



ONS(ONC(SC))98/01

ONE NUMBER CENSUS STEERING COMMITTEE

One Number Census Project Report

1. This paper is a project report for the One Number Census. Section 2 gives an overview of progress since the last ONC Steering Committee meeting on 27 November 1997. Section 3 outlines details of the research programme for the next 6 - 9 months.
2. **The Steering Committee are asked to:**
 - a) **note the paper and the progress made;**
 - b) **endorse the research programme and**
 - c) **provide any comments at the meeting on the 27 April 1998, or in writing by 10 May 1998.**

**Marie Cruddas
Census Division
Office for National Statistics**

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

March 1998

Project Report for the One Number Census

Marie Cruddas and Lisa Buckner

1. Introduction

This report documents progress on the One Number Census (ONC) project since the ONC Steering Committee meeting on 27 November 1997. Information is also included on the research programme for the next 6 - 9 months.

2. Update on key areas of research

Consultation

The One Number Census consultation paper has been prepared and is being distributed to all members of the Census Advisory Groups for the April/May round of meetings. A shortened version of the overview section is being published in Census News along with information on how to obtain the complete consultation paper. It is also being put on the ONS census webpage. It is hoped that this will encourage discussion on the ONC and feedback on the proposed methodology.

A ONC workshop will take place from 12-13 May in Leeds as part of a Joint Workshop on Census Issues. This has been arranged by Professor Phil Rees from Leeds University and Mr Steve Simpson from the University of Manchester. Members of the ONC project team will be giving presentations on the key areas of the methodology to facilitate discussion. All ONC Steering Committee members were sent details of the workshop.

James Brown presented a paper at the Population Association America conference in Chicago from 2 - 4 April 1998. This paper, 'Statistical models to estimate underenumeration in the 2001 Census of England and Wales' (ONS(ONC(SC))98/06), outlines the proposed methodology for the ONC focusing on the use of spatially smoothed random effects models for micro level adjustments to the census database.

On 7 April Andy Teague gave a presentation on the One Number Census to the Market Research Society.

A paper has been submitted to the Royal Statistical Society Series A for publication (ONS(ONC(SC))98/05).

Census Coverage Survey

We plan to conduct another Census Coverage Survey (CCS) pilot in order to develop further the practical aspects of undertaking a postcode based CCS. This will take forward the lessons learnt during a small pilot survey of 20 postcodes carried out in Brent following the 1997 Census Test. More information on this proposed survey is given in ONS(ONC(SC))98/04.

Ian Diamond, Lisa Buckner and Marie Cruddas met with Patrick Heady (Methods and Quality Division, ONS) and Peter Lynn (SCPR) to discuss the issues and problems in managing a large survey. These meetings proved extremely useful and details of resulting ideas are outlined in ONS(ONC(SC))98/04, the paper also addresses other practical issues that arise in undertaking the CCS following the 2001 Census.

Matching

Work is underway to develop a strategy for matching the CCS data to that gathered by the census. Data obtained during the Brent Census Coverage Survey pilot has been matched by hand to that from the 1997 Census Test. The variables used in matching were postcode, address number, name, date of birth and sex. The process was repeated without using name to examine the effect on the number of matches.

The next stage is to develop computer algorithms to reproduce this hand match.

Lisa Buckner met with Dr Fiona O'Brian from Information and Statistics Division (ISD) in Edinburgh. ISD make extensive use of matching techniques for linking health records. This meeting proved very interesting and useful and will enable the development of computer algorithms for matching to proceed more rapidly.

Local Authority District Estimates

Following the last meeting of the Steering Committee, where it was suggested that census-based population estimates by age and sex would be required for Local Authority Districts (LADs), the focus of the CCS design has changed. In the design, large Local Authority Districts will be treated as separate strata and small Local Authority Districts will be grouped together into pseudo 'counties'. (Note: that in this context 'large' refers to estimated population size and not area covered). Census-based population estimates by age and sex can then be obtained for these large LADs and groups of smaller LADs directly using a combination of standard estimation techniques.

The regression estimator developed previously to estimate the population by age and sex assumes that the CCS captures everyone. However, as there are likely to be some people who are missed it is now proposed to use a combined Dual-System and regression estimator. Dual System estimation will be used to estimate the total population for an area including those people thought to have been missed by both the census and CCS. This estimate will then be used in the regression estimator. This approach is described in detail in Section 4 of ONS(ONC(SC))98/05.

Demographic Analyses in Support of a One Number Census

See ONS(ONC(SC))98/08 for proposed further research on demographic analyses in support of a ONC.

Small Area Estimation

Work has taken place to extend the simulations on the small area modelling to include another explanatory variable (Primary Activity Last Week) in addition to age-sex and hard to count index. An additional random effect at the enumeration district level has also been added. The inclusion of these are detailed at the start of Section 5 in ONS(ONC(SC))98/06. The results presented in Section 5.3 of the paper show that the fixed effect approach to the modelling continues to perform well reducing the bias across all the variables. The results in Section 5.4 show that the first work with random effects and spatial smoothing has not given the expected efficiency gains to the models in terms of further bias reduction. Initial investigations have revealed that in a few cases the random effects model fits the data very poorly. The exact reasons for this are currently under investigation.

Extending the simulation has also highlighted the need to consider different models for different age groups: a variable such as Primary Activity Last Week is an appropriate explanatory variable for people aged 16+ but not for those under 16. This is true for other variables such as marital status. In the paper three separate models have been fitted, one for people aged 16+, one for children aged 5-15 at the household level, and one for children aged 0-4. This is detailed in Section 5.2 of ONS(ONC(SC))98/06.

Imputation and weighting

See ONS(ONC(SC))09/07 for an illustrative example of imputation and weighting and also the pros and cons of using each method.

3. Programme of further work

Management issues

- Developing a CCS model developing the ideas in (ONS(ONC(SC))98/04). John Dixie will be taking forward this work.

Practicalities of CCS

- Undertake additional pilot (September 1998) to take forward the lessons learnt during Brent CCS pilot 1997 and further develop practical aspects of undertaking a postcode based CCS.
- Preparations for 1999 Dress Rehearsal:

CCS design

- Further develop Hard to Count (HtC) index.
- Estimate national (England and Wales) sample size required for the CCS following the 2001 Census based on designing survey at level of large LADs and groups of smaller LADs.
- Extend design to include Scotland and Northern Ireland and calculate sample size.
- Further develop combined Dual System and regression estimator.

Matching

- Develop matching strategy and probability matching algorithms
- Develop and test computer algorithms for matching
- Investigate whether name significantly improves ability to match

Small area modelling

- Increase complexity of model for excluding people from census in simulations
- Further develop spatially smoothed random effects models

Imputation and weighting

- Develop imputation methods for putting individuals into counted households
- Develop techniques for weighting the database and producing suitable output