

The methodological approach to the 2001 Census



Contents

1. Introduction
 2. Achieving complete coverage of the population: The One Number Census approach
 3. The Census Response Rate
 4. Precision of the One Number Census Estimates
 5. Dependency
 6. Achieving complete coverage of the population: Imputation of response for missing values
- Annex A - One Number Census and Edit and Imputation system results
- Annex B - Application of the One Number Census and Edit and Imputation methodologies to the Census question on religion and religion brought up in.

1. Introduction

Government, local authorities, health and education boards, commercial businesses and the professions need reliable information on the number and characteristics of people and households if they are to conduct many of their activities effectively. This need is currently met by conducting a census every ten years covering the whole of the population. Population estimates are updated every year between censuses using data from the registrations of births and deaths and estimates of migration. Over time these population estimates accumulate inaccuracies and a regular census is necessary to provide information for revising the annual population estimates.

The government needs this kind of information to form policy, to plan services for specific groups of people and to distribute resources effectively such that they are directed to where they are needed. The information must be authoritative, accurate and comparable for all parts of the country. Only a Census can provide the information on a uniform basis both about the country as a whole and about small areas and sub-groups of the population, in relation to one another.

The primary objective of the Census is thus to collect and provide access to a high quality dataset to enable informed decision making by all. In particular, the consultation process prior to the 2001 Census identified the need to have results that were both complete and consistent. A major request by users was that two issues that had affected previous UK Censuses would be addressed:

- the adjustment of Census results for people who were missed by the Census or failed to return a Census form; and
- the adjustment of Census results for respondents who either failed to answer a question, answered inconsistently or answered incorrectly.

The decision to address these issues was noted in the White Paper preceding the Census (Cm 4253, published 1999).

2. Achieving complete coverage of the population: The One Number Census approach

Every effort was made to ensure that everyone was counted in the 2001 Census and a number of initiatives were introduced to maximize coverage. It is widely acknowledged however that no enumeration will ever count everyone. In most countries where a Census like those in the UK is taken, it is standard practice to measure the level of Census underenumeration - that is the number of households and people not counted - either by a post enumeration survey and/or by comparison of Census counts with aggregate data from other, mostly administrative, sources. Historically, this has led to the official population estimates (adjusted for Census underenumeration) being different from the Census count (not adjusted for underenumeration). An aim of the 2001 Census in the UK has been to produce detailed robust estimates of underenumeration and to adjust the Census database, and hence Census counts, for the estimated underenumeration. The process to achieve complete coverage of the population in the Census has been termed the One Number Census (ONC) approach.

The primary source of information in estimating the level of underenumeration in 2001 was the Census Coverage Survey (CCS) – a large postcode based representative sample of approximately 10,000 households, drawn from all areas in Northern Ireland.

The One Number Census process involved a number of stages:

- the Census Coverage Survey was designed and conducted independently of the Census during May and June 2001. Further information on the Northern Ireland Census Coverage Survey can be found at www.nisra.gov.uk/census/censusevaluation/timetable.html
- for those geographical areas where the CCS was conducted, records from the CCS were matched with those from the 2001 Census;

- the populations (adjusted for Census underenumeration) of CCS areas were estimated using dual system estimation techniques - this enabled the population estimates to include persons missed by both the Census and the CCS;
- for CCS areas, statistical models to estimate the characteristics of the adjusted population from the unadjusted census counts were determined;
- the models from CCS areas were applied to the unadjusted census counts for the rest of Northern Ireland to estimate the population of Northern Ireland, adjusted for underenumeration;
- households and persons estimated to have been missed by the Census were then imputed to produce a fully adjusted Census database; and
- all population estimates produced were quality assured using demographic analysis and comparison with aggregate level administrative data.

Further information on the ONC methodologies can be found at

www.nisra.gov.uk/census/censusmethodology/Onenumpercensus.html

www.statistics.gov.uk/census2001/IntroOneNumber.asp

Further information on the ONC Quality Assurance Strategy can be found at

www.statistics.gov.uk/census2001/pdfs/oncinfopaper.pdf

3. The Census Response Rate

It has been estimated from the Census Coverage Survey that households identified by enumerators, but from which a form was not returned, accounted for 3.0 per cent of the population. Some households were also missed by enumerators, or some people were not included in Census returns. The Census Coverage Survey has estimated that this represents a further 1.8 per cent of the population. It is thus estimated that 95.2 per cent of the population in Northern Ireland responded to the 2001 Census. The response rate in England and Wales was 94 per cent while in Scotland it was 96 per cent.

The Census and the Census Coverage Survey were designed to produce robust estimates of underenumeration and to incorporate these in the final Census output. The results presented thus provide 100 per cent coverage of the population. It is estimated that 95.2 per cent of the population responded to the census. Considering the population in private households, the overall response rate was 95.3 per cent as further described below.

Area	Census Response Rate (people in households)	Percentage imputation for households identified by enumerators, although no completed Census forms returned	Percentage imputation for persons missed and persons in households not identified by the Census
Northern Ireland	95.3	3.0	1.7

The response rates varied by Local Government District and by population subgroup, such as five-year age group. Further details are shown in Annex A.

4. Precision of the One Number Census Estimates

The estimates of underenumeration, and thus the Census results, are based upon a sample survey (the Census Coverage Survey) and are therefore subject to sampling error. Standard statistical techniques have been used to calculate these error levels and produce confidence intervals for the One Number Census results. The error levels associated with the ONC estimates are mainly determined by the magnitude of the estimated underenumeration and the sample size of the CCS. The resulting 95 per cent confidence interval for the Northern Ireland population is +/- 0.7% or about +/- 12,000.

It should be noted that as with virtually all statistical analyses of precision, these calculations do not capture all sources of variation. There will also be, for example, response, capture and coding errors. These issues will be described in full in a forthcoming 2001 Census Quality Report.

5. Dependency

Within CCS areas, the One Number Census process estimates the true population of an area through combining the results of the Census with those of the CCS and estimating the number of people missed both by the Census and the CCS. The estimation of the number of people missed by both the Census and the CCS, through a method called dual estimation, requires that the statistical dependence between the Census and the CCS be determined. A simpler estimation process can be determined by assuming independence of the Census and the CCS; in practice this assumption means that the probability of a given person being identified by the CCS is independent of the probability of their being identified by the Census. Steps were taken to minimize the dependence between the Census and the CCS but it is acknowledged that, in practice, the assumption of complete independence is difficult to maintain. The final One Number Census estimates did not assume independence between the Census and the CCS, and the level of dependence was estimated. This issue will be described in full in the forthcoming 2001 Census Quality Report.

6. Achieving complete coverage of the population: Imputation of response for missing values

The adjustment of Census results for respondents who either failed to answer a question, answered inconsistently or answered incorrectly was made possible using an Edit and Donor Imputation System (EDIS) that was devised for the 2001 Census. The system was created to fill in a number of gaps in the records for enumerated people and households. At a later stage in processing the database was adjusted using the One Number Census process described above.

EDIS contained four initial components, these were:

- **Multi-tick rules** when more than one box was ticked but only one option was allowed;
- **Range checks** to prevent answers being outside an acceptable range;
- **Filter rules** to resolve some inconsistencies and to decide which fields should be set to 'No Code Required' where questions were answered but should not have been; and
- **Edit rules** to deal with missing items or responses which appeared to be in error or inconsistent when compared with other data. Edit either set a specific value or left it to imputation to determine a value.

After the application of these components the Imputation component was applied. The basis for the Imputation component is to search for a single “donor” person to supply all the missing variables for a recipient person. The method searched for a donor person who was similar using a number of other Census variables. A series of criteria were drawn up to determine what was meant by ‘similar’. A suitable selection of variables known as Primary Matching Variables was defined to match on for each missing item. Values were copied from the donor person to fill the missing values on the record of the recipient person.

If more than one suitable donor person was found a donor was selected from a similar household. This was based on the age, sex, marital status and relationship between the people in the household. For the Community

Background, Ethnicity, Language, Address one year ago and Country of birth variables, the system also considered the responses given by the rest of the household. If there was still more than one suitable donor the person in the geographically closest household was picked.

A similar method was applied for household variables (e.g. tenure) and people living in communal establishments. If several people in a household had missing responses or some of the responses to the household questions were missing the system tried to select all the donors from the same household in order to preserve household structure.

An initial paper which details the EDIS methodology more fully can be found at

<http://www.statistics.gov.uk/census2001/pdfs/ag0013.pdf>

It should be noted that this paper details the methodology as proposed in August 2000 and some small changes in application occurred since. This issue will also be described in full in the forthcoming 2001 Census Quality Report.

The application of the EDIS system means that missing responses have been catered for in all Census topics (except a person's current religion). The system was designed to remove bias that would otherwise have been created in the final statistics by missing responses.

The application of the edit and imputation and the One Number Census processes for the question on religion and community background: religion or religion brought up in are described in Annex B.

Census Office

September 2002

Revised and updated January 2003

Annex A: One Number Census and Edit and Imputation system results



1. As described in the main paper the published census results have been adjusted for underenumeration and missing information within returned forms. This paper quantifies the effect of underenumeration. For a number of key statistics, the tables below show the distribution observed solely among those who were on returned census forms and the distribution in the published estimates adjusted for underenumeration. The tables given are:

Table 1 – Marital status

Table 2 – Ethnic Group

Table 3 – Religion

Table 4 – Community Background

Table 5 – Limiting long term illness

Table 6 – Gender

Table 7 – Age

Table 8 – Area of residence – Local Government District

2. For example, table 1 shows that on returned census forms 32.2% of the population gave their marital status as single, whereas the adjusted estimate was 33.1% of the population. This shows that those people missed by the census were disproportionately more likely to be single, or equivalently that the census response rate for single people was lower than the average throughout the population.
3. The distributions given below for census respondents are those following adjustment for cases where respondents omitted that particular question. This paper will be updated in the near future with information on the proportion of respondents failing to answer each topic.

Table 1: NI Distribution of Census Output and Respondents (Marital Status)

	<u>Census Output</u>	<u>Census Respondents(*)</u>
Single (never married)	33.1%	32.2%
Married	48.5%	49.5%
Re-married	2.7%	2.7%
Separated (but still legally married)	3.8%	3.7%
Divorced	4.1%	4.0%
Widowed	7.8%	7.9%

(*) For those people who failed to respond to this question their information was derived using EDIS. The method used was as noted in section 6 of the main paper. The primary matching variables used were Relationship to Person One on the form, Age, Sex and Highest Qualification.

Table 2: NI Distribution of Census Output and Respondents (Ethnic Group)

	<u>Census Output</u>	<u>Census Respondents(*)</u>
White	99.15%	99.20%
All Other Ethnic Groups	0.85%	0.80%

(*) For those people who failed to respond to this question their information was derived using EDIS. The method used was as noted in section 6 of the main paper. The primary matching variables used were Country of Birth, Age, Marital Status and Religion. For this variable the Ethnic Group of the other people in the household was also taken into account.

Table 3: NI Distribution of Census Output and Respondents (Religion)

	<u>Census Output</u>	<u>Census Respondents(*)</u>
Catholic	40.3%	39.9%
Presbyterian	20.7%	21.1%
Church of Ireland	15.3%	15.5%
Methodist	3.5%	3.6%
Other Christian	6.1%	6.1%
Other Religion and Philosophy	0.3%	0.3%
No Religion and Not Stated	13.9%	13.5%

(*) For those people who failed to respond to this question their information was not adjusted by EDIS thus there is a not stated category in the Religion output.

Table 4: NI Distribution of Census Output and Respondents (Community Background)

	<u>Census Output</u>	<u>Census Respondents(*)</u>
Catholic	43.8%	43.3%
Protestant and Other Christian (including Christian related)	53.1%	53.8%
Other religions and philosophies	0.4%	0.4%
None	2.7%	2.5%

(*) For those people who failed to respond to this question their information was derived using EDIS. The method used was as noted in section 6 of the main paper. The primary matching variables used were Irish Language, Ethnic Group and Age. For this variable the Community Background of the other people in the household was also taken into account.

Table 5: NI Distribution of Census Output and Respondents (Limiting Long Term Illness)

	<u>Census Output</u>	<u>Census Respondents(*)</u>
Yes - Has limiting long-term illness	20.4%	20.4%
No limiting long-term illness	79.6%	79.6%

(*) For those people who failed to respond to this question their information was derived using EDIS. The method used was as noted in the EDIS paper. The primary matching variables used were Activity Last Week, Age group and Company Size.

Table 6: NI Distribution of Census Output and Respondents (Gender)

	<u>Census Output</u>	<u>Census Respondents(*)</u>
Male	48.7%	48.5%
Female	51.3%	51.5%

(*) For those people who failed to respond to this question their information was derived using EDIS. The method used was as noted in section 6 of the main paper. The primary matching variables used were Activity Last Week, Relationship to Person One on the form, Marital Status and Occupation group.

Table 7: NI Distribution of Census Output and Respondents (Age)

	<u>Census Output</u>	<u>Census Respondents(*)</u>
0-4	6.8%	6.7%
5-9	7.3%	7.3%
10-14	7.9%	8.0%
15-19	7.7%	7.6%
20-24	6.5%	6.3%
25-29	6.8%	6.6%
30-34	7.6%	7.5%
35-39	7.7%	7.7%
40-44	7.0%	7.0%
45-49	6.1%	6.1%
50-54	5.8%	5.9%
55-59	5.3%	5.4%
60-64	4.4%	4.5%
65-69	3.9%	4.0%
70-74	3.4%	3.5%
75-79	2.8%	2.8%
80-84	1.8%	1.8%
85+	1.4%	1.4%

(*) For those people who failed to respond to this question their information was derived using EDIS. The method used was as noted in section 6 of the main paper. The primary matching variables used were Activity Last Week, Relationship to Person One on the form, Marital Status and Occupation group.

Table 8: NI Distribution of Census Output and Respondents (Local Government District)

	<u>Census Output</u>	<u>Census Respondents(*)</u>
Antrim	2.9%	2.9%
Ards	4.3%	4.4%
Armagh	3.2%	3.2%
Ballymena	3.5%	3.5%
Ballymoney	1.6%	1.6%
Banbridge	2.5%	2.5%
Belfast	16.5%	15.9%
Carrickfergus	2.2%	2.3%
Castlereagh	3.9%	4.0%
Coleraine	3.3%	3.3%
Cookstown	1.9%	1.9%
Craigavon	4.8%	4.8%
Derry	6.2%	6.1%
Down	3.8%	3.8%
Dungannon	2.8%	2.9%
Fermanagh	3.4%	3.4%
Larne	1.8%	1.9%
Limavady	1.9%	1.9%
Lisburn	6.4%	6.5%
Magherafelt	2.4%	2.4%
Moyle	0.9%	0.9%
Newry and Mourne	5.2%	5.2%
Newtownabbey	4.7%	4.8%
North Down	4.5%	4.6%
Omagh	2.8%	2.8%
Strabane	2.3%	2.2%

(*) Locational information is embedded into census processing and therefore EDIS was not applied here.

Detailed results of the extent of application of the EDIS system will given in a future paper. However results of the application of EDIS for Community Background are given in Annex B attached.

Annex B: Application of the One Number Census and Edit and Imputation methodologies to the Census question on religion and religion brought up in



1. The Census question on religion has been analysed using the same principles as all other topics. The form of the religion question in the 2001 Census in Northern Ireland is given in Figure 1 below.

Figure 1: 2001 Census Question on Religion

The image shows a screenshot of the 2001 Census form for Northern Ireland, specifically questions 8, 8a, and 8b. Question 8 asks if the respondent regards themselves as belonging to any particular religion, with options for Yes (Go to 8a) and No (Go to 8b). Question 8a asks for the respondent's current religion, with options for Roman Catholic, Presbyterian Church in Ireland, Church of Ireland, Methodist Church in Ireland, and Other (please write in). Question 8b asks for the religion the respondent was brought up in, with the same options as 8a, plus a 'None' option. Each 'Other' option includes a two-row grid for writing in the response. Navigation arrows point to the next question in the sequence.

8 Do you regard yourself as belonging to any particular religion?

Yes **▶ Go to 8a**

No **▶ Go to 8b**

8a What religion, religious denomination or body do you belong to?

Roman Catholic

Presbyterian Church in Ireland

Church of Ireland

Methodist Church in Ireland

Other, *please write in*

▶ Go to 9

8b What religion, religious denomination or body were you brought up in?

Roman Catholic

Presbyterian Church in Ireland

Church of Ireland

Methodist Church in Ireland

Other, *please write in*

None

2. The responses to the first part of the question (Q8 and Q8a) have been used to report 'Religion'. Those who did not have a religion or did not

respond to the question have been categorised as “No Religion or Religion Not Stated”.

3. The Religion output was augmented where applicable by the response to Q8b on religion brought up in to produce information on ‘Community Background – religion or religion brought up in’. The utility of such data is discussed in more detail at paragraph 8 below.
4. Two key processes were applied in the analysis of the 2001 Census. These are the adjustment of Census results for people who were missed by the Census or failed to return a Census form and the adjustment of Census results for respondents who failed to answer a question, answered inconsistently or answered incorrectly. The processes used to achieve this were called One Number Census (ONC) and Edit and Imputation (EDIS) respectively. The processes are described in detail in the main paper. The application of these methods to the Religion and Community Background output is detailed in the remainder of this annex.

Religion Output

5. As described in section 2 of the main paper, and indicated in the 2001 Census White Paper (cm 4253, published 1999), all Census outputs are designed to provide complete coverage of the population. This has been achieved by using information from both the Census and the Census Coverage Survey that gave information on the coverage by the Census of households and people within households. Religion was one of the demographic characteristics collected in the Census Coverage Survey, enabling Census coverage to be determined separately for different religious groups.
6. Table 1 below shows how the Census figures for the total population were constructed from returned forms and estimates for people missed by the Census.

Table 1: Religion with ONC imputation

	Data from returned forms	%	People missed by the Census as determined by the Census Coverage Survey	%	Complete population	%
Catholic	640,025	39.9	38,437	47.1	678,462	40.3
Protestant and other Christian	741,940	46.3	25,984	31.8	767,924	45.6
Other religions and philosophies	4,662	0.3	366	0.4	5,028	0.3
No religion or Not Stated	217,013	13.5	16,840	20.6	233,853	13.9
Population	1,603,640	100	81,627	100	1,685,267	100

7. The reporting of the religion question is different from other Census questions in one aspect. The Census White Paper (cm 4253, paragraph 155) described how omissions would be resolved before releasing output. This is a better solution for the interpretation and utility of the resultant data than simply coding the response “Not Stated”. However, for the Census output on religion the EDIS process was not applied and thus the “Not Stated” category has been retained and responses given in the table on religion are as provided (or omitted) by respondents. A further breakdown of the “No religion or religion not stated” category is described below in paragraph 13.

Community Background Output

8. The Census White Paper also included proposals for collecting additional information on religion brought up in where respondents indicated they had ‘no religion’. The resultant community background data are of use in informing employment and other equality monitoring.

9. Those stating at Q8 that they did not belong to a religion were asked in which religion they were brought up in (Q8b). Some 9.1% of the total population answered this question (or 66% of those who did not state a current religion), the majority of who were brought up as Protestant or in no religion (“None”).

10. Over 95% of the population thus indicated either the religion they currently belonged to or, if they did not currently belong to a religion, the religion in which they were brought up. As outlined in the Census White Paper the remaining individuals who did not identify either a current religion or the religion in which they were brought up - 4.8% of the population - were allocated to a community background using the standard statistical methods used for all other Census questions. The outcome of this process is detailed in table 2. The Community Background distribution of these individuals is not markedly different from the distribution in the population as a whole.

Table 2: EDIS imputation of community background where neither religion nor religion brought up in was indicated

Community Background	Number whose community background was imputed	% of total whose community background was imputed	% of total population
Catholic	31,906	39.6	1.9
Protestant and other Christian	44,773	55.5	2.7
Other religions and philosophies	466	0.6	0.0
None	3,525	4.4	0.2
All backgrounds	80,670	100	4.8

11. Combining the information from those who answered Q8a (current religion), those who answered Q8b (religion brought up in) and the imputed community background of those who gave no information, as noted in paragraph 10 above, gives the final published figures for Community Background.

Table 3: The three component parts of Community Background

	Gave current religion	% of total population	Gave religion brought up in	% of total popn	Not stated - Community Background Imputed	% of total popn
Catholic	678,462	40.3	27,044	1.6	31,906	1.9
Protestant and other Christian	767,924	45.6	82,680	4.9	44,773	2.7
Other religions and philosophies	5,028	0.3	1,075	0.1	466	0.0
None	-	-	42,384	2.5	3,525	0.2
All backgrounds	1,451,414	86.1	153,183	9.1	80,670	4.8

12. These results were combined to give the headline reported results presented in December and repeated in table 4. In addition a further column is shown which details the Community Background distribution of those people who indicated a religion or religion brought up in.

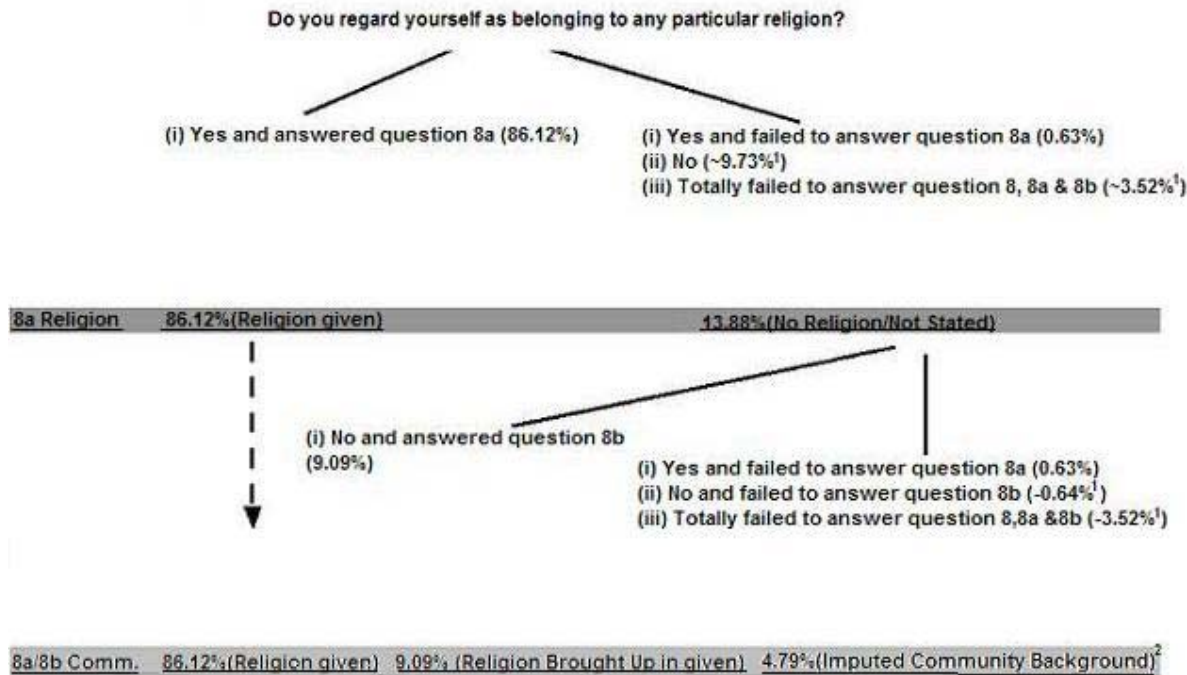
Table 4: Community background estimates (reported and based solely on those who gave answer)

	Reported		Those who gave answer	
	Number	Percentage	Number	Percentage
Catholic	737,412	43.8	705,506	44.0
Protestant and other Christian	895,377	53.1	850,604	53.0
Other religions and philosophies	6,569	0.4	6,103	0.4
None	45,909	2.7	42,384	2.6
All backgrounds	1,685,267	100	1,604,597	100

Processing Flow

13. The flow diagram below shows how the Census cases have been distributed among the religion and community background categories. It also provides estimates of the number of people who did not complete any part of question 8 and the number who said “No” at the filter part of question 8. Note that these particular numbers are estimates because these two groups were merged for processing purposes at an early stage of the analysis.

Figure 1: Flow Diagram for Religion and Community Background Results from 2001 Census in NI



1. Based on the pre-Output Census database, the percentage of persons who are "FullNot Stated" is estimated to be 3.52% and the percentage of persons who stated that they had no religion but didn't answer question 8b is estimated to be 0.64%.
 2. For these variables imputation was only applied to Community Background.

Conclusion

14. In conclusion, it should be noted that in reporting this material Census Office has not changed any given response to the religion question provided by any respondent. Those who indicated that they do not belong to a religion, religious denomination or body have been counted as such in the Census output table on religion. Similarly, those who indicated they belonged to no religion and were not brought up in any religion are reported as 'none' in the output table on community background. In no case were respondents allocated to, respectively, another religion or community background.

15. Whilst the Census collects information about each person and household it is not concerned with facts about individuals – the purpose is to provide facts about the community, and groups within the community, as a whole. In this context the purpose of imputing a response to a particular question, left blank in a Census return, is to make the best estimate of the population distribution for that Census question. The inclusion of 'not stated' replies may make interpretation of the output more difficult, and where the distribution of the 'not stated' group is skewed in a particular way can result in misinterpretation. However, the use of imputation is solely for the purposes of producing aggregate Census statistics. Microfilm images of individual Census forms are retained securely as originally completed under closed conditions. Statistics are released for population groups and geographical areas only and precautions are taken such that published tabulations and abstracts do not reveal information about identifiable individuals or households with special precautions applied to statistical outputs for small areas.