

# POPULATION AND MIGRATION STATISTICS COMMITTEE (SCOTLAND)

## Improving the GROS migration data for Council Areas – proposals for 2002

### Introduction

This short paper and the attached supporting information build on the work described in previous CHI and SAPE (Small Area Population Estimates) working group papers, which investigated the Community Health Index (CHI) as a source of improved Council level migration estimates. As a result of this work, a methodology was developed to improve the mid 2000 – mid 2001 migration data, and this method was implemented in the production of GROS' provisionally revised<sup>1</sup> 2000 mid year estimate (MYE). We now propose to use this method in the production of the 2002 MYE in Spring 2003.

The recommendations put forward in this paper have been agreed by the SAPE working group, and PAMS is now asked to give their endorsement of the 2002 migration methodology for council areas.

The results of the 2001 Census have brought to light certain deficiencies in the estimation of migration, particularly out-migration, and work is required to improve the overall Scotland-level and Health Board-level migration. However, this paper is only concerned with improvements to migration estimation for councils.

### Proposed Methodology – A summary

The approach used for estimating council migration involves matching currently registered CHI patient records from extracts taken at two different points in time, i.e. a year apart for the mid-year estimates. Patient records with matching CHI numbers and where the patient's postcode of address differs between the two extracts are considered to be migrants.

This matching of two extracts, say A and B, will create three sets of patients:

- A set of patients in extract A and extract B
- A set of patients in extract A but not in extract B
- A set of patients in extract B but not in extract A.

The set of patients in both extracts whose postcodes were not the same in each extract can be considered as within Scotland migrants, i.e. moved from one address in Scotland to another within Scotland between the dates of the two extracts. The remaining two sets of patients are either migrants into or out of Scotland; babies born between the two extracts; deaths between the dates of the two extracts; movements to/from the Armed Forces; and a small number of records that are for the same patient but have different CHI numbers because they could not be matched when the GP registration was processed.

A very basic version of this methodology was used in the 2000 Mid Year Estimates to estimate within-Scotland migration at local authority level. This method has now been refined and was tested using data for 2001. We propose to use the method summarised below to produce migration estimates for councils for 2002. The proposed method is also outlined graphically in Appendix A (PAMS (03) 10 - supplement, p16).

---

<sup>1</sup> Rolled back from 2001 mid year estimate

### **Step 1 – validate postcodes and match CHI extracts (PAMS (03) 10 - supplement, page 1)**

We start with two CHI extracts taken at end September 2001 and end September 2002. These are used to calculate migration between mid 2001 and mid 2002 because we assume a three month time lag between moving house and re-registering with a GP in the new area.. Each extract contains around 9 million records, which constitute the majority of the records ever placed on the CHI. These records must therefore be filtered to give an estimate of the number of people registered on the CHI in each of the two years. The filters that are applied are outlined in the supporting information on page 1. The postcodes of the CHI records on each extract are then validated against the GROS postcode index. Mapping both extracts to the same version of the postcode index avoids creating false migrants due to postcodes being introduced or deleted between the dates of the two extracts. Finally, the two CHI ‘currently registered’ populations are matched by encrypted CHI number.

### **Step 2 – identify potential migrants (PAMS (03) 10 - supplement, page 2)**

After matching the 2 CHI extracts we are left with three types of record.

1. Records present in 2001 extract, but not in 2002 extract. If any exist, they can be discarded.
2. Records present in 2002 extract, but not in 2001 extract. This contains babies born between July 2001 and June 2002, and moves onto the CHI from outside Scotland, or from the Armed Forces.
3. Records present in both extracts. This contains all within-Scotland moves of people aged over 0, including people who die in the mid01 – mid02 period. It also contains people who transfer off the CHI, ie. moves out of Scotland or to the Armed Forces. This is because, although the person is no longer registered on the CHI, his record is still present on the extract, but with a ‘transferred out’ date.

### **Step 3 – impute missing postcodes and area codes (PAMS (03) 10 - supplement, page 10)**

Missing postcodes are imputed in 5 stages, using a donor imputation process. Donors are sought from the set of complete records that have as many characteristics in common with the recipient records as possible. Because the imputation is carried out on matched extracts, we can use the origin and destination information to improve upon the imputation methodology used in the production of SAPEs. For example, if a record is missing an origin postcode, we look for a donor with the same CHI health board (HB) of origin indicator, HB of destination, GP practice of origin and destination, age group, sex, council, and destination postcode, and randomly choose a postcode from the set of possible donors. If a donor is not found, the matching criteria are gradually relaxed until a donor is found.

### **Step 4 – control CHI migration to HB/NHSCR migration (PAMS (03) 10 - supplement, page 12)**

It is recognised that the GROS migration data derived from the NHSCR is the most reliable data available at health board level, so estimates from the CHI should be controlled to agree with the NHSCR data for moves across a health board boundary.

CHI migration records for moves between health boards, moves to and from Scotland, and moves to and from the Armed Forces are controlled to agree with the NHSCR-derived HB migration data by randomly duplicating or deleting CHI records until the desired total is achieved. This method controls the CHI within-Scotland moves between HBs to the NHSCR within-Scotland moves. CHI moves between Scottish HBs and outside Scotland are controlled to a combined NHSCR-derived dataset as a group – no distinction is made between cross-border moves and international moves, or moves to and from the Armed Forces.

The controlling process produces a CHI dataset that is fully consistent with NHSCR-derived migration at HB level, by origin, destination, age and sex. No controlling has been carried out on moves within a health board, so in the mid 2002 migration data, the CHI will be used directly to estimate moves between building bricks within health boards. The building bricks used in the mid 2001 data are the 36 council part-areas, i.e. councils split by health board.

## **Changes agreed by SAPE Working Group**

The SAPE Working Group agreed the following changes to the basic CHI-based migration methodology as used in 2000. The PAMS group is now asked to agree that these changes be implemented in the production of the council migration data for use in the 2002 Mid Year Estimates:

### **1. Postcode validation**

Previously, postcodes on each CHI extract were mapped to postcodes on the most recent version of the GROS postcode index.. This would mean that the September 2001 extract is validated against the 2001/v2 index, and the September 2002 extract is validated against the 2002/v2 index. This could give rise to false migrant records if there have been changes to the index between years. For 2002, we propose to validate the postcodes on both CHI extracts against the 2002/v2 index.

### **2. Births and deaths**

In the 2000 MYE, the CHI was used in an ad hoc way to estimate within-Scotland, building brick level migration. However, the CHI was not used to estimate the migration of zero year olds or of people who died before the second extract. In the 2002 data we propose to estimate migration of people who were born between the dates of the two extracts by matching CHI data against Vital Events births data, and estimates of people who died between the two extracts will be obtained simply by not excluding the records of such people from the CHI 'currently registered' population.

### **3. Imputation of missing postcodes**

In the estimation of 2000 migration, any CHI records with a missing or invalid postcode were discarded. However, a method has been developed to impute postcodes on the set of matched CHI records. Because each record contains both origin and destination information, there is more information available on which to base the search for a suitable donor record, and so this imputation method is an improvement on the method used in the production of SAPEs. For this reason, the postcodes imputed by the migration data processing should be fed back into the CHI data for use in the production of SAPEs.

### **4. Migration outside Scotland**

Using data from the NHSCR and ONS, we derive health board flows by age and sex for all moves - both within and outwith Scotland. In 2000, CHI flows for moves within Scotland only were used to apportion this health board data to building bricks, using an age/sex distribution from the 1991 Census. However, by using information on transfers on and off the CHI we can produce estimates of migration into and out of Scotland at Council level. For 2002, flows at health board level will remain unchanged, but CHI data for all moves, not just those within Scotland, will be used to apportion the health board flows to building bricks. Specifically:

- From 2002, all transfers onto the CHI, except those from another CHI consortium in Scotland will be counted as moves into Scotland from outside Scotland.
- Further work should be undertaken to investigate the reasons for the observed discrepancies between the CHI and NHSCR data, particularly in Fife and Forth Valley Health Boards, with a view to producing separate migration flows for moves to/from the rest of the UK and moves to/from overseas.
- From 2002, only those transfers off the CHI with a destination area code that represents a move outside Scotland will be counted as migration moves.

## **5. Age/Sex distribution**

In 2000, data from the 1991 Census was used to apply an age/sex distribution to the building brick migration flows. From 2002, the CHI age/sex distribution will be used directly, replacing the 1991 Census distribution.

## **6. Controlling between-HB moves**

In 2000, NHSCR-based migration totals for Health Boards were apportioned to building bricks using CHI totals, and age/sex details were calculated separately. From 2002, we propose to use a method of controlling the CHI migration dataset to the NHSCR migration dataset that randomly duplicates or deletes CHI records so that the CHI dataset becomes consistent with the NHSCR data. The method is described in detail in the Supporting Information paper, page 13.

## **7. Further work**

Further work is required to improve the quality of the data on transfers on and off the CHI, in order to be able to produce migration data for moves between Scotland and 'rest of UK' and Scotland and 'overseas'. For 2002, we do not propose to split the flow of migrants to and from Scotland into these components at council level, only at Health Board level.

Further work is also required to validate a method used by ONS to control within-HB moves. The method is based on the assumption that the scaling required to control the CHI data to the NHSCR data for moves between HBs will be the same as for moves within HBs. 2001 Census migration data will be useful for evaluating this method, with a view to implementing it in the 2003 migration data.