



ONS(ONC(SG))97/05

ONE NUMBER CENSUS STEERING COMMITTEE

Research into the use of data from administrative sources

1. The attached paper reviews the research into the use of data from administrative sources as part of the One Number Census. Key potential sources have been identified, and the type of information held has been examined.
2. Further work is planned to examine the availability of data to the Census Offices, and to obtain and analyse extracts of data where this is possible. Data have already been received from two sources. Progress made and plans for further work are outlined in Section 4 of the paper. Annex 1 contains summary tables detailing the type of information held on certain key data sources.
3. **The Steering Committee are asked to:**
 - a) **note the paper**
 - b) **provide any comments (at the forthcoming meeting or in writing by 30 June) on the proposed plans for further research.**

**David Thorogood
Census Division
Office for National Statistics**

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

May 1997

Research into the use of data from administrative sources

1. Background

1.1 This paper reviews progress made to investigate the potential use of data from administrative sources as part of the One Number Census project. Current work aims to identify potential data sources and to reach preliminary conclusions about their suitability for use as part of the Census. It is intended to highlight key sources to which access will be pursued. Further work will entail obtaining access to extracts of the data to more fully evaluate their usefulness, and to develop procedures to obtain and use regular supplies of administrative data. In the case of two sources - Family Health Service Authority (FHSA) registers and Council Tax Dwelling Lists - access has already been obtained.

1.2 The use of administrative sources is also being considered as part of the development of a Data Quality Management System for the 2001 Census, and as an additional source of data in the planning of the enumeration process. This work will inform research into the use of these sources as part of the One Number Census.

1.3 It is envisaged that administrative data will be of use at both the individual record level and as aggregate counts. Obtaining individual record data from certain sources may be problematic due to confidentiality concerns or because the administrative system does not support the production of extracts of records. For certain data sources, only aggregate counts may be available.

1.4 Aggregate use may provide an estimate at national or sub-national level of the whole population or of a particular population sub-group. This will be closely linked to the work on demographic analysis at national or sub-national level.

1.5 At individual record level, it may be possible to match records from different administrative sources with Census and Census Coverage Survey records, allowing the use of capture/recapture estimation techniques. This will provide additional information for the production of estimates at county level and below. Experiments with the use of FHSA data in capture/recapture analyses are detailed at Section 4 below.

1.6 A possible additional use of administrative data would be to assist the imputation of missing persons and households to create a corrected micro-level database which would be used for outputs under a full One Number Census. Basic demographic characteristics of persons estimated to have been missed by the enumeration could be copied from an administrative source, possibly providing information which could be used to help impute other characteristics not obtainable from the administrative source. This advanced use of administrative data may present significant problems - technical, data quality, data protection, public perception - and may therefore be unfeasible.

2. Criteria for evaluation

2.1 A series of criteria has been developed to guide investigation into potential sources.

2.2 Type of data held

- Do individuals, households or addresses form the basis of the information?
- Basic identifiers - NHS number etc.
- Are data held for only some sub-groups of the population?
- Geographic detail
- Variables stored
- Is the definition of variables the same as for the Census?
- Are the data stored centrally or in many local databases?
- Are data held for all areas of the country?
- Timing of collection of data. Does it coincide with Census date?
- Data coding

2.3 Data sources can be broadly grouped into two categories: those providing a source of person information; and those relating to households, addresses or dwellings. A particular problem is that definitions used, such as that of households or dwellings, vary from those used in the Census.

2.4 Certain of the sources examined offer information on all or most of the population. Other sources are believed to offer particularly good information on population sub-groups, such as children or the elderly.

2.5 Where data are held on a number of local databases (possibly owned and managed by separate organisations), there can be practical difficulties in finding out about the type and quality of the data held. It is likely that this will also have implications for the ease of arranging access to the data.

2.6 Certain of the data sources examined cover only England and Wales. Similar registers exist in the other parts of the UK, but separate arrangements may be necessary to access these. Where there are differences in the type of data held, systems to make use of the administrative data will need to be designed to take account of this.

2.7 Data quality

- Frequency of updating
- Checks on accuracy
- Ease of removing 'deadwood' - would this already be done by keepers of the register?
- Frequency of missing data
- Are there reasons for individuals to avoid registration, or to fraudulently appear?
- Consistency with other registers in timing and information.

2.8 Certain aspects of data quality can be assessed by contacts with owners of the data source. More detailed quality assessments will require access to extracts of the data to be compared with or linked to other sources, or to be examined for internal consistency.

2.9 Particular problems relate to the frequency of updating the information. Information about a person may be present on a register, but may contain address or other information which is no longer correct.

2.10 Use of administrative data at individual level generally requires the data to be of greater quality than use at aggregate level. At aggregate level, compensating inaccuracies reduce the net errors - eg. a person moving from an area not being de-registered balancing out a person moving to that area but not registering there. In evaluating the suitability of a data source for Census use, it is necessary to consider whether it can be seen as 'fit for purpose'. If data from a particular source are likely to be used (or may only be available) at aggregate level, a higher level of gross error may be acceptable.

2.11 Accessibility

- Legal restrictions - Data Protection Act and other legislation
- Precedents for releasing data
- Are data held electronically and easy to extract?
- What computing systems are used?
- Public perception of linking register with Census data
- Is there likely to be a continuity of supply?

2.12 Current understanding of UK and European law is that the use of administrative data for a statistical purpose such as the Census is allowed. Detailed legal advice will be obtained to confirm this, and to advise on likely legislative developments.

2.13 The issue of public perception will require careful treatment to ensure that the public fears about confidentiality are not to be increased. A review of public beliefs and attitudes on this is planned.

2.14 UK and International precedent

- Is there experience of using this type of register for a similar purpose in the UK or elsewhere?

2.15 The types of data held, and the administrative, legal and social environment controlling the use of administrative data vary greatly between countries. However, international experience of developing the use of administrative sources provides pointers to issues which need to be considered.

3. Sources considered

3.1 The following sources have been considered as likely to be the most suitable for use for Census purposes:

Person level sources

- NHS Central Register (NHSCR)
- Family Health Service Authority (FHSA) registers
- DSS Retirement Pension lists
- DSS Child Benefit lists

Household or dwelling sources

- Council Tax Dwelling Lists
- Council Tax Billing Lists

Details of the information held on these sources is given in Annex 1.

3.2 The following sources have been identified but will require further investigation to look at the type of data held.

- Other DSS data - inc. National Insurance contributions,
- Data on students held by universities and colleges, and by the Higher Education Statistics Agency.
- Public utilities
- Electoral registers

4. Progress and plans for further research

Health Service registers

4.1 1991 Census and FHSA register records for the Hillingdon area of London have been linked using sex, date of birth, and postcode as matching variables. A matching FHSA record was found for around 70% of Census records. A relatively large number of cases were found to have a record only on the Census or the FHSA file. It is probable that these related to people who were recorded by the Census or by the FHSA systems as resident in another area. This high proportion of cases found on one, but not both, of the data sources meant that the estimate of the population produced by the capture/recapture experiment was implausibly high. The delay which can occur in people reporting new address information is believed to be a particular problem with FHSA registers.

4.2 This matching exercise will be repeated using data from the 1997 Census Test. The quality of the data on FHSA registers is known to have improved since 1991, and the extent of this improvement will be examined. It is also planned to link FHSA records with their corresponding NHSCR record. This will provide a combined record containing both name and postcode, as well as demographic details. Names will be captured for the first time in the

1997 Test, and these should allow records to be matched using names as well as other variables. The costs and benefits of capturing names will be evaluated using the 1997 Test. It is probable that the cost of capturing names from all census forms in 2001 would be prohibitive, although it may be worthwhile doing this for a sample of forms corresponding to the CCS areas.

4.3 Health Service data are also being examined as a potential source of aggregate counts at national and sub-national level. The extent to which problems due to list inflation (where persons who have died or moved away are not removed from the registers) can be overcome will be significant in deciding whether these sources can be used in this way.

Council Tax lists

4.4 Arrangements have been made to receive copies of the Dwelling Lists for the areas in the 1997 Census Test. This will serve as a list which is independent to that used by the enumerators. Although the Dwelling List is compiled separately for each local authority district or unitary authority, it is owned by a single agency - the Valuation Office Agency of the Inland Revenue.

4.5 Attempts will be made to negotiate access to the Billing List in an area covered by the 1997 Test. A possibility may be to retrospectively look at the extent to which 1997 Test enumerators correctly identified dwellings which had no usual residents. Unlike the Dwelling List, the Billing List data are owned and maintained by each local authority. It is likely that the difficulties in accessing a data source held by many separate local authorities will mean that, in 2001, the Billing List would be used in only a few sample areas, if at all.

Department of Social Security data

4.6 Arrangements are being made with the DSS Analytical Services Division to obtain extracts of data from the Retirement Pension and Child Benefit systems. These will be based on the samples (currently 5% of Pensioners, 1% of Child Benefit claims) which are already extracted from the administrative systems run by the DSS Benefits Agency for use by the Analytical Services Division. Further information will be collected on data quality and on developments in the type of data extracts available. In parallel, further research is necessary into other DSS data sources such as National Insurance contribution records.

Student data

4.7 Sources of information on the number of students within an area will be examined further. It is believed that more up to date information may be available by approaching the educational institution directly, rather than relying on the Higher Education Statistics Agency which compiles information on behalf of the Department for Education and Employment. However, as with the Council Tax Billing Lists, it may prove to be impractical to access data from many individual bodies.

Public Utilities

4.8 It has been proposed that customer or supply point information held by electricity and water companies may provide a list of dwelling information. Some contact has taken place

with these companies but more is required to assess the potential use of these sources. There are particular problems which may restrict the use of data from utility companies. In particular, there appears to be no precedent for the release of data to external organisations, and this may be prohibited by specific clauses in the utilities privatisation legislation.

Electoral Registers

4.9 There are known to be significant problems of coverage and data quality in the use of Electoral Registers. It is therefore proposed that research into this source should take a low priority.

Scotland and Northern Ireland

4.10 Further research will be undertaken to examine differences in the availability of administrative data in Scotland and Northern Ireland. In many cases, similar information will be held, but it is necessary to establish whether different access arrangements may need to be made.

Annex 1

Initial investigation of potential data sources

Sources of data on individual persons

Data source	Variables held (which may be available to Census Offices)	Population coverage	Advantages	Disadvantages	Comments
NHS Central Register (NHSCR)	<ul style="list-style-type: none"> •Name, •Date of Birth, •Sex, •FHSA of residence, •NHS number, •Date of Registration with current FHSA. 	<ul style="list-style-type: none"> •All persons registered with an NHS GP in England & Wales. •Students may be recorded as resident at their term-time <u>or</u> parental address. 	<ul style="list-style-type: none"> •Data are relatively easy to access - stored on a single system administered by ONS. •Currently used as a migration data source •No disincentive to register. 	<ul style="list-style-type: none"> •Poor geographical detail - equivalent to knowing only the county or borough of residence. •Registration at current address is not compulsory - may be delayed until medical treatment is needed. •Counts are inflated due to non-removal of deaths and emigrants. 	<ul style="list-style-type: none"> •Holds similar information to the FHSA registers - main difference is lack of geographical detail. •Separate but similar registers are in use in Scotland and Northern Ireland.

Data source	Variables held (which may be available to Census Offices)	Population coverage	Advantages	Disadvantages	Comments
Family Health Service Authority (FHSA) registers	<ul style="list-style-type: none"> •Date of Birth •Sex •Postcode •NHS number •Date of Registration with GP 	<ul style="list-style-type: none"> •All persons registered with an NHS GP in England & Wales. •Students may be recorded as resident at their term-time <u>or</u> parental address 	<ul style="list-style-type: none"> •Good geographical detail •No disincentive to register. • Counts are inflated due to non-removal of deaths and emigrants. 	<ul style="list-style-type: none"> •Registration at current address is not compulsory - may be delayed until medical treatment is needed. 	<ul style="list-style-type: none"> •Data are not held centrally, but ONS Population & Vital Statistics Division have developed procedures to obtain regular extracts of data as a potential migration data source, and are reviewing the use of these data. •Scottish Area Health Boards provide a similar function to FHSAs.

Data source	Variables held (which may be available to Census Offices)	Population coverage	Advantages	Disadvantages	Comments
DSS Retirement Pension	<ul style="list-style-type: none"> •Age (calculated from Date of Birth) •Sex •<u>UK</u> country of residence 	<ul style="list-style-type: none"> •Persons eligible for, and claiming, the State Retirement Pension - estimated as 95% of: males aged 65+ females aged 60+ 	<ul style="list-style-type: none"> •Good information on the elderly. •Good incentive to register 	<ul style="list-style-type: none"> •Aggregate figures only are currently available - based on a 5% sample of Pensioned records •Poor geographical detail •Address details can be out-of-date - or may relate to where a claimant's bank account is held. 	<ul style="list-style-type: none"> •The administrative system is controlled by the Benefits Agency which supplies sample data to the DSS Analytical Services Division (ASD) •Data available may be postcoded by end-1997
DSS Child Benefit	<ul style="list-style-type: none"> •Number of families and children by - age of child/ren (calculated from Date of Birth) - size of family •<u>GB</u> country of residence 	<ul style="list-style-type: none"> •Estimated as 96% of families with children. 	<ul style="list-style-type: none"> •Good coverage of mothers and children (information on fathers may be absent or not updated). •Good incentive to register. 	<ul style="list-style-type: none"> •Aggregate figures only available - based on a 1% sample of claims. •Poor geographical detail currently. 	<ul style="list-style-type: none"> •The administrative system is controlled by the Benefits Agency which supplies sample data to the DSS ASD. •Postcode and sex of child may be added as variables by end-1997.

Household or dwelling level sources

Data source	Variables held (which may be available to Census Offices)	Coverage	Advantages	Disadvantages	Comments
Council Tax Dwelling Lists	<ul style="list-style-type: none"> •Address •Postcode •Whether combined residential / commercial property •Date added to register •No person variables are held 	<ul style="list-style-type: none"> •Estimated to be in excess of 99% of dwellings in England and Wales 	<ul style="list-style-type: none"> • Several routes exist by which the registers are updated with information on building, conversion, and demolition. • Flats and bed-sits (which may share a postal address and a single street entrance) are generally listed separately in the Dwelling List. •Lists are owned and administered centrally by the Valuation Office - part of central rather than local government. This means that access need only be negotiated with one organisation. 	<ul style="list-style-type: none"> • No person or household information is available. •Empty properties are included. 	<ul style="list-style-type: none"> • Dwelling List are open to public inspection. Census Offices access to this data source has already been agreed. •Computing developments should allow a single national (England & Wales) file to be output by 1998. •Similar but separate lists are available in Scotland.

Data source	Variables held (which may be available to ONS)	Coverage	Advantages	Disadvantages	Comments
Council Tax Billing Lists	<ul style="list-style-type: none"> • Property reference to link to Dwelling List. • Name of person(s) liable to pay the Tax . • Whether dwelling occupied by a lone occupant • Whether dwelling is empty, a holiday home, a second home, or student accommodation. 	<ul style="list-style-type: none"> • Likely to vary between different authorities - but will be similar to Dwelling List. It is in the Authority's interests to ensure that the list is accurate. 	<ul style="list-style-type: none"> • May help identify dwellings with no usual residents - a particular problem for enumerators. 	<ul style="list-style-type: none"> • Separate Billing Lists are held and used by each local authority. A separate approach may be needed to be made to a large number of data owners. 	<ul style="list-style-type: none"> • The Billing List (unlike the Dwelling List) is not normally available for public inspection. LAs may be more willing to release aggregate figures - eg. number of holiday dwellings by ward or postcode.