

One Number Census Evaluation, Scotland, 2001

Project Objective

1. The One Number Census (ONC) provided a method to identify and adjust for the households and people missed by the 2001 Census.
2. The Census Coverage Survey (CCS) covered about 40,000 dwellings. It provided the data to make the ONC adjustment. It measured the rate of under enumeration among households and people and characterised people missed by the Census. Sample stratification was based on 1991 Census data.
3. The final One Number Census (ONC) results provided a new 2001 base for the General Register Office for Scotland's (GROS) population estimates for the following decade. This report evaluates the methodology, processes and operations. The results of the ONC were also used to impute records for missed households and people for inclusion in the results from the Census. The effects on any particular Census variable are documented at <http://www.gro-scotland.gov.uk/census/censushm/censcr02/data-quality/census-variables/index.html>
4. The ONC project was implemented across the UK with some design changes in Scotland to reflect Scottish differences. This report evaluates the success of the process in Scotland only. The project was largely designed by ONS and Southampton University. This report should therefore be read in conjunction with the ONS report¹

Background

5. Every effort is made to ensure everyone is counted in a Census. However, no Census is perfect and some people are missed. This under enumeration is not usually uniform across all geographical areas or sub-groups of the population. Non-response to censuses and surveys is increasing in much of the developed world so assessing the scale and type of under enumeration is increasingly important². Therefore, it is accepted best practice to assess the extent of under enumeration, in the UK in 2001 with a post-enumeration survey.

¹ http://www.statistics.gov.uk/census2001/onc_es.asp

² <http://www.unece.org/stats/documents/ces/2003/28.e.pdf> (para 20)

6. In the 1991 Census in Great Britain, 3.8% of the population was missed overall (2.2% were completely missed (see next paragraph) and 1.5% were missed but imputed³).
7. In 1991, households absent on Census night were invited to return a Census form voluntarily. Where they did not, records were imputed. This process also imputed records for present households identified by enumerators as refusing to complete a form. It was found that under enumeration was greater in certain types of areas and age-sex groups. However, the 1991 Census post-enumeration survey, the Census Validation Survey (CVS) focused on both the accuracy of responses to specific questions and coverage; and it did not succeed in identifying the full extent and distribution of under enumeration.
8. Adjustments for under enumeration for the 1991 Census took time to resolve and four different resident population counts were published (for Census day). Initially 'raw' Census counts were published, followed by provisional estimates (for 30 June), which raised the Census count in line with the CVS. Subsequently a decision was made in Scotland to base the national population estimates on demographic estimates, rather than Census counts, and to produce consistent counts at Local Authority (LA) level using a mathematical model. This third set of counts was then subject to a final revision following further analysis.
9. At geographic levels below local authority, no official estimates of under enumeration were produced, although those produced by the Economic and Social Research Council funded project 'Estimating with Confidence' are widely used⁴. Census tabulations therefore were inconsistent with population estimates for reasons other than different reference dates.
10. During the consultation prior to the 2001 Census, Census users were clear that they wanted one definitive set of Census figures consistent with population estimates. The ONC project was designed to estimate under enumeration and adjust Census responses appropriately so that all published Census statistics would add to the national estimate of the population.
11. It was a further step from methods in previous Censuses:
 - 1981: The introduction of automatic editing and imputation of missing variables;

³ Undercoverage in Great Britain. 1991 census User Guide No 85, OPCS and GROS, January 1994

⁴ 'Estimating with Confidence (Office of Population Censuses and Surveys (OPCS) 1993, 1994; Simpson, Cossey and Diamond, 1997)

- 1991: The absent household enumeration plus imputation in the 1991 Census (imputing records for households we knew existed from the Census operation);
- 2001: The addition of individual records for imputed households and their imputed people, and imputed people into Census households.

Methodology

12. The key stages or processes used to achieve the aims are summarised below. Full details of the development of the methodology can be found on the ONS website⁵.
13. Underpinning the ONC methodology was the assumption of independence between the count of population given by the Census and CCS. For the ONC to be accurate, there must be no systematic relationship between the chance of a household or individual being enumerated in the Census and of being captured by the CCS.

Key stages of One Number Census methodology

14. Scotland was divided into 8 areas, each with an average population of about 625,000 people. These areas are known as Estimation Areas (EAs) (or, synonymously, Design Groups) and are made up of whole Health Board Areas - into which most Council Areas nest.

Census Coverage Survey (CCS)

15. A clustered, stratified sample of postcodes was drawn from each EA. CCS interviewers visited and tried to interview every household in the sampled postcodes. Under enumeration has historically been found to be distributed unevenly. Therefore the sample was stratified according to a Hard to Count (HtC) index and the areas thought to be hardest to count were over sampled. The HtC index was constructed from 1991 Census variables thought to be associated with under enumeration.
16. The CCS began three and a half weeks after Census Day. To maintain independence, CCS interviewers were not provided with the Census address lists for their areas. Maps of the CCS postcodes were supplied to interviewers and they attempted to interview every household at

⁵ <http://www.statistics.gov.uk/census2001/ONC%20Steering%20Committee>

every address within the postcode, in effect re-enumerating the area. Comprehensive training was provided for the CCS interviewers.

17. The ONC estimation rests on the assumptions that people were equally likely to be found by the Census and the CCS, and that the Census and the CCS were methodologically independent. Considerable efforts were made to ensure that the Census and CCS were independent:

- The postcode sample was confidential;
- Census Managers were not able to work on the CCS;
- CCS managers were not allowed to work on the Census; and
- Census Enumerators could not be CCS interviewers in their Enumeration Districts. They did not know their CCS postcodes until Census fieldwork finished.

Matching the CCS and the Census

18. The CCS records were matched with those from the Census using automated and clerical matching. The matching process had to be as accurate as possible because the number of mismatches has a direct impact on the final estimates. The five key stages of the matching process for each unique postcode were: Exact Matching; Probability Matching; Clerical Resolution; Clerical Matching and Quality Assurance (including a double matching strategy). Further detail of the matching algorithms can be found in the ONC Steering Committee paper⁶.

Estimation of populations for each EA and LAD

19. Populations for each EA, by age and sex, were estimated using a combination of standard statistical techniques. Estimates of the total population in postcodes covered by the CCS were based on a methodology known as Dual System Estimation (DSE).

20. It was inevitable that some households and people would be missed by both the Census and CCS. DSE provides an estimate of the number missed by both, using the known numbers of people counted in both the Census and the CCS, those counted only by the Census and those counted only by the CCS.

21. Having established the relationship between the Census counts and the dual system estimates in the postcodes covered by the CCS, it is

⁶ <http://www.statistics.gov.uk/census2001/pdfs/sc9814.pdf>

then possible to estimate EA under enumeration. These estimates were made for each age-sex group in each HtC group⁷.

22. DSE requires a number of conditions to be met to ensure that error in the estimates is minimised. In particular, for an unbiased estimate, independence is required between the Census and CCS. As described above, the Census and CCS were made operationally independent.
23. Simulation work was undertaken to examine the impact if there did prove to be some 'dependence'; these simulations demonstrated that, even for quite extreme levels of dependence, the impact on the population estimates is small provided both have high response rates⁸.
24. Small area estimation techniques were then used to estimate Council Area populations by age and sex. Many EAs consist of more than one Council Area. With the exception of a few large Council Areas, Council Areas did not contain sufficient CCS postcodes to enable accurate direct estimates of under enumeration to be made (during the development stage it had been decided that it was not practically feasible to undertake a CCS of the massive size that would be required to provide direct estimates at Council Area level) (see foot note 6).

Imputation of records for households and individuals

25. Records for households and individuals estimated to have been missed by the Census were imputed to produce a complete set of responses for each household and person, whether directly responding or imputed by the adjustment for under enumeration.
26. The ONC imputation process had three main stages:
 - Imputation of missed households (and the individuals within them); imputed households were geographically placed into either an occupied dwelling identified by Census Enumerators from which no Census response had been received, or into areas where similar households already existed;
 - Imputation of missed individuals, i.e. individuals missed from households counted by the Census. Donor individuals were then selected and used to impute individuals into the types of households that were likely to have missed people from their

⁷ <http://www.statistics.gov.uk/census2001/pdfs/sc0003A.pdf>
<http://www.statistics.gov.uk/census2001/pdfs/sc0003b.pdf>

⁸ <http://www.statistics.gov.uk/census2001/pdfs/SC0103.pdf>

Census return. This process added people to real households; and

- Calibration to estimates of the population. This ensured that the overall distribution of imputed individuals and households was the same as the ONC estimates of households and individuals missed by the 2001 Census.

27. The result was an individual level database that represented the best estimate of what would have been collected had the 2001 Census not been subject to under enumeration. Tabulations derived from this database automatically include compensation for under enumeration measured by the CCS, for all variables and all levels of geography⁹.

EA/Council Area Quality assurance

28. Our quality assurance process judged whether the overall ONC population estimates were coherent, consistent with other relevant sources, and of the right overall size. This involved a series of quality checks for each EA, aided by data, grouped by age, sex and geography, drawn from the annual mid-year population estimates and aggregate level administrative data.

29. The administrative data were NHSCR patient counts, mid-year estimates, school and student populations.

30. The various data sources were used to calculate a range of plausible values for the number of people of each sex within five-year age groups in each geographical area. The ONC population estimates, with confidence intervals, were compared with these 'diagnostic ranges'.

31. A range of descriptive information also gave a fuller picture of the area under consideration, for example information about the fieldwork for the Census and the CCS and feedback received from Councils commenting on past mid-year population estimates from GROS. Demographic ratios (of males to females) were also calculated. All this information was presented to the Quality Assurance Panel At GROS. It considered the evidence for each EA and Council Area before either accepting or rejecting estimates.

32. There was a predetermined contingency strategy in the event of the ONC estimate for any Council Area being rejected. Information from similar Council Areas which had already passed the QA process was used to make adjustments. This process was called 'borrowing

⁹ <http://www.statistics.gov.uk/census2001/pdfs/sc9908.pdf>

strength'. This was never used. Where it might have been, local factors were considered - leading to exclusion of data, e.g. where a number of households were moved pending refurbishment.

Dependency

33. Underpinning the ONC methodology is the assumption of independence between the count of population given by the Census and CCS. For the ONC to work well, there must be no systematic relationship between the chance of a household or individual being enumerated in the Census and of being captured by the CCS.
34. The QA system indicated a degree of dependence sufficient to warrant a further adjustment. Plausible target sex-ratios, or other demographic information would be the easiest to use. However, there have been consistent changes in the observed sex ratios over time in the 1971, 1981, and 1991 Censuses and migration figures are poorly known, so assuming say a 50:50 sex ratio would perhaps mean only adjusting males upwards, while female populations too would need adjusting.
35. Therefore it was not appropriate to base an adjustment on an assumed sex ratio. Alternative methods of adjustment were developed and implemented. The key principle of these further adjustments were that the evidence had to come from the Census. Two types of adjustments were made – household level dependency and within Census household individual level dependency.
36. The first adjustment was based on the Census Enumerators occupied-household count (from some of which no return was received). If the ONC adjusted household count was below the enumerator household count at the Council Area level, and the person-level count was below the mid year population estimates (suitably adjusted) then a further person-level ratio adjustment was made and the ONC imputation system re-run. This countered household level dependency.
37. The second adjustment countered within household dependency. There is much discussion about whether a census is more likely to miss people in wholly missed households or to miss people in enumerated households, e.g. where a single parent does not admit to the presence of a partner. Within the ONC results we found there was poor consistency between LAs of the ratio of the types of missing people, so that in Edinburgh one might find that proportionately many more people were imputed into Census (i.e. enumerated) households than in Glasgow.

38. Therefore we calculated a 'Scotland average ratio' of 2:1 missed people in missed households to missed people in Census households and adjusted Local authority population estimates up to this 2:1 ratio.
39. What was not realised was that the rate of imputation of people into types of households was probabilistic so we could not hold the ratio to a predetermined level (even crudely by predetermining the number of households to be imputed). Therefore there was only one reiteration of this adjustment and then only in areas where the ratio had been higher than 2:1 (e.g. 4:1).
40. Table 1 shows the adjustments made. Appendix 1 is a technical document explaining and presenting in detail the adjustments made

Assessment and Lessons Learnt

Operational and organisational

41. ONS and Southampton University developed three key statistical elements (matching, estimation and imputation). The ONC was a new and ground-breaking methodology (both here or anywhere in the world). All census results were successfully adjusted for **measured** under enumeration.
42. The ONC process added around three months elapsed time to the original timetable. The overwhelming view from users during the consultation phase was that they were content to wait.

Table 1: Comparison of ONC and 2:1 and Final Census Population count.

Council Area/Bailery	Census Count	ONC Estimate ¹⁰	2:1 estimate	Final Population or re-estimation	Source
Aberdeen City	206,894	212,136	211,968	212,136	ONC
Aberdeenshire	223,454	226,863	226,423	226,863	ONC
Angus	107,298	108,396	109,312	108,396	ONC
Argyll & Bute	85,196	89,193	87,173	89,193	ONC
Clackmannanshire	46,893	48,079	47,982	48,079	ONC
Dumfries & Galloway	142,959	147,758	147,045	147,758	ONC
Dundee City	137,899	145,671	143,806	145,671	ONC
East Ayrshire	115,434	119,280	120,234	120,234	2:1
East Dunbartonshire	105,457	107,219	108,245	108,245	2:1
East Lothian	88,112	89,205	90,095	90,095	2:1
East Renfrewshire1	23,209	24,164	23,951	24,164	ONC
East Renfrewshire2	63,223	64,453	65,118	65,118	2:1
Edinburgh, City of	421,288	441,421	448,615	448,615	2:1
Eilean Siar	26,268	26,500	26,682	26,500	ONC
Falkirk	138,930	142,772	145,198	145,198	2:1
Fife	335,590	344,704	349,421	349,421	2:1
Glasgow City	531,483	561,507	577,867	577,867	2:1
Highland	204,977	208,913	209,314	208,913	ONC
Inverclyde	81,670	84,172	84,417	84,172	ONC
Midlothian	78,395	80,942	79,649	80,942	ONC
Moray	84,365	85,436	86,036	85,436	ONC
North Ayrshire	132,059	135,364	135,819	135,819	2:1
North Lanarkshire1	16,642	17,037	16,894	17,037	ONC

¹⁰ After occupied household adjustment

North Lanarkshire2	295,880	303,685	304,028	304,028	2:1
Orkney Islands	19,059	19,246	19,256	19,246	ONC
Perth & Kinross	132,197	134,950	134,001	134,950	ONC
Renfrewshire	167,437	172,973	175,491	172,973	ONC
Scottish Borders	104,395	105,839	106,764	106,764	2:1
Shetland Islands	21,523	21,988	21,685	21,988	ONC
South Ayrshire	109,012	111,815	112,104	112,104	2:1
South Lanarkshire1	52,545	53,780	55,089	55,089	2:1
South Lanarkshire2	240,643	244,853	247,123	247,123	2:1
Stirling	82,352	86,212	85,114	86,212	ONC
West Dunbartonshire1	45,375	47,811	47,498	47,811	ONC
West Dunbartonshire2	43,374	44,875	45,465	45,465	2:1
West Lothian	151,545	158,718	157,560	158,718	ONC
Scotland	4,863,032	5,017,930	5,052,444	5,058,343¹¹	

43. In the event the publication of first results was delayed by a further month. This was mainly due not to the ONC but to considerably longer delays to the basic processing timetable. However, there were unforeseen delays within the ONC process.
44. The causes of these delays were inappropriate 'spikes' (1 'donor' donating many times) and the need to adjust for dependency; donors were wrongly re-used only after imputation had used all the available occupied dwelling slots (see para 26). These spikes were relatively rare in Scotland where imputation into occupied dwelling slots was relatively high. These delays limited ONS' ability to catch up on the output production timetable.
45. One benefit of the delays was that the trained in-house matching staff team at GROS were able to be used to correct other quality issues within main Census processing. It also allowed time to repeat the ONC matching process twice for every EA and then to investigate fully all discrepancies. It is vital for a small organisation to build enough staff into its plan so that it retains flexibility if a contingency arises. If things go smoothly then quality can be improved.

Statistical

46. The ONC was a natural progression from advances made in previous censuses: from the introduction of automatic editing and imputation of missing variables in 1981, and absent household imputation in the 1991 Census to the addition in 2001 of imputed households that the

¹¹ There are later armed forces and minor imputation adjustments

ONC estimated. There would have been much bigger question marks about the Census results in 2001 if they had not been adjusted for the 200,000 people that the ONC estimated were missed.

47. There are some caveats though, about whether the ONC accounts for all under enumeration. It certainly accounts for the under enumeration measured by the CCS and for occupied household under enumeration (Assuming that Enumerators can distinguish identify such households successfully – which might be quite difficult in holiday areas for example). However it is difficult to make an objective adjustment for within Census household under enumeration if there is dependency between the CCS and the Census. In the final analysis though, the lack of any nationally credible third source of population count means it is extremely difficult to say whether all Census under enumeration was accounted for.
48. The ONC also had to make an assumption that the population imputed because of the extra dependency adjustments was the same as the population missed by the Census but found by the CCS. This is similar to extending a regression beyond the data points.
49. A better assumption is that the dependency population (i.e. the population assuming dependency) is different from the DSE population, but we lacked any evidence about that population unless we had used an expected male:female ratio for each age group to drive the extra dependency imputation. IN that case we may well have 'cleaned up' the male female ratios which later led to the adoption of an anecdotal 'Ibiza' effect. Without form migration statistics, and with the principle of deriving all evidence from the Census, the characteristics of the people in the final dependency adjustment was not driven by any demographic analysis.
50. Though the ONC was innovative, other countries have different approaches – which arguably give more confidence to the Census results. Ireland enumerates on a population present basis (actually usual residents + visitors) and uses a much more intensive enumeration methodology – and visitors are checked for form completion by the field staff at their usual residence. Australia has good birth, death and migration statistics and so under enumeration is the population estimates *minus* the Census results. So though the ONC was a brave methodological step forward, it is not the only way forward – though given the amount of under enumeration in the UK, there may be no turning back until another population estimate source is authoritative.

Quality Assurance and dependence

51. Following the three adjustments (CCS-based, occupied household and 2:1) there were no substantial gaps between Council population estimate expectations and the ONC expectations, except perhaps in Glasgow. This led to concerns about the lack of confidence in the ONC results. Even councils who were happy about the Census results were concerned about the lack of methodological transparency.
52. The ONC relied initially on the CCS as the source of information on under enumeration and subsequent work showed this to be a shortcoming. In practice there was little alternative; investigations during the development found that none of the other sources which might have been used, for example in a triple system estimator (Census, CCS and a third national population source) was reliable across the country as a whole.
53. However, using the enumerator estimate of occupied households to drive the ONC estimation population totals and allow the CCS to characterise those populations by type seems to have been a correct judgement that was easy to implement at the time.
54. The assumption this rests on, that the enumerator can correctly distinguish occupied properties, is being tested in the 2006 Census Test, where enumerator supervisors are having to go back to all habitable vacant properties. It is hoped that the ONC process will thereby be put on a firmer foundation if it is implemented in a similar way in 2011.
55. The CCS was designed as practicably as possible to be independent; all planned operational measures put in place were implemented. The CCS also achieved high response rates in the majority of areas. However dependence was found. Intuitively it is likely that the chance of being missed by the CCS is higher if the person in question was missed by the Census. The measurement of such 'dependence' is extremely difficult; internationally there has been some theoretical work but nobody else, to our knowledge, has attempted the practical application of measuring 'dependence'. In hindsight, ONS should have done more in contingency planning to consider the impact on the ONC of the lack of any third national source of adequate quality.

Conclusion

56. It was a huge achievement by ONS and Southampton University to carry out the ONC. This should never be underestimated. GROS

improved on the initial design by implementing further adjustments based on a realistic perception of the high quality of enumeration achieved in Scotland.

57. There are some lessons to be learnt and some further studies to be undertaken, as detailed above, but the ONC approach provides a platform on which to build for the future. It will face particular challenges with large semi-resident population migration flows – these populations are typically high under enumeration groups because they are deprived, alienated and migratory; the challenges may lie in creating good community links and looking for a third way to create population estimates rather than rely on the ONC which, especially in England, had to be adjusted in an ad hoc way to create a realistic picture.

Appendix 1: The ONC Adjustment for Scotland.

Section 1. Introduction

1. Along with the enumerated households and people, the final 2001 Census output included synthetic households with synthetic people and real households with added synthetic people. They were added to the Census database to adjust for under enumeration in the Census.
2. The adjustments were based on results from the Census Coverage Survey (CCS). The CCS re-enumerated all households in a sample of about 2,400 postcodes. It collected identification data such as name and address - used for matching - and data related to under enumeration - such as age, sex and employment status. The Survey respondents were matched against Census respondents. The number and characteristics of the Survey respondents missed by the Census were used to estimate under enumeration using the Dual System Estimation (DSE) method by five year age and sex groups.
3. The results were modelled up to Estimation Areas (see below) and then down to constituent Council Areas to provide control totals for 5 year age sex groups. Finally, individuals and households were imputed into postcodes. The final Census output was thereby fully adjusted for under enumeration at an individual level. The adjusted Census was known as the One Number Census (ONC) since only 1 set of numbers was published and all tables added up to the same national total. ¹²
4. The ONC DSE method relied on the assumption that the Census and the CCS were 'independent'; that is that all people had an equal chance of

¹² <http://www.statistics.gov.uk/census2001/onc.asp>

being captured by the Census and the CCS. Evidence suggested that this assumption did not hold in all areas. People missed by the Census were more likely to be missed by the CCS. Therefore, in some areas, a 'dependency' adjustment method was used. This method was called the '2:1' method. It was applied directly at Council Area Level.

- 5. ONS developed another method of adjusting for dependency - the Odds Ratio (OR) method. This was applied at the higher level Estimation Area. In both methods, once control totals were estimated, the original ONC imputation system was used to place people at the household level.
- 6. In this paper, there is a background section followed by a description of the 2:1 method. GROS also investigated the application of the ONS OR method and the findings are presented in Section 4. This is followed by a discussion of the problems with the GROS method.

Section 2: Background

- 7. Initially CCS results were used to adjust Census results for under enumeration in all Council or part Council (Baileries) Areas. There are 32 Councils in Scotland. 4 are split by Health Board Area boundaries (Table 1a). The set of 36 Council and split Council Areas are referred to as 'Baileries'. In 21 Baileries the CCS/ONC result was accepted. In 15 Baileries the 2:1 method was used. No area was adjusted using both methods.

Table 1a: Split Council Areas (Baileries) and their constituent parts

ZF	North Lanarkshire2	Cumbernauld & Kilsyth, Monklands and Motherwell Districts
ZF	South Lanarkshire2	Clydesdale, East Kilbride and Hamilton Districts
ZG	East Dunbartonshire	

ZG	East Renfrewshire2	Eastwood
ZG	Glasgow City	
ZG	North Lanarkshire1	Chryston
ZG	South Lanarkshire1	Rutherglen
ZG	West Dunbartonshire2	Clydebank
ZH	East Renfrewshire1	Barrhead part area
ZH	West Dunbartonshire1	Dumbarton District - Helensburgh part area

8. For Estimation processes, Bailerries were dragooned into Health Board Areas which themselves were marshalled into Estimation Areas. Once the population of an Estimation Area had been derived, the Council Area population was calculated. The organisation of areas is shown in Tables 1a and 1b.

Table 1b: The Health Board Areas within Estimation Areas

Estimation Area	Health Board Areas	Council/Bailery Area
ZA	Dumfries & Galloway	Dumfries and Galloway
	Ayrshire & Arran	East Ayrshire North Ayrshire South Ayrshire
	Borders	Scottish Borders
ZB	Lothian	City of Edinburgh East Lothian Midlothian West Lothian
ZC	Forth Valley	Clackmannanshire Falkirk Stirling
	Fife	Fife
ZD	Grampian	Aberdeen City Aberdeenshire Moray

ZE	Tayside	Angus Dundee City Perth & Kinross
	Highland	Highland
	Orkney	Orkney
	Shetland	Shetland
	Western Isles	Western Isles
ZF	Lanarkshire	Clydesdale SL2 Cumbernauld NL2
ZG	Greater Glasgow	Chryston NL1 Clydebank WD2 East Dunbartonshire Eastwood ER2 Glasgow city Rutherglen SL1
ZH	Argyll & Clyde	Argyll & Bute Barrhead part ER1 Dumbarton WD1 Inverclyde Renfrewshire

Section 3. The GROS 2:1 Adjustment Method

9. The evidence used in the 2:1 ratio method was:
- The number of Absent and Refusal households identified by Census enumerators.
 - The average ratio of synthetic people in synthetic households to synthetic people in Census households in Scotland.
10. In outline GROS used the following simple procedure to make the 2:1 population adjustment:

Equation 1

$$X = \frac{\text{ONC synthetic household people} * \text{Absent and Refusal household count}}{\text{ONC synthetic households}}$$

Where X = the number of synthetic people in synthetic households according to the number of Absent and Refusal Households.

Equation 2

$$Y = X + X/2$$

Where Y = the total number of synthetic people in Synthetic households and Census households after the dependency adjustment. $X/2$ is the assumed total number of synthetic people in enumerated households.

11. Equation 1 implies that the Census Enumerator identified the correct number of occupied household spaces. Therefore the number of synthetic household people was multiplied up proportionately by the number of appropriate dummy forms. This was done at a Council Area level and the imputation process used the location of Absent and Refusal Households to place synthetic households.
12. Equation 2 implies a constant 2:1 ratio between synthetic household people and synthetic people in Census households. Indirectly it assumes that in some areas the CCS did not find enough missing people in partially enumerated census households. The national average ratio of synthetic household people to synthetic Census household people was about 2:1. In Glasgow it was about 4:1. We interpreted that to mean that dependency was higher in Glasgow than elsewhere. The assumptions here are investigated in Section 5.
13. The calculation for every Bailery is shown in Table 2. (Here a , b and c are the quantities that feature in Equation 1). The choice of whether to use the CCS/ONC population or the 2:1 method was not always straightforward and where the obvious rule was not followed (see 1a and 1b below) was based on 4 factors

- 1a We made no negative adjustments except. If the result of using the 2:1 method was negative compared to the ONC result, the ONC result was kept.
- 1b Inversely, if the 2:1 figure was larger we accepted it rather than the ONC result.

But

1. If the ONC and 2:1 results were substantially the same, we 'preferred' the ONC result;
 2. If the ONC outcome was close to the 1991 based mid year estimate adjusted to Census day (CDMYE) then preferred the ONC result;
 3. If the number of ONC imputed households was close to the Census enumerator count of Absent and Refusal house holds then the ONC was accepted.
 4. If the ONC results for the Estimation Area were correct, then an assumption was made that the method worked for all Bailerries within the EA.
14. The decision was straightforward in most areas using rule 1 or 2. The areas where further evidence determined the outcome are shown in the below:
- **Moray:** The CDMYE was 84,480. The ONC estimate was 85,436. The ONC estimate was higher than the 2:1 ratio estimate in the other two Council Areas in the Estimation Area.
 - **Angus:** The CDYME was 108,430 and the ONC estimate was 108,396. The ONC estimate was higher than or about the same as the 2:1 ratio estimate in the other six Council Areas in the Estimation Area.
 - **Highland:** The number of A and R households was 1,720. The number of ONC households was 1,692.

- **Inverclyde:** The ONC estimate (84,172) was higher than the CDMY Estimate (84,090).
- **Renfrewshire:** The ONC estimate for the other 4 Council Areas in the EA were accepted.
- **Eilean Siar:** The ONC estimate for the other 6 Council Areas in the EA were accepted.

15. Once the age/sex by Council Area population totals were re-estimated (where necessary), they were fed back into the ONC system for modelling and individual and household imputation using the original ONC systems.

16. The Census count, the ONC and 2:1 adjustments and final populations are shown in Table 3. This shows the population total as 5,058,344. The final ONC population of Scotland was in fact 5,062,011. The difference of 3,667 is explained by an armed forces adjustment. That adjustment is outside the scope of this paper.

Table 2: The '2:1' Calculation Method for all Baileries and the Source of the Final Population Total.

Bailery	Census	ONC				Dependency Adjustment			Source of final population
Measure	Absent & Refusal Dummies	Imputed holds	Imputed people into synthetic hhlds	Imputed people into Census hhlds	Total imputed people	Imputed people into synthetic hhlds (Equation 1)	Imputed people into Census hhlds	Total imputed people (Equation 2)	
Column	a	b	c	d	e	f	g	h	
Source	Census	ONC			c+d	C*a/b	f/2	f+g	
Dumfries & Galloway	1,203	1,444	3,270	1,529	4,799	2,724	1,362	4,086	ONC
North Ayrshire	1,102	1,056	2,402	903	3,305	2,507	1,253	3,760	Dependency
South Ayrshire	902	883	2,018	785	2,803	2,061	1,031	3,092	Dependency
East Ayrshire	1,382	932	2,158	1,688	3,846	3,200	1,600	4,800	Dependency
Scottish Borders	723	547	1,195	249	1,444	1,579	790	2,369	Dependency
East Lothian	640	378	781	312	1,093	1,322	661	1,983	Dependency
Midlothian	440	894	1,698	849	2,547	836	418	1,254	ONC
Edinburgh, city of	9,944	6,662	12,205	7,928	20,133	18,218	9,109	27,327	Dependency
West Lothian	1,879	1,773	3,784	3,389	7,173	4,010	2,005	6,015	ONC
Falkirk	1,972	1,410	2,988	854	3,842	4,179	2,089	6,268	Dependency
Stirling	857	874	1,878	1,982	3,860	1,841	921	2,762	ONC
Clackmannanshire	328	434	961	225	1,186	726	363	1,089	ONC
Fife	4,398	3,438	7,208	1,906	9,114	9,221	4,610	13,831	Dependency
Aberdeen City	2,288	1,591	2,352	2,890	5,242	3,382	1,691	5,073	ONC
Aberdeenshire	1,039	842	1,604	1,805	3,409	1,979	990	2,969	ONC
Moray	607	359	659	412	1,071	1,114	557	1,671	ONC
Dundee City	2,312	2,567	4,372	3,400	7,772	3,938	1,969	5,907	ONC

Bailiery	Census	ONC				Dependency Adjustment			Source of final population
Measure	Absent & Refusal Dummies	Imputed holds	Imputed people into synthetic hhlds	Imputed people into Census hhlds	Total imputed people	Imputed people into synthetic hhlds (Equation 1)	Imputed people into Census hhlds	Total imputed people (Equation 2)	
Column	a	b	c	d	e	f	g	h	
Source	Census	ONC			c+d	C*a/b	f/2	f+g	
Angus	780	402	692	406	1,098	1,343	671	2,014	ONC
Perth & Kinross	684	1,205	2,119	634	2,753	1,203	601	1,804	ONC
Highland	1,720	1,692	2,844	1,092	3,936	2,891	1,446	4,337	ONC
Orkney Islands	84	80	125	62	187	131	66	197	ONC
Shetland Islands	66	185	302	163	465	108	54	162	ONC
Eilean Siar	163	82	139	93	232	276	138	414	ONC
South Lanarkshire 2	2,102	1,555	3,196	1,014	4,210	4,320	2,160	6,480	Dependency
North Lanarkshire 2	2,720	2,724	5,440	2,365	7,805	5,432	2,716	8,148	Dependency
South Lanarkshire 1	1,025	698	1,155	80	1,235	1,696	848	2,544	Dependency
North Lanarkshire 1	95	208	368	27	395	168	84	252	ONC
Glasgow City	16,105	12,272	23,563	6,461	30,024	30,923	15,461	46,384	Dependency
East Renfrewshire 2	482	319	836	394	1,230	1,263	632	1,895	Dependency
East Dunbartonshire	854	566	1,232	530	1,762	1,859	929	2,788	Dependency

Bailery	Census	ONC				Dependency Adjustment			Source of final population
Measure	Absent & Refusal Dummies	Imputed holds	Imputed people into synthetic hhlds	Imputed people into Census hhlds	Total imputed people	Imputed people into synthetic hhlds (Equation 1)	Imputed people into Census hhlds	Total imputed people (Equation 2)	
Column	a	b	c	d	e	f	g	h	
Source	Census	ONC			c+d	C*a/b	f/2	f+g	
West Dunbartonshire 2	706	634	1,252	249	1,501	1,394	697	2,091	Dependency
East Renfrewshire 1	230	267	574	381	955	494	247	741	ONC
West Dunbartonshire 1	769	609	1,121	1,315	2,436	1,416	708	2,124	ONC
Inverclyde	924	1,168	2,315	187	2,502	1,831	916	2,747	ONC
Renfrewshire	2,558	2,118	4,446	1,090	5,536	5,370	2,685	8,055	ONC
Argyll & Bute	701	1,480	2,782	1,215	3,997	1,318	659	1,977	ONC
Scotland	64,784	54,348	106,034	48,864	154,898	126,395	63,197	189,592	Mixed

Table 3: Comparison of ONC and 2:1 Estimate, Final population and Source.

Council Area/Bailery (see foot note on next page)	Refusal and Absent hhlds	No of synthetic hhlds imputed	Census Count	ONC adjust-ment	2:1 adjust-ment	Final Pop	Source
	Census	ONC	Census	ONS	GROS		
Aberdeen City	2,288	1,591	206,894	5,242	5,074	212,136	ONC
Aberdeenshire	1,039	842	223,454	3,409	2,969	226,863	ONC
Angus	780	402	107,298	1,098	2,014	108,396	ONC
Argyll & Bute	701	1,480	85,196	3,997	1,977	89,193	ONC
Clackmannanshire	328	500	46,893	1,186	1,089	48,079	ONC
Dumfries & Galloway	1,203	1,669	142,959	4,799	4,086	147,758	ONC
Dundee City	2,312	2,567	137,899	7,772	5,907	145,671	ONC
East Ayrshire	1,382	1,206	115,434	3,846	4,800	120,234	2:1
East Dunbartonshire	854	853	105,457	1,762	2,788	108,245	2:1
East Lothian	640	591	88,112	1,093	1,983	90,095	2:1
East Renfrewshire1	230	267	23,209	955	742	24,164	ONC
East Renfrewshire2	482	489	63,223	1,230	1,895	65,118	2:1
Edinburgh, City of	9,944	9,460	421,288	20,133	27,327	448,615	2:1
Eilean Siar	163	82	26,268	232	414	26,500	ONC
Falkirk	1,972	2,187	138,930	3,842	6,268	145,198	2:1
Fife	4,398	5071	335,590	9,114	13,831	349,421	2:1
Glasgow City	16,105	20,094	531,483	30,024	46,384	577,867	2:1
Highland	1,720	1,692	204,977	3,936	4,337	208,913	ONC
Inverclyde	924	1,168	81,670	2,502	2,747	84,172	ONC
Midlothian	440	845	78,395	2,547	1,254	80,942	ONC
Moray	607	359	84,365	1,071	1,671	85,436	ONC
North Ayrshire	1,102	1,288	132,059	3,305	3,760	135,819	2:1
North Lanarkshire1	95	213	16,642	395	252	17,037	ONC
North Lanarkshire2	2,720	3,224	295,880	7,805	8,148	304,028	2:1
Orkney Islands	84	80	19,059	187	197	19,246	ONC

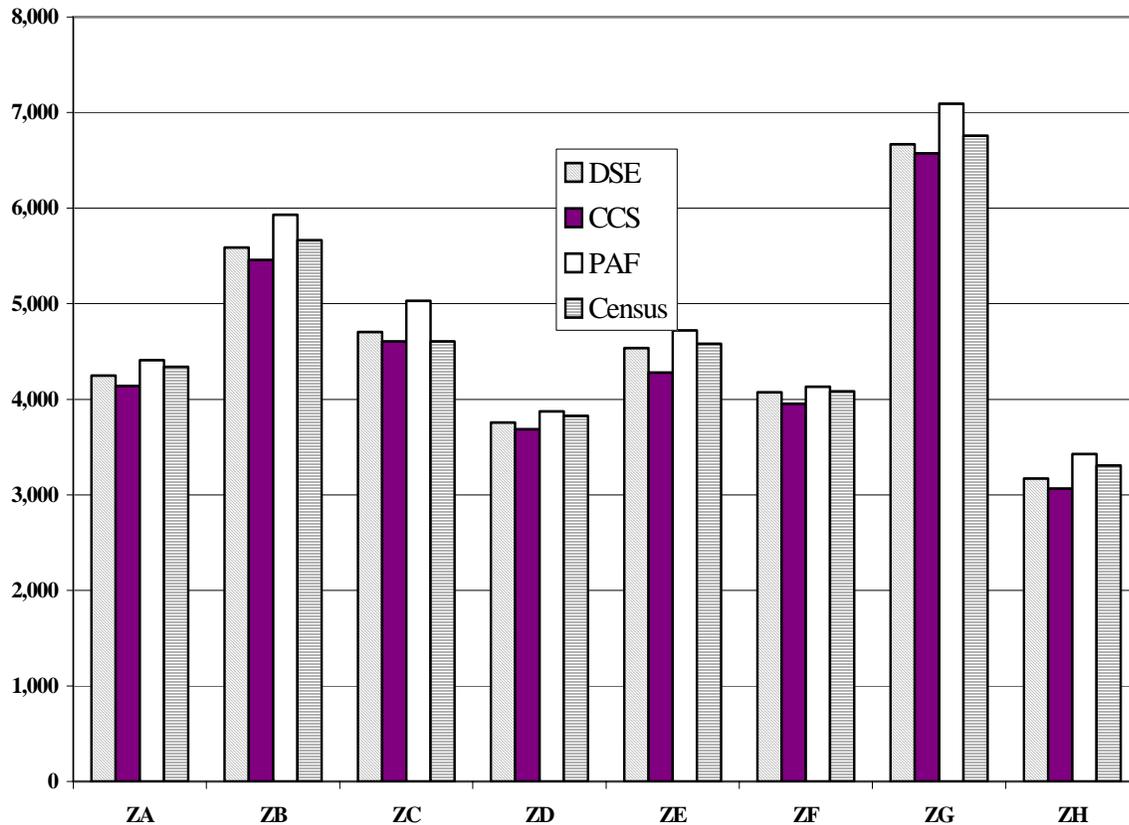
Perth & Kinross	684	1,205	132,197	2,753	1,804	134,950	ONC
Renfrewshire	2,558	2,118	167,437	5,536	8,054	172,973	ONC
Scottish Borders	723	769	104,395	1,444	2,369	106,764	2:1
Shetland Islands	66	185	21,523	465	162	21,988	ONC
South Ayrshire	902	1,079	109,012	2,803	3,092	112,104	2:1
South Lanarkshire1	1,025	1,368	52,545	1,235	2,544	55,089	2:1
South Lanarkshire2	2,102	2,146	240,643	4,210	6,480	247,123	2:1
Stirling	857	1,510	82,352	3,860	2,762	86,212	ONC
West Dunbartonshire1	769	609	45,375	2,436	2,123	47,811	ONC
West Dunbartonshire2	706	880	43,374	1,501	2,091	45,465	2:1
West Lothian	1,879	2,336	151,545	7,173	6,015	158,718	ONC
Scotland	64,784	72,425	4,863,032	154,898	189,412	5,058,344	Mixed

Section 4: The Odds Ratio Method

17. As mentioned ONS developed an 'Odds Ratio' (OR) method to adjust Census results for dependency.¹³ For a discussion of the technical issues please refer to ONS papers.
18. GROS applied the OR method to estimate Estimation Area populations and Scotland total adjustments (The OR method was applied at a higher level of geography than the 2:1 method). Broadly, the Household-level OR was estimated taking the average of suitably adjusted Census figures and Postal Address File (PAF) figures for the number of households. Figure 1 shows the count of households by the Census, PAF, the Dual System Estimate and the CCS.

¹³ http://www.statistics.gov.uk/census2001/pdfs/dependency_paper.pdf

Figure 1: Comparison of DSE, CCS, PAF and Census Total Households by Estimation Area



19. The household level OR was then converted into a person level OR. To do this an assumption was made about the relative level of dependency in wholly missed households and Census households. ONS assumed a 50:50 ratio. However, the adjustment itself only covered dependency of people within wholly missed households. It did not cover people missed within enumerated households¹⁴. The ONS and GROS methods produced similar total results. These are shown in Table 4.

¹⁴ Page 15 <http://www.lga.gov.uk/Documents/Publication/onenumpercensus.pdf>

Table 4: Comparative Population Estimates, Census Count and Odds Ratio Estimate

EA	Census day rolled forward estimate	Census Count	Odds Ratio
ZA	626,210	603,859	624,715
ZB	786,690	739,340	779,595
ZC	629,610	603,765	622,274
ZD	520,990	514,713	527,530
ZE	660,000	649,221	669,642
ZF	561,890	536,523	551,050
ZG	907,720	812,724	858,749
ZH	420,600	402,887	422,237
Scotland (sum)	5,113,710	4,863,032	5,055,792
Scotland level OR estimate			5,062,515

20. Table 5 shows various population totals as a percentage of the 2001 rolled forward population estimates (prior to any 2001 Census influence). The population estimates for EAs ZD, ZE and ZH are greater than the rolled forward population estimates. ZA, ZB, ZC and ZF are below the population estimates, while ZG (Greater Glasgow) is distinctly so (See also Figure 2). The OR made the greatest % difference in ZB (Lothian) and ZG (Greater Glasgow) but not appreciably more than elsewhere. The adjustment was not sensitive to the geography of Census under enumeration.

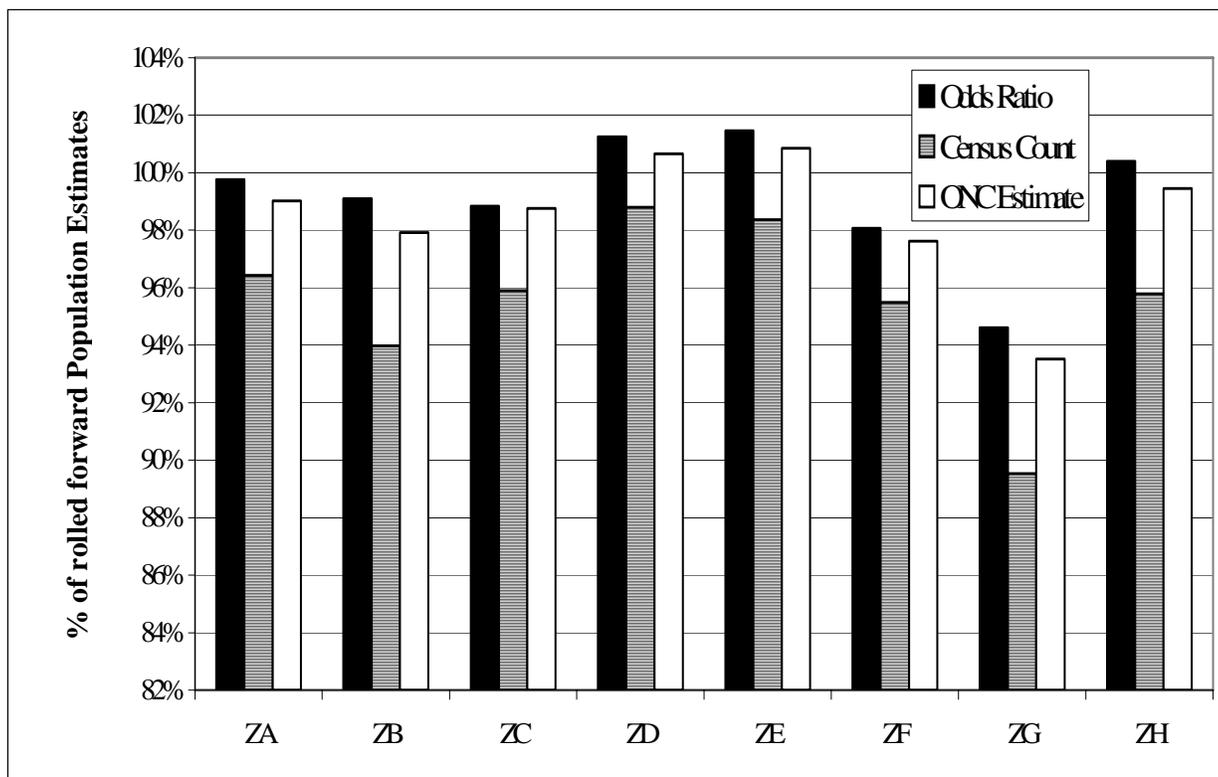
Table 5: The % difference between the Census, the ONC and the Odds Ratio results compared to the rolled forward Census Day Estimate and the net % amount of change.

Est Area	Census Count	ONC Estimate	Odds Ratio	% difference between ONC and OR method	% of gap filled between census count and CDE
ZA	-3.6%	-1.0%	-0.2%	0.7%	93.3%
ZB	-6.0%	-2.1%	-0.9%	1.2%	85.0%
ZC	-4.1%	-1.2%	-1.2%	0.1%	71.6%
ZD	-1.2%	0.7%	1.3%	0.6%	204.2%
ZE	-1.6%	0.9%	1.5%	0.6%	189.5%
ZF	-4.5%	-2.4%	-1.9%	0.4%	57.3%
ZG	-10.5%	-6.5%	-5.4%	1.1%	48.4%
ZH	-4.2%	-0.5%	0.4%	0.9%	109.2%

21. GROS judged that the fit between the OR adjustment and the Census evidence of missed households was improved by direct reference to, and only to, the number of occupied households missed by the Census, i.e. Census Absent and Refusal households. The GROS Council Area 2:1 adjustment did not use the PAF.

22. The Census enumerator started with PAF and supplemented that address list during the geography check and the Census form delivery phase. During the field work they also assessed the property status. GROS felt that the enumerator work was likely to be comprehensive, up-to-date and reflect contemporary occupation. Therefore GROS used Census evidence to drive the adjustments within Council Areas.

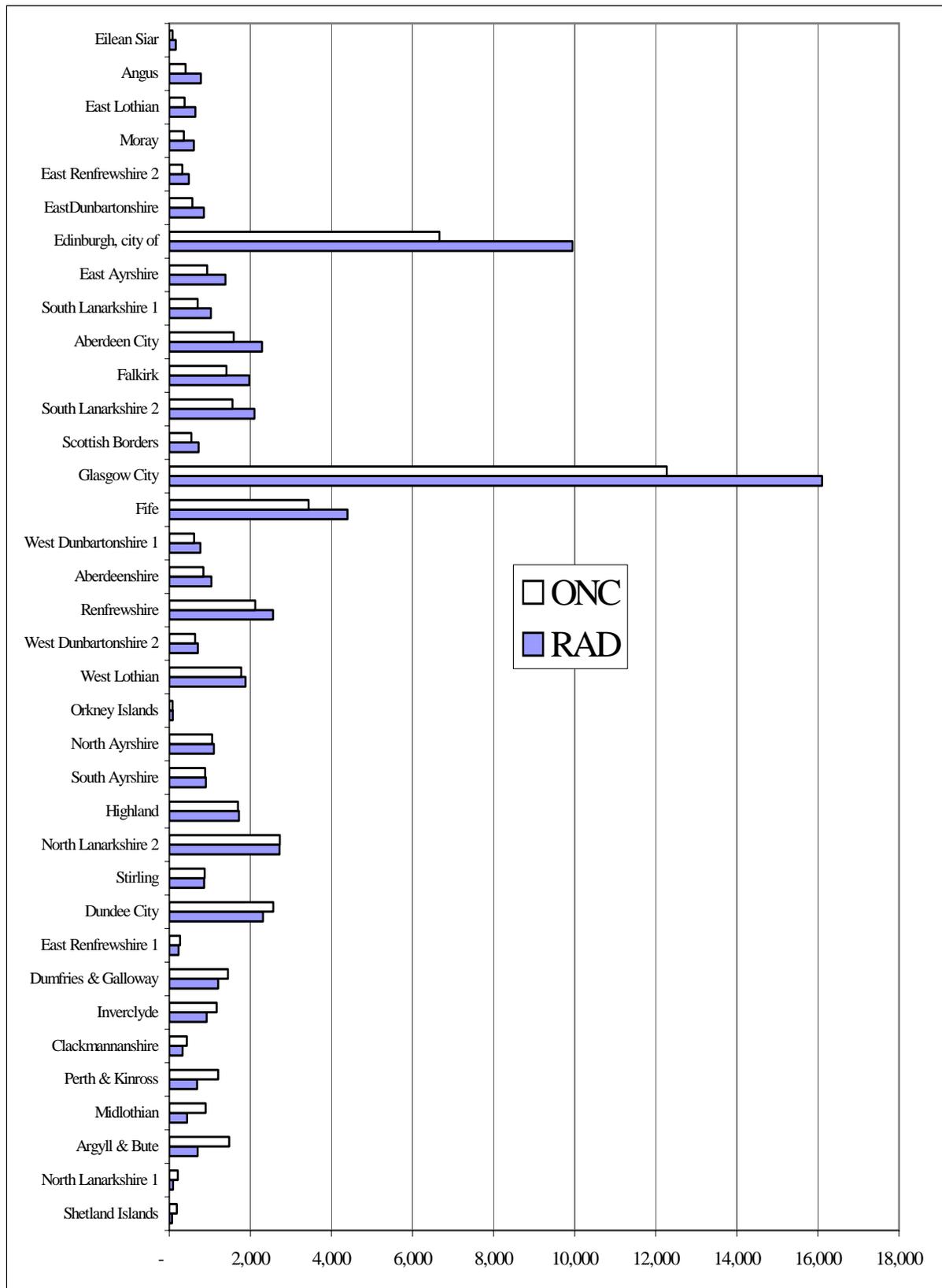
Figure 2: Different Counts by EA compared to Rolled Forward Population Estimates (=100%)



Section 5: Discussion

23. Once it became apparent that dependency was a problem, there were a number of ways of dealing with it. It can be estimated by reference, for example, to :
- Previous population estimates;
 - Administrative estimates;
 - Demographic analysis such as sex ratios and dependency ratios.
24. An aim of the ONC was to remove reliance on previous estimates and demographic analysis assumptions. Therefore GROS worked with the principle that the Census was the best source of evidence available. In some areas the ONC/CCS results were used; in other areas we made a dependency adjustment that was firstly based on the number of occupied household dwelling spaces found by Census enumerator.. The effect of this adjustment can be judged from Figure 3.

Figure 3: Refusal/Absent (RAD) Households and ONC Households sorted by Ratio



25. In the 2:1 adjustment method the CCS evidence was still used to characterise the imputed population, although using a donor from a real census record. Once dependency has been established, you have either to use the population characteristics of the people missed by the Census and the CCS or make assumptions about their characteristics. GROS made the assumption that the people missed by both the Census and the CCS had the same characteristics as those missed by the Census and found by the CCS. Arguably it may have been better to make any adjustment for dependency using external evidence of age sex ratios.
26. However there was uncertainty about previously adjusted 1981 based population estimates and intervening migration flows. This problem is discussed by the LGA¹⁵. We felt that the best assumption, and following the principle of using Census data, was to assume that the final adjustment population had the same characteristics as the ONC adjustment population.
27. Demographic Analysis indicated that perhaps the final adjustment should have been predominantly males but this would leave us with a certain amount of demographic guess work about these people's other characteristics. Our imputation was at least based on real people. The Census had primacy and the method avoided repeating previous assumptions.
28. There was also the question of the level of dependency within Census households. The enumeration household count gave a basis to adjust for household dependency. The only Census evidence we could find to adjust for within Census household dependency was the relative number of people imputed into the categories of new households and census households in

¹⁵ <http://www.lga.gov.uk/Documents/Publication/onenumpercensus.pdf>

each Council Area. Those figures are available within Table 2 and repeated as Table 6 below.

29. The evidence that was of critical importance, beside the difference between the number of Census occupied household spaces and ONC created household spaces, was the number of synthetic people placed in Census households. In Edinburgh and Glasgow respectively the numbers are 7,928 and 6,461, i.e. more synthetic people were imputed into Census households in Edinburgh than in Glasgow.

Table 6: The number and ratio of synthetic people in synthetic households and in Census households by Bailery from the ONC result.

Bailery	Imputed people			Bailery	Imputed people		
	In synthetic hhlds	In Census hhlds	Ratio		In synthetic hhlds	In-Census hhlds	Ratio
Dumfries & Galloway	3,270	1,529	2.1	Highland	2,844	1,092	2.6
North Ayrshire	2,402	903	2.7	Orkney Islands	125	62	2.0
South Ayrshire	2,018	785	2.6	Shetland Islands	302	163	1.9
East Ayrshire	2,158	1,688	1.3	Eilean Siar	139	93	1.5
Scottish Borders	1,195	249	4.8	South Lanarkshire 2	3,196	1,014	3.2
East Lothian	781	312	2.5	North Lanarkshire 2	5,440	2,365	2.3
Midlothian	1,698	849	2.0	South Lanarkshire 1	1,155	80	14.4
Edinburgh, city of	12,205	7,928	1.5	North Lanarkshire 1	368	27	13.6
West Lothian	3,784	3,389	1.1	Glasgow City	23,563	6,461	3.6
Falkirk	2,988	854	3.5	East Renfrewshire 2	836	394	2.1
Stirling	1,878	1,982	0.9	East Dunbartonshire	1,232	530	2.3
Clackmannanshire	961	225	4.3	West Dunbartonshire 2	1,252	249	5.0
Fife	7,208	1,906	3.8	East Renfrewshire 1	574	381	1.5
Aberdeen City	2,352	2,890	0.8	West Dunbartonshire 1	1,121	1,315	0.9
Aberdeenshire	1,604	1,805	0.9	Inverclyde	2,315	187	12.4
Moray	659	412	1.6	Renfrewshire	4,446	1,090	4.1
Dundee City	4,372	3,400	1.3	Argyll & Bute	2,782	1,215	2.3
Angus	692	406	1.7	Scotland with Glasgow	106,034	48,864	2.2
Perth & Kinross	2,119	634	3.3	Scotland - no Glasgow	82,471	42,403	1.9

30. Therefore, having determined the number of people who should form synthetic households, we then determined that the ratio of such people to

Census household synthetic people should be 2:1 - about the average of Scotland. Bailerries with a larger ratio would have their population increased - therefore, in Glasgow for example, the original ratio was 3.6:1 and it became 2:1 by raising the number of synthetic people within Census households.

31. This meant in effect that in Bailerries where within-Census household dependency was smaller than average, we increased it to the national average. This seems to be a very conservative estimate. In Australia for example ¹⁶ there was a 1:5 ratio within the Post Enumeration Survey showing that under enumeration largely occurs within enumerated households. It might imply that that is where most dependency lies, However local factors could play a role in this difference.
32. Therefore a criticism of the GROS method was that the Odds Ratio or Dependency for within Census household people is unknown. In fact so much so that there are conflicting views on the matter. They vary from their being little likelihood of within Census household dependency (because of the high quality of the CCS design and fieldwork) to it being the largest cause of dependency because whole houses are easy to enumerate (see Australian reference) and follow-up. In these circumstances, we felt the adoption of a national average where the Bailery was below the average was a safe assumption on which to base a further adjustment for a particular part of overall dependency.
33. However one can still question the choice of the national average on the grounds that the stability of the figure obtained was unknown. In this work,

¹⁶ Appendix 1 in

[http://www.ausstats.abs.gov.au/ausstats/free.nsf/Lookup/6C9F76A932B2F054CA256CFB0004962D/\\$File/29400%5F2001.pdf](http://www.ausstats.abs.gov.au/ausstats/free.nsf/Lookup/6C9F76A932B2F054CA256CFB0004962D/$File/29400%5F2001.pdf)

it was assumed that the ONC ratio resulting from apportioning synthetic people into synthetic households and Census households would be reasonably stable.

34. Ultimately household formation was driven by the choice of donor household and its size. In one sense then it would be more sensible to adopt a stable assumption when estimating within census household dependency. We had strong evidence for within Census household dependency, but to an unknown extent this could have been an artefact of the imputation programme rather than the CCS.

GROS Census Division

March 2007