



ONS(ONC(SC))00/12

ONE NUMBER CENSUS STEERING COMMITTEE

The One Number Census in Northern Ireland

This paper describes how NISRA will apply the principles of the One Number Census Process in Northern Ireland, and research conducted to estimate the quality of the resulting estimates.

The Steering Committee is invited to note the paper.

**Robert Beatty
NISRA
McAuley House
2-14 Castle Street
Belfast
BT1 1SA**

**James Brown
University of Southampton
Department of Social Statistics
Highfield
Southampton
SO17 1BJ**

June 2000

The One Number Census in Northern Ireland

1. Introduction

In recent decades the Census of Population in Northern Ireland has followed broadly the methodology and questionnaire content of the other Censuses conducted in the rest of the UK. However, separate output has traditionally been produced for the Northern Ireland Census with no outputs for the UK as a whole. At the start of the planning process for the 2001 Census of Population, NISRA decided to become more fully integrated with the Censuses proposed for England and Wales and Scotland, with the ultimate aim of producing UK-level output. Accordingly, NISRA plan to apply the One Number Census principles, adjusting 2001 Census outputs for estimated under-enumeration, and producing Census output which is conceptually consistent with the mid-year population estimates.

The One Number Census methodology in Northern Ireland will broadly follow that in Great Britain, as described in many papers already seen by the Steering Committee. This paper will take the previous methodology papers as given and concentrate on any differences which will apply in Northern Ireland; it is believed that these differences are marginal.

2. Is there a need for a One Number Census approach in Northern Ireland?

It is generally accepted that a successful census was conducted in Northern Ireland in 1991. However, even with a 'successful' census, the follow-up Census Validation Survey to the 1991 Census identified under-enumeration of 0.66 per cent although the survey was too small to allow separate estimates of under-enumeration by age (1991 Northern Ireland Census General Report, NISRA). Accordingly, at that time the mid-year estimates (MYEs) of population took the 1991 Census base and grossed-up by 0.66 per cent (over and above correction factors for HM Forces, moving students to term-time address and changes between Census Day and 30 June).

Later analyses which retrospectively examined administrative data sources suggested that there were a number of problems, not identified by the Census Validation Survey, for, in particular, very young children and the very elderly (NISRA Occasional Paper Number 12, 1999). It is now believed that total under-enumeration in 1991 was of the order of just over 1 per cent but concentrated in certain age-groups.

NISRA's view of the One Number Census is to accept that some under-enumeration is inevitable and to plan accordingly. Our aim remains good coverage in the Census and the One Number Census is essentially our contingency for not achieving full coverage.

3. Differences in the 2001 Census in Northern Ireland compared to GB

The Census methodology in Northern Ireland will be very similar to that in the rest of the UK and NISRA participates in a range of formal Working Groups and Steering Groups, in addition to ad hoc liaison with ONS and GROS colleagues, to ensure consistent practice throughout the UK. Differences in fieldwork practices are minor; some examples are that Northern Ireland will be using an advance letter, and the split of duties between enumerators and Team Leaders may vary between countries. Details are still being finalised, but NI may conduct more stringent completeness checks than England and Wales. These differences are considered to be marginal.

The questionnaire in Northern Ireland is very similar to that in GB. The main differences are: Northern Ireland has a religion question (as it has always had); The ethnic question in Northern Ireland is less detailed reflecting local need; Northern Ireland has an Irish language question, similar to questions in Scotland and Wales; An additional household question on the number of floors occupied.

3.1 The Census Coverage Survey

As in GB, ONC methodology will be based on a large scale face-to-face household survey – the CCS - which will commence approximately three weeks after Census Day. CCS fieldwork dates are similar in Northern Ireland and GB.

3.1.1 CCS - questionnaire

For ONC purposes, the religion question is viewed as more important in Northern Ireland than the ethnic origin question as numbers of people from ethnic minorities are small in Northern Ireland. Accordingly, the CCS questionnaire in Northern Ireland will ask about religion and not ask about ethnic origin. Apart from this the Northern Ireland questionnaire is similar to the GB questionnaires.

3.1.2 CCS - fieldwork

One of the key differences between the Northern Ireland CCS and its GB counterparts is the use of an experienced fieldforce in Northern Ireland. NISRA's equivalent of Social Survey Division (SSD) in ONS, the Central Survey Unit (CSU), maintains a fieldforce of over 100 interviewers to conduct the Northern Ireland elements or equivalents of the Labour Force Survey, the General Household Survey, the Family Expenditure Survey and a range of other regular and ad-hoc sample surveys. NISRA used the CSU fieldforce to conduct the 1997 Census Test, the follow-up Census Test Evaluation Survey and the 1999 Census Rehearsal. NISRA have decided to use the CSU fieldforce to conduct the CCS.

Fieldwork procedures will generally be similar to those in GB, and joint training is being organised with colleagues in GROS. The biggest difference in the fieldwork procedure (compared to England and Wales) is likely to be that NISRA CCS interviewers will not share workloads in pairs.

3.1.3 CCS - sample size and design

A fuller description of the proposed Northern Ireland design is given in the technical annex, the content of which is summarised below.

In GB, the CCS is designed as a series of Design or Estimation Areas (EA), where each EA has a population of about 500,000 and each EA has a separate CCS. This 0.5m building block principle has been applied in Northern Ireland (total population 1.7m) resulting in 3 EAs. For administrative purposes, Northern Ireland is split into 26 Local Government Districts (LGDs) - the equivalents of LAs in GB although Northern Ireland's LGDs are generally much smaller (typical population of about 50,000) than LAs in GB. Each EA in Northern Ireland is the aggregate of a number of LGDs, similar to the aggregation of LAs in GB.

Belfast is the largest LGD in Northern Ireland (1998 MYE 287,500) but surrounded by dormitory towns, which are socio-economically different to Belfast itself. Outside of greater Belfast, most of Northern Ireland is very rural by British standards. Northern Ireland's three EAs are thus Belfast (single LGD, population 287,500), East of Northern Ireland (12 LGDs surrounding Belfast, population 764,400) and West of Northern Ireland (the remaining 13 LGDs, population 636,600).

For each EA, the CCS is being designed in a similar manner to GB, with 1991 EDs being stratified and then used as primary sampling units from which unit postcodes are selected. The CCS will consist of complete re-enumeration of the sampled postcodes. The main difference in the Northern Ireland approach is the method by which EDs are stratified. The stratification of 1991 EDs in GB is based on a Hard to Count index derived from research into under-enumeration in 1991. Some key GB indicators such as multi-occupancy dwellings and ethnic minorities have very low incidence levels in Northern Ireland and there is no research to support similar effects in Northern Ireland. The stratification of 1991 EDs in Northern Ireland will be based on observed response rates to Northern Ireland's (voluntary) 1997 Census Test, which was designed as a fractional replicate of a 2x2x3 experiment where EDs were the sampling units, classified by predominant religious background (3 levels), urban/rural and deprived/non-deprived. The 12 design strata from the 1997 Census Test have been collapsed to 3 levels of stratification for the CCS, merging strata with similar response rates.

A design has been produced which has the same overall sampling fraction as England and Wales (approximately 3.6 per cent of 1991 EDs as primary sampling units); in Northern Ireland this corresponds to 134 EDs. Within each ED, a sample of postcodes, consisting of on average 5 postcodes and 75 households, will be selected for re-enumeration, giving a total sample of 670 postcodes involving 10,000 households. At this stage, the ONS approach (5 random postcodes) and the GROS approach (continue to sample postcodes until a workload is achieved) are both being considered. The experience of the 1999 rehearsal, which demonstrated a very skewed distribution of postcode size (by number of households), makes the GROS approach attractive.

3.1.4 Efficiency of the proposed CCS design

The efficiency of the proposed design has been examined through a simulation exercise based on the 1991 census database for Northern Ireland. Under the assumptions of census coverage of 97.25 per cent, CCS coverage of 90 per cent (households) and 98 per cent (people within households) and independence for the dual estimator, the simulations suggest a relative RMSE of 0.46 per cent for the Northern Ireland population estimate. Further details including analysis by EA are given in the Technical Annex.

4. ONC analysis

The ONC statistical analysis (matching, use of dual estimators, regression analysis and so forth) is planned to be similar to that in GB except that religion will take the place of ethnic origin. The analysis software is being developed by ONS and discussions are underway about how the ONC analysis will actually be delivered, for example can NISRA run the processes remotely from Belfast?

Associated with being relatively small in population terms, the LGDs have very limited powers and NISRA do not anticipate the extent of interest expected from LAs in GB regarding Census estimates of their population. Accordingly, NISRA do not anticipate demand for separate under-enumeration estimation for each LGD. Each EA is however composed of an aggregation of NUTS level 3 units (the NUTS classification is the standard aggregation method used when making statistical returns to the European Union); each of the East and West EAs are the aggregate of 2 NUTS areas. NISRA intend to make estimates for the NUTS areas within each EA in a similar manner to ONS's plan for LAs within EA, that is allowing different patterns of under-enumeration for different NUTS areas within an EA. Estimates for LGDs will then be produced using standard synthetic estimation for each LGD within the NUTS unit.

NISRA will require age x sex population estimates for each LGD by summer 2002 for the purpose of producing the 2001 mid-year population estimates. These LGD figures must of course be consistent with those that will be produced in census outputs in 2003. Therefore, the 'pruning and grafting' step will be required at LGD level.

5. ONC contingency

Northern Ireland anticipates using similar procedures to those being developed in the rest of the UK. Work is ongoing to develop tolerance estimates by which MYEs and adjusted Census counts can be compared.

One area that requires further work is the idea of borrowing strength from other similar areas should the ONC results look suspicious in one area. In a GB context, inner-city Liverpool could, for example, borrow strength from inner-city Newcastle. There is very limited scope for such borrowing strength for EAs in Northern Ireland.

Implementing the CCS design in Northern Ireland

1. Design Groups

The Northern Ireland census count in 1991 was about 1.6 million in 3,729 EDs and the population is projected to be about 1.7 million by 2001, suggesting the use of 3 Estimation Areas for 2001 CCS purposes. The country is split up into 26 Local Government Districts (LGDs) that are similar to Local Authority Districts (LADs) in England and Wales. However, they differ in that they are generally smaller in population terms than those in England and Wales (some are very small, Moyle has a population of 15,000) and in general they do not have as much political power as LADs in England and Wales.

Belfast is easily the largest LGD in population terms, with a population of about 300,000. The LGD boundary defines the urban area very tightly, and many areas that might be considered as suburbs of the city are in neighbouring LGDs such as Castlereagh and Newtownabbey. A consequence of this is that the population of the LGD area is declining as people move to the suburbs. However, it is still important and likely to differ in terms of underenumeration from the rest of Northern Ireland. Therefore Belfast is considered as a design group on its own. The rest of the LGDs are grouped into the standard East and West classification as shown in Table 1.

Table 1: Classification of LGDs (excluding Belfast) into Two Design Groups (Number of EDs is based on the 1991 Census)

East		West	
LGD	Number of EDs	LGD	Number of EDs
Antrim	88	Armagh	153
Ards	162	Ballymoney	71
Ballymena	132	Coleraine	120
Banbridge	92	Cookstown	89
Carrickfergus	64	Derry	182
Castlereagh	118	Dungannon	133
Craigavon	155	Fermanagh	214
Down	146	Limavady	64
Larne	76	Magherafelt	90
Lisburn	202	Moyle	47
Newtownabbey	147	Newry and Mourne	229
North Down	145	Omagh	137
		Strabane	106
Total	1527	Total	1635

This classification means that using 1991 Census figures the population of the Belfast design group is 279,215, the population of the East design group is 700,364, and the population of the West design group is 598,111. This variety in the population size will have consequences when choosing sample sizes.

2. Enumeration District Type (EDT) Index for Northern Ireland

In England and Wales, EDs have been classified using a Hard to Count Index which incorporates local indicators, such as multi-occupancy housing and ethnic minority populations, which research has demonstrated to be related to underenumeration. For Northern Ireland no similar research is available, and at least some of the indicators have very low incidence levels.

The stratification of 1991 EDs in Northern Ireland will be based on observed response rates to Northern Ireland's (voluntary) 1997 Census Test, which was designed as a fractional replicate of a 2x2x3 experiment where EDs were the sampling units, classified by predominant religious background (3 levels), urban/rural and deprived/non-deprived. Further details on the classification methods used in the 1997 Census Test and the observed response rates can be found in NISRA Occasional Paper Number 13 (1999).

The religious background classification was reduced from 3 levels to 2 on the basis of similar response rates, giving eight initial strata for the EDs, as shown in table 2 where the strata are ranked according to observed response rates (highest first).

Table 2: Ranking of the EDT Index

Level of Index	Religion	Location	Deprivation Status
1	Protestant	Rural	Not Deprived
2	Protestant	Rural	Deprived
3	Catholic & Mixed	Rural	Not Deprived
4	Protestant	Urban	Not Deprived
5	Protestant	Urban	Deprived
6	Catholic & Mixed	Urban	Not Deprived
7	Catholic & Mixed	Rural	Deprived
8	Catholic & Mixed	Urban	Deprived

While it is desirable to spread the sample over all the eight categories it will not be possible to estimate independently in all eight. Therefore, estimation will use a three level categorisation that combines the categories in Table 2. Levels 1 to 5 will form an 'easy to count' group containing about 33 per cent of the population, levels 6 and 7 will form a middle group containing about 50 per cent of the population, with level 8 forming a 'hard to count' group. The distribution of the EDT index across the design groups is given in Table 3 for both the full and collapsed categorisation.

Table 3: Distribution of the EDT Index by Design Group

Level of Index	Number of EDs			
	Belfast	East	West	Northern Ireland
1	-	135	61	196
2	-	59	100	159
3	-	149	145	294
4	80	251	38	369
5	128	63	18	209
Estimation 1	208	657	362	1227
6	180	524	275	979
7	-	166	733	899
Estimation 2	180	690	1008	1878
8	179	180	265	624
Estimation 3	179	180	265	624
All	567	1527	1635	3729

3. Allocating the Sample at Stage One

The approach that is being used in England and Wales forms the basis of the allocation, with some specific differences. The number of Northern Ireland EDs sampled is specified using the same sampling fraction as for England and Wales (approximately 3.6 per cent). For Northern Ireland this implies a first-stage sample of 134 EDs. An initial allocation to the design group by collapsed EDT index is made proportional to the population sizes within the groups. The use of population at this stage rather than number of EDs reflects the fact that the East design group is the largest in terms of population but has less EDs than the West design group. If any allocation is less than eight EDs this is forced to equal eight and the proportional allocation is repeated for the remaining groups. This is to guarantee sufficient sample within each collapsed EDT index category for estimation. The specified sample of EDs is then proportionally allocated (by number of EDs) to the full EDT index. This will ensure that although the sample is designed for estimation using the specified collapsed categories the sample will be spread across all eight categories and allow estimation within a different set of collapsed categories.

The ONS design assumes a second level of stratification below EDT index based on population size. The problem is choosing which age-sex ED counts to use as the size variable. In England and Wales this has been solved by constructing a design variable based on the first three principal components derived from six age-sex groups (males 0-4, females 0-4, males 20-24, males 25-29, males 30-34, females 85+) that suffered high underenumeration in 1991. The within index stratum boundaries are then defined using minimum variance cluster analysis on the three principal components. Optimal allocation based on the design variable is used to allocate a pre-specified within index sample to the size strata such that the relative standard error (RSE) for the estimate of the design variable total is minimised. (RSE is the standard error of the estimated total expressed as a percentage of the total and is also called the ‘coefficient of variation’.)

The same approach has been taken in Northern Ireland. However, proportional allocation, rather than optimal allocation, is used to allocate the specified sample to the within EDT index size strata such that the relative standard error (RSE) for the estimate of the design variable total is minimised across the whole design. This approach is not necessarily as ‘efficient’ as the ‘optimal’ allocation approach but it does have one advantage, it spreads the sample evenly across all the strata. This is important for two reasons; firstly it is robust when you have little information about the expected pattern of underenumeration, secondly it looks fair to the user if they perceive there to be little information about the expected pattern of underenumeration.

Table 4 presents an indicative design for Northern Ireland based on a simulation dataset constructed from the 1991 Census data for Northern Ireland. It demonstrates the structure of the final design for the CCS. The total first-stage sample is 130 EDs plus four EDs included as outliers based on the size of the six population groups. (This is consistent with the approach used by ONS.) The lower sample in the West reflects the less dense population in that design group compared to the East but in general the proportional allocation has led to a sample that is spread evenly over the province.

Table 4: Indicative First-Stage Design for Northern Ireland

Group	EDT Index	Total Number of EDs	Number of Size Strata	ED Sample Size
Belfast	4	80	2	4
	5	128	2	5
	6	180	2	9
	8	179	4	7
	All levels	567	10	25
East	1	134	2	4
	2	59	2	2
	3	148	1	5
	4	251	3	7
	5	63	1	2
	6	523	7	21
	7	166	3	6
	8	180	3	8
	All levels	1524	22	55
West	1	61	1	2
	2	100	2	2
	3	145	2	4
	4	38	1	1
	5	18	1	1
	6	274	4	7
	7	733	6	20
	8	265	5	13
	All levels	1634	22	50
Northern Ireland	All levels	3725	54	130

4. Allocating the Sample at Stage Two

In England and Wales, simulation work has suggested that a random selection of five postcodes from each sampled ED (or less in situations where the ED does not contain five postcodes) is a good compromise between clustering for cost efficiency and spreading the sample of postcodes for statistical efficiency. The expected household sample will be approximately 75 households but this is subject to considerable variation. The variation then needs to be handled by the field management systems of the CCS.

An alternative approach being investigated by GROS draws a sample of postcodes from within the sampled ED until a target number of households is achieved based on the PAF data on number of addresses. With the target household approach care must be taken to:

- a) guarantee that those carrying out the CCS fieldwork are not aware of expected numbers of households
- b) account for the fact that the PAF is number of addresses and in areas where multiple-occupancy is expected the number of households will be higher
- c) decide what to do when a sampled ED does not contain the target number of households
- d) handle large postcodes, some of which contain over the specified target, and the situation where a randomly selected postcode is 'too big' based on the previous draws.

If these problems can be satisfactorily taken care of, the approach has some attractions for management of workloads in the CCS. Northern Ireland's experience in the 1999 rehearsal was that the distribution of postcodes (in terms of population size) was very skewed with many postcodes having very small numbers of households. The two approaches to second stage sampling are currently being considered for Northern Ireland, but the end-point will be ED samples containing, on average, 5 postcodes and 75 households.

5. Preliminary Simulation Results

The 1991 census database has been used to simulate the outcomes of using the ONS process to estimate the true population following a census with less than 100 per cent coverage. The outcomes from using the CCS sample size and design described above have been simulated using the 1991 census data, with the main results summarised in table 5.

Table 5: Results of simulation exercise

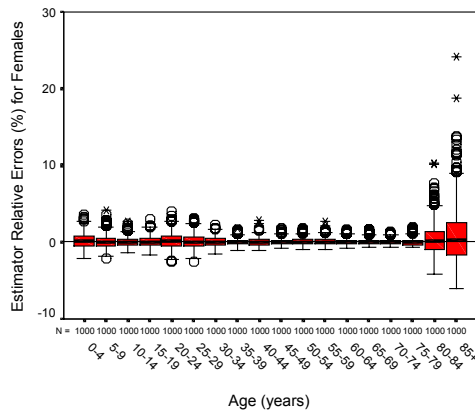
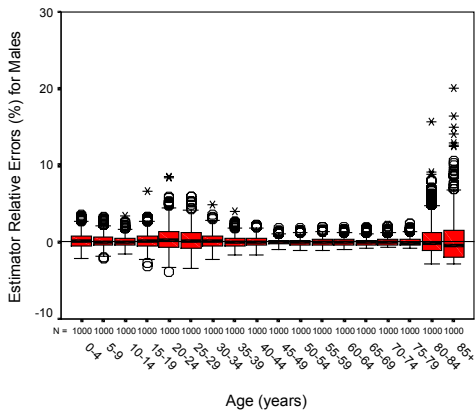
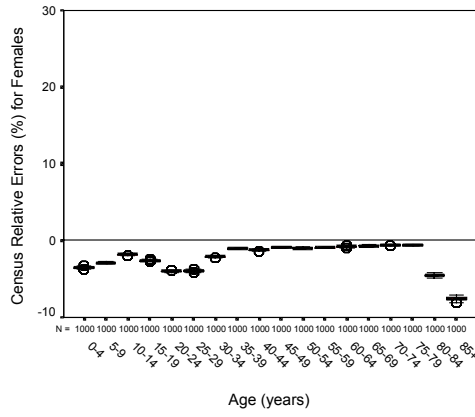
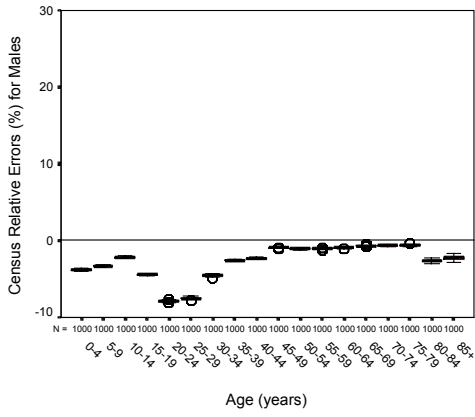
Properties of Estimated Population Total				
Design Group	Census Coverage (%)	Relative Bias (%)	Relative RMSE (%)	Z-Value for Bias
Belfast	92.51	0.19	1.74	3.55
East	98.68	-0.02	0.39	-1.63
West	97.81	0.24	0.75	10.71
Northern Ireland	97.26	0.12	0.46	8.37

The outcome of the simulation is of course largely driven by the underlying assumptions. For the above simulation, the assumptions were

- overall census coverage of 97.25 per cent
- census coverage assumed worst in Belfast
- CCS coverage of 90 per cent of households
- CCS coverage of 98 per cent of individuals within covered households
- Census and CCS independence for dual estimator purposes.

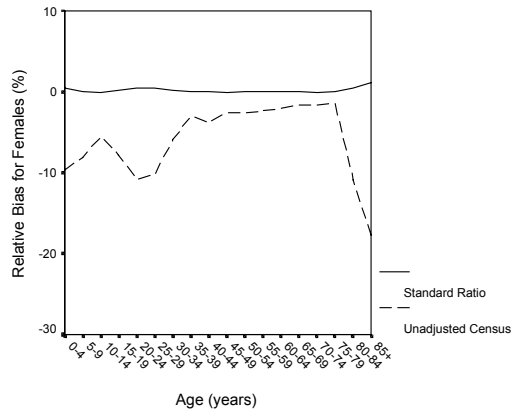
At the Northern Ireland level, the simulation is suggesting a relative RMSE of 0.46 implying a 95 per cent confidence interval of less than 1 per cent. For Belfast, the low census coverage has been recovered (final bias of 0.19 per cent) although with quite a high relative RMSE.

The charts below summarise further results from the simulation. The first charts show how the simulated census underenumeration was distributed by age-group, with high underenumeration concentrated in people in their 20s and the very elderly. The following charts demonstrate how the ONC process has recovered this underenumeration. The following charts then show the original (census) and final (adjusted) relative bias and relative RMSE for each Design Group (EA) separately.

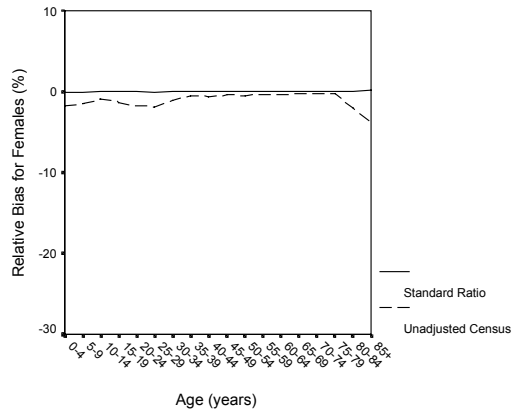
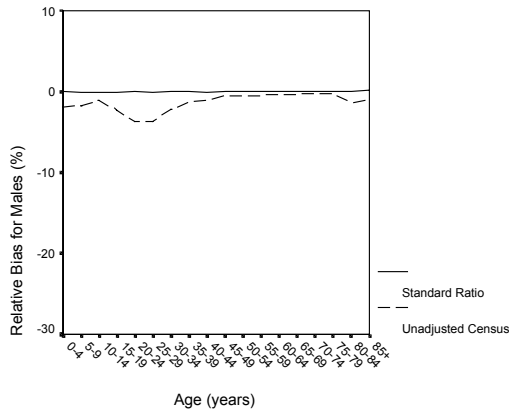


Relative Bias by Design Group

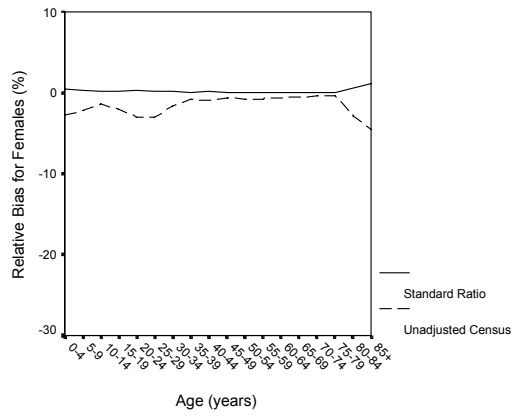
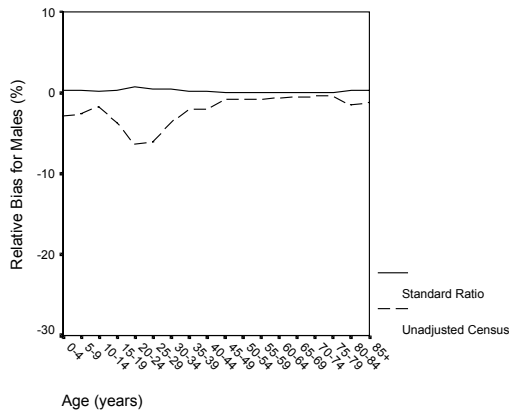
Belfast



East

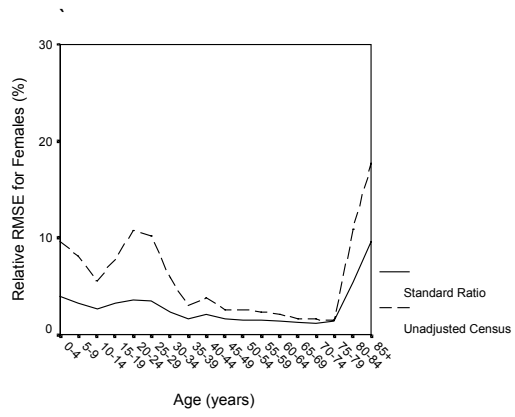
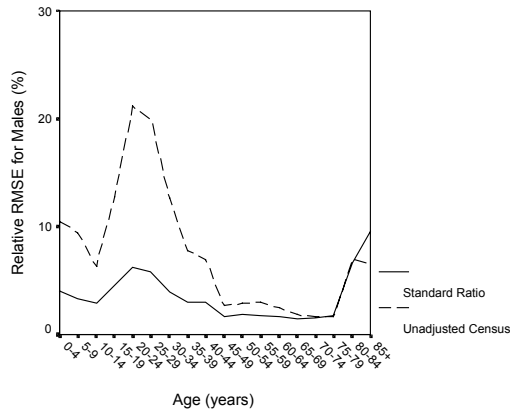


West

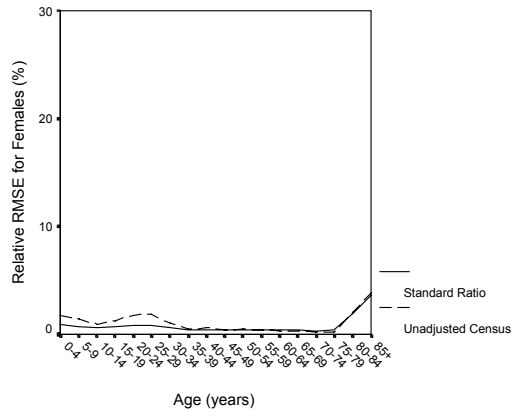
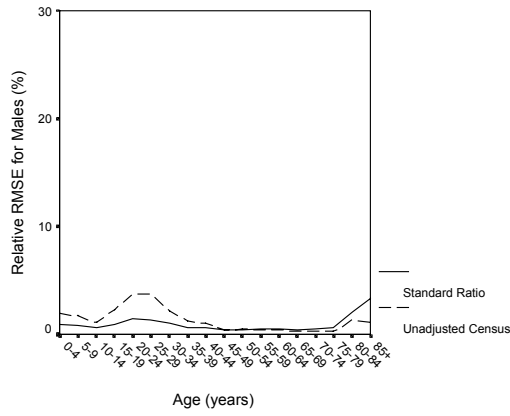


Relative RMSE by Design Group

Belfast



East



West

