



ONS(ONC(SC))97/14

ONE NUMBER CENSUS STEERING COMMITTEE

Using administrative records in the One Number Census

1. This paper reviews the work underway to evaluate the potential of using data from administrative sources as part of the One Number Census.
2. The paper proposes that resources should be concentrated on assessing whether data from NHS sources are of adequate quality to use as an adjunct to Census and CCS data, to improve the accuracy of capture-recapture estimates.
3. **Members of the Steering Committee are asked to:**
 - **note the paper; and**
 - **provide any comments (at the forthcoming meeting or in writing by 10 December 1997) on the proposals for further research.**

**David Thorogood and Wayne Codd
Census Division
Office for National Statistics**

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

November 1997

Using administrative records in the One Number Census

David Thorogood, Wayne Codd

Introduction

In an ideal world the Census backed up by the Census Coverage Survey would obtain information from every individual in the areas sampled by the coverage survey. However, in reality one typically finds that a number of individuals are missed both in the census and in the coverage survey. It is important to identify the number of these individuals in order to be able to estimate efficiently the total number of people who were not covered by the census. In the years following the second world war the development of dual-system estimators led to a number of countries making estimates of those missed both by the census and the coverage survey using relatively simple strategies based on a two by two contingency table. However, after some success, a number of papers suggested that the accuracy of these estimates was open to question.

This paper concentrates on the potential use of NHS administrative sources data at the individual record level, as an adjunct to Census and CCS data in capture-recapture models. A key issue to address now is whether data from NHS and other sources are of sufficient quality to allow use at this individual record level.

There is also scope for the use of administrative data at aggregate level, to provide an estimate at national or sub-national level of the whole population or of a particular population sub-group.

Identification and assessment of potential sources

The first stage of this work has been to identify and review the potential data sources. Work now is concentrating on assessing more fully the suitability of selected key sources and on obtaining extracts of data for analysis. As discussed in paper *ONS(ONC(SC))97/05*, presented to the Steering Committee in May 1997, a series of criteria were developed to guide the investigation into the suitability of data from different administrative sources. These criteria can be grouped under the following broad headings:

- coverage and type of data held;
- data quality; and
- accessibility.

Legal advice obtained by ONS indicates that the use of administrative data for a statistical purpose such as the Census is allowed under UK and European law. However, using administrative data in this way would be a sensitive issue, and this will require careful treatment to ensure that the public fears about confidentiality are not increased. For example, it could be feared incorrectly that Social Security

information was being linked to census records to check on a person's entitlement to benefits, leading to a possible loss of public co-operation with the Census. Ways of assessing public beliefs and attitudes on this are being considered.

Data sources

The results of the work so far suggest that the following sources are of the greatest potential use in the Census:

- the Family Health Services Authority and NHS Central Register (NHSCR) used in the administration of the NHS General Practitioner (GP) service;
- the Department of Social Security (DSS) Child Benefit and Retirement Pension; and
- the Council Tax Dwelling Lists.

Of these, the NHS sources are seen as being most suited for use in capture-recapture models, and are discussed in more detail below. These appear to be the only administrative data which are both available at individual record level and cover all or most of the population. There are, however, serious concerns about whether the data are accurate enough for this type of use.

The DSS sources could be useful at an aggregate level as an adjunct to demographic estimates work. DSS data have been used previously by the ONS Population Estimates unit in re-basing the population estimates following the 1991 Census. Extracts of DSS data are being obtained.

The Council Tax Dwelling Lists may help in the planning and assessment of how well the Census and CCS identifies all dwellings in an area. Data at individual dwelling level have been obtained for the local authority areas which include the 1997 Census Test areas, and are currently being compared by ONS Census Geography with the address lists used by the Census Test enumerators. The Dwelling List data are currently held by the Valuation Office Agency on separate databases for each local authority area, and on computer systems which are not readily compatible with those used by ONS. The problems caused by this should not occur in later years as the data will shortly be transferred to a single database with updated computer hardware.

Further work is needed on some of the other sources listed in ONS(ONC(SC))97/05 - in particular information on students, and data available from public utility supply address lists. However, it is proposed that this work should take a low priority. New and additional data sources are still being brought to the attention of the project team. More generally, the research into administrative sources needs to continue to ensure that the work keeps abreast with changes in data availability.

Using data from National Health Service sources

Administration of the NHS in England and Wales involves patient registers at two hierarchical levels. At the higher level, the National Health Service Central Register (NHSCR) includes details of all persons who are registered as a patient with a doctor (GP). Registration with a GP in the current area of residence is not compulsory, but must normally occur if a person requires non-emergency treatment. There are also around 100 local level registers roughly corresponding to counties or metropolitan boroughs (Family Health Service Authorities - FHSAs) which are linked to individual GP practices and to NHSCR.

Entries on the NHSCR include the name, sex, date of birth and FHSA of current residence of patients. FHSAs hold similar information, but additionally, hold the postcode for each patient. Although the NHSCR offers far less geographical information, access is far easier to arrange as it is administered by ONS and has an established function of supplying migration data for the production of annual population estimates. ONS Population and Vital Statistics are investigating the possible use of FHSA data to estimate local migration, and have been able to obtain annual extracts from the FHSA databases. Names are held by FHSAs, they are not included in the information currently copied to ONS.

The NHSCR and FHSAs are thought to offer almost universal coverage of the population, although there are problems of patient count inflation - where old registrations are not removed - and non-reporting of new information such as change of address. These problems have been thought to be more severe in certain areas of the country - particularly those with relatively mobile populations - and for certain age/sex groups.

A number of recent administrative and computing changes are believed to have led to great improvements in the quality of both NHSCR and FHSA data. There have been improvements in the communication between GPs, FHSAs and NHSCR, with much registration information now being rapidly transferred by electronic means. Combined with the introduction of a new NHS patient number system, this has required the reconciliation of patient lists held by different registers leading to a reduction in the number of duplicate or inaccurate records. Finally, changes to the contracts by which GPs are paid now mean that it is less in the GP's interests to have a patient list containing patients who have died or moved away.

Examination of the FHSA data collected by ONS Population and Vital Statistics for the years 1994-6 suggest that there has been a decrease in the numbers of records with a missing or grossly incorrect postcode - from 3.5% of records in 1991 to 1.7% in 1996. There is also some evidence that there has been an improvement in the quality of other data held on the records, and that the number of duplicate records has decreased. The extent to which the data available from NHS sources has improved will also be assessed by matching to data from the 1997 Census Test.

Certain aspects of the data may be better in deprived urban areas where Census enumeration is worst. In such areas, there is generally a greater need for health care resulting in people having a greater propensity to be registered at their correct address.

The greater demand for health care is reflected in the GP payment system, with patients resident in deprived areas attracting a higher payment for the GP. Therefore, more effort is made by inner-city GPs and FHSA's to check that the patient postcode is accurately recorded. However, there may still be a problem with young adults, particularly males, who form a very mobile section of the population and who, being generally healthy, do not bother to register with a GP. Unfortunately, these young adults have historically been the group most likely to be missed by the Census.

Linkage of 1991 and 1997 Census Test data

A trial linkage of 1991 Census records with FHSA patient register data was reported to the Steering Committee in May 1997. This work is being repeated using data from the 1997 Census Test to assess the degree to which register quality has increased since 1991.

The first of three linkage stages will be to match the FHSA and NHSCR records using the NHS number which is common to records from both sources. This first process will provide a composite record containing both name and postcode, as well as demographic details.

The second stage of the process will link these combined NHS records to the Census Test records using date of birth, sex and postcode information. This will use the same matching system developed for the work using 1991 data.

A final matching stage will use the name information on the NHS and Census Test records. A system to do this is being developed within ONS Census Division but will draw on external experience of record linkage using names. In particular, it is important that a linkage system can adequately cope with differences in the spelling or ordering of names. By matching on name, the dependence on the accurate and timely reporting of address will be reduced. This should allow an increased record matching rate. The cost of capturing names from all Census forms in 2001 may be prohibitive, although it may be worthwhile doing this for a sample of forms corresponding to the CCS areas.

Timetable for NHS data experiments

August - October 97:	NHSCR data received for Census Test areas in England
November 97 - February 98:	Receipt of FHSA data for mid-1997 Develop systems to match NHS records using NHS no. Matching of records using DoB, Sex, Postcode
March - April 98:	Development of linkage system using Names. Comparison of linkage rates achieved with / without Names
May 98:	Recommendation on use of NHS records.

Issues and recommendations

Resources should be concentrated on investigating the use of data from NHSCR and FHSAs, including the collection of extracts of data and matching exercises as detailed above. The practicalities of using these data as a third source, together with Census and CCS data, in capture-recapture analyses should be examined. An issue to be considered is whether the quality of the NHS data is now sufficient to add strength to the capture-recapture model used.

It is important that this work is undertaken closely with Population and Vital Statistics who are supplying the FHSA data. P&VSD plan to decide by June 1998 on the potential of using FHSA data to estimate migration. Although there are differences, many of the factors which P&VSD will need to consider will also be of relevance to the ONC decision on the use of these data.

ONS has been made aware of the Coldicott Committee which is preparing a report on the confidentiality and security of individual level records from NHS sources. This could impact upon the ways in which these data could be used in the One Number Census. The implications of the report will be considered when its recommendations are known.

The use of DSS and Council Tax Dwelling List data should also be pursued. The use of the DSS data should be evaluated as part of the on-going work to assess and improve the demographic estimation methods used. The best ways of using the Dwelling List as an independent source of address information should be considered in consultation with Census Geography.

The availability of data on Students, and from Public Utility sources should be examined further, but this is of lower priority.

The availability of administrative data in Scotland and Northern Ireland is being investigated by GRO(Scotland) and the Northern Ireland Statistics and Research Agency. In many cases, similar information will be held, but it is necessary to establish whether different access arrangements need to be made.

On-going research will be needed into likely changes in the availability and quality of data. In particular, advice will be sought from the Department of Health about any long term changes which may be made to the structure of the NHS registers, and the availability of data from them. This risk, that an administrative or technical change may lead to data from a particular source becoming unavailable, is common to all administrative sources.