Scotland mid-1991 and mid-2001 population estimates: age, sex and ethnic group. Documentation
"Scotland Council Areas population estimates with ethnic group Documentation Dec 07.doc"

Ludi Simpson, CCSR (Cathie Marsh Centre for Census and Survey Research, University of Manchester) ludi.simpson@manchester.ac.uk

Background

1. This document (a) describes the status of the population estimates (b) comments on the relationship between census and population estimates, and (c) describes the methods used to derive the population estimates.

2. These estimates have been prepared for a research project funded by the Leverhulme Trust, *Race, Migration and Population Dynamics* (http://www.ccsr.ac.uk/research/mrpd). The work has been carried out as a preliminary to estimating the contributions of births, deaths and migration to population change during the decade 1991-2001. The population estimates complement equivalent estimates for England and Wales prepared by Albert Sabater[1]. For England and Wales the estimates include sub-District geography of Census Output Areas, while in Scotland the estimates are limited to Council Areas, since at present the census data for smaller areas have not been released.

3. The population estimates from CCSR with an ethnic group dimension are consistent with the latest population estimates from GRO(S) for 1991 and for 2001, without an ethnic group dimension. The full set of estimates includes single year of age, sex, and the ethnic group classification used in 1991 and 2001, for each of the 32 Council Areas defined in the 2001 Census output.

Status of the population estimates

4. The estimates have been prepared using the methods described below, for which code can be supplied on request. The methods and estimates have been seen by GRO(S) census and demography sections, and local authority staff in Glasgow and Falkirk. This consultation exercise during 2007 did not identify any improvements for the methods, but the responsibility for the data remains with CCSR. The population estimates are intended to be of general use and form a time series to which further population estimates with ethnic group dimension can be added.

[1] Sabater, A and Simpson, L (2007) Enhancing the population census: a time series for sub-national areas with age, sex and ethnic group dimensions in England and Wales, 1991-2001, CCSR Working Paper 2007-11, University of Manchester.

Commentary on the relationship between census and population estimates

5. These CCSR population estimates are consistent with the mid-year population estimates published by GRO(S), and therefore include allowances for census non-response not estimated within Census procedures in 1991, and refer to mid-year rather than census day. As in GRO(S) mid-year estimates, students are included at their term-time address rather than their vacation address as in the 1991 Census.

6. For these reasons, the population estimates are different from the equivalent census output, particularly in 1991. For example, the population estimates for Glasgow and Edinburgh are approximately 20,000 higher than 1991 Census counts, due both to a net gain of students during term-time and to non-response not already estimated in the census outputs. On the other hand, the 1991 population estimate for the Scottish Borders and Highland is slightly less than the census count because a net loss of students studying elsewhere during term-time outweighs the impact of non-response and changes between census day and mid-year. For 2001, students were counted at their term time address already in the census, and GRO(S) considered that all non-response had been estimated already within the adjustments made before census output was released. The differences between census and population estimates in 2001 therefore reflect only births, deaths and migration in the ten weeks between census and June 30th 2001; they are always relatively small and are sometime positive and sometimes negative.

7. The change in population between 1991 and 2001 can be different from that shown by census counts, because of the adjustments described in the previous paragraph. The comparison of census and population is shown for the population totals of each Council Area on sheet "Summary", and is taken directly from GRO(S) output. The population for each ethnic group and its change between 1991 and 2001, given in sheets "1991pop" and "2001pop" and summarised in 'Summary eth' is estimated by CCSR and is consistent with the GRO(S) totals across ethnic group, for each age-sex group.

Scotland total resident population: Census and population estimate: 1991, 2001 and decadal change

	1991	2001	1991-2001
Census output	4,998,567	5,062,011	+63,444
Population mid-year	5,083,330	5,064,200	-19,130
Enhancement (pop-cen)	+84,763	+2,189	-82,574

Method summary

8. In summary for 2001, the difference between GRO(S) population estimates and census output was relatively small, and was distributed to ethnic groups in proportion to their population size.

9. In summary for 1991, the difference between GRO(S) population estimates and census output was larger and where the population estimate was greater than the

census count, a differential for each ethnic group was taken from the work of the *Estimating with Confidence* project, reflecting lower non-response for the White group than for other groups.

10. In both 1991 and 2001, population estimates when summed across the ethnic groups are consistent with the GRO(S) population estimates for each single year of age, and sex. The assumptions made in creating these detailed estimates are sometimes crude, but no further data are known to be available with which to improve them.

Method detail

11. Notation:

P: Population estimate from or consistent with most recent GRO(S) mid-year estimates
P*: GRO(S) population estimate for 1991 before revision
C: Census count
d01: 2001 Districts: 32 Council areas.
S: Scotland
a18: 18 quinary age groups to 80-84 and 85+.
a20: As a20, with 15, 16-17 and 18-19 separately.
a91: single year of age (syoa) 0, 1, …, 89, 90+.
e: ethnic group (ten categories in 1991, 14 in 2001)
s: sex

If a suffix is missing, its total is implied. For example, if e is missing, the dataset is not differentiated by ethnic group.

12. Datasets:

Specially commissioned census tables provided by GRO(S): sex-ethnic group with 20 or 91 age categories for Districts or Scotland respectively:

C91d01a20se	Districts as at 2001.
C91Sa91se	Scotland.
C01d01a20se	Districts as at 2001.
C01Sa91se	Scotland.

ONC data
Imputation rate for persons, by eth, from 2001 Census ONC: Scotland: from Alan Fleming in pdf file. Not directly used.

Rebased population estimates from GRO(S)

P91d01a91s	1991 revised single year of age for 2001 Council areas. "Scotlandrebased-mye-council-1991-2000.xls"
P01d01a91s	2001 single year of age for 2001 Council areas. "scotlandcae01a.xls" dated 24Oct03 Crown Copyright 2002.

EWCPOP-SOCPOP population estimates

P*91p/w91a18s	1991 EwC unrevised population estimates for postal sectors and wards. Scotplp.sys, scotwlp.sys (not used)
P*91d91a18se	1991 EwC SOCPOP population estimates (census day, students at home) scotde.csv These incorporate an estimate of non-response which is consistent with that estimated by GRO(S) in 1993-4, and an ethnic group differential consistent with research in the 1990s (which gives a greater-than-average rate of non-response to all non-White groups, particularly Black groups, as described in 'The Census Data System', Wiley, 2002, edited by Phil Rees, David Martin, Paul Williamson).

Methods

13. Preliminary checks: compute (P91-C91)d01a20s and (P01-C01)d01a20s and explore.

Implemented in 'Check P-C no eth.sps' and 'Check P-C no eth.spo'.

- Usually positive in 1991 (unless students studying away in 1991or move to mid-year acts to reduce census more than non-response adds to it – Highlands and Islands for example).
- Age and area pattern shows young-adult and urban gradients in 1991.
- In 2001 there was no adjustment from the ONC-based population estimate. So P01-C01 reflects only ageing, births, deaths, and migration.

14. Mid-1991 population estimates

Step 1. SOCPOP and census91, for d91a18se: explore their difference and decide how to represent non-response differentials for 1991.

The SOCPOP for Scotland appears to have age errors (probably due to the 'timing' adjustment being separate from ageing, unlike in OPCS files: 0-4 is negative while older ages are all positive).

[Census and SOCPOP by age and sex are compared using 'Create c91d91a20se.sps' and 'Create SOCPOP91d91a18se.sps', then compared to derive non-response rates on 'SOCPOP-C91d91a18se.xls'.]

For this reason the SOCPOP estimates are not used for each age and sex. Instead the SOCPOP differentials between ethnic groups' non-response are used in Step 2.

These differentials are as follows, estimated from 1991 Census CVS and other sources. For any district-age-sex sub-group of the population, the ratios (census+missing)/census are:

White: 1+x

Black groups: 1+6x

S Asian groups: 1+4x

Other groups: 1+3x

The value of x varies, so that the total non-response is that in the statistical agencies' estimate of non-response for each district-age-sex sub-group (Simpson and Middleton, 1999).

Step 2. Distribute the Population-Census discrepancy to ethnic groups, at each sex-age-district, using ethnic group non-response differentials used in SOCPOP as above.

Apply SOCPOP differential rates to the 1991 census counts for districts at each age and sex and ethnic group, c91d01a20se, by applying the SOCPOP differential rates (as above: 1:6:4:3 for White:Black:Asian:Other). Compute the ethnic group distribution of this initial non-response, separately for each age-sex-district combination.

Attach GRO(S) 1991 population estimates without ethnic group dimension, for age-sex-district groups and compute the Pop-Cen discrepancy with no ethnic dimension.

(a) When the Population estimate is more than Census, as occurs whenever non-response is significant, then distribute the Pop-Cen discrepancy according to the non-response distribution estimated from SOCPOP

(b) When the Population is less than Census, as occurs when students leave the area to study elsewhere, or migration is away from the area and more than non-response, then spread the Pop-Cen discrepancy to ethnic groups in proportion to their population.

[Implemented with p91d01a20se.sps]

Exploration of the final Pop-Cen results for District-age-sex-ethnic groups shows a plausible pattern:

i. The greatest absolute adjustments to the census are to the White population.

ii. Population less than Census. All such negative adjustments of 3 or more people are to the White population, mostly at ages 18-19 and in the least urban areas, the most extreme ones to ages 18-19 in the Highlands (losses of about 350 to each of males and females). These reflect mainly student losses to term-time addresses elsewhere.

iii. Population more than Census. All positive adjustments above 500 are to the White population, mostly at ages 20-29 and in the most urban areas, the most extreme to men aged 20-24 in Glasgow and Edinburgh (gains of about 4,000, around 20%). These reflect non-response, as well as some student gain). The largest gain to a minority ethnic group is to Pakistani men aged 20-24 in Glasgow city which adds 424 or 80% to the census count of 532.

Step 3. Single year of age (syoa).

Iterative Proportional Fitting to add syoa within each District-sex-quinary-age-ethnic group category. The pattern is taken from the 1991 census tabulation for Scotland.
Margins: P91d01a20se, P91d01a91s
Initial values: C91SCa91se

The fitting maintains consistency with the two margins (population estimates for Districts' age groups with an ethnic group dimension, and for District's single year of age without an ethnic group dimension), but fits the detailed internal pattern on the initial values (the census table for Scotland as a whole of ethnic group for single year of age).

15. Mid-2001 population estimates

Step 1. Give to District census tabulations (C01d01a20se) syoa from national pattern (C01SCa91se), to derive C01d01a91se.

Therefore the same single age pattern within an a20 age*sex*eth group, for every District. Could do something that may be slightly more locally accurate by smoothing the five-year age groups separately for each local District then constraining to national, but not thought worth the effort.

Step 2. Population from census: same proportional adjustment to each ethnic group. P01d01a91s as constraint for C01d01a91se.

Could only do better with estimates of births, deaths and migration for each group, but these not available. Likely to slightly underestimate non-White groups whose growth from both immigration and natural change will be greater than the White groups, in the 9 weeks between census and mid-year.

[Implemented in 'P01Sd01a91se.sps' and 'P01Sd01a91se.spo'.]