

## NHS Central Register (NHSCR) Governance Board Meeting 24th October 2013

### Minority Ethnic Names in the Linkage of the NHS Central Register Extract to the 2011 Census

#### 1. Purpose

In 2012 a major project was undertaken to link records in the NHS Central Register (NHSCR) extract to those in the 2011 Census. The NHSCR extract file which was used in the procedure contained 5.67 million records of people who, according to their extract data, were alive and registered to a Scottish health board on Census Day (27 March) 2011. This note reports the results of a specific piece of work in which these records were partitioned into two sets – those which had been linked to a census record with a probability of over 0.50 and those which had been linked with a probability of 0.50 or less or had not been linked at all. These two sets were then examined to investigate the relative frequency within them of names characteristic of people of Minority Ethnic (ME) origin.

#### 2. Method

Defining ME is inevitably an imprecise matter. Any definition will incorrectly include people who are not in the target group and will also incorrectly exclude people who are in the target group. For present purposes, it is more important that the group identified should be 'pure' (i.e. include few records not in the target group but still be large enough to be representative of the target group) than that it should be exhaustive (i.e. exclude few records which are in the target group). This is because in Scotland, all minorities form very small proportions of the total population so the incorrectly excluded records will have little effect on the baseline population not included. A 'loose' definition however will pull in enough records from outside the target group to affect materially the results of the calculation. For example, the appearance of the letter Z will include many Polish names but will also pull in MacKenzie, Menzies, Dalziel, Frizell and other majority ethnic names. For this reason, 'tight' definitions were adopted.

A second methodological problem is the identification of the ME groups to use in the research. This also is arbitrary but, for the purposes of the present research, the main thing is that each group selected should be sufficiently large to be identifiable using a tight definition but that the groups taken together should be sufficiently diverse to be representative of minority ethnicity in general. For the present work, five groups were used. These were; Chinese, Mahomed, Other South Asian, Sikh and East European.

To derive definitions for each of these groups, the 940,000 NHSCR extract records which had not been linked to a census record at all were isolated and frequency distributions were compiled of all first names and last names occurring in these records at least 300 times. These distributions were then scrutinised manually to identify names which appeared to be associated with an ME group. The criterion used for each group is given below.

**Chinese:**

Either first name or last name is CHAN, CHEN, HO, LIU, TAN, WANG, LI, WONG, ZHANG, LIN, LIM or WU.

**Mahomed:**

The definition here was slightly different as this group is a name in itself. There is a multiplicity of ways of spelling it, most of which involve vowel variations. Therefore the procedure adopted was to remove all the vowels and to include the record if the compressed version of either the first or last name was MHMD, MHMMD, MHMT or HMD. This would pick up most of the variations of Mahomed and also related names such as Mehmet, Ahmed and Mahmood.

**Other South Asian:**

Either first name is ABDUL or SYED or last name is HUSSAIN, KHAN, ALI, KUMAR, AKHTAR, IQBAL, HASSAN, BEGUM, BIBI, SHAH or SHARMA.

**Sikh:**

First name, middle name or last name is either SINGH or KAUR.

**East European:**

First name is KATARZYNA, PIOTR, MONIKA, MAGDALENA, TOMASZ, AGNIESZKA, PAWEL, LUKASZ, KRZYSZTOF or MAREK.

**3. Results**

For each of the five groups, the 5.67 million NHSCR records were partitioned twice. The first partition was into those in the ME group and those outside it; the second was into those with a link probability greater than 0.50 and those without. The resulting 2x2 contingency table was then used to calculate the probabilities that a record in the group, and a record outside the group, will have a link status (i.e. will have been linked to a census record with a confidence of at least 50%).

The probability for records outside the group is almost the same for all groups since the removal of small ME groups has little effect on the remaining majority. It is about 0.841. The interest lies in the probability of link status for the various ME groups. The estimates of these probabilities, and their standard errors, are given in table 1.

ME group	Number of records	Estimate	Standard error
Chinese	14,830	0.537	0.004
Mahomed	23,969	0.534	0.003
Other South Asian	19,481	0.605	0.003
Sikh	7,011	0.564	0.006
East European	12,497	0.681	0.004

**4. Discussion**

It can readily be seen that for all five of the ME groups used, the probability that a record which falls within the definition will have link status is much smaller than that for the majority population. The figures for the four Asian groups are comparable, all

falling between 0.53 and 0.61. That for the Eastern European group is larger though still well short of the value of 0.84 for the majority group.

The following grounds can be identified to explain why a record in the NHSCR extract cannot be linked to one in the census.

- (i) The person was not in Scotland on Census Day and so was not included on any return.
- (ii) The person was in Scotland on Census Day but nevertheless was not included on any return.
- (iii) The person was in Scotland on Census Day and was included on a return but was not linked due to record linkage difficulties which are more likely to occur for ME names than they are for others.

Without further information it is not possible to attribute the patterns in table 1 to these three possible causes. The fact that the Eastern European group lies between the Asian group and the British majority suggests that there may be a cultural effect, though the word 'cultural' may have many different interpretations. These range from the political (some immigrants come to Scotland from countries where registration with the authorities is not always viewed in a wholly positive light) to the alphabetical (European names, however complicated, were originally written using the Roman alphabet while Asian names are transcriptions from a quite different original script). More detailed information would be required in order to distinguish between these explanations but they are of some intrinsic interest in the light they throw on this aspect of the difference between linked and non-linked NHSCR extract records.