ONS(ONC(SC))97/08

ONE NUMBER CENSUS STEERING COMMITTEE

An overview of the One Number Census process

1. This paper gives an overview of the main features of the proposed methodology for the One Number Census and outlines proposed further work.

2. The steering committee are asked to:

a) note the paper; and

b) provide any comments on the currently proposed strategy and planned further work at the meeting, or on writing by 10 December 1997.

Marie Cruddas
Census Division
Office for National Statistics

Room 4200W
Segensworth Road
Titchfield
Fareham
HANTS
PO15 5RR

November 1997

An overview of the One Number Census process

This paper first gives an overview of the proposed methodology the One Number Census. The aim is to describe the main features of the process more detail can be found in the individual papers prepared for the One Number Census Steering Committee.

The paper goes on to outline proposed further work.

1. Overview of the ONC methodology

The process outlined below represents the proposed ONC methodology. The ONC process can best be considered as consisting of the four main stages summarised below and illustrated by the Figure in Annex A.

The first two stages will be to produce the best estimate of the population by age and sex at national and county (average 1 million population) levels. These stages are essential for rebasing the population estimates at the National and County level. The third stage will produce estimates for lower levels of geography (generically referred to as "Ward level" in Annex A and below, but could be any suitable geographical level. In particular Local Authority level will be required for the next stage of rebasing population estimates) and for other characteristics of people and households. The final stage is to impute both for households that are estimated to have been missed and for people estimated to have been missed from counted households. This last stage would allow all statistics based on the 2001 Census to aggregate to 'One Number'. These last two stages perhaps have less relevance for central government distribution of resources but are essential for many users, for example for local authorities who need to allocate resources at a local level. In addition, it is of course important that users know the implications of any undercount on figures for small population groups and geographical areas.

Stages 1 and 2 - National and County level estimates

The first stage in the ONC process, following the successful completion of the Census is to estimate the population by age and sex at the county level. Counts from the 2001 Census will be adjusted for estimated net underenumeration using a post enumeration survey to be known as the *Census Coverage Survey (CCS),* perhaps combined with data from administrative records using *capture/recapture analysis* techniques. The aim here is to estimate those missing from all sources and thus to reduce the error attached to the county level estimates. The CCS will be particularly innovative in that it will be concerned only with coverage and not with quality as in past censuses. The questionnaire will be short and will cover only those characteristics which are likely to be associated with underenumeration.

The county level estimates will then be aggregated to produce a *national Census-based estimate* and compared with *the national demographic estimate* of the population (the estimate of the population rolled forward from the previous Census).

Steering Committee paper 97/10 'Design of the Census Coverage Survey' describes the methodology for the CCS. It proposes a design based on the re-enumeration of a sample of whole postcode units, stratified by county and an index based on difficulty to enumerate. The paper applies the methodology to Hampshire, and shows that designing to achieve a relative

standard error of 1% for the county totals would require a sampling fraction of 10% of enumeration districts from each of which it is proposed to sample four postcodes. At a national level a simple extrapolation would infer that a 1% RSE for each county would require a sample of 40,000 postcodes (approximately 600,000 households) although, of course, a larger RSE would be commensurate with a smaller sample size.

Although not available at the design stage, auxiliary information in the form of the 2001 unadjusted Census counts will be available for use in regression estimation of the County population by age-sex totals. Simulations presented in the paper show that the regression estimator is working well.

The practicalities of the CCS were assessed by a pilot postcode survey carried out in the Brent area of London following the 1997 Census Test. Steering Committee paper 97/11 'Report on the Census Coverage Survey Pilot' presents the methodology and discusses the lessons learnt. A short questionnaire was used to collect information on characteristics believed to be associated with underenumeration. The simplicity of the questionnaire and the fact that sampling whole postcodes makes efficient use of interviewer time, makes a much larger sample size possible than was the case for the 1991 Census Validation Survey. Important lessons learnt were that a postcode based CCS is possible and that it needs to be carried out very soon after the census and have as short a duration as possible in order to capture information on the mobile groups thought to be associated with underenumeration.

The conclusions drawn from work on the design of the CCS and pilot survey lead to an estimated cost of £6 million (see Steering Committee paper 97/18 'Census Coverage Survey - management, resources and costs).

However it is inevitable that the CCS will fail to find all the missing people. Consideration has been given of ways to improve the CCS estimate by linking data from administrative sources to the Census and CCS records at the individual level using capture-recapture techniques. Possible sources of administrative data have been examined (Steering Committee paper 97/14 'Using administrative data in the One Number Census') and the only ones that are strong candidates for provision of suitable individual level information are those from National Health Service sources - the Family Health Service Authority registers and the NHS Central Register. However it must be stressed that the potential use of these records depends greatly on their not overcovering the population. Current research is investigating the accuracy of these lists.

Steering Committee Paper 97/12 'Capture-recapture estimation in a One Number Census' outlines the proposed methodology for using administrative data with the Census and CCS data. Simulation results are presented based on the use of two and three sources of data under varying assumptions about the degree of dependence between sources and the over representation of individuals in the lists. The key question is whether an administrative list really adds to the accuracy or whether a regression estimator based on the CCS can provide an acceptable estimate. Simulations are underway to address this. The results show that when the coverage of the Census is not high the degree of dependence between sources is important and that over representation of individuals is always crucial.

Steering Committee paper 97/15 'Demographic analysis for a One Number Census' summarises work underway to optimise the methodology used to produce the demographic estimates which will be used as a check on the Census plus CCS based estimates. This work,

using a cohort approach independent of census data, suggests that the 1981 Census provides the best base from which to roll forward population estimates to produce a demographic estimate of the population in mid 2001. Migration is identified as the key source of error in the population estimates and the paper gives an analysis of the size of the error. Future work is identified to improve the migration estimates and to examine the potential use of more administrative data. Some of the issues surrounding the production and release of demographic population estimates are discussed in Steering Committee paper 97/17 'Demographic Based Estimates for 2001'.

Stages 3 and 4 - Small area estimation and imputation

The production of adjusted census counts for small areas (and ultimately to adjust the Census database) represent the final goals of the ONC process. The aim is first to estimate the number of people underenumerated in each 'small area' who are either

a) in households which were enumerated; or
b) in households which were missed in the Census.

This is achieved by using a multilevel multinomial logistic regression estimator (described in Steering Committee paper 97/13 'Modelling to small areas - a full One Number Census').

These people are allocated to individual households using

- an adaptation of donor imputation for a) , and
- iterative proportional fitting for b) above.

The initial results of simulations to assess the efficiency of this approach are extremely encouraging.

2. Plans for further work

At present substantial progress has been made in a very short period of time. However research continues at all levels of the programme. Current and future research includes:

a) Extending the CCS design to all counties so as to finalise the information on the sample sizes required for particular RSEs. As a result firmer information will be available on costs and timings.

b) Developing the assessment, both substantively and methodologically, of the use of administrative records. At a methodological level the work includes further work on loglinear capture-recapture models

- to assess the benefits of three as opposed to two lists, and
- to assess the impact of dependence between lists.

On the substantive side the work involves an assessment of the accuracy of National Health Service registers.

c) Extending the research which imputes individual records from the county estimates. The work includes:

- spatial smoothing of multilevel unobserved variances across hard to enumerate areas in order to make regression based estimates of small area underenumeration,
- simulations of the accuracy of donor imputation and iterative proportional fitting to estimate individual level records.

ANNEX A

The stages of a One Number Census