation strategy. It is intended, if time permits, to use this for the 1999 Rehearsal estimation.

4.3 Differences in Methodology Specific to the Rehearsal

Due to the voluntary nature of the 1999 Census Rehearsal, low levels of coverage were experienced. Since the estimation processes are designed for the high levels of coverage expected in 2001, the population estimates produced in the Rehearsal may not be precise. Therefore it is not appropriate to evaluate the feasibility of the Rehearsal population estimates.

4.4 Evaluation Topics

The evaluation of the estimation processes will focus on the following four areas:

- The design group estimation software;
- The timing of the design group estimation;
- The LAD estimation software;
- The timing of the LAD estimation.

4.4.1 Design Group Estimation Software

Objective: *To evaluate the prototype design group estimation system and determine an appropriate software package for the development of the 2001 system.*

Areas of Evaluation:

- Was the software package used to produce the prototype design group modelling system appropriate?
- Is the system easy to use?
- Can the prototype system be adapted for 2001 purposes?

4.4.2 Timing of Design Group Estimation

Objective: *To determine the time required to perform the design group estimation in 2001.*

Areas of Evaluation:

- Speed of the system.
- Estimation of the time taken to check the resulting estimates.
- Whether any changes in methodology would impact on the time taken to derive estimates.

4.4.3 LAD Estimation Software

Objective: *To evaluate the prototype LAD estimation system and determine an appropriate software package for the development of the 2001 system.*

Areas of Evaluation:

- Was the software package used to produce the prototype LAD modelling system appropriate?
- Is the system easy to use?
- Can the prototype system be adapted for 2001 purposes?

4.4.4 Timing of LAD Estimation

Objective: *To determine the time required to perform the LAD estimation in 2001.*

Areas of Evaluation:

- Speed of the system.
- Time taken to check the resulting estimates.
- The effect of any additional functionality or changes in methodology on the timing.

5. DEMOGRAPHIC ESTIMATES AND ADMINISTRATIVE RECORDS

5.1 Background

The 2001 Census-based population estimates will be compared to demographic estimates produced by Population & Vital Statistics Division. The strategy for the Quality Assurance of ONC estimates is given in ONS(ONC(SC))00/04.

5.2 Rehearsal Methodology

The key use of demographic estimates will be at the level of the national estimates. The 1999 Census Rehearsal does not provide opportunity to investigate estimates at the national level.

As the Census Rehearsal does not cover complete local authority areas, it is not possible to disaggregate the demographic estimates into the required format. Therefore a synthetic dataset will be derived from FHSA data, treating all Census Rehearsal areas in England and Wales as one design group. This will allow development and evaluation of the proposed processes up to the comparison of census based estimates to an alternative independent population estimate.

5.3 Evaluation Topics

The evaluation will focus on:

- Testing interfaces between processes, in particular the mechanisms for data flow;
- Quality of data from administrative records.

5.3.1 Design Group Estimates

***Objective:** To evaluate the processes of determining whether the Design Group estimate is within a plausible range.*

Areas of Evaluation:

The sub-national demographic research and strategy has yet to be developed. A synthetic demographic estimate with synthetic uncertainty intervals will be generated for use in 1999. This will be based on FHSA registers. The process of collecting the data needed for comparisons and providing data to those carrying out comparisons will be evaluated.

5.3.2 Quality of Data from Administrative Records

***Objective:** To assess the quality and reliability of data from administrative records to evaluate whether they are of sufficient quality for making comparisons with the ONC estimates.*

Areas of Evaluation:

A full assessment of the quality and reliability of data from Administrative Records will not be made as part of this evaluation as work is being continued elsewhere. However, the evaluation will report on this previous work and evaluate whether each source that has been considered in the Census Rehearsal is appropriate.

6. IMPUTATION

6.1 Background

The final stage of the ONC process adjusts the census database at micro-level for underenumeration in three main steps:

- Matched Census and CCS data is used to model the probability of being counted in the Census in terms of characteristics of individuals and households. The models are calibrated to the agreed LAD estimates and applied to all individuals and households counted by the Census in order to calculate their 'census coverage' probabilities.
- A donor imputation system uses the coverage weights to create records for non-responding households and individuals who were missed from counted households.
- Finally, the imputation of individuals and households is controlled to agreed LAD totals for household size and age-sex structure.

To date the imputation methodology has been evaluated using simulation studies. The imputation process is described in detail in the paper ONS(ONC(SC)99/08).

6.2 Changes in Methodology

There have been no changes to the imputation methodology outlined in the above paper.

6.3 Differences in Methodology Specific to the Rehearsal

The voluntary nature of the 1999 Census Rehearsal has led to a significantly lower response rate to that expected for the 2001 Census. Although we are confident that the adjusted Rehearsal database will be in agreement with the LAD estimates we should be aware of the data structure when evaluating the system.

6.4 Evaluation Topics

The evaluation of the imputation process will focus on the following six areas:

- Household and individual coverage weights;
- Household imputation;
- Individual imputation;
- Constraint of marginal totals;
- Imputation software;
- Timing.

An important component of the evaluation will be to ensure that the imputation system interfaces with other systems in a live environment.

6.4.1 Household and Individual Coverage Weights

***Objective:** To assess whether household and individual coverage weights created are appropriate.*

Areas of Evaluation:

- Does the model determine household and individual weights which are in agreement with the LAD estimates?

- Is the statistical model appropriate?
- Are optimum household and individual characteristics used in the model?
- Are the weights calculated appropriate (i.e. do they allow imputation of households and individuals with the correct characteristics)?

6.4.2 Selection of Donor Households

Objective: *To evaluate whether the method selects appropriate donor households.*

Areas of Evaluation:

- Does the system select the correct number of the correct types of donor households?
- Do the donor households selected have the correct characteristics?
- Can the method be improved?

6.4.3 Imputation of Households

Objective: *To evaluate whether the system successfully adds those households selected for imputation into the database.*

Areas of Evaluation:

- Does the system place the donors correctly into the database?
- Does the system assign postcodes within the correct ED?
- Can the method be improved?

6.4.4 Selection of Donor Individuals

Objective: *To evaluate whether the system selects appropriate donor individuals.*

Areas of Evaluation:

- Does the system select the correct number of the correct types of donor individuals?
- Do the donor individuals have the correct characteristics?
- Can the method be improved?

6.4.5 Imputation of Individuals

Objective: *To evaluate the suitability of recipient households*

Areas of Evaluation:

- How close is the age-sex structure of the recipient households?
- How close are the recipient households on HtC Index, tenure and ethnicity?
- Can the method be improved?

6.4.6 Constrain Marginal Totals

Objective: To evaluate whether pruning and grafting process constrains margins.

Areas of Evaluation:

- Do the total number of households by size agree with LAD estimates;
- Do the total number of people by age and sex agree with LAD estimates.

6.4.7 Imputation Software

Objective: To identify what changes are required to the current prototype to develop a fully automated imputation system.

Areas of Evaluation:

- Are all variables, active and passive, maintained throughout the process?
- Are the current run times acceptable?
- Can the run times be improved?
- Is the current language the most suitable for purpose?
- Is it feasible/practicable to port the system to another language?

6.4.8 Timing

Objective: To determine the timing and resources required to perform the imputation in 2001.

Areas of Evaluation:

- How long will the imputation take in 2001?
- How much hardware will be required?
- How many staff will be needed?
- Is there potential for further time savings?

7. ONC OVERVIEW

7.1 Background

This section considers aspects of the One Number Census that do not fit within the previous project areas. One important part of the evaluation involves 'taking a step back' from the individual processes and understanding the effect that the ONC as a whole has on the Census data.

Another key issue is that of Communal Establishments. The methodology for estimating the underenumeration within Communal Establishments is an area that has yet to be addressed in detail. This methodology will impact on all ONC processes, as well as other Census areas.

Interfaces with other Census project areas are also covered in this section. The ONC is highly dependent on work undertaken by other areas such as Field, CCS, IS, Processing, Data Quality and Outputs and it is therefore essential that proper communication channels be established. It is also important that GRO(S) and NISRA are kept fully informed of ONC work to ensure a successful UK One Number Census.

Some data quality issues of particular relevance to the ONC are evaluated in this section.

Finally, ONC timetables and staffing plans for 2001 will need to be revised as a result of the evaluation findings. Any revisions will need to fit in with the overall Census timetable and budget.

7.2 Evaluation Topics

The overview evaluation will focus on the following four areas:

- Measurement of change of data through ONC processes;
- Communal Establishments;
- Project interfaces;
- Data quality;
- Timetable and staffing.

7.2.1 Measurement of Change of Data through ONC Processes

Objective: To measure and understand the changes in the dataset as it passes through the ONC processes and back to the Census database.

Areas of Evaluation:

- How are the data being changed?
- What is happening to the distributions?
- Are the changes acceptable?
- How do we decide if this is realistic?

7.2.2 Communal Establishments

Objective: To determine an appropriate method for estimating the underenumeration within Communal Establishments in 2001.

References:

An initial paper on the treatment of Communal Establishments was tabled at the ONC Project Board meeting held on 11 December 1998.

Areas of Evaluation:

- Do we need to sample a range of Special Enumeration Districts e.g. student halls of residence, hospitals, prisons?
- Would the resulting data be sufficient to make estimates to the required level of accuracy?
- How could underenumeration within communal establishments be estimated if not included in the CCS?
- What about transient groups such as circuses, travellers and persons sleeping rough?

7.2.3 Project Interfaces

Objective: *To ensure that interfaces are in place with other areas of Census to allow co-operation between different areas and organisations.*

Areas of Evaluation:

- Are other Census areas fully aware of the aims and objectives of the ONC?
- Are the requirements of the ONC communicated effectively?
- Are the requirements of the ONC met?

7.2.4 Data Quality

Objective: *To assess the quality of the question responses and dummy form information collected in the Census and CCS rehearsals.*

Areas of Evaluation:

- How do the responses to questions differ between matched Census and CCS records?
- How accurate is the information collected on the Census dummy forms?
- How should dummy person details be imputed?
- What were the response rates to the questions in the Census and CCS?

7.2.5 Timetable and Staffing

Objective: *To ensure that any timetables and staffing plans for the ONC in 2001 are achievable, and to revise these in line with evaluation findings as necessary.*

Areas of Evaluation:

- Is the current ONC timetable for 2001 achievable?
- Does this timetable need revising, and if so, how does this fit in with/impact on the timetables for other Census areas and the Census as a whole?
- What are the implications of the evaluation for ONC staffing in 2001?

APPENDIX A: TIMETABLE**KEY DATES:**

1999	April 25th	Census Rehearsal Day
	May/June	CCS Rehearsal fieldwork
	Nov 25th	Census order laid before Parliament
2000	February	ONC Steering Committee meeting
	March	Agree Hard to Count index for 2001
	March	Agree sampling frame for 2001
	March	Agree content of CCS questionnaire
	April 14 th	Census regulations come into effect
	June	ONC Steering Committee meeting to agree overall ONC methodology
	June	Decide on capture of names and/or addresses
	October	Select CCS Sample
2001	April 29th	Census Day
	May/June	CCS fieldwork

EVALUATION DATES:

EVALUATION TOPIC	RESPONSIBILITY	DEADLINE
<i>CCS Sample Selection</i>	Owen Abbott	
2.4.1 Sampling Frame		30 Apr 2000
2.4.2 Hard to Count Index		30 Apr 2000
2.4.3 Size Stratification		30 Apr 2000
2.4.4 Prototype Sampling System		31 Mar 2000
2.4.5 Timing		
<i>Matching</i>	Jennet Woolford	
3.4.1 Matching Feasibility		16 Jun 2000
3.4.2 Derived Variables		16 Jun 2000
3.4.2 Matching Variables		16 Jun 2000
3.4.3 Input Data		16 Jun 2000
3.4.4 Matching Techniques		16 Jun 2000
3.4.5 Matching Software		31 Dec 2000
3.4.6 Clerical Matching Training		31 Jul 2001
3.4.7 Calculating / Updating Matching Weights		31 Jul 2000
3.4.8 Timing		31 Jul 2000
<i>Estimation</i>	Owen Abbott	
4.4.1 Design Group Estimation Software		31 Dec 2000
4.4.2 Timing of Design Group Estimation		31 May 2000
4.4.3 LAD Estimation Software		31 Dec 2000
4.4.4 Timing of LAD Estimation		31 May 2000

EVALUATION TOPIC	RESPONSIBILITY	DEADLINE
<i>Demographic Estimates and Administrative Records</i>		
5.3.1 Design Group Estimates	Heather Wagstaff	31 May 2000
5.3.2 Quality of data from Administrative Records	Roma Chappell	31 May 2000
<i>Imputation</i>		
6.4.1 Household and Individual Coverage Weights	Daniel Howell	16 Jun 2000
6.4.2 Selection of Donor Households		16 Jun 2000
6.4.3 Imputation of Households		16 Jun 2000
6.4.4 Selection of Donor Individuals		16 Jun 2000
6.4.5 Imputation of Individuals		16 Jun 2000
6.4.6 Constrain Marginal Totals		16 Jun 2000
6.4.7 Imputation Software		31 Dec 2000
6.4.8 Timing		31 Dec 2000
<i>ONC Overview</i>		
7.2.1 Measurement of Change of Data through ONC Processes	Heather Wagstaff	31 May 2000
7.2.2 Communal Establishments	Emma Wright	31 May 2000
7.2.3 Project Interfaces	Emma Wright	Ongoing
7.2.4 Data Quality	Jennet Woolford	31 May 2000
7.2.5 Timetable and Staffing	Emma Wright	Ongoing

APPENDIX B: DATA SOURCES

1999 Census Rehearsal Load data
1999 Census Rehearsal Data
1999 CCS Load data
1999 CCS Data
1999 Rehearsal response rates
1999 Rehearsal Enumerator feedback
1999 CCS Interviewer feedback
1999 Clerically matched Census Rehearsal and CCS data
1999 Matched Census Rehearsal and CCS data
1999 ED centroids

Contiguous postcode lists
ONS Geography Central Postcode Directory Data

1997 Test Data
1997 Test response rates
1997 Test Enumerator feedback

1991 Census Data

Southampton Test Data
Southampton Test Interviewer feedback

1997 ED to 1991 ED lookup file
1999 ED to 1991 ED lookup file

APPENDIX C: EVALUATION AND QUALITY ASSURANCE TEAMS

EVALUATION TOPIC	EVALUATION TEAM	QA TEAM
<i>CCS Sample Selection</i>	Owen Abbott Marie Cruddas Jacqui Jones	Ian Diamond Emma Wright
<i>Matching</i>	Jennet Woolford Heather Wagstaff	Ian Diamond Steve Kendrick
<i>Estimation</i>	Owen Abbott Marie Cruddas	Ray Chambers James Brown Emma Wright
<i>Demographic Estimates and Administrative Records</i>	Heather Wagstaff Roma Chappell Pam Spicer	Ian Diamond Emma Wright
<i>Imputation</i>	Daniel Howell Heather Wagstaff	Ray Chambers Fiona Steele James Brown Faith Anderson
<i>Overview</i>	Heather Wagstaff Marie Cruddas Emma Wright Jennet Woolford	Ray Chambers Ian Diamond

BIBLIOGRAPHY

CCS(PB)99/13 *'Evaluating the 1999 CCS Rehearsal'*; Jones, J., Gardner, D.

ONS(ONC(SC))98/12 *'Census Coverage Survey: The Precision of Population Estimates for Different Sample Sizes and Design Areas'*; Abbott, O., Brown, J., Chambers, R., Diamond, I., Dixie, J.

ONS(ONC(SC))98/14 *'One Number Census Matching'*; Baxter, J.

ONS(ONC(SC))99/08 *'A Donor Imputation System to Create a Census Database Fully Adjusted for Underenumeration'*; Steele, F., Brown, J. and Chambers, R.

ONS(ONC(SC))00/01 *'One Number Census Methodology'*; Diamond, I.

ONS(ONC(SC))00/03A *'Estimation Strategy for Design Group Estimates by Age and Sex for the Census Coverage Survey'*; Brown, J., Chambers, R., Cruddas, M.

ONS(ONC(SC))00/03B *'One Number Census Local Authority Estimation'*; Abbott, O., Brown, J., Chambers, R., Cruddas, M.

ONS(ONC(SC))00/04 *'A Quality Assurance and Contingency Strategy for the One Number Census'*;

ONS(ONC(SC))00/10 *'Design Groups for 2001'*; Wright, E.

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