

POPULATION AND MIGRATION STATISTICS COMMITTEE (SCOTLAND)

Census Alternatives

Introduction

1. The Alternative Sources branch was set up in October 2004 to take forward the use of non-Census sources in the production of population statistics. Our role is to investigate the availability of other data sources, negotiate access for GROS and evaluate their potential to contribute to the production of population statistics. Although the long term aim of this work is to help supplement and eventually replace traditional census-taking with an alternative system, the shorter term benefits include support for the production of small area population estimates, neighborhood statistics and census activities.

2. In broad terms, our research falls into two main categories, sources and methods.

Sources

3. We are currently focusing on 4 administrative data sources which can provide time series data at datazone level to support ratio estimates work for the SAPE programme. Work has progressed to various degrees, main obstacles being legal barriers to access.

- a. School Census (ScotXed). We have obtained the 2003 extract of the data and compared it with the mid-year population estimates at national and LA level. An illustration of the Scotland level comparison is given at Figure 1. The level of agreement between the two data sources varies across LAs and this needs further investigation. The ScotXed database does not cover comprehensively all school-age children, the most notable omission being independent schools pupils. We are considering ways of using Census to supplement the sub-national comparisons with estimates for independent schools. 2004 data is due shortly and this will enable us to begin to look at change/stability over time at the datazone level.
- b. Electoral Roll. Work here is currently focused on obtaining access for GROS as current regulations preclude our use of the full register of electors. We are looking into options to amend regulations for future years and obtain commercially available copies of the register for previous years.
- c. Child benefit. We are due to receive shortly copies of 2003 datazone level data on child benefit recipients from HMRC. We are also working with SE to commission further work to generate the equivalent extracts for earlier years and expect to have this available later in the year.
- d. Old persons database. This is a database compiled by DWP from benefit records and covers persons aged 65 and over. We are working with SE and DWP to commission the production of datazone level aggregates for 2001 and subsequent years.

4. We also have an interest in surveys, most importantly the SHS, the LFS and the newly planned Continuous Population Survey, in terms of their potential role in a future integrated system of population statistics.

5. To explore the feasibility of substituting individual data items now collected by Census, we have obtained 3 extracts from the DVLA database of registered vehicles, relating to 2000, 2001 and 2002. The postcode information on these extracts has been cleaned and initial comparison with Census counts carried out. Table 1 illustrates the initial high level results for Scotland.

Methods

6. In addition to evaluating individual data sources, we are also responsible for developing knowledge and expertise in GROS about methods and techniques for using data across a number of sources. GROS is involved in a number of studies which involve data linkage across sources and we are aiming to learn lessons from this experience for the development of an integrated system of population statistics. Two of these projects are worth mentioning in this context.
 - a. The Coronary Heart Disease/Census linkage study is carried out by a team led by Prof R. Bhopal of Edinburgh University. The study has implemented an individual level linkage between census and health data in order to allow the analysis of the incidence and outcomes of CHD by ethnicity. 94% of census records were successfully matched to the Community Health Index system. The research team has now completed the analysis and is finalizing the report from the study which is due to be published in the next few months. We have begun to evaluate the variations in match rates across the population in order to understand how well the technique works for different groups.
 - b. The Scottish Longitudinal Study (SLS) is a project modeled on the ONS Longitudinal Study (LS) that has existed in England and Wales for about 30 years. The SLS is led by Prof Paul Boyle of St. Andrew's University and involves the linkage of records from the 1991 and 2001 Census, as well as information on vital events (marriages, births and deaths) occurring in the intervening and subsequent years. The SLS database covers a sample of 5.5% of the Scottish population. It is currently being developed and is due to be launched in early 2006. GROS is working with the research team to provide a system of controlled access to the data for external researchers and promote its use.
7. As part of our work in this area we are also looking to learn from international experience, review current methodologies and tools for integrating data across sources and set up new data linkage experiments. Most notably we are discussing with ONS and DWP the feasibility of conducting a linkage exercise involving DWP, NHSCR and Census records to simulate an alternative Census. This is at a very early stage and there are considerable difficulties still to be overcome.
8. We would welcome views and suggestions from PAMS members for any additional data sources, either national or local, projects or areas of work which we should seek to include in our programme of work for the coming years.

Ganka Mueller
GROS: Alternative Sources
27 May 2005

Annex:

Figure 1.

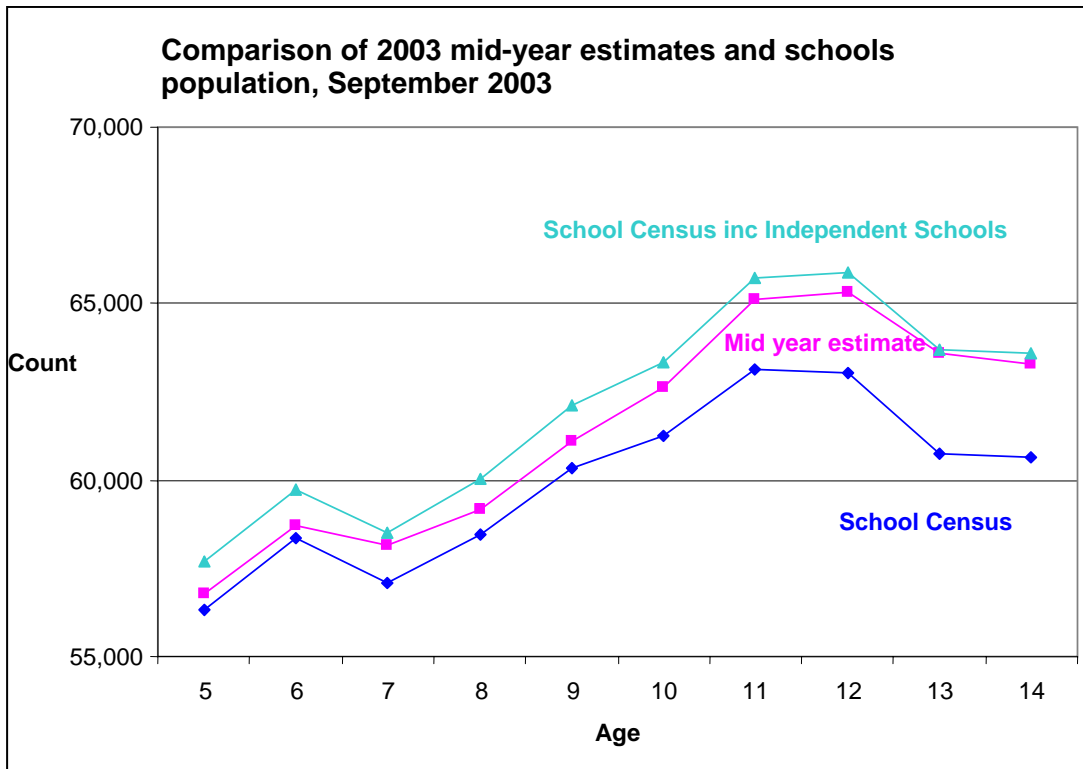


Table 1.

Source	Number of vehicles	With valid postcodes
2000 DVLA	1,835,628	1,820,168
2001 DVLA	1,938,800	1,926,647
2002 DVLA	1,993,390	1,983,678
Census 2001	2,044,018	