



**ONS(ONC(SC))99/04**

**ONE NUMBER CENSUS STEERING COMMITTEE**

**1999 Census Rehearsal: ONC Evaluation Issues**

1. This paper outlines the proposed One Number Census evaluation of the 1999 Census Rehearsal.
2. **Members of the Steering Committee are asked to:**
  - a) **note this paper;**
  - b) **comment and advise on the areas of evaluation outlined, either at the meeting on 1 July 1999 or in writing by 15 July 1999.**

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

**June 1999**

# 1999 CENSUS REHEARSAL: ONC EVALUATION ISSUES

## TABLE OF CONTENTS

|  |           |
|--|-----------|
| <b>1. INTRODUCTION.....</b>                                      | <b>3</b>  |
| <b>2. CCS SAMPLE SELECTION.....</b>                              | <b>4</b>  |
| 2.1 SAMPLING FRAME.....  | 4         |
| 2.2 HARD TO COUNT INDEX.....                                     | 5         |
| 2.3 SIZE STRATIFICATION .....                                    | 6         |
| 2.4 PROTOTYPE SAMPLING SYSTEM.....                               | 6         |
| 2.5 TIMING.....  | 7         |
| <b>3. CCS PRACTICALITIES .....</b>                               | <b>8</b>  |
| 3.1 CONTACT AND COVERAGE.....                                    | 8         |
| 3.2 RECRUITMENT, PAY AND ADVERTISING .....                       | 8         |
| 3.3 TRAINING.....  | 8         |
| 3.4 FIELDWORK.....   | 9         |
| 3.5 PLANNING AND INTER PROJECT LIAISON .....                     | 10        |
| <b>4. MATCHING THE CCS AND CENSUS REHEARSAL DATA .....</b>       | <b>11</b> |
| 4.1 MATCHING FEASIBILITY .....                                   | 11        |
| 4.2 MATCHING VARIABLES.....                                      | 11        |
| 4.3 INPUT DATA.....  | 12        |
| 4.4 MATCHING TECHNIQUES.....                                     | 12        |
| 4.5 MATCHING SOFTWARE .....                                      | 13        |
| 4.6 CLERICAL MATCHING TRAINING.....                              | 13        |
| 4.7 CALCULATING / UPDATING MATCHING WEIGHTS .....                | 13        |
| 4.8 TIMING.....  | 14        |
| <b>5. DESIGN GROUP ESTIMATION .....</b>                          | <b>15</b> |
| 5.1 MODELLING SOFTWARE.....                                      | 15        |
| 5.2 TIMING.....  | 15        |
| <b>6. LOCAL AUTHORITY DISTRICT ESTIMATION.....</b>               | <b>16</b> |
| 6.1 MODELLING SOFTWARE.....                                      | 16        |
| 6.2 TIMING.....  | 16        |
| <b>7. DEMOGRAPHIC ESTIMATES AND ADMINISTRATIVE RECORDS .....</b> | <b>17</b> |
| 7.1 DESIGN GROUP ESTIMATES.....                                  | 17        |
| 7.2 QUALITY OF DATA FROM ADMINISTRATIVE RECORDS .....            | 17        |
| <b>8. IMPUTATION .....</b>                                       | <b>18</b> |
| 8.1 HOUSEHOLD AND INDIVIDUAL COVERAGE WEIGHTS.....               | 18        |
| 8.2 DETERMINATION OF NUMBER OF IMPUTED HOUSEHOLDS .....          | 18        |
| 8.3 SELECTION OF DONOR HOUSEHOLDS .....                          | 19        |
| 8.4 HOUSEHOLD IMPUTATION.....                                    | 19        |
| 8.5 IMPUTATION OF INDIVIDUALS IN MISSED HOUSEHOLDS .....         | 19        |
| 8.6 AGREEMENT WITH MARGINAL TOTALS .....                         | 19        |
| 8.7 IMPUTATION SOFTWARE .....                                    | 20        |
| 8.8 TIMING.....  | 20        |
| <b>9. ONC OVERVIEW .....</b>                                     | <b>21</b> |
| 9.1 MEASUREMENT OF CHANGE OF DATA THROUGH ONC PROCESSES .....    | 21        |
| 9.2 COMMUNAL ESTABLISHMENTS .....                                | 21        |
| 9.3 PROJECT INTERFACES.....                                      | 22        |
| <b>10. RESPONSIBILITIES AND DEADLINES .....</b>                  | <b>23</b> |

## 1. INTRODUCTION

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

This document outlines the proposed areas for 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 these purposes. 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 key stages of the ONC evaluation are as follows:

- CCS sample selection;
- CCS practicalities;
- Matching the Rehearsal CCS and Census data;
- Design Group estimation;
- Local Authority District estimation;
- Demographic estimates and administrative records;
- Imputation;
- ONC Overview.

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

- Evaluation objective;
- Assessment criteria;
- Data sources.

It should be noted that the assessment criteria are not decision-making criteria, but rather issues to be considered when performing the evaluation. A detailed evaluation plan will be produced once the general areas of evaluation have been agreed. The teams involved in the evaluation process will produce detailed plans for each area of the evaluation. These will be quality assured by members of the ONC team not involved in the specific area of evaluation.

## 2. CCS SAMPLE SELECTION

The primary aim of the ONC is to estimate the population on Census Day, by age and sex, for all Local Authority Districts (LADs). LADs will be grouped together to form 'Design Groups', each with a population of approximately half a million people, to which the CCS design is applied separately.

Information from the 1991 Census is used to select 1991 Enumeration Districts (EDs) from within a Design Group. The EDs are stratified according to a Hard to Count (HtC) index, constructed from census variables associated with underenumeration. In the design for the 1999 Census Rehearsal the HtC index is calculated using 1991 Census data for multi-occupancy, private rented accommodation and young migrants. This index categorises the EDs into five groups.

Within each of the five HtC categories the 1991 EDs are clustered on the basis of their age-sex distribution. A sample of EDs is randomly selected from each of these clusters. This provides an efficient design and spreads the sample with respect to the size of certain key age-sex groups.

Finally, five postcodes are selected at random from each sampled ED. These selected postcodes are those that will be re-enumerated in the CCS.

This strategy has been applied to the 1999 Rehearsal areas with three main differences:

- In order to allow the CCS Team to evaluate the practicalities of co-ordinating a large-scale survey, the CCS must involve a large sample of postcodes. Therefore, a significantly higher sampling fraction than in 2001 was used. The rehearsal CCS covers approximately 18,000 households.
- The 1999 Rehearsal areas are aggregated 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 harder to enumerate areas. Spreading the sample in this way ensures that data is available for all HtC categories within all rehearsal areas.

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

- Sampling frame;
- Hard to Count index;
- Size stratification;
- Sampling system;
- Timing.

### 2.1 Sampling Frame

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

The CCS sampling frame is based on two sources of information:

- 1991 Census information from the Census Geography Database.

- Royal Mail address information from the ONS Geography Central Postcode Directory (CPD).

Both of these data sources are subject to error and will have been out of date by the time the fieldwork was undertaken.

***Assessment Criteria:***

- Proportion of sampled postcodes that no longer exist.
- Proportion of sampled postcodes that do not contain any residential addresses.
- Proportion of sampled postcodes that have changed location in the field.
- Number of postcodes added after the sample had been selected (containing addresses with a zero probability of selection).
- Number of 1991 EDs erroneously included in the sample (e.g. SEDs).
- Number of 1991 EDs erroneously omitted from the sample.
- Were postcodes linked correctly to 1991 EDs?
- Accuracy of Delivery point, Addresses and Business Addresses counts on the CPD
- Accuracy of Residential Household counts on Address-point

***Data Sources:***

- Feedback from CCS interviewers
- 1999 CCS Data
- 1999 Rehearsal Data
- Address-point Data
- ONS Geography CPD Data

## 2.2 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.

The Hard to Count Index (HtC) uses characteristics thought to be associated with underenumeration to provide an indication of the expected level of coverage. The HtC index is calculated at the 1991 Enumeration District level and is based on 1991 Census information.

***Assessment Criteria:***

- Are the variables used when calculating the HtC index appropriate?
- Is 1991 data suitable at ED level?
- How well does the HtC index reflect actual levels of underenumeration at the ED and postcode levels?
- Is the assumption that all postcodes have the same characteristics within a 1991 ED valid?

***Data Sources:***

- 1991 Census Data
- 1997 Test Data
- 1997 Test Enumerator feedback
- 1997 ED to 1991 ED lookup file

- 1999 ED to 1991 ED lookup file
- 1999 Rehearsal Data
- 1999 Rehearsal Enumerator feedback
- 1999 CCS Interviewer feedback
- Southampton Test Data
- Southampton Test Interviewer feedback

### 2.3 Size Stratification

**Objective:** *To evaluate the data used to implement the size stratification.*

The CCS sampling methodology uses a size stratification to spread the sample with respect to number of people, by age and sex, contained within each ED. The stratification is primarily based on 1991 Census age/sex counts and is carried out within each Hard to Count group.

**Assessment Criteria:**

- Are the 1991 age/sex population counts an accurate proxy for the 2001 age/sex populations?

**Data Sources:**

- 1999 CCS Data
- 1999 Rehearsal Response rates
- 1991 ED Age/Sex counts
- 1997 Test Data
- 1999 Rehearsal Data
- 1997 ED to 1991 ED lookup file
- 1999 ED to 1991 ED lookup file
- 1997 Test Response Rates

### 2.4 Prototype Sampling System

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

The 1999 CCS Rehearsal postcode sample was selected using a prototype sampling system.

**Assessment Criteria:**

- Is the system easy to use?
- Is the system flexible enough to allow reselection?
- Is the output summarising the sample selection adequate?
- Did the system operate at an acceptable speed?
- Does SAS have the capabilities necessary for the 2001 system?
- Would another programming language be more appropriate for the 2001 system?

**Data Sources:**

- Feedback from system developer
- Feedback from ONC team

**2.5 Timing**

**Objective:** *To determine the time and number of people required to perform the sample selection for 2001.*

**Assessment Criteria:**

- How long will the sample frame construction take for 2001?
- How long will the sampling take for a single Design Group for 2001?

**Data Sources:**

- Timings for 1999 Sample frame construction
- Timings for 1999 sample selection

### 3. CCS PRACTICALITIES

This section outlines the evaluation plan for the practicalities of conducting the CCS.

The evaluation of CCS practicalities will focus on the following five areas:

- Contact and coverage;
- Recruitment, pay and advertising;
- Training;
- Fieldwork;
- Planning and inter project liaison.

#### 3.1 Contact and Coverage

*Objective: To evaluate the numbers and types of households and communal establishments contacted and interviewed in the CCS Rehearsal.*

**Assessment Criteria:**

- Comparison of addresses found by CCS Rehearsal Interviewers to those on the Postal Address File/ Address point and those found by enumerators.

**Data Sources:**

- LM CCS database.
- Postcode address file.

#### 3.2 Recruitment, Pay and Advertising

*Objective: To evaluate the Team Manager and Interviewer recruitment process.*

**Assessment Criteria:**

- Rates of pay
- Analysis of sources of applications
- Performance of field staff
- Analysis of staff feedback
- Number of listed households and contacts
- Analysis of workloads
- Number of Team Managers and Interviewers required in 2001

**Data Sources:**

- Team Manager / Interviewer job application forms
- Team Manager's and Interviewers debriefing questionnaire
- Qualitative analysis from the Team Manager's and Interviewers debriefing day
- Management Information system
- Team Manager and Interviewer hours worked logs
- Query logs

#### 3.3 Training

*Objective: To evaluate the training given to Team Managers and Interviewers.*

**Assessment Criteria:**

- Review home study video
- Review workbook
- Review timetable / content
- Review instructions
- Qualitative analysis from Team Managers and Interviewers debriefing

**Data Sources:**

- Team Manager's debriefing questionnaire
- Interviewer's debriefing questionnaire
- Team Manager's training debriefing questionnaire
- Interviewer's training debriefing questionnaire
- Observation by HQ staff during training sessions and fieldwork
- Logs of queries

**3.4 Fieldwork**

**Objective:** *To evaluate the CCS Rehearsal fieldwork. In particular:*

- *CCS Survey Forms*
- *Property listing, maps, Response Card, Prompt Card and Pocket Guide*
- *Self-completion post-back form*
- *Calling strategies*
- *Allocation of workloads*
- *Proxy information*
- *Population subgroups less likely to be found*
- *Progress reports*

**Assessment Criteria:**

- Compare Census and CCS responses to questions
- Percentage of forms posted back
- Response rates
- Interviewer feedback
- Team Manager feedback

**Data Sources:**

- Team Manager's debriefing questionnaire
- Team Manager's training debriefing questionnaire
- Interviewer's debriefing questionnaire
- Interviewer's training debriefing questionnaire
- LM CCS database
- Matched Census & CCS dataset
- Team Manager and Interviewer hours worked logs
- Logs of queries
- Observations of HQ staff during fieldwork

### **3.5 Planning and inter project liaison**

*Objective: To evaluate the management and operational processes*

*Assessment Criteria:*

- Value of Hard to Count index when planning workloads
- Comparison of address lists and Postal Address File
- Management information requirements
- Whether all supplies were sent and received according to schedule
- Working arrangements with other projects

*Data Sources:*

- Field staff debriefing questionnaire
- Logs of queries
- LM CCS database
- Property listing
- Census listing
- Postal Address File
- HQ debriefing sessions

#### 4. MATCHING THE CCS AND CENSUS REHEARSAL DATA

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.

Prior to the analysis described here, it will be necessary to match a sample of data clerically. This 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.

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

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

##### 4.1 Matching Feasibility

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

***Assessment Criteria:***

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

***Data Sources:***

- Clerically matched dataset
- 1999 Rehearsal data
- 1999 CCS data

##### 4.2 Matching Variables

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

***Assessment Criteria:***

- 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?

***Data Sources:***

- Clerically matched dataset
- 1999 Rehearsal data
- 1999 CCS data

### 4.3 Input Data

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

Census and CCS data will be available in many forms. In particular:

- Raw data, flagged for inconsistencies
- Edited data, prior to item-imputation
- Post item-imputation data

***Assessment Criteria:***

- Perform automated matching on each of the above data sets. Use the clerically matched data to assess which of the matching processes has been most accurate.
- Investigate the technical issues involved with each data set. For example, item imputed data is required for the modelling stages of the ONC process and is therefore required as the output of the matching process. How can this be achieved if an earlier form of the data is used for matching?
- Investigate timing implications of each data set.

***Data Sources:***

- Raw 1999 Rehearsal data
- Raw 1999 CCS data
- Post edit 1999 Rehearsal data
- Post edit 1999 CCS data
- Post item-imputation 1999 Rehearsal data
- Post item-imputation 1999 CCS data
- Contiguous postcode lists

### 4.4 Matching Techniques

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

Once the data has been analysed and practical experience gained of matching census data, it is sensible to review the high level matching strategy to determine whether it is indeed the most appropriate.

**Assessment Criteria:**

- Consider structure of match; is matching households and then individuals within households appropriate?
- Consider emphasis on address information; is this justified?
- Are contiguous postcode lists necessary?

**Data Sources:**

- 1999 Rehearsal data
- 1999 CCS data
- Contiguous postcode lists

#### 4.5 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.

**Assessment Criteria:**

- Is the system easy to use?
- Does the system operate at an acceptable speed?
- Does the system output satisfy requirements?
- Could the system be easily written in any other language?

**Data Sources:**

- Clerical feedback

#### 4.6 Clerical Matching Training

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

**Assessment Criteria:**

- Was the training easy to understand?
- 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?

**Data Sources:**

- Clerical feedback

#### 4.7 Calculating / Updating Matching Weights

**Objective:** *To produce starting weights for 2001 automated matching process and determine methodology to update 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.

***Assessment Criteria:***

- 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?

***Data Sources:***

- 1999 Matched dataset
- Clerically matched dataset

**4.8 Timing**

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

***Assessment Criteria:***

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

***Data Sources:***

- Clerically matched dataset
- 1999 Rehearsal data
- 1999 CCS data

## 5. DESIGN GROUP ESTIMATION

Estimates of the Design Group populations will be based on a methodology known as dual system estimation. This assumes that a certain proportion of the population will be missed by both the Census and the CCS, but that this number can be estimated by considering the relative numbers of the people observed by both the Census and CCS; by the Census but not the CCS; and by the CCS but not the Census.

The dual system estimates will be used to construct regression models to predict the level of underenumeration for each age-sex group in each Hard to Count classification. These estimates can be summed across HtC groups to provide age-sex populations for the Design Group in question. The error associated with these estimates can also be calculated.

Due to the voluntary nature of the 1999 Census Rehearsal, low levels of coverage are expected. The estimation processes will therefore not produce sensible estimates and it is not appropriate to evaluate the methodology here. The evaluation of the design group estimation will therefore focus on two areas only:

- Modelling software
- Timing

### 5.1 Modelling Software

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

The 1999 Rehearsal Design Group estimation will be carried out by a prototype system.

***Assessment Criteria:***

- Was the system user friendly?
- Did the system produce the appropriate estimates?
- Was the system flexible enough?
- Was the system fast enough?
- Were appropriate diagnostics produced?
- Did the system meet the specified requirements?

***Data Sources:***

- Feedback from system developer
- Feedback from ONC Team

### 5.2 Timing

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

***Assessment Criteria:***

- How long will the estimation take in 2001 for a single design group?

***Data Sources:***

- Matched Census and CCS dataset

## 6. LOCAL AUTHORITY DISTRICT ESTIMATION

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. Furthermore, the accuracy of the resulting LAD population estimates can also be calculated.

At present, the prototype LAD estimation system uses a simple synthetic estimator. The results of current methodological research will determine how the prototype system should be revised.

As for the Design Group estimates, it is expected that the estimation processes will not produce sensible population estimates. Therefore only the prototype modelling software and timing issues are evaluated here.

### 6.1 Modelling Software

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

**Assessment Criteria:**

- Was the system user friendly?
- Did the system produce the appropriate estimates?
- Was the system flexible enough?
- Was the system fast enough?
- Were appropriate diagnostics produced?
- Did the system meet the specified requirements?

**Data Sources:**

- Feedback from system developer
- Feedback from ONC Team

### 6.2 Timing

**Objective:** *To determine the time and numbers of people required to perform the LAD estimation in 2001.*

**Assessment Criteria:**

- How long will the LAD estimation take in 2001 for a single design group?

**Data Sources:**

- Matched Census and CCS dataset

## 7. DEMOGRAPHIC ESTIMATES AND ADMINISTRATIVE RECORDS

The 2001 Census based estimates will be compared to demographic estimates produced by P&VS using the rolled forward 1981 Census data and aggregate administrative records.

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, to allow development and evaluation of the proposed processes up to the comparison of census based estimates to an alternative independent population estimate.

The evaluation will focus on:

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

### 7.1 Design Group Estimates

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

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.

### 7.2 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.*

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

## 8. IMPUTATION

The final stage of the ONC process is to adjust the census database at microlevel for underenumeration in three main steps:

- First the matched Census and CCS data is used to model the probability of being counted in the Census in terms of the characteristics of individuals and households. These models are applied to all individuals and households counted by the Census in order to calculate their ‘census coverage’ probabilities.
- A donor imputation system then uses the coverage weights, together with information from dummy forms, to create records for non-responding households and individuals that were missed from counted households.
- The imputation of individuals and households is controlled to the agreed LAD totals.

To date the imputation methodology has been evaluated using simulation studies. The 1999 Census Rehearsal provides opportunity to test and evaluate both the imputation system and methodology in a live environment

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

- Derivation of household and individual coverage weights;
- Determination of number of imputed households;
- Selection of donor households;
- Household imputation;
- Imputation of individuals in counted households;
- Extent of agreement with marginal totals;
- Ease of use, adequacy and flexibility of imputation software – the prototype system has been written in SAS;
- Running time for processing.

### 8.1 Household and Individual Coverage Weights

*Objective: To evaluate whether household and individual coverage weights are created appropriately.*

**Assessment Criteria:**

- Does the multinomial logistic regression (mlr) perform well?
- Are appropriate household and individual characteristics used in mlr?
- Are the derived household and individual weights appropriate?

**Data Sources:**

- Census Database
- Matched Census + CCS database
- Estimated LAD totals

### 8.2 Determination of Number of Imputed Households

*Objective: To evaluate whether an appropriate number of households have been determined at the LAD level*

***Assessment Criteria:***

- Were a realistic number of households determined for imputation at the LAD level?
- Was the geographic information used appropriate?

***Data Sources:***

- Census Database
- Calibrated household weights
- Geographical Information

**8.3 Selection of Donor Households**

***Objective:*** *To evaluate whether appropriate donor households were selected*

***Assessment Criteria:***

- Did the donor households have the correct characteristics?
- Were the households sorted and cumulated correctly?
- Can the method be improved?

***Data Sources:***

- Census Database

**8.4 Household Imputation**

***Objective:*** *To evaluate the household imputation*

***Assessment Criteria:***

- Did the search for a donor household work well?
- Can the method be improved?

***Data Sources:***

- Census Database

**8.5 Imputation of Individuals in Missed Households**

***Objective:*** *To evaluate the imputation of individuals within missed households*

***Assessment Criteria:***

- Was the recipient household appropriate?
- Was the recipient household in the optimum location?
- Has consistency been maintained?
- Can the method be improved?

***Data Sources:***

- Census Database

**8.6 Agreement with Marginal Totals**

***Objective:*** *To establish whether imputation process obtains results which agree with the fixed marginal totals*

***Assessment Criteria:***

- Did the marginal totals agree with the LAD totals?
- Was amount of pruning and grafting as small as possible and did it take the correct/minimum time to do it?
- Can the method be improved?

***Data Sources:***

- Census Database

**8.7 Imputation Software**

***Objective:*** *To evaluate the prototype computer system*

The synthetic population estimated by earlier ONC processes will be placed in the Census database using a prototype imputation system.

***Assessment Criteria:***

- Is the system easy to use?
- Does the system operate at an acceptable speed?
- Does the system output satisfy requirements?
- Is SAS the optimum strategic language for the ONC imputation?

***Data Sources:***

- Census Database
- ONC Imputation System

**8.8 Timing**

***Objective:*** *To determine the staffing resources to perform the imputation in 2001.*

***Assessment Criteria:***

Within a single design group, how long does it take to:

- calculate the weights?
- impute the required number of households?
- impute the required number of individuals?
- prune and graft individuals and households?
- hence, complete the whole process?

***Data Sources:***

- Matched dataset
- LAD totals from Estimation System
- Census Database

## 9. ONC OVERVIEW

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.

Also covered in this section are interfaces with other Census project areas. The ONC is highly dependent on work undertaken by other areas such as Field, CCS, IS, Processing, Data Quality and Outputs. Therefore it is essential that proper communication channels be established.

### 9.1 Measurement of Change of Data through ONC Processes

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

**Assessment Criteria:**

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

**Data Sources:**

- Census Database
- Matching System output
- Estimation Systems output
- ONC Imputation System output

### 9.2 Communal Establishments

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

**Assessment Criteria:**

- 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?

**Data Sources:**

- Census Database

- Matched Census + CCS database

### 9.3 Project interfaces

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

**Assessment Criteria:**

- 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?

**Data Sources:**

- Feedback from other areas of Census
- Feedback from ONC Team

**10. RESPONSIBILITIES AND DEADLINES**

| <b>Evaluation</b>  | <b>Responsibility</b> | <b>Deadline</b> |
|--|-----------------------|-----------------|
| <b>2. CCS sample selection</b>                             | Owen Abbott           | April 2000      |
| 2.1 Sampling frame   |                       |                 |
| 2.2 Hard to Count index                                    |                       |                 |
| 2.3 Size stratification                                    |                       |                 |
| 2.4 Sampling system  |                       |                 |
| <b>3. CCS Practicalities</b>                               | Jacqui Jones          | September 2000  |
| 3.1 Contact and coverage                                   |                       |                 |
| 3.2 Recruitment, pay and advertising                       |                       |                 |
| 3.3 Training   |                       |                 |
| 3.4 Fieldwork  |                       |                 |
| 3.5 Planning and inter project liaison                     |                       |                 |
| <b>4. Matching</b>   | Jennet Baxter         | August 2000     |
| 4.1 Matching feasibility                                   |                       |                 |
| 4.2 Matching variables                                     |                       |                 |
| 4.3 Input data   |                       |                 |
| 4.4 Matching strategy                                      |                       |                 |
| 4.5 Matching software                                      |                       |                 |
| 4.6 Clerical matching training                             |                       |                 |
| 4.7 Calculating / updating probability weights             |                       |                 |
| 4.8 Timing   |                       |                 |
| <b>5. Design Group Estimation</b>                          | Owen Abbott           | August 2000     |
| 5.1 Modelling software                                     |                       |                 |
| 5.2 Timing   |                       |                 |
| <b>6. LAD Estimation</b>                                   | Owen Abbott           | August 2000     |
| 6.1 Modelling software                                     |                       |                 |
| 6.2 Timing   |                       |                 |
| <b>7. Demographic Estimates and Administrative Records</b> | Heather Wagstaff      | August 2000     |
| 7.1 Design Group estimates                                 |                       |                 |
| 7.2 LAD estimates  |                       |                 |
| 7.3 Quality of data from administrative records            |                       |                 |
| <b>8. Imputation</b>                                       | Daniel Howell         | August 2000     |
| 8.1 Household and individual coverage weights              |                       |                 |
| 8.2 Determination of number of households to be imputed    |                       |                 |
| 8.3 Selection of donor households                          |                       |                 |
| 8.4 Imputation of individuals in missed households         |                       |                 |
| 8.5 Imputation of individuals in counted households        |                       |                 |
| 8.6 Constrain marginal totals                              |                       |                 |
| 8.7 Imputation Software                                    |                       |                 |
| 8.8 Timing   |                       |                 |
| <b>9. ONC Overview</b>                                     |                       |                 |
| 9.1 Measurement of change of data through ONC processes    | Heather Wagstaff      | August 2000     |
| 9.2 Communal Establishments                                | Marie Cruddas         | August 2000     |
| 9.3 Project Interfaces                                     | Lisa Buckner          | August 2000     |