



General Register Office for Scotland
information about Scotland's people

Statistical Evaluation of the Census 2006 Test In Scotland

2nd May 2007

Table of Contents

1.	Introduction	6
1.1	Purpose	6
1.2	What is covered in this report.....	6
1.3	What is not covered in this report.....	6
1.4	Structure of report	6
2.	Summary of Key Conclusions and Recommendations	8
2.1	Key Conclusions	8
2.1.1	Enumeration Method.....	8
2.1.2	Income Question	8
2.1.3	Form Design.....	8
2.1.4	Second Residency and Visitors.....	8
2.2	Key Recommendations	8
2.2.1	Data Capture.....	8
2.2.2	Multi-ticking	9
2.2.3	Income Question	9
2.2.4	Past Work Question	9
2.2.5	Marital/Civil Partnership Status Question	9
2.2.6	Ethnic Group	9
2.2.7	Respondents' Views on Questions.....	9
2.2.8	Placeholder Forms	9
3.	Background to 2006 Census Test.....	11
3.1	Purpose.....	11
3.2	Design.....	11
4.	Data	12
4.1	Forms Delivered.....	12
4.2	Data Received.....	14
4.3	Quality Checking and Editing	15
4.4	Data Completeness.....	18
4.4.1	Household data	18
4.4.2	Person data.....	22
4.4.3	Placeholder data	29
4.4.4	Communal Establishment Data.....	30
5.	Analysis	32
5.1	Household Form Return Rates by Delivery Method and Income Question 32	
5.2	Responses to the Income Question	36
5.3	Item response – Household Questions (excluding income)	44
5.4	Household Questions Filter	45
5.5	Item Response – Person Questions.....	45
5.5.1	Single tick questions	46
5.5.2	Multi-tick questions.....	48

5.5.3	Write-in questions	49
5.6	Person Question Filters.....	51
5.7	Form Completeness.....	56
5.7.1	Completeness from Household Questions	56
5.7.2	Completeness from Person Questions.....	58
5.7.3	Overall Completeness from Household Form	59
5.7.4	Breakdown by Census District of Completeness Ratios	61
5.7.5	Breakdown of Completeness by Income and Enumeration Method.....	62
5.7.6	Breakdown of Person Questions Completeness by Age and Sex.....	62
5.8	Ethnic group	64
5.9	Marital and Civil Partnership Status	65
5.10	Second Residency	66
5.11	Visitors	67
5.12	Respondents' Views on the Questions.....	70
5.13	Placeholder Forms	73
5.14	Communal Establishments.....	74
5.15	Test Area Profiles.....	76
5.15.1	Deprivation	76
5.15.2	Age and Sex of all respondents	83
5.15.3	Person One on the Household Questionnaire.....	87
5.15.4	Further Work	92
	It would be informative to also have an ethnicity profile for each Census Test District..	92
6.	Conclusions and Recommendations.....	93
6.1	Conclusions.....	93
6.1.1	Enumeration Method.....	93
6.1.2	Income Question	93
6.1.3	Item Response	93
6.1.4	Form Completeness	94
6.1.5	Civil Partnerships	94
6.1.6	Second Residency and Visitors.....	94
6.1.7	Controversial Questions	95
6.2	List of Recommendations.....	95
	Annex A -The Scottish 2006 Census Test: How well did respondent complete the questionnaire?	98
	Annex B - 2006 Census Test Follow Up Survey.....	99
	Annex C - Frequency counts for Test questions.....	100
	Annex D - Cross-tabs of Respondents' Views with Responses to Questions.....	101

1. Introduction

1.1 Purpose

To inform GROS of the outcome of the statistical evaluation of the Census 2006 Test held in five Test areas in Scotland on 23rd April 2006.

1.2 What is covered in this report

This report details principally the findings of the Statistical Methodology and Geography (SMAG) branch of GROS Census Division with regard to the statistical evaluation of the Census 2006 Test in Scotland. Although the work of other branches involved in the Test is covered elsewhere, this report does touch on the interaction of branches throughout the statistical Test evaluation work, as the output analysis is essentially the last stage of the Test, and issues and problems which occur earlier in the process do affect the final outcome at this stage of the work.

An evaluation plan, covering the entire process of the Census Test, was presented to the Census Programme Board (CPB) at various times by Ian Maté, who was responsible for designing the Test. The latest version was version 3, presented to the CPB in March 2005, reference SCPB 05/022. It was written before the questionnaire was finalised and hence refers to some questions that were not on the form used in 2006. From this plan SMAG determined what statistical evaluation was possible given the form used in the Test. These include items listed under 2d - income question (see sections 5.1, 5.2 and 5.12 of this report, and Annexes B and D) and 2e - other questions (sections 5.3, 5.5, 5.8, 5.9 and 5.12, and Annexes B, C and D). This report focuses principally on results from the household and person questions. Results of the follow-up survey are presented in Annex B.

1.3 What is not covered in this report

This report paper does not detail any in depth evaluation of the questionnaire form design. However ONS Data Collection Methodology staff examined a random sample of 170 completed Test Household questionnaires. The results of their investigation are found in a report in Annex A. The ONS report looks at particular questions which differed from the English and Welsh Test questions and provides a pointer for further work that GROS should consider undertaking in finalising the Scottish Census Household form.

1.4 Structure of report

A summary of key conclusions and recommendations is contained in section 2 below. Section 3 contains background to the purpose of the Test and the design approach. Section 4 describes what data was received from Census IT and how

this data was quality checked and edited before analysis. Section 5 presents the results of analysis of the Test data. Finally, section 6 presents the conclusions reached and brings together the recommendations made throughout the report.

2. Summary of Key Conclusions and Recommendations

This section presents the most significant conclusions and recommendations coming out of this evaluation. A fuller description of conclusions is found at section 6.1, and a comprehensive list of all the recommendations in the report is given in section 6.2.

2.1 Key Conclusions

2.1.1 Enumeration Method

More forms were returned when they were delivered by hand than when they were sent through the post. However it is possible that hand delivery had a negative effect on response to some individual questions; more work needs to be done on this.

2.1.2 Income Question

There is conflicting evidence on this. More forms that had an income question were returned than those without one. However the question had a relatively high non-response rate and a significant level of dissatisfaction among respondents.

2.1.3 Form Design

Most problems with individual questions were due to the layout of the form and ordering of the questions. In particular, some questions should have had a filter applied as they were not relevant to all sections of the population, and the design of the language question was not successful.

2.1.4 Second Residency and Visitors

The quality of address information for second residences was not good, with less than half providing a valid postcode. This puts a question mark over the usefulness of this information in a Census - it is possible that a postcode could be derived from the address information, but this is resource-intensive and work has yet to be carried out on the quality of addresses provided. Visitor information was better, with about two-thirds giving a valid postcode for their usual residence, although it is not possible to say how many visitors were omitted from the form as visitor information was not covered in follow-up work.

2.2 Key Recommendations

2.2.1 Data Capture

Better communication is needed between IT and statistical staff throughout the process. In particular, statistical staff need to obtain a clearer idea of what data to

expect, including its format and volume, and the timetable for its delivery to statistical staff.

2.2.2 Multi-ticking

More work is required on the reasons for an apparent preponderance of multiple responses to certain questions, in particular ethnicity.

The data capture system needs to be able to distinguish between marks that should be captured and crossings out. It is incorrect for an amended response to be captured as a multiple response.

2.2.3 Income Question

It is unclear how useful the data obtained from this question is. Work is needed to identify suitable alternative sources of income data and compare them with responses to the Test to determine whether they accurately reflect the income pattern of the areas covered. Work should also be done to compare the effectiveness of an individual income question, as being used in the Test being carried out by ONS and NISRA in 2007, with the household question used in Scotland.

2.2.4 Past Work Question

The question "have you ever worked?" caused problems in the Test. Further analysis is needed into why this was the case, with particular reference to differing response by age, and what lessons can be learned for 2011.

2.2.5 Marital/Civil Partnership Status Question

This question needs to come earlier on the form so that it can form part of the validation rule for individual responses.

2.2.6 Ethnic Group

More in-depth analysis is needed of this question.

2.2.7 Respondents' Views on Questions

More investigation is needed of these: the written comments need to be analysed, and an analysis also needs to be carried out of individual responses to questions that respondents objected to.

2.2.8 Placeholder Forms

Work is needed to determine the usefulness of these forms compared to other options. In particular, late returns can be used to assess the quality of

enumerator estimates of such information as vacant properties and number of usual residents.

3. Background to 2006 Census Test

3.1 Purpose

The purpose of the 2006 Census Test was to test a range of strategies, procedures and instructions to inform future decisions on how to conduct the 2011 Census. Census Tests have been found to be valuable exercises in defining the scope and methodology of the full Census, and this is expected to be the case again for the 2011 exercise.

3.2 Design

The 2006 Census Test was sub-divided into five areas, known as Census Districts: Breadalbane, Lochaber, West Dunbartonshire, North Glasgow and South Glasgow. The selection of areas was purposive rather than a random sample selection, which means that results cannot necessarily be generalised to Scotland as a whole.

The areas were selected to cover urban, rural and semi-rural locations, and each was chosen for the particular challenges that it provides to enumeration. The two urban areas were both in Glasgow. The South Glasgow district in particular contained areas of high ethnic diversity. The North Glasgow district was known to contain high numbers of asylum seekers. West Dunbartonshire had a particularly low enumeration rate in 2001, which has been attributed to poor housing stock and deprivation. The rural areas, Breadalbane and Lochaber, have large numbers of crofts and holiday homes. In addition, Lochaber is believed to have a high concentration of Gypsy/Traveller sites.

The Census Test took place on Sunday 23rd April 2006. The date was chosen to avoid holidays so that the results closely approximated to those for the usually resident population. The Test was voluntary and the response rates were expected to be much lower than for a compulsory Census.

The Census Test was a two-factor design with five blocks (one block for each Census District), using purposive sampling. The first factor, enumeration methodology, had two levels – post out- post back/collection and delivery- post back/collection; and the second factor was income, with half the forms having an income question, the others not. The two different enumeration methods were trialled in order to explore ways of reducing costs while maintaining coverage and quality, following on from the suggested link between postal methods and reduced coverage and quality¹.

¹ House of Commons Treasury Select Committee Report HC310 “The 2001 Census in England and Wales”, 2001-02 Session

4. Data

4.1 Forms Delivered

Prior to the receipt of data we required to know how many Census Test forms were delivered to each area, which method of enumeration was used and the split of the income/no income forms. No detailed record exists of how many Household forms field staff and Royal Mail actually attempted to deliver. However from the address list provided to the printers we can determine how many forms we initially planned to deliver, though these figures omit any new addresses discovered during enumeration to which forms were subsequently delivered.

Table 1 details the breakdown of the number of Household forms targeted to be delivered by planned delivery method and whether an income question was included.

Table 1 - Targeted Delivery of Test Forms

Census District	Delivery	Income/ No Income	Number of forms	Total
01 - North Glasgow	Hand delivered	No Income	2914	11994
	Hand delivered	Income	2886	
	Post out	No Income	3067	
	Post out	Income	3127	
02 - South Glasgow	Hand delivered	No Income	3816	15532
	Hand delivered	Income	3854	
	Post out	No Income	3898	
	Post out	Income	3964	
03 - West Dunbartonshire	Hand delivered	No Income	2797	11340
	Hand delivered	Income	2713	
	Post out	No Income	2937	
	Post out	Income	2893	
04 - Lochaber	Hand delivered	No Income	1479	6276
	Hand delivered	Income	1669	
	Post out	No Income	1588	
	Post out	Income	1540	
05 - Breadalbane	Hand delivered	No Income	1595	6521
	Hand delivered	Income	1638	
	Post out	No Income	1638	
	Post out	Income	1650	
			Total	51663

Figure 1 to Figure 5 show the above information in chart form.

Figure 1 - Target Delivery North Glasgow

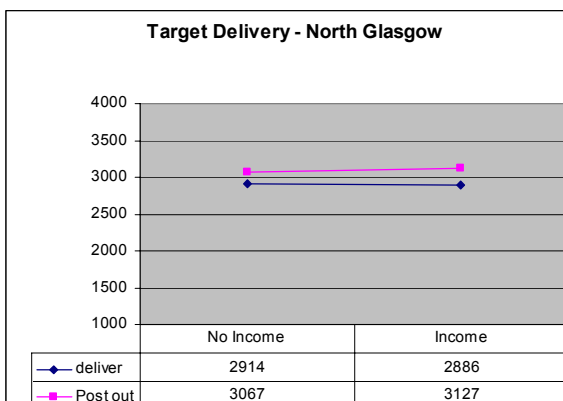


Figure 2 - Target Delivery South Glasgow

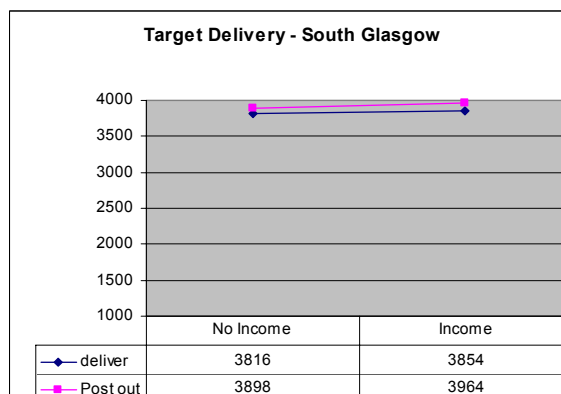


Figure 3 - Target Delivery West Dunbartonshire

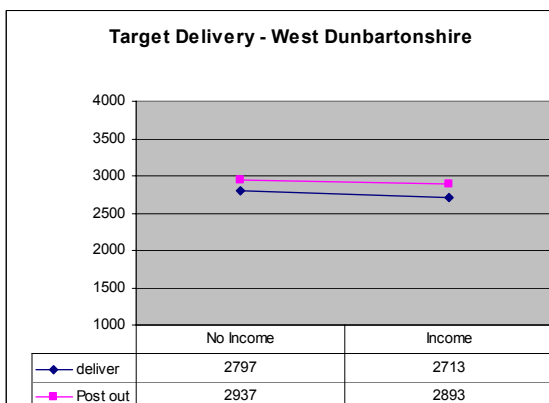


Figure 4 - Target Delivery Lochaber

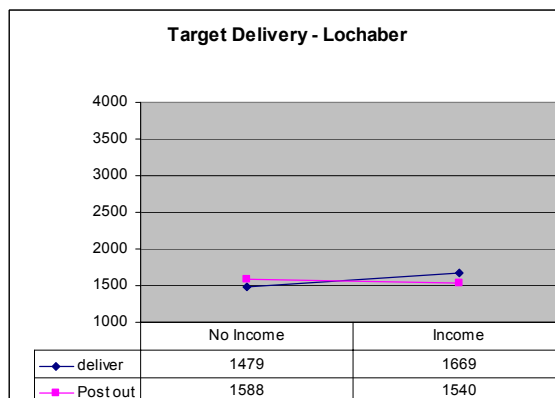
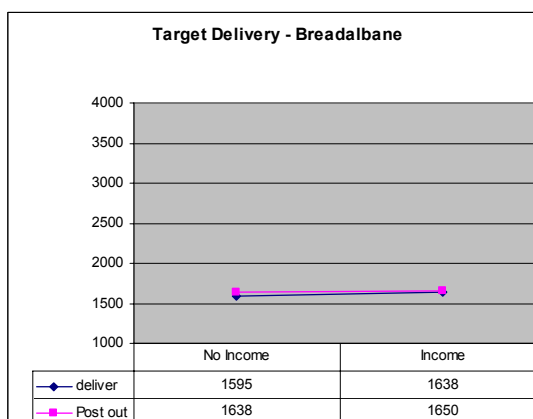


Figure 5 - Target Delivery Breadalbane



Note that in almost all areas there were more forms targeted to be posted out than hand delivered. The split between income and non-income was almost even within each area.

In some cases an additional delivery of forms had to be made. These included forms for new addresses that were not on the original list, addresses that had been missed by enumerators and householders who had lost their original form. In these cases, the new form sent out always had the income question present. Table 2 gives a breakdown of the 398 income forms delivered in areas where a non-income form was originally targeted to be delivered. The numbers are small compared to the total number of forms and so unlikely to have had an impact on the results of the Test.

Table 2 - Income forms delivered in non income areas

Census District	Number of income forms delivered
01 -North Glasgow	93
02 –South Glasgow	85
03 – West Dunbartonshire	136
04- Lochaber	47
05 – Breadalbane	37
	398

Recommendation 1

For future work it is important that data on the number of forms actually delivered should be readily available to analysts.

4.2 Data Received

The data was received in SAS dataset format from Census IT on 7th November 2006. A further version of the Household data was received on 9th November 2006 following some revision to the variable detailing visitor numbers. The number of records remained the same.

The data was provided as 8 SAS datasets as detailed in table 3. The date at the end of each dataset name reflects the date of the final delivery of the data in question.

Table 3 - Number of Records in each SAS dataset

SAS dataset	Contents	Number of records
COMMUNAL1NOV	Communal Establishment data	56
HOUSEHOLD9NOV	Household data	23,905
PERSNOHH1NOV	Data for persons with no associated household	7
PERSON1NOV	Data for persons in households and communal establishments	48,129
PERSONREMOVED8NOV	Persons removed by the 2 of 3 rule (see section 4.4.2)	828
PLACEHOLDER1NOV	Data from Placeholder forms	28,697
VIEWS1NOV	Respondent views	23,896
VISITORS1NOV	Visitor data	892

The structure and content of these datasets were difficult to work with. The original specification was inadequate for the eventual purposes. In some cases it was necessary to recreate variables and so time and effort were duplicated.

Recommendation 2

It is recommended that more time and effort is spent on providing an output specification that accurately represents the format of the data as delivered.

4.3 Quality Checking and Editing

Statisticians were not provided with any report on the quality of the data or issues that arose during processing. They were also not supplied, at the time of delivery, with details of the number of form images supplied, which would have provided a check against the number of records in the datasets.

Table 4 details the actual number of images provided by the contractor.

Table 4 - Number of forms provided as images

Form type	CD01	CD02	CD03	CD04	CD05	Total
Household Individual	12	90	27	42	46	217
Communal Establishment Individual	28	286	146	200	300	960
Placeholder	8203	9479	5322	2590	3160	28,754
Household (Non-income/Post Out)	856	1452	1539	931	855	5,633
Household (Income/Post Out)	893	1499	1411	869	956	5,628
Household (Non-income/Hand Delivery)	927	1591	1495	848	896	5,757
Household (Income/Hand Delivery)	1156	1698	1666	1236	1143	6,899
Communal Establishment	2	11	9	9	26	57
Total	12,077	16,106	11,615	6,725	7,382	53,905
<i>Total Household forms</i>	<i>3,832</i>	<i>6,240</i>	<i>6,111</i>	<i>3,884</i>	<i>3,850</i>	<i>23,917</i>

This information highlights the problem of form reconciliation at a high level. Table 1 implies that, besides Communal Establishment and Individual forms, images should have been received for 51,663 households (either a Household form where a response was received or a Placeholder form in the case of a non-response). From Table 4 we can see that there were in fact 23,917 Household forms and 28,754 Placeholder forms, making a total of 52,671. Placeholder forms were used when no response was received for any reason (e.g. derelict or vacant properties, commercial premises or refusals by the householder), and the total includes non-responses from communal establishments as well as households. Another reason for the discrepancy is that about 3,000 Placeholders were created for households that subsequently returned a form; the data from these was not captured although their images were stored and are included in the totals in Table 4.

The information on Placeholder forms, which includes an estimate of the number of residents, is potentially useful in coverage adjustment, and it would be a valuable exercise to compare the information on the 3,000 Household late returns with that on the Placeholders that had been completed for those addresses. This would give information about the quality of the Placeholder information given by the enumerator.

Recommendation 3

It is recommended that the accuracy of answers given by enumerators on the Placeholder forms is assessed by comparing them to the corresponding Household Form.

From the information in Table 4 there would appear to be images for 23,917 Household forms, but the SAS dataset contained only 23,905. Census IT had

created other SAS datasets containing data from forms with combinations of missing Census District (CD), Enumeration District (ED) or line number, which are essential to linking the Individual and Household or Individual and Communal Establishment forms. These datasets were re-supplied on 21st November 2006.

Table 5 - Number of records with no Geography

SAS dataset	Contents	Number of records
NOCOMGEOGRAPHY20OCT	Communal Establishments missing CD, ED or line number	1
NOHHGEOGGRAPHY20OCT	Households missing CD, ED or line number	21
NOPERGEOGRAPHY20OCT	Persons missing CD, ED or line number	62
NOPLACEGEOGRAPHY20OCT	Placeholder forms missing CD, ED or line number	57
NOVISITORSGEOGRAPHY20OCT	Visitor information missing CD, ED or line number	21
NOVIEWSGEOGRAPHY	Respondent views missing CD, ED or line number	21

By appending this data to the supplied data we had established receipt of all the data as captured by ADS. Details of issues identified on closer examination of the data are supplied in each section below.

(It should be noted that although these files are named "NOGEOGRAPHY.." these refer to the lack of either CD and ED and or line number on the actual front page of the Census forms. By looking at the image of the form and the contents within the form in some cases if a postcode could be obtained, and by cross referencing to the Geography postal address file and cross checking the Enumerator Record Book (ERB), we were able to locate the right CD/ED and line number to allow further analysis and matching. This was extremely time consuming and had to be done form by form. Details of all modifications to analysis files are held by SMAG within Excel spreadsheets.)

During the Test Lessons learnt sessions it was discovered that around 800 forms posted out by the Royal Mail were returned to GROS HQ as "undeliverable" during the fieldwork phase. These forms were subsequently hand delivered by HQ staff. There is no way of determining which of the post out forms these 800 were. Further investigations have been unable to determine exactly why these forms were classed as "undeliverable". Later, at the end of fieldwork, around 160 forms were returned to HQ as undelivered. Presumably these were a subset of the 800, which once again had failed to be delivered, but this is not certain.

Recommendation 4

It would be beneficial if a data quality report or incident log were passed over from the data capture team to the statisticians to highlight the reasons for apparent discrepancies in the data. It would also be helpful to receive an incident log from the fieldwork team detailing major issues occurring in the field.

An unfortunate effect of the data capture system used was that any mark in a tick box was captured as a response. This meant that, if a respondent ticked one category but subsequently changed his or her response, it was captured as a multi-tick response, which for most questions was invalid. The only way to identify these was to view the form image, which, due to the time and resources involved, was only carried out for the income question (see section 5.2).

Recommendation 5

Deleted responses should be identified at data capture so that they are not incorrectly identified as images, leading to an invalid multi-tick response. It will not be practical to check images manually in the Census itself.

4.4 Data Completeness

4.4.1 Household data

The household data includes all the responses to the household questions as well as the unique identification information: Census District (CD), Enumeration District (ED), line number and postcode. Each record within the household SAS dataset ought to have been a uniquely identified household, corresponding to the information recorded within a line of the Enumerator Record Book (ERB). This was to have been achieved by a unique combination of the Census District and enumeration district and the form line number of each Household form. In total we received 23,905 records within the household dataset, each record representing a household. However we were able to trace some of the missing identification information, as detailed below, which allowed us to increase the number of records in the Household dataset to 23,770.

The total number of households covered by the Test was approximately 52,000. There are approximately 2.2 million households in Scotland, so this represents 2.4% of the total. Issues arising during the Test might therefore be expected to be multiplied in scale by approximately 40-fold in the 2011 Census if it were carried out in exactly the same way as the Test.

No Line Number on the Household Form

In 9 cases the Household form contained no line number. However by cross referencing the geography file listing addresses for the print contractor, in some

cases we were able to infer what this number should have been, which we did for 5 cases. However if no address appeared on the form we could infer nothing, which was the case for the remaining 4 households in this category.

No CD, ED or Line Number on the Household Form

There were 8 cases where no CD, ED or line number was present within the data supplied. However on looking at the forms, in some cases finding a postcode allowed us to cross reference to the geography postal address file. In 5 cases we were thus able to determine a unique CD, ED or line number. Three cases were unresolved, although responses were provided within the census form both at household and person level. However with no CD, ED or line number it was impossible to link these cases with persons data or determine where the forms came from.

No CD or ED on the Household Form

Four cases were provided with no CD or ED but with line numbers. Only one of these cases contained a postcode which matched with the geography address file so only in that case could we determine CD and ED. One other case provided an address outwith the Test area although we were able to identify it as in West Dunbartonshire from the postcode. The remaining 2 cases were unresolved.

Duplicate Combinations of CD/ED/line number on Household Forms

There were 49 cases of what were thought to be "duplicate" households, these were Household forms with the same CD, ED and line number. However there were a number of ways a "duplicate" household occurred.

In 23 cases, although the Household forms had the same CD, ED and line number they were in fact physically different households. These were due to written forms being supplied with identical CD, ED and line number information to pre-printed forms for physically different addresses. Further, in some cases enumerators had over-written details on the pre-printed forms leading to duplicated line numbers being used.

In 10 cases a Household form had been supplied instead of an Individual form. In some cases the respondents repeated their response to the household questions and in other cases they only answered person questions, leaving the household questions blank.

In 2 cases a hotel, which should have been given a Communal Establishment form, instead returned several Household forms. From the details, it looks as if the people included were foreign workers employed by the two establishments.

In other cases, each member of a couple filled in a separate Household form not acknowledging the other's existence at the address. There were also at least 2

cases where an older person had filled out the form twice, although there were some differences in responses between the two forms.

The default position was to keep the duplicate with the most information or, in cases where information conflicted, to retain the copy which was the original pre-addressed form.

Postcode or CD/ED Information Conflicted on Household Form

There were 7 cases where the postcode of the household and the geography of the census district and enumeration area conflicted. These were all hand written forms, or forms where enumerators had over written the printed information.

By printing off the forms and cross referencing the address and postcode with the geography address file, we were able to obtain the correct CD, ED and postcode.

No Response on Household Form

There were a number of Household forms containing no responses to any of the household questions (n=199). In some cases respondents had not even signed the form (n=76). Of the 199 forms, 78 supplied person details to person questions and 121 (detailed below) supplied no household or person details. It is worth noting that the combination "hand delivered/income" had almost as many forms of this type as all others put together.

Table 6 - Non-Response to Household or Person question

	Number of Forms
Post out Non - Income	29
Post out – Income	26
Hand delivered – Non Income	24
Hand delivered - Income	42
Total	121

These 121 cases which supplied no information about the household or persons were defined as "refusals" irrespective of whether signatures were captured or not, as ERBs indicated "refusal" to comply.

The 78 cases who supplied no household information but who supplied person data were defined as a household non response.

Table 7 - Non-Response to Household Question only

	Number of Forms
Post out Non - Income	12
Post out – Income	17
Hand delivered – Non Income	14
Hand delivered - income	35
Total	78

The household data can be seen as being processed in two stages. The household data as received from the contractor ADS is described in section 4.2. As this represents the forms returned to field staff, it is known as “returned” data.

The second stage is the household data cleaned by the removal of duplicates and refusals; and recoding of postcode, CD, ED and line number, as detailed in the preceding sections. The remaining data is known as household “response” data, and includes the 78 cases where respondents supplied no answers to household questions but did respond to person questions.

Household refusals

As detailed in Table 6, we have 121 cases where the respondent returned a form with no answers to either household or person questions. These cases should not have occurred as enumerators were to check that forms were completed. Table 8 details how these cases broke down by CD and delivery type. Once again the combination of an income question and hand delivery shows a significantly higher total than any other combination, although in the case of Census District 05 (Breadalbane) income/post out is by far the highest.

Table 8 - Number of Non-Response Refusals by Census District

Form type	CD01	CD02	CD03	CD04	CD05	Total
Household (Non-income/Post Out)	3	10	2	7	7	29
Household (Income/Post Out)	1	6	3	1	15	26
Household (Non-income/Hand Delivery)	2	12	6	2	2	24
Household (Income/Hand Delivery)	3	15	7	9	8	42
	9	43	18	19	32	121

In cases of refusal Part C of the Placeholder form should have been completed. Table 9 details the numbers of Placeholder refusals broken down by CD and delivery type. This does not show such a clear pattern as Table 7 and Table 8, with income/hand delivery showing the highest number of refusals in only 3 of the 5 districts.

Table 9 – Number of refusals recorded on Placeholder Form by Census District

Form type	CD01	CD02	CD03	CD04	CD05	Total
Household (Non-income/Post Out)	141	688	126	58	52	1065
Household (Income/Post Out)	147	685	141	41	38	1052
Household (Non-income/Hand Delivery)	189	654	123	61	52	1079
Household (Income/Hand Delivery)	138	787	127	84	62	1198
Communal Establishment	0	0	0	2	10	12
	615	2814	517	246	214	4406

Final Household Data

The target number of Household forms to be delivered was 51,663. Returns were received from 23,917 households (a return rate of 46.3%). Following removal of refusals and edits for duplicates we were left with 23,770 unique household records, meaning that the final return rate was 46.0%. A breakdown of these records by Census District, enumeration method and presence or absence of the income question is given in Table 12 in section 5.1.

A form reconciliation exercise was carried out by enumerators at the end of fieldwork, prior to delivery of boxes to the Data Capture contractor. The evidence from the Data Capture evaluation report is that the system did not work well in this case. However, if it is improved it could address some of the issues which are described in the preceding sections. This information could also be fed into a report as described in Recommendation 4.

Recommendation 6

It is recommended that research is carried out into the form reconciliation system that was used in the Test to identify a better system for 2011.

Ideally, an analysis would have been carried out into the number of follow-up visits carried out by enumerators to determine the optimal number to achieve a response. Unfortunately this was not considered at the design stage, which meant that the ERBs containing this information were not designed to be scanned. A full analysis would therefore involve physically working through all the ERBs used in the Test, which would not be an efficient use of time or resources.

4.4.2 Person data

The person data hold all the responses to the individual person questions. These include all people who were captured on household forms, those who filled in a Household Individual form (either for privacy reasons or because the household

contained more than 5 people) and people resident in communal establishments, all of whom filled in a Communal Establishment Individual form.

The data includes a form barcode ID, the CD, ED and line number of the household or communal establishment (which are needed in the case of Individual forms to link the person to the relevant Household or Communal Establishment form), the postcode of residence and a person number, which should be unique for each person within a particular household or communal establishment.

In total we received 48,129 person records with all identifying information, and a further 62 records which were missing at least one of CD, ED and line number. In all there were 48,191 records, each record representing an individual. There were a number of problems with the data.

No Line Number on the Associated Test Form

Most of the cases where a person record contained no line number were as a consequence of there being no line number on the Household form with which it was associated, and these were resolved at the household level (see section 4.4.1). In one case a Household Individual form was missing a line number. The corresponding Household form was located by searching for households in the same postcode with the same family name, and the line number of this form was inserted into the person record to restore the link.

No CD, ED or Line Number on the Associated Test Form

After the cases resolved at household level were taken out, there remained 4 person records with no CD, ED or line number. All of these were Communal Establishment Individual forms, with some form of the address written on the front but no other identifying information. Only one of them had a postcode. These were resolved by viewing the images of forms with a neighbouring form ID to locate those belonging to the same communal establishment, and (after checking the image of the communal establishment form) using the numbers from these forms.

No CD or ED on the Associated Test Form

There were no person records falling into this category apart from those already resolved at household level.

No CD, ED, Line Number or Person Number on the Associated Test Form

There was one person record where all four identifying numbers were missing. This was a Household Individual form where nothing was written on the front page apart from the postcode. The corresponding Household form was located by searching for a household within the same postcode where the members had

the same surname, and the CD, ED and line number were taken from this form. The correct person number was found by looking at the household members' table of the form.

No Line Number or Person Number on the Associated Test Form

There were 2 person records with a CD and ED but no line number or person number. Both of these were Household Individual forms with only the CD and ED written on the front. In each case, the corresponding Household form was located by searching for a household within the same CD and ED where the members had the same surname. The line number was taken from the front of the Household form and the person number from the household members' table.

No Person Number on the Associated Test Form

There were 136 person records where the CD, ED and line number were present but the person number was missing. All of these were Household Individual forms, and in 131 cases the person number had been omitted from the relevant space on the front of the form. This was possibly because the space was within the instructions to the respondent and it may not have been obvious that a response was required. In another 4 cases the number had been written correctly on the form but for some reason did not appear on the SAS data. In all of these 135 cases, an attempt was made to locate the relevant Household form using the CD, ED and line number, and in 128 cases this was successful. There was also one case where the line number had been miscaptured from the Individual form, which was corrected after visual inspection of the form image.

In 3 cases, it was clear from inspecting the Individual and Household form images that the wrong line number had been entered on the Individual form. The correct line number was found by searching the relevant CD and ED for a household whose members had the same surname. There were a further 3 cases where the CD, ED and line number on the form did not relate to a household on the database, and it was not possible to locate the correct Household form by any other means. Since they could not be linked to a household, these records were deleted.

Finally, there was one Individual form which could be linked to a household but was clearly a duplicate of a person already listed on the Household form. This record was deleted.

Duplicate Person Identifiers on the Associated Test Form

After the "duplicate" households had been deleted, there remained 127 combinations of CD, ED, line number and person number that were not unique. Most of these were Communal Establishment Individual forms. For communal establishments, the person number was not taken from the form but rather assigned to each person during processing, with the intention that the person

numbers in each communal establishment should form an unbroken sequence starting at 1 with no repetition (e.g. 1 to 14 if there were 14 residents). However, in 26 of the 56 communal establishments enumerated in the Test there was a failure of the processing instructions, resulting in one or more duplicate person numbers. This accounted for 119 of the 127 duplicates, which were resolved by amending the person numbers to make them unique within each communal establishment.

In another 3 Communal Establishment Individual forms, the line number had been miscaptured. This led to a clash with a person in another household or communal establishment, which was resolved by correcting the line number.

3 cases were Household Individual forms with the wrong person number written on the form (usually 1), which were resolved by finding the correct person number in the household members' table of the Household form.

In one household there were, according to the household members' table on the Household form, 7 members. As expected, 2 Individual forms were returned along with the Household form. According to the table, person 1 had completed an Individual form, but no form in this name was found. Person 2 on the table had completed the space for person 1 in the Household form, person 3 had completed person 2's questions and so on, down to person 6 who had been filled in as person 5. An Individual form was returned for person 7. The second Individual form had the same name as person 6, but all other details the same as person 7. A decision was taken to treat this as a mistake and delete the spurious person record.

Finally, one duplicate appears to have been created due to a processing error. Usually the form ID given in the SAS data is the barcode from the front page of the form (ending in 01), but in this case a person record has been created for person 3 using the barcode of the first page of person 3's responses (ending in 10). As person 3 has also been recorded in the normal way, this additional record was deleted.

Duplicate Persons on one Test Form

There were 113 known cases found where the same person had apparently filled in the form two or more times. These were detected as multiple records within the same household (CD, ED line number) considered to belong to the same person as the details had the same first name, surname, gender, date of birth and marital status.

In some cases the respondent had left details blank so it was impossible to determine whether these were the same individuals. However by looking at a sample of the forms one was able to see easily that in most cases the person has continued to complete the details for persons 2 to 5 in addition to person 1. Elderly respondents also appear to struggle with the person details and there

were cases of married couples filling in the person details with the husband's name but the wife's details. These were not included in the 113 cases above as gender and date of birth differed. In some cases where respondents fill in the same details multiple times, the details they fill in are not entirely the same each time. In the Census itself, these issues would be corrected by downstream processing edit and imputation rules. However in the Test, with limited time, the decision was simply to report on the issue as there was no systematic way of determining how to deal with this issue. These records were therefore left in the data. They are not expected to make a significant difference to the analysis.

The 2 of 3 Rule

A rule was applied to filter out spurious data records. This could for instance include a case where a respondent has scored through pages that were not applicable but has inadvertently crossed through a tick box, creating a person record with only one piece of information (e.g. a religion). Three key questions were selected as indicators of a valid response and it was agreed that at least two of these should be answered for a responses to be considered a genuine person. These three questions were name, sex and age. If two of these question were coded as missing or "error" the person was removed from the valid person dataset.

A similar rule was applied in 2001 to verify valid individual responses, using four key questions of which two were required to be answered. The four questions were name, age, sex and marital status. For the Census Test it was decided not to use marital status as a determining question since the response rate tends to diminish towards the end of the questionnaire. The marital status question was positioned later in the form at question 7 of the individual questions in this Test, compared with the earlier position of question 4 in the 2001 Census. Another reason for not using this question for the person rule is that the marital status question had changed significantly from the 2001 question and the response to the new form of this question was unknown. However it would have been preferable to have the marital status question early on the Test form.

There were 828 people removed by the 2 of 3 rule. These "removed persons" were provided as a separate dataset by Census IT (see Table 3). SMAG investigated the contents of these responses, and 8 were determined to have been removed incorrectly. Six were removed because they had corrected their answer to the question on sex or ticked both male and female. Corrected answers were not identified during data capture. Two people were removed because they only entered a year of birth, rather than day, month and year.

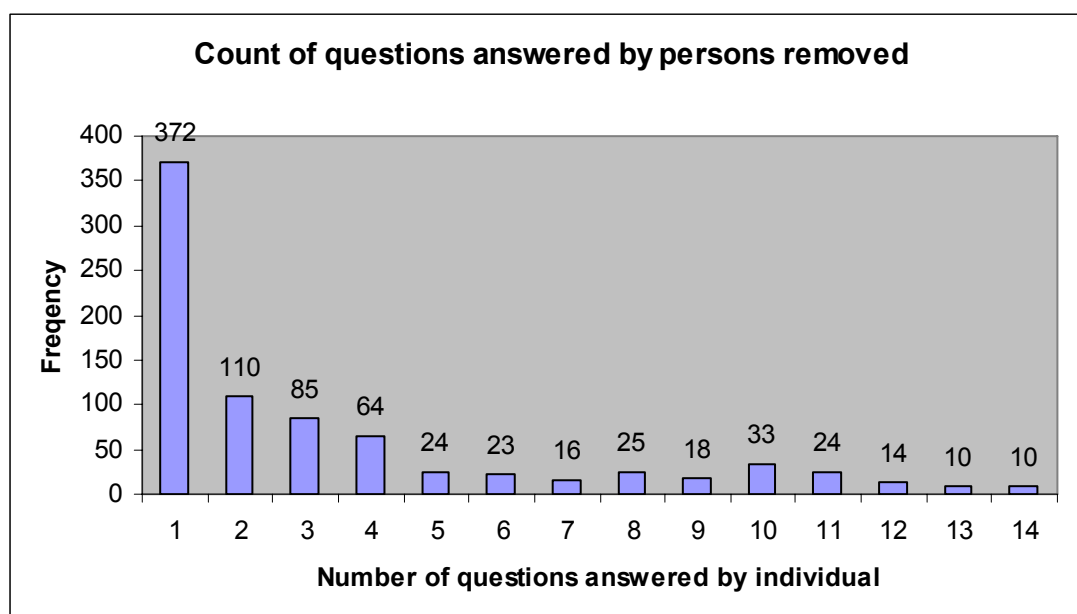
It was decided to investigate the response to individual questions for "removed persons". These were person records that failed the 2 of 3 rule and were therefore removed from the main dataset, but were retained in a separate dataset for analysis. Since the persons dataset contained data for both people living in

households and communal establishments, only questions that applied to everyone were considered. The 16 indicator questions were:

- Name
- Sex
- Date of birth
- Student
- Country of birth
- Marital/civil partnership status
- Health
- Long term limiting illness
- Confidence going out at night
- Discrimination
- Current religion
- Religion of upbringing
- National identity
- Ethnic identity
- Activity last week; and
- Language.

Any questions that could be bypassed by routing instructions or only appeared on one form type were omitted from this investigation.

Figure 6 - Count of Questions Answered by Persons Failing 2 of 3 Rule

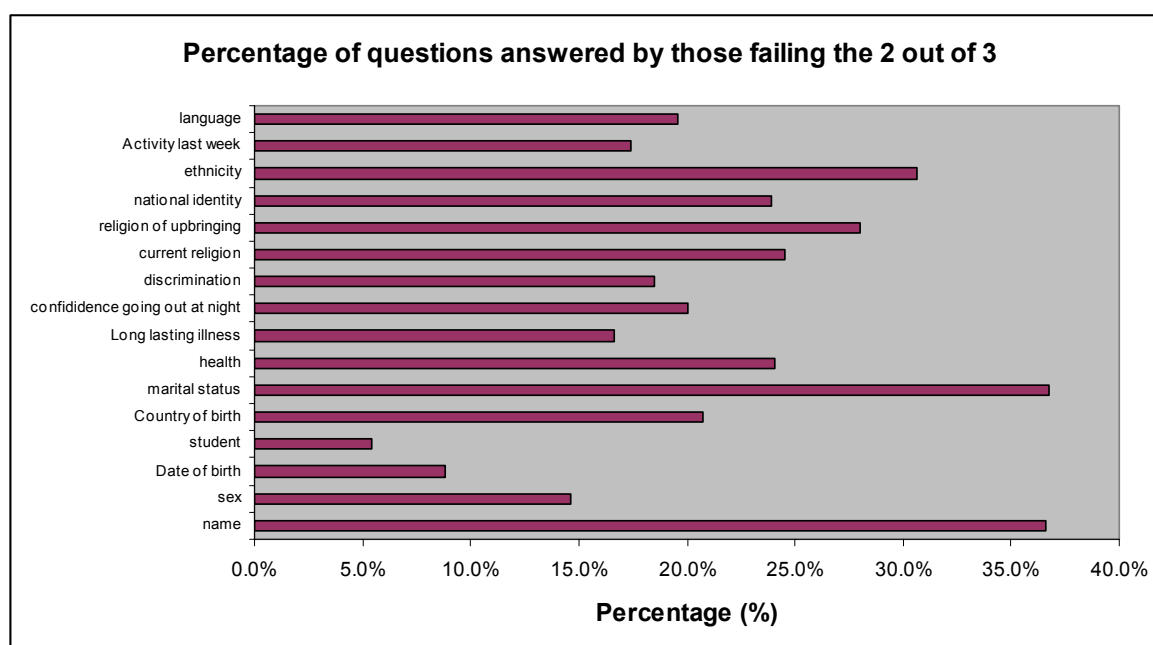


For this investigation, a question was considered to be answered if some attempt had been made to complete the question regardless of multi-ticking. The chart below shows the number of questions answered by people removed from the persons dataset. It shows that 10 of the 828 people removed answered 14 out of 16 common individual questions, 10 answered 13 common individual questions,

14 answered 12 individual questions, through to 372 who answered 1 question. This means that 16% (134 of 828) of apparently spurious person records included answers to more than half the questions (i.e. 8 or more).

The validity of the 2 of 3 rule was also explored by examining the proportion of those who failed the rule which answered each question. Figure 7 shows that the Sex, Date of Birth and Full Time Education questions are poorly answered compared to the other questions. The Full Time Education question may have been poorly answered among older respondents who did not consider it applicable to them. The two best answered questions are name and marital status. This would suggest that most people failed the 2 of 3 rule for not answering Sex and Date of Birth and that inclusion of marital status in the rule would have captured more people.

Figure 7 - Percentage of those failing the 2 of 3 rule who answered each question



Final Person Data

After the edits detailed in the preceding sections there remained 48,126 unique person records. These included three different types of form, namely Household forms, Household Individual forms and Communal Establishment Individual forms (which account for all people counted in communal establishments). Those who returned a Household Individual form can be further subdivided into those who used an Individual form because there was insufficient space on the Household form and those who used an Individual form for privacy reasons. Since there was space for 5 people on the Household form, it has been assumed that those listed among the first 5 people on the form would only have used an Individual form to keep their responses private from other household members. Those who did not fit on the Household form (person numbers 6 and above) are assumed to have

used the Individual form purely for this reason, although of course it is possible that some of them also wished to keep their responses private.

The breakdown is as follows:

Table 10 - Number of Individuals

Form type	Number of individuals
Included on Household form	46,959
Household Individual form (lack of space)	197
Household Individual form (privacy)	11
Communal Establishment Individual form	959
Total	48,126

Recommendation 7

It is recommended that the marital/civil partnership status question comes before any filter questions on the questionnaire. A "2 of 4" rule can then be used, similar to that in 2001, to define a valid response, with any 2 questions out of name, sex, date of birth and marital/civil partnership status needing to be answered.

4.4.3 Placeholder data

Placeholder forms were completed by enumerators in cases where it was not possible to obtain a response from a household or communal establishment. Issues of time and resource prevented any detailed investigation into the integrity of Placeholder data as was carried out for household and person data (see sections 4.4.1 and 4.4.2). The number of placeholder forms for which data was received is detailed in Table 4.

In a number of cases a Household form was received after a Placeholder form had been completed for that address. Since an assessment had been made by the enumerator as to the status of the address (e.g. vacant, household present but no contact, etc.) it was suggested that these Placeholders should be investigated, in the light of the Household forms later received, to determine whether or not enumerators generally made a correct assessment. Since this would involve viewing a large number of forms individually, it has not yet been possible to carry out this work.

In most cases Placeholder data should have been removed from the delivered datasets when a Household form was received from the same address, but there were 28 combinations of CD, ED and line number that corresponded to both a Household and Placeholder form. These have not yet been investigated.

Placeholder forms replaced the Dummy forms used in 2001. The principal difference is that Dummy forms were not completed for certain types of address, e.g. derelict properties and non-existent addresses. Work needs to be carried out to investigate the usefulness of the Placeholder form compared to the Dummy

form. It has also been suggested that, instead of a specific Placeholder form, it may be possible to use a Household form for this purpose with additional tick boxes added for enumerator use. This would save on printing and scanning of additional forms and may aid form reconciliation.

Recommendation 8

It is recommended that work is carried out to investigate the usefulness of Placeholder forms as compared to other options.

4.4.4 Communal Establishment Data

A communal establishment is a residential address which does not contain households, and is usually some form of managed accommodation. Examples of communal establishments are hotels, hospitals and student halls of residence.

Each communal establishment covered by the Test was given a Communal Establishment (CE) form. The data taken from these forms include the same identification and geographical data as from the Household form, along with characteristics of the establishment (e.g. type of establishment, characteristics of residents) and the number of Individual forms issued and collected. The number of establishments for which data were received was 56, broken down as follows:

Table 11 - Number of Communal Establishment Forms

Test Area	Number of CE forms
01 - North Glasgow	2
02 - South Glasgow	11
03 - West Dunbartonshire	9
04 - Lochaber	8
05 - Breadalbane	26
Total	56

Due to problems with the definition of a communal establishment, instructions were changed part way through the field operation to exclude certain types of establishment from the Test. It is therefore difficult to ascertain the number of communal establishments targeted by the Test in each CD.

In the small amount of communal establishment data, no data integrity problems (such as duplicate or missing form IDs or line numbers) were found. However it is known that some hotels were enumerated using Household forms and therefore not captured as communal establishments.

Many types of communal establishment, including care homes, hospitals, ships, prisons, oil rigs etc. have their own particular enumeration difficulties. This will be further complicated if, as is the present intention, visitors to communal establishments are included in the population base for enumeration at their

location on Census night as well as their home address. This may mean that airports and ferries have to be enumerated.

Recommendation 9

It is recommended that communal establishments are more clearly defined by HQ so that they can be easily distinguished by enumerators, and clearer instructions given to aid the collection of data from such establishments.

5. Analysis

This section presents various analyses of the data as it stood once the data cleaning described in section 4 had been carried out. It includes the effect on response rates of the three factors being tested (geographical area, enumeration method and income question); a more detailed look at responses to the income question; analyses of item response to individual questions, and overall form completeness levels; and an introductory analysis of some specific questions of particular interest. Recommendations for future work are included throughout.

5.1 Household Form Return Rates by Delivery Method and Income Question

Table 12 details the data available for analysis by CD, enumeration method and income question.

Table 12 - Number of received Test Forms by CD, delivery method and income

Census District	Enumeration Method	Income/ No Income	Target Delivery	Forms Received	%
01 -North Glasgow	Hand delivered	No Income	2914	925	31.7
	Hand delivered	Income	2886	1140	39.5
	Post out	No Income	3067	853	27.8
	Post out	Income	3127	892	28.5
02 -South Glasgow	Hand delivered	No Income	3816	1579	41.4
	Hand delivered	Income	3854	1682	43.6
	Post out	No Income	3898	1442	37.0
	Post out	Income	3964	1493	37.7
03 -West Dunbartonshire	Hand delivered	No Income	2797	1489	53.2
	Hand delivered	Income	2713	1655	61.0
	Post out	No Income	2937	1537	52.3
	Post out	Income	2893	1408	48.7
04 - Lochaber	Hand delivered	No Income	1479	846	57.2
	Hand delivered	Income	1669	1222	73.2
	Post out	No Income	1588	924	58.2
	Post out	Income	1540	868	56.4
05 - Breadalbane	Hand delivered	No Income	1595	894	56.1
	Hand delivered	Income	1638	1127	68.8
	Post out	No Income	1638	848	51.8
	Post out	Income	1650	941	57.0
All districts	Hand delivered	No Income	12601	5733	45.5
	Hand delivered	Income	12760	6826	53.5
	Hand delivered	Total	25361	12559	49.5
	Post out	No Income	13128	5604	42.7
	Post out	Income	13174	5602	42.5
	Post out	Total	26302	11206	42.6
	Total		51663	23765	46.0

This data is presented graphically below in Figure 8 to Figure 12.

Figure 8 - Response Rate North Glasgow

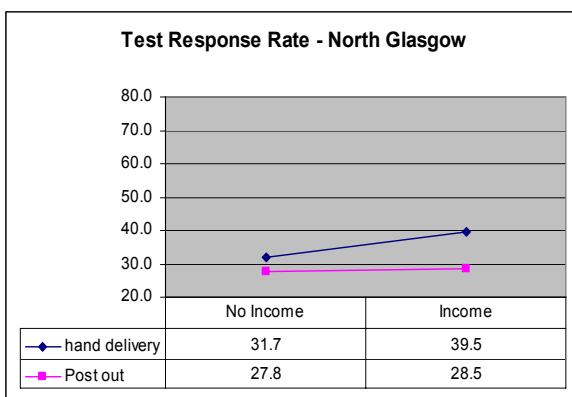


Figure 9 - Response Rate South Glasgow

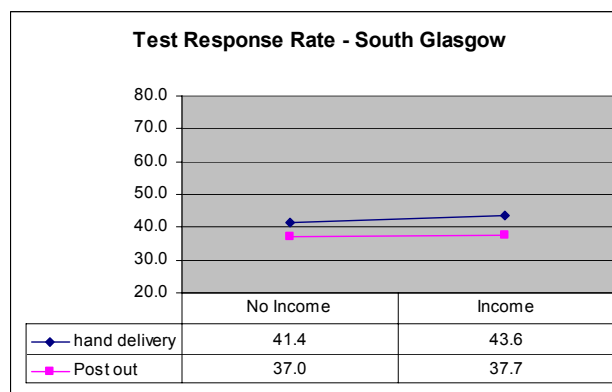


Figure 10 - Response Rate West Dunbartonshire

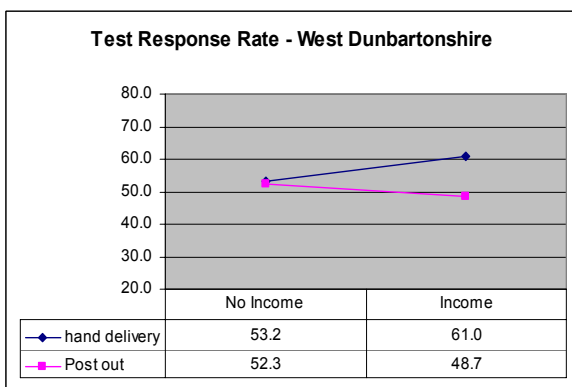


Figure 11 - Response Rate Lochaber

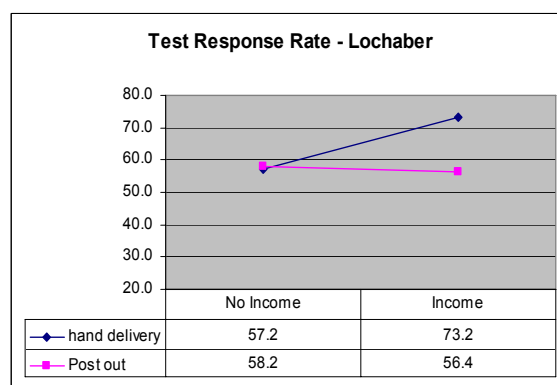
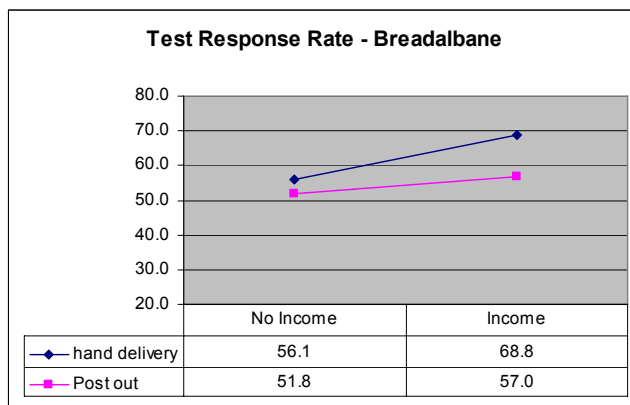


Figure 12 - Response Rate Breadalbane



The figures show that hand delivery generally appears to yield a higher response in forms received irrespective of area or whether income was asked or not. The only exception to this is in Lochaber, where post-out yields a slightly higher response where the income question is not present.

Further the presence of an income question appears to yield a higher response from household respondents. The two exceptions to this were where the forms were posted out in West Dunbartonshire and Lochaber. In West Dunbartonshire there was a 3.6% decrease in the return rate where the income question was included and in Lochaber there was a 1.8% decrease. But generally for post out delivery the presence or absence of income appears to have little or no effect on response rates.

For all 5 areas the income question yields a higher response when hand delivered. This is also the general pattern where there is no income question, although it is not universal in this case.

Using logistic regression to model the response data, fitting main factors for income, area and delivery yielded results that were statistically significant for all main effects. However the 3 way and 2 way interactions between these three main factors were also statistically significant. This means that the interpretation of the data is not as straightforward as each of the 3 factors having an independent effect on the data - the effect varies under different combinations of factors. Therefore it would be wrong simply to conclude that in general hand delivery is better. It may at first not seem surprising that the results are different across the five areas, as these areas were chosen because they are very different localities presenting particular features. A follow up exercise which had included interviewing respondents about their responses or lack of responses might have provided some indication of why the results have turned out as they have.

Conclusion

The results presented here suggest that the presence of a household income question does not detrimentally affect response rates. Hand delivery of forms by enumerators, rather than sending them through the postal service, also appears not to have a detrimental affect. However it is not possible to make definitive statements for several reasons, including:

- the Test areas are not representative of Scotland as a whole
- the Test was voluntary - we have no information on how those who did not respond to the Test would have behaved in a compulsory Census
- interactions between the three factors mean the picture is not straightforward

Section 5.15 looks in more detail at the areas selected for the Test.

5.2 Responses to the Income Question

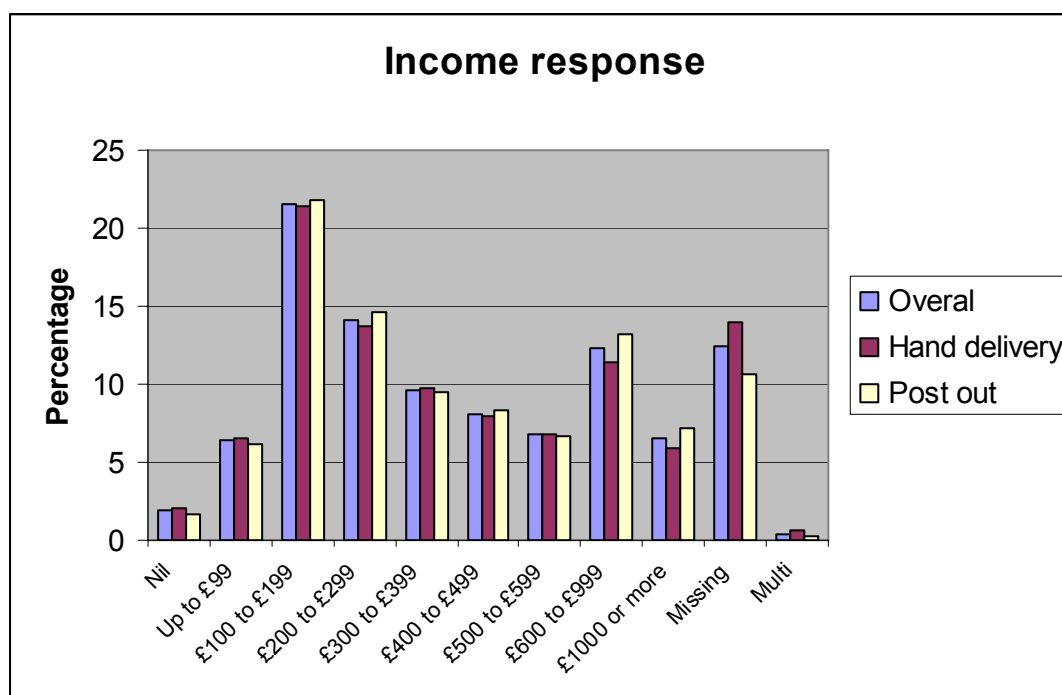
Of the 23,770 households from which we received a household response, 12,433 were asked the income question, comprising 52% of all responses received. The target number of forms to be delivered with the income question was 25,934, giving an overall response rate of 48% for these households.

Of the 12,433 forms returned with an income question, 6,831 (55%) were hand delivered and 5,602 (45%) were posted out. Of the 25,934 income forms targeted for delivery, 13,174 (51%) were to be posted out and 12,760 (49%) were to be hand delivered. The response rate was 46% for post out and 53% for hand delivery.

In some cases respondents had attempted to revise their original answer to the income question. The Optical Mark Recognition (OMR) system recognised any mark as a response, causing these responses to be coded as multiple ticks and therefore invalid. Due to the importance of the income question in the Test we investigated all responses coded as invalid and recoded those that were clearly an amended response. All analysis in this section refers to these recoded data. Details are given at the end of this section.

Below is a bar chart of the response to the income question for all 12,433 responses overall and by enumeration method. There does not appear to be any significant difference between the distribution of responses for hand delivery or post out forms, although there is a higher level of missingness for forms delivered by hand which is discussed later in this section. If possible, this distribution should be compared to one or more external sources to determine whether it is an accurate reflection of income levels in the Test areas. However, no such sources have yet been identified.

Figure 13 - Response to income by enumeration method



After the editing mentioned above, there were 53 cases where respondents selected more than one response. Thirteen of these occurred for post out and 40 for hand delivery. Of the 13 post-out multi selections, 11 selected 2 choices, one selected 6 choices and one selected 9 choices. Of the 40 hand delivery multi selections, 27 selected 2 choices, 7 selected 3 choices, one selected 7 choices, 2 selected 8 choices and 3 selected 9 choices.

Below is a bar chart of the income responses by Census Districts. It shows that a very high proportion of households responding in North Glasgow stated that they were on a low income (under £200 per week). The distributions of the other four areas are fairly similar, although South Glasgow has a significantly higher proportion of households in the top bracket (£1000 or more per week) than any other area. South Glasgow also had the smallest proportion of households that did not reply to this question.

Figure 14 - Income responses by Census District

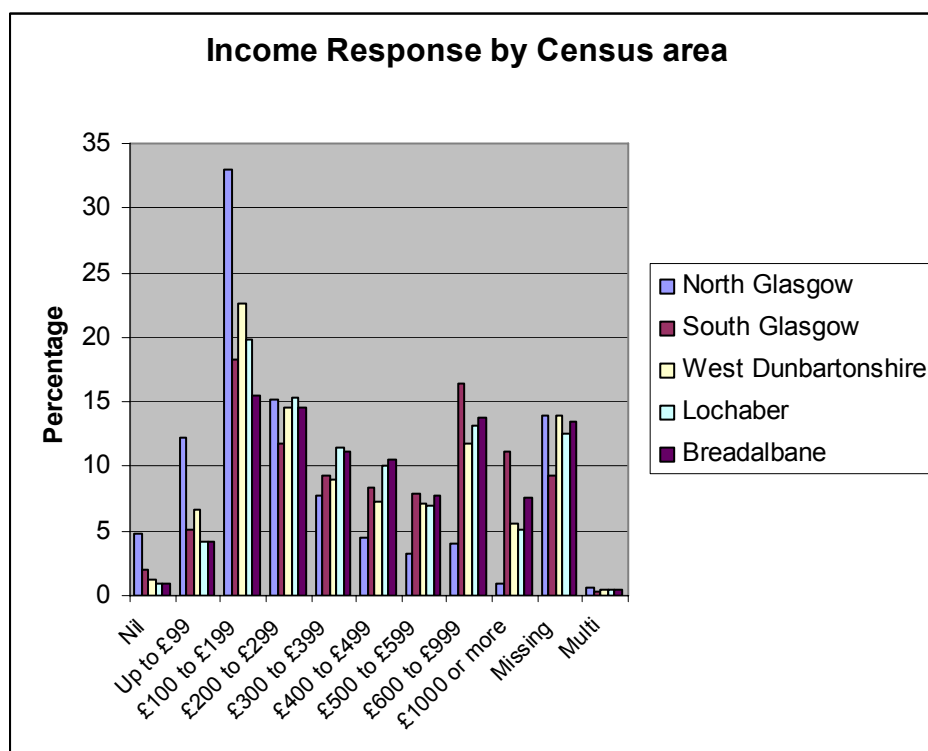


Table 13 gives the number of forms with a response to the income question (called the item response rate), broken down by Census District and delivery method and given as a proportion of all forms received with an income question. For these purposes, an “error” code (i.e. two or more responses to a single-tick question) has been treated as a response, since the respondent apparently attempted to respond to the question. It appears in some of these cases that the respondent attempted to treat it as an individual, rather than household, question, putting a tick for each earning member of the household.

Table 13 - Item response to the income question by CD, delivery method and income

Census District	Delivery	Forms Received with Income Question	Forms with Response to Income Question	%
01 - North Glasgow	Hand delivered	1140	971	85.2%
	Post out	892	779	87.3%
	Total	2032	1750	86.1%
02 - South Glasgow	Hand delivered	1682	1511	89.8%
	Post out	1493	1367	91.6%
	Total	3175	2878	90.6%
03 - West Dunbartonshire	Hand delivered	1655	1393	84.2%
	Post out	1408	1244	88.4%
	Total	3063	2637	86.1%
04 - Lochaber	Hand delivered	1222	1054	86.3%
	Post out	868	775	89.3%
	Total	2090	1829	87.5%
05 - Breadalbane	Hand delivered	1127	948	84.1%
	Post out	941	840	89.3%
	Total	2068	1788	86.5%
Grand Total		12428	10882	87.6%

The item response rates for the Test areas as a whole were 86.1% for hand delivered and 89.3% for post out, which was found by a chi-square test to be a statistically significant difference at the 5% significance level. The results therefore suggest that the income question is more likely to be answered on forms that are posted out than those hand delivered. For individual areas, the item response rate to the income question was found to be significantly higher for post out at the 5% level in West Dunbartonshire, Lochaber and Breadalbane, but the difference was not significant in North or South Glasgow.

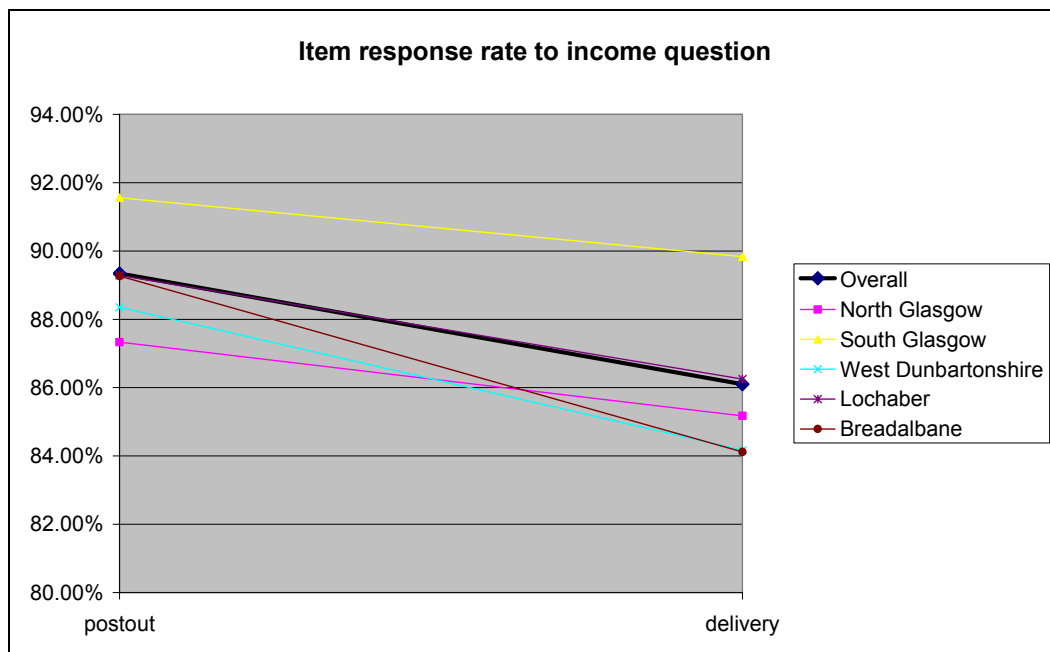
Recommendation 10

It is recommended that item response rates for other questions, besides income, are investigated to determine whether they are affected by the enumeration method.

Table 13 also shows the overall item response rate for the income question in each area, and in the Test areas as a whole. The differences between the five areas were also found to be significant at the 5% level, suggesting a real area effect on item response to the income question, with North Glasgow and West Dunbartonshire the areas with the lowest item response and South Glasgow the area with the highest.

The figures are illustrated graphically in Figure 15.

Figure 15 - Item response rate to income question



As will be seen in later sections, the item response rate to the income question was the lowest of any household question, but was by no means the lowest on the Test form as a whole.

Conclusion

The item response rate for the income question is not low compared to some of the individual questions on the form (see section 5.5) but it is possible that the level of non-response varies between households of differing income levels, reducing the value of the data obtained. In addition, it will be seen at section 5.12 that a very large number of respondents expressed dissatisfaction with this question. If this is the case in a voluntary test, where people who object to the question can choose not to return the form, including an income question in a full-scale Census might prejudice the results.

Recommendation 11

It is recommended that more work be done looking at the quality and usefulness of the response to the income question. It is also recommended that we work closely on this with NISRA and ONS, who are testing a person-level income question in 2007.

As was mentioned earlier, in a number of cases the income response was captured as a “multi-tick”. There were 217 such responses (2.0% of all responses to this question), of which 164 were recoded to one valid response after inspecting the form image. A summary of the type of invalid responses observed is shown in Table 14.

Table 14 - Reasons for an invalid responses to the Income Question

Reason for response captured as invalid	Frequency
Respondent revised original response to different response	164
Respondent revised original response to different response then back to original response	8
Respondent ticked two boxes	17
Respondent spoilt the question (e.g. strike through or crosses all boxes)	16
Cross extend to adjoining boxes	2
One box ticked and others crossed	1
Bad image	2
Respondent revised original response but did not correct according to instructions given	7
Total	217

Figure 16 is a plot of the original answer given against the corrected answer. The area of each circle represents the number of occurrences of that correction. The numbers on the axes represent the income response categories as shown in Table 15.

Table 15 - Coding for Income responses

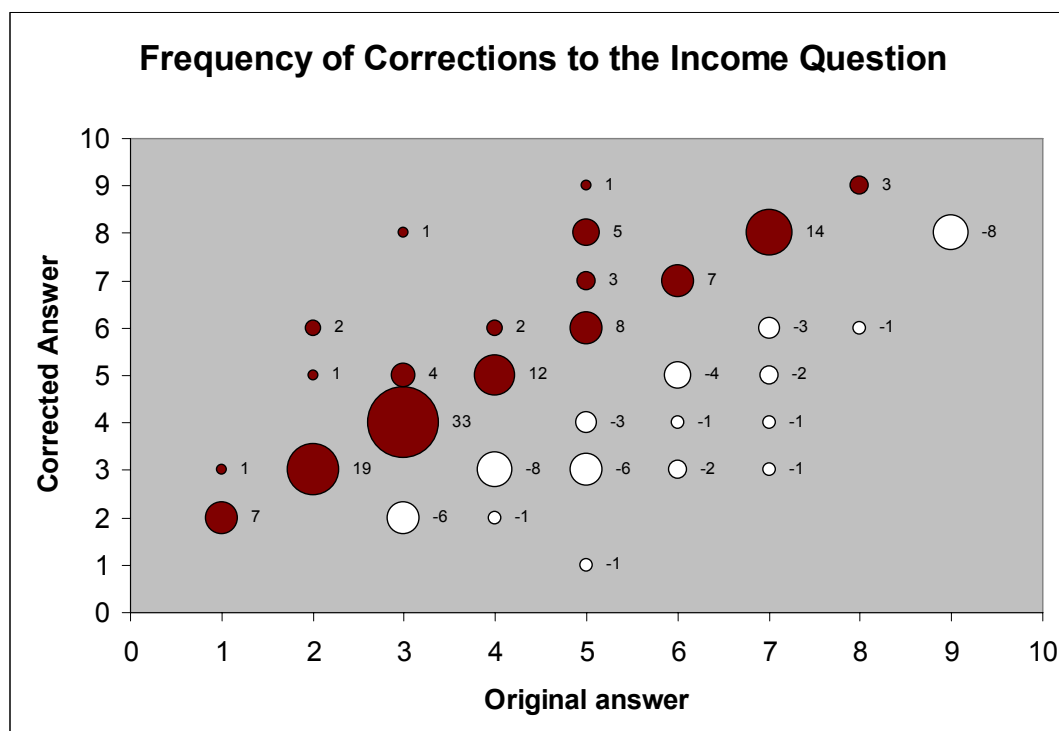
Income Code	Income Band
1	Nil
2	Up to £5,199
3	£5,200 to £10,399
4	£10,400 to £15,599
5	£15,600 to £20,799
6	£20,800 to £25,999
7	£26,000 to £31,199
8	£31,200 to £51,999
9	£52,000 or more

It is immediately clear from this plot that more positive corrections were made from lower income bands to higher income bands than negative corrections. The majority of positive corrections increased their response by one income band. The most frequent correction was from “£5,200 to £10,399” to “£10,400 to £15,599”, occurring 33 times. The high proportion of corrections involving the “£5,200 to £10,399” category is not unexpected since this is the most common income band for the Census Test, see Figure 13.

The largest correction between income bands was from “£5,200 to £10,399” to “£31,200 to £51,999”, a difference of 5 income bands. Interestingly no corrections were made involving the highest income band “£52,000 or more”.

Without cognitive studies it is not possible to say why people made these corrections but it does give some indication of the error associated with responses.

Figure 16 - Frequency of Corrections to the Income Question



Recoding these multi-tick responses was a manual process which involved visually inspecting each image, and was therefore very time-consuming. Ideally this process would be carried out for all questions, or at least those where there were a particularly large number of multi-tick responses, which include accommodation type, ethnicity and travel to work/study (see Annex C).

Recommendation 12

It is recommended that the images of all multi-tick responses are examined to determine which were genuine multi-ticks and which were corrections. This should be done for as many questions as possible, but especially for those where multi-ticking is thought to be a particular issue.

5.3 Item response – Household Questions (excluding income)

In the analysis that follows, “error” is treated as a separate category from “non-response”. An error in this case is a multiple response to a single-tick question. There are several household questions where multiple ticking is allowed, and these do not have an “error” category. In addition, question H4 “how many rooms do you have for use only by your household?” is a write-in question which also does not have an “error” category.

There is only one filter question in the household section of the form, namely question H6, “does your household own or rent the accommodation?” Those who reply “owns outright” or “buying with a mortgage or loan” are directed to skip question H7, “who is your landlord?” All other households should answer this question, and all other questions should be answered by all households. In this analysis, respondents who omitted question H6 are assumed to own (or be buying) their home, and were therefore not required to answer H7. Three percent (7,991) of households that omitted H7 had also omitted H6, whereas only 945 had answered that they did not own their house but did not state who was the landlord, so it seems reasonable to assume that most of those who omitted both questions did in fact own the property.

With the filter rule applied as above, the number of households who should have answered H7 is 8,056, compared to 23,770 for all other household questions. Table 16 shows the response rate for each question.

Table 16 - Item response to household questions

Question	Response	Non-response	Error	% response
H1. Accommodation type	22678	303	789	95.4%
H2. Repairs	22526	1244	n/a	94.8%
H3. Heating	23464	306	n/a	98.7%
H4. Number of rooms	23421	349	n/a	98.5%
H5. Self-contained	23325	401	44	98.1%
H6. Tenure	22814	635	321	96.0%
H7. Landlord	7930	61	65	98.4%
H8. Crofting	22207	1497	66	93.4%
H9. Access to garden	23132	638	n/a	97.3%
H10. Eat together	23132	638	n/a	97.3%

As can be seen, the question with the lowest item response rate is H8, “does your household live on or work a registered croft?”, with a response rate of 93.4%. The majority of households in the Test lived in urban areas and may have assumed that the question was not applicable to them. The next lowest proportion of valid responses was 94.8% to H2, “does your house need any repairs or adaptations?” possibly due to respondents feeling that the question did

not require an answer if no repairs or adaptations were needed. Question H1, “what type of accommodation does your household occupy?”, had a response rate of only 95.4% but also had a high error rate of 3.3%. This appears to be mainly due to people in flats ticking categories for both the flat and the building it is in (e.g. both “terraced” and “flat in a purpose-built block”).

Item response to the income question was dealt with in section 5.2. For comparison, if income had been considered on the same basis as the other household questions (without any editing for multi-ticking), the item response rate would have been 85.8%, significantly lower than any other household question.

5.4 Household Questions Filter

As previously stated, only one question in the household section was affected by a filter, namely the landlord question which should only have been answered by households that had indicated in the previous question that they did not own, and were not buying, their home. 15,714 households fell into this category, of which 1366 (8.7%) ignored the routing instruction and gave a response to this question. Of these, 151 said that they were buying the property with a mortgage or loan and that their landlord was the council, so it is possible that some of these were in the process of buying a council house. It is believed, though, that most were simply errors.

5.5 Item Response – Person Questions

This analysis only considers people counted within households, i.e. people in communal establishments are not included. This is because the question set asked of people in communal establishments was slightly different from that asked of people in households. All individuals in households are included, whether they were captured on a Household form or a Household Individual form.

In addition, only valid person records are considered; those who were removed due to failing the “2 of 3” rule (i.e. they had not completed at least 2 of the questions name, sex and date of birth) are excluded. After analysis by SMAG there were 820 such records (see section 4.4.2). The total number of records included in this analysis is 47,167.

This section looks at item response rates for the individual questions on the Census Test form. For each question, those who should have responded (taking account of filter rules) are divided into valid responses, non-responses and errors. Particularly high incidences of non-response and multi-ticking are investigated, with possible reasons suggested. Some recommendations for further work are also given.

Since there were a large number of person questions, they have been divided for these purposes into the three response types: single tick, multi-tick and write-in. There were five filters in the person section, which accounts for the different

numbers of people expected to answer each question. The details of the filters are given in the next section.

The total number of questions in the person section was 28, but they were captured as 36 separate variables (for instance, question 1 (name) was captured as two fields - first name and surname). They are treated in this analysis as 36 separate responses.

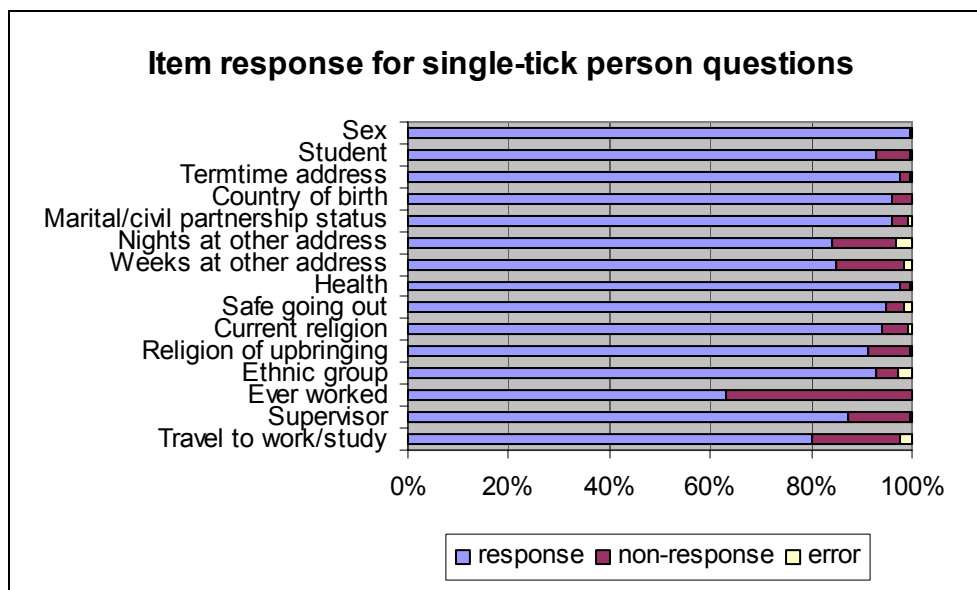
5.5.1 Single tick questions

Of the 36 fields captured, 15 expected a single-tick response. Some of these also included one or more text boxes that should have been filled in for a particular response (e.g. "other, please state"), but only the tick box responses are considered here. Table 17 and Figure 17 summarise the response to these questions ("error" in this case means ticking multiple boxes when only one was expected):

Table 17 - Item response to single-tick person questions

Question	Response	Non-response	Error	% response
2. Sex	46934	145	88	99.5%
4. Full-time education filter	43799	3258	110	92.9%
5. Term time address filter	6703	138	16	97.8%
6. Country of birth	44730	1735	29	96.2%
7. Marital/civil partnership status	44679	1511	304	96.1%
10. Nights at other address	1403	207	54	84.3%
11. Weeks at other address	1412	226	26	84.9%
12. Health	45473	827	194	97.8%
15. Safe going out	44088	1682	724	94.8%
17. Current religion	43696	2467	331	94.0%
18. Religion of upbringing	42498	3767	229	91.4%
20. Ethnic group	43161	2099	1234	92.8%
22a. Ever worked	29376	17071	47	63.2%
24. Supervisor	34268	4846	88	87.4%
27. Travel to work/study	37319	8103	1072	80.3%

Figure 17 - Item response for single-tick person questions



As can be seen, by far the poorest response rate in this group of questions was to question 22a, “Have you ever worked?” (63.2%). However this masks a considerable variation by age, with particularly low response rates in the age groups 40-49 (57.4%) and 50-59 (54.27%), both groups where one would expect the majority of respondents to be currently working. Possibly some respondents not unreasonably assumed that the question “have you ever worked?” did not need to be answered if you had already said that you were working currently - only 57.0% of those who said they were currently in some form of employment answered this question, against 68.7% of those who were not. There is also some evidence that people have written in the year they last worked without first ticking “yes” to “have you ever worked?”, but since the response rate to this question among those who have written in a year is 66.8%, as against 56.7% for those who have not, this does not account for the low overall response to this question.

The ONS report in Annex A on a random sample of completed Test forms looked at this question and states

“There was no routing onto this question. This meant that *everybody*, including children, elderly people, and those currently working, answered this question. In contrast, in the 2001 England and Wales Census only people aged between 16 and 74 years who were not currently working, answered the ‘ever worked’ question. Several respondents wrote comments on the questionnaire suggesting that they didn’t feel these labour market questions were relevant to them, for example, because they were elderly. This may explain why a significant minority (more than 10 per cent) of people left this question blank.

Additionally, general completion errors were also fairly common. For example, some people crossed the ‘Yes’ box, but did not write in the year

they last worked. This may be because they were currently working and didn't know how to report the year they *last* worked. Other respondents wrote in a year, but did not cross the 'Yes' box. This problem is seen in other questions and may be related to the layout of the question".

It would be beneficial to look at this question in more depth.

There are three questions with a particularly high incidence of multi-ticking. 3.2% of those who should have responded to question 10, "on average, how many nights a week do you stay at this other address?", ticked more than one category. This appears to be mainly due to a misunderstanding - some people have ticked a number of nights per week (e.g. 7) but have also ticked "I do not stay there every week", indicating that they stay 7 nights a week when they are at their second address but that they are not there every week. They should only have only ticked "I do not stay there every week" and then gone on to question 11, "how many weeks in the last 12 months did you stay at this address?" Some have in fact realised this and crossed out their first response, but the data capture process failed to recognise the crossing out.

2.7% of respondents ticked more than one box in question 20, "what is your ethnic group?" This appears to be mainly due to people wishing to identify themselves as both Scottish/British and as a member of another ethnic group (e.g. Pakistani). It is also recommended that we investigate this question in more depth particularly looking at multiple responses and text provided by respondents (see also section 5.8).

Question 27, "how do you usually travel to your main place of work or study?", had a multi-tick rate of 2.3%. This appears to be mainly due to people whose journey involves more than one method of transport failing to apply the instruction "tick the box for the longest part, by distance, of your usual journey to work or study".

Recommendation 13

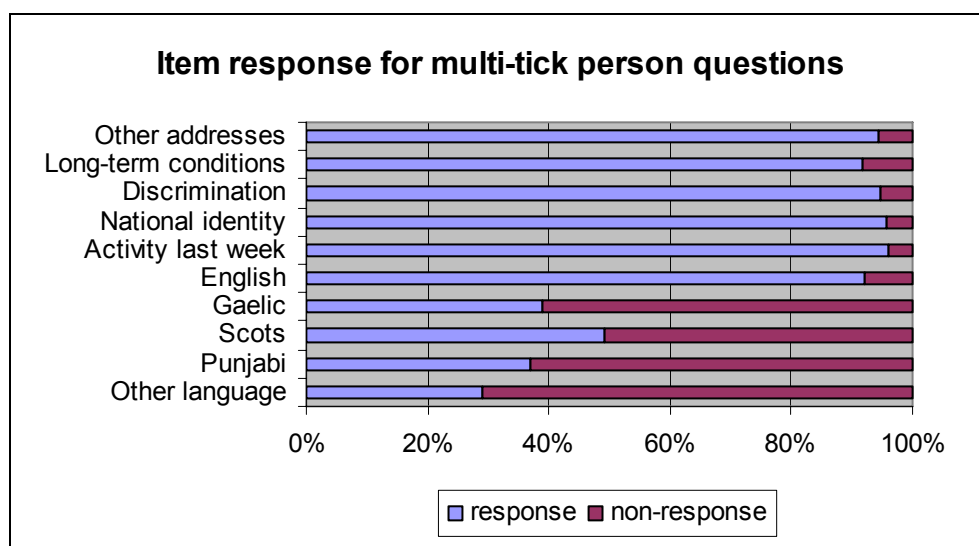
It is recommended that investigative work is carried out into question 22, "have you ever worked?", to determine reasons for the high item non-response and how this information could be better obtained.

5.5.2 Multi-tick questions

Ten tick box questions allowed multiple responses. Table 18 and Figure 18 present the response rates to these questions. No attempt has been made to determine the number of invalid combinations (e.g. in question 13, ticking both "A learning difficulty" and "None of the above"). As multiple ticking was allowed, there is no "error" category for these questions.

Table 18 - Item response to multi-tick person questions

Question	Response	Non-response	% response
8. Other addresses	43935	2559	94.5%
13. Long-term conditions	42688	3806	91.8%
16. Discrimination	44083	2411	94.8%
19. National identity	44567	1927	95.9%
21. Activity last week	44604	1890	95.9%
28a. English	42886	3608	92.2%
28b. Gaelic	18105	28389	38.9%
28c. Scots	22883	23611	49.2%
28d. Punjabi	17227	29267	37.1%
28e. Other language	13473	33021	29.0%

Figure 18 - Item Response for Multi-tick person questions

By far the poorest response was to the non-English language questions, with none of these achieving a rate of over 50%. This is probably due to people omitting the row entirely if they had no ability in a particular language, rather than ticking the “no ability” box. In addition, the “other language” section included a write-in box to allow the respondent to specify the language, and 1,933 respondents gave a written response but did not tick any of the boxes. Including these as a response to the question would have increased the item response rate for question 28e to 33.1%, still the poorest item response to any question on the questionnaire.

5.5.3 Write-in questions

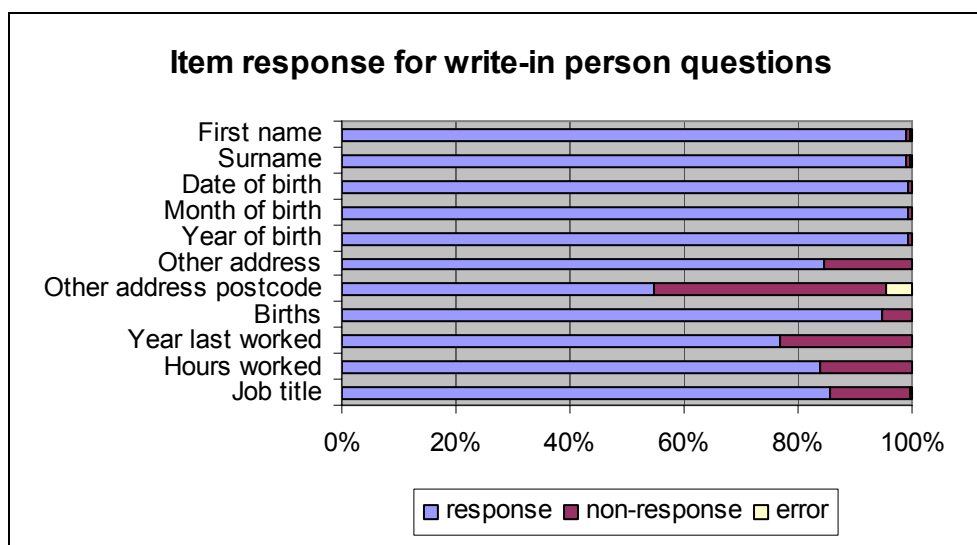
The remaining 11 questions required a write-in response. Table 19 and Figure 19 present the response rates to these questions. In this case, “error” usually refers to an illegible response. Question 14, “how many children have you given birth to?”, is slightly different in form: it has a write-in section and a tick box for “none”,

so a number greater than zero **and** a tick in the box, in combination, are considered an error.

Table 19 - Item response to write-in person questions

Question	Response	Non-response	Error	% response
1a. First name	46690	342	135	99.0%
1b. Surname	46629	404	134	98.9%
3a. Date of birth	46812	316	39	99.2%
3b. Month of birth	46812	355	0	99.2%
3c. Year of birth	46811	355	1	99.2%
9a. Other address	1406	258	0	84.5%
9b. Other address postcode	908	681	75	54.6%
14. Births	23674	1327	31	94.6%
22b. Year last worked	30188	9014	0	77.0%
25. Hours worked	32809	6393	0	83.7%
26. Job title	33506	5551	145	85.5%

Figure 19 - Item Response for write-in person questions



The poorest response rate to this set of questions was to the postcode part of question 9, “what is your other address?” Of those who stated that they had another address, only 54.6% gave a legible postcode (with another 4.5% making a mark in the box that could not be deciphered). Note that a response in this case simply means that something legible was written; it was not necessarily a valid, or complete, postcode. A full analysis of these responses is given in section 5.10. There were 430 respondents who gave an address but no postcode; many of these gave an address abroad, but there were also many whose second address was within the UK but who apparently did not know the postcode. The text section of question 9 was also relatively poorly completed, with only 84.5% of those who said they had a second address actually stating what it was.

5.6 Person Question Filters

Five filters were applied to the person questions. Unless otherwise stated, a non-response to the filter question has been assumed to mean that the following questions **should** have been answered. The first came at question 4, “are you a schoolchild or student in full-time education?” A response of “yes” to this question meant that question 5 should have been answered, whereas a response of “no” should have caused the respondent to go straight to question 6. In fact, of the 36,942 respondents who answered “no” to question 4, there were 1,108 (3.0%) who nevertheless went on to answer question 5.

Question 5 was itself a filter question, and a “no” response should have caused the respondent to omit all the remaining person questions. This filter will be considered in detail at the end of this section, but in the discussion of the remaining filters it is assumed that the respondent has not already been filtered out at question 5.

Question 8, “do you live at other addresses for part of the week or year?”, was the next filter question. A response of “no” or “yes, but I have no other fixed addresses” should have caused the remaining second address questions (9-11) to be omitted. In this case it was assumed that, if question 8 was omitted altogether, the respondent did not have a second address. Of those households that omitted question 10, “on average, how many nights a week do you stay at this other address?”, 2,399 had omitted to answer question 8 while only 207 had said they had a second address. The figures for question 11, “how many weeks in the last 12 months did you stay at this other address?”, were 2,477 and 226 respectively. It therefore seems reasonable to assume that the majority of those who omitted question 8 did not in fact have a second address.

Following this rule, Table 20 gives the proportion of those who should have been filtered out at question 8 who went on to answer the following questions in error.

Table 20 - Errors following second address filter

Question	Non-response	Error	% error
9a. Other address	44513	317	0.7%
9b. Other address postcode	44604	226	0.5%
10. Nights at other address	44089	741	1.7%
11. Weeks at other address	44466	364	0.8%

Of the 741 respondents who answered question 10 despite apparently not having another address, 589 (79.5%) stated that they stayed at their other address 7 nights a week. This suggests that the bulk of the erroneous responses to this question were people mistakenly answering it for their main address - it is possible that they missed the word “other” in question 10 and so were answering the question “on average, how many nights a week do you stay at this address?”

The next filter only applied to question 14, “how many children have you given birth to?”, which should only have been answered by women. The number of people who said at question 2 that they were male, and were not filtered out at question 5, was 21,462. Of these, 5,246 (24.4%) nevertheless answered question 14. As may be expected, most of these said that they had not had any children, but there were 482 (2.2% of all males in this category) who gave a positive number. Presumably these men failed to see the instruction “this question is for women only” and mistakenly entered the number of children that they had fathered.

Question 22a, “have you ever worked?”, was the next filter question. Only those who answered “yes” to this question should have gone on to answer the remaining work-related questions, 22b-26. Table 21 gives the proportion of those who should have been filtered out at question 22a who went on to answer the following questions in error.

Table 21 - Errors following work filter

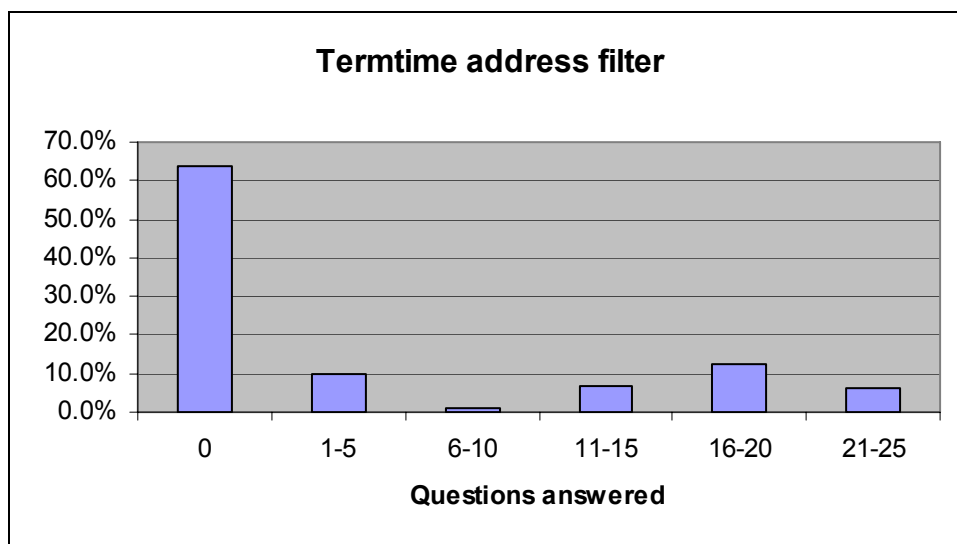
Question	Non-response	Error	% error
22b. Year last worked	7276	16	0.2%
24. Supervisor	7031	261	3.6%
25. Hours worked	7261	31	0.4%
26. Job title	7244	48	0.7%

Of the 261 who incorrectly answered question 24, “do (did) you supervise any other employees?”, all but 5 answered “no”, which would be correct if they have never worked as they stated at question 22a. Of the 5 who answered “yes”, it is clear from the answers to other questions that all but one have in fact worked but answered question 22b incorrectly.

The filter at question 5 applies to all subsequent questions, which should not have been answered if the respondent answered “yes” to question 5 (i.e. he/she is a student or schoolchild who is not currently at their term time address). 377 people stated at question 4 that they were a student or schoolchild in full-time education, and at question 5 that they lived elsewhere during term time. Table 22 shows the number of questions (out of 25) following question 5 that were answered in error by these people. Figure 20 shows the information graphically.

Table 22 - Number of errors following term time address filter

Questions answered	Frequency	%
0	241	63.9%
1	18	4.8%
2	12	3.2%
3	3	0.8%
4	2	0.5%
5	2	0.5%
9	1	0.3%
10	3	0.8%
12	2	0.5%
13	8	2.1%
14	7	1.9%
15	8	2.1%
16	7	1.9%
17	9	2.4%
18	12	3.2%
19	10	2.7%
20	8	2.1%
21	10	2.7%
22	3	0.8%
23	5	1.3%
24	5	1.3%
25	1	0.3%

Figure 20 - Termtime Address Filter

As can be seen, 36.1% of those who should have been filtered out at question 5 nevertheless went on to answer at least one subsequent question. It should be

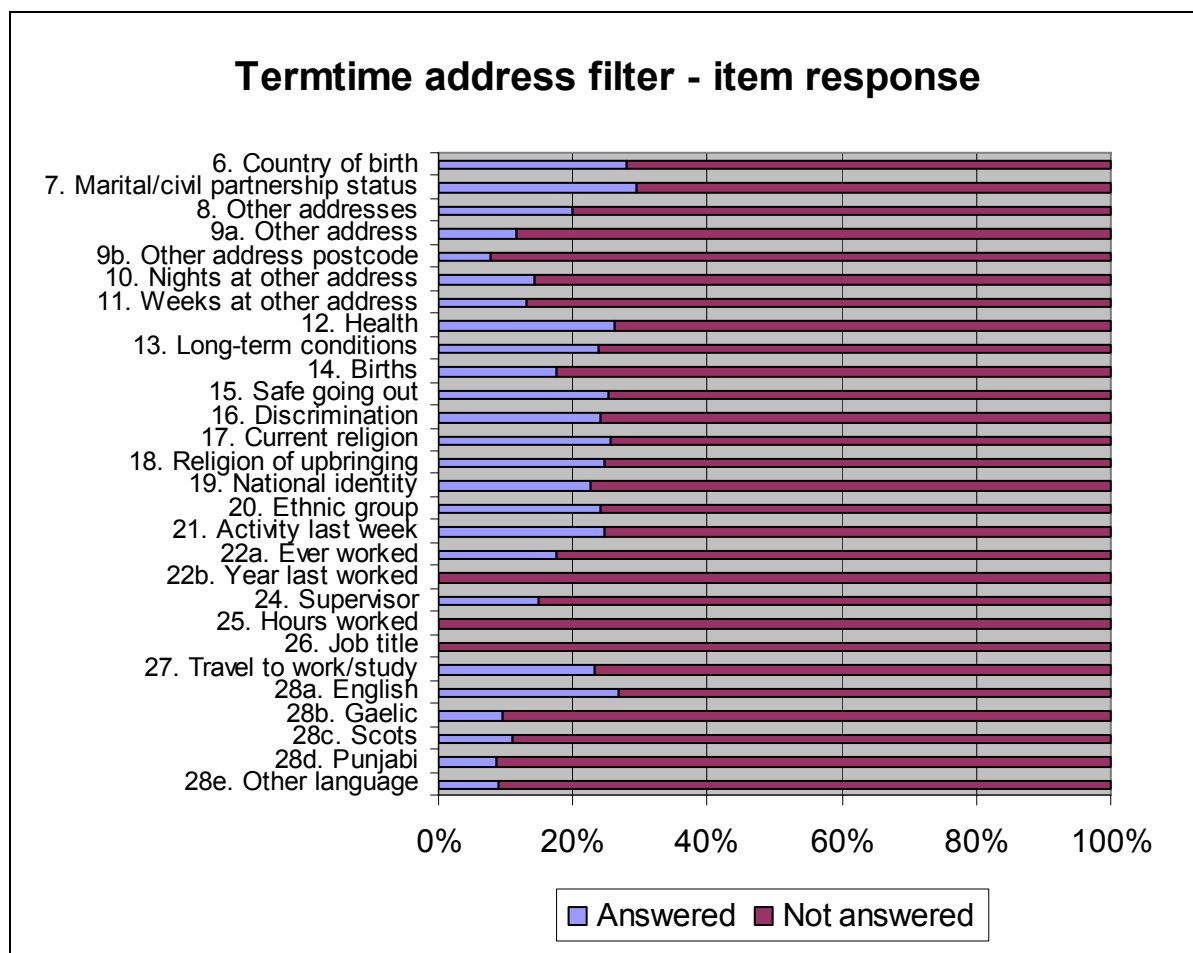
borne in mind that some people who missed the filter at question 5 may nevertheless have correctly followed one or more later filters, so although they did not complete all 28 questions following question 5 they may still believe that they have completed the form correctly. For instance, a respondent who has no other address may correctly follow the filter at question 8 and skip questions 9a, 9b, 10 and 11. The total number of questions erroneously answered would then be 24.

Table 23 and Figure 21 show the number of times each individual question was answered by those who missed the filter at question 5.

Table 23 - Questions answered in error after term time address filter

Question	Frequency	%
6. Country of birth	105	27.9%
7. Marital/civil partnership status	111	29.4%
8. Other addresses	75	19.9%
9a. Other address	44	11.7%
9b. Other address postcode	29	7.7%
10. Nights at other address	54	14.3%
11. Weeks at other address	49	13.0%
12. Health	99	26.3%
13. Long-term conditions	90	23.9%
14. Births	66	17.5%
15. Safe going out	95	25.2%
16. Discrimination	91	24.1%
17. Current religion	97	25.7%
18. Religion of upbringing	93	24.7%
19. National identity	85	22.5%
20. Ethnic group	91	24.1%
21. Activity last week	93	24.7%
22a. Ever worked	66	17.5%
22b. Year last worked	0	0.0%
24. Supervisor	56	14.9%
25. Hours worked	0	0.0%
26. Job title	0	0.0%
27. Travel to work/study	87	23.1%
28a. English	101	26.8%
28b. Gaelic	36	9.5%
28c. Scots	41	10.9%
28d. Punjabi	33	8.8%
28e. Other language	34	9.0%

Figure 21 - Term time Address Filter - Item Response



The two questions with the highest response rates are those immediately following the filter, which is not surprising - some people may have got that far on the form before realising that they should not be completing this part of it, and in other cases the form may have been completed in their absence by someone who only knew this basic information.

More surprising is the number of people who completed the English language question (and, in some cases, the other language questions but no others). This appears to be due to the form of the filter instruction at question 5. Those who live elsewhere during term time are instructed to “go to 29”, but 29 is in fact not a question but an instruction, telling the respondent to continue to the next set of person questions if there are any more people in the household. It may be that some respondents expected to be directed to a question rather than an instruction, and either not correctly reading the instruction, or else assuming it to be a mistake, filled in the final question, number 28, which is the language question.

Recommendation 14

It is recommended that non-response by age to the question "have you ever worked?" is investigated.

5.7 Form Completeness

A completeness ratio was calculated by dividing the number of questions with a valid response by the number of questions that should have been answered, using the filter rules described in sections 5.4 and 5.6. Questions that were answered in spite of the filters (e.g. a man answering the question "how many children have you given birth to?") were ignored for these purposes. The income question was not included in this analysis, as it was not asked of all households and was part of the design of the Test. However the completeness ratio has been analysed according to the presence or absence of the income question (see section 5.7.5).

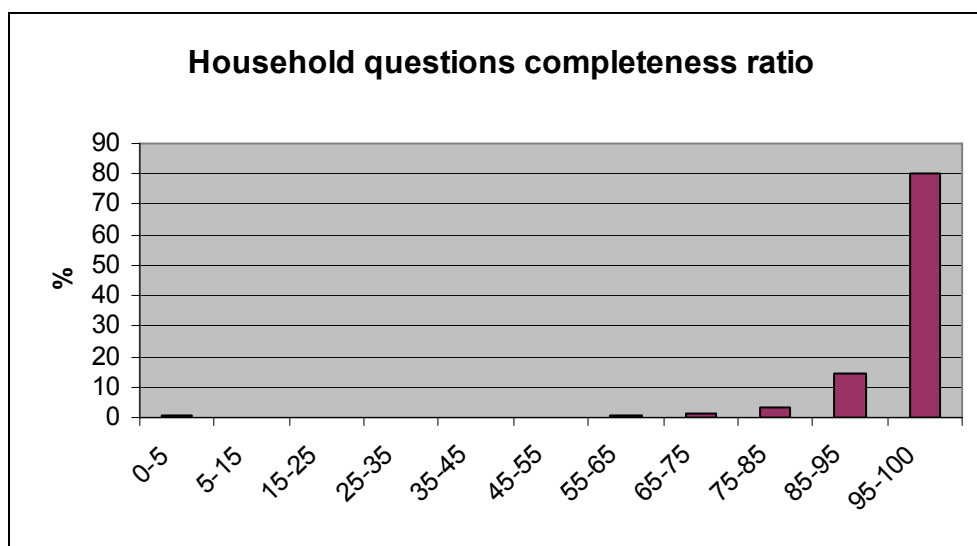
5.7.1 Completeness from Household Questions

The household section of the form contained 10 questions (excluding the income question), of which only one (landlord) was dependent on a filter question. The total number of questions to be answered was therefore either 9 or 10, depending on whether or not the household owned their accommodation.

The mean ratio was 96.6%, and the distribution is shown in Table 24 and Figure 22. The ratios shown reflect the number of questions that should have been answered - for instance, a household that should have answered 9 questions but only answered 4 has a completeness ratio of 44%, while a household answering 6 questions out of 10 has a ratio of 60%.

Table 24 - Completeness Ratio for Household Questions

HH completeness ratio (%)	Frequency	% of total	Cumulative %
0	78	0.33	0.33
11	22	0.09	0.42
20	1	0.00	0.42
22	8	0.03	0.46
30	2	0.01	0.47
33	20	0.08	0.55
40	6	0.03	0.58
44	39	0.16	0.74
50	18	0.08	0.82
56	76	0.32	1.14
60	40	0.17	1.30
67	167	0.70	2.01
70	113	0.48	2.48
78	459	1.93	4.41
80	329	1.38	5.80
89	1998	8.41	14.20
90	1371	5.77	19.97
100	19023	80.03	100.00
Total	23770	100.00	

Figure 22 - Household Questions Completeness Ratio

As can be seen, the vast majority of households completed all the questions required and less than 1% answered under half of the questions. Of the 3369 households that omitted only one question, the question most commonly missed was "Does your household live on or work a registered croft?", which was not answered by 725 (21.5%) of these households. This was also the question most

missed by all households (see section 5.3), and other questions highly likely to be the only question missed were also consistent with the overall item response rates.

5.7.2 Completeness from Person Questions

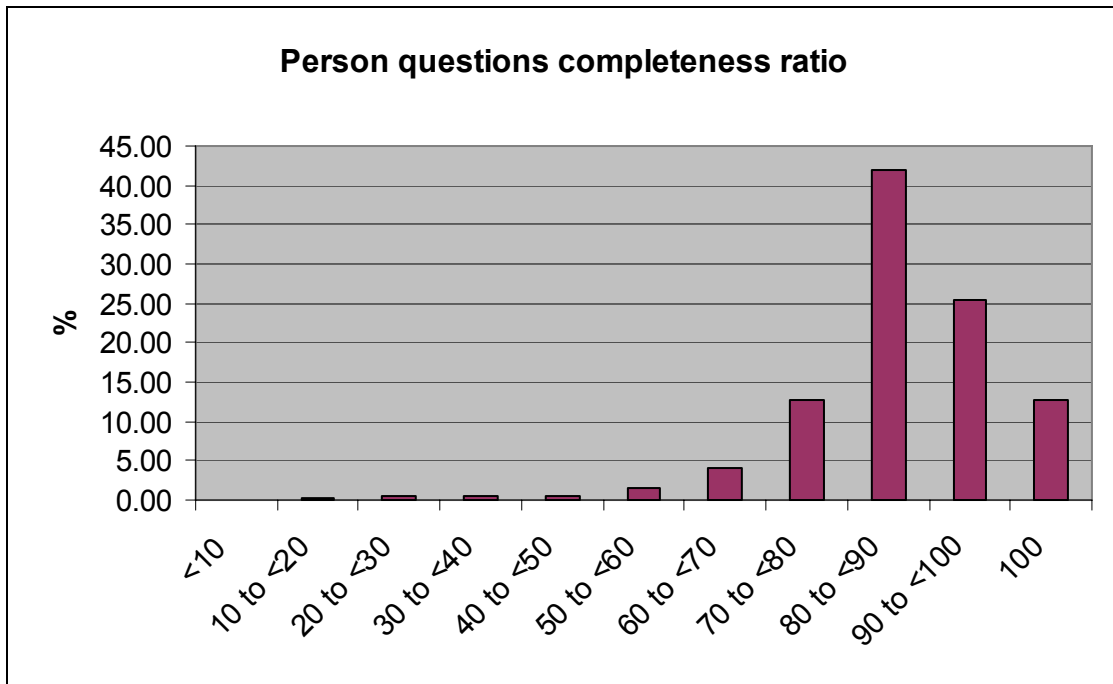
A total of 36 separate question responses were captured for each person on the form. Due to the various filters, the number of questions that should have been answered varied between 7 and 36, although all but about 1.4% of respondents should have completed at least 26 questions. The results are summarised in the table below. Although only people in households (not communal establishments) are included in this analysis, it is carried out by person rather than by household; a three-person household would therefore carry three times as much weight as a person living alone. Individual forms are included, but people who failed the “2 of 3” rule (see section 4.4.2) are excluded, regardless of how many questions they answered. There are no people with a completeness ratio of zero, as a completely blank person section would not be recognised as indicating a person.

The mean ratio was 85.4%. Table 25 and Figure 23 show the distribution.

Table 25 - Completeness Ratio for Household Questions

Person completeness ratio (%)	Frequency	% of total	Cumulative %
<10	23	0.05	0.05
10 to <20	114	0.24	0.29
20 to <30	255	0.54	0.83
30 to <40	183	0.39	1.22
40 to <50	276	0.59	1.80
50 to <60	697	1.48	3.28
60 to <70	1927	4.09	7.37
70 to <80	5960	12.64	20.00
80 to <90	19755	41.88	61.89
90 to <100	11944	25.32	87.21
100	6033	12.79	100.00
Total	47167	100.00	

Figure 23 - Person Questions Completeness Ratio



It is immediately apparent that, compared to the household questions, very few people (under 15%) have completed the entire individual section. This is partly due to the language questions which, as was noted at section 5.5.2, were very poorly completed. If the four non-English language questions are omitted from the analysis, the number completing all questions rises to 15,558 (32.98%). However 80% of people completed at least 80% of the form, and only 1.8% completed less than half.

Of the 37,732 people completing at least 80% of the form, the poorest item response rate after the language questions was 64.7% for postcode of second address, although this question only applied to 3.4% of the people in this group (those who said they had a second address). Of those questions that applied to the majority of the population the next lowest item response was to the question “have you ever worked?” (69.4%) which was also the most poorly-completed question for the population as a whole; other poorly-completed questions for this group also corresponded to those for the rest of the population.

5.7.3 Overall Completeness from Household Form

An overall completeness ratio was calculated for each household by counting the total number of questions, both household and person, that should have been completed. All questions on the form (other than income, which is excluded) carry equal weight, and hence a person who should have completed all 36 person questions carries about 4 times as much weight as the household section (9 or 10 questions), and 5 times as much weight as a person who is only required to answer the minimum 7 questions.

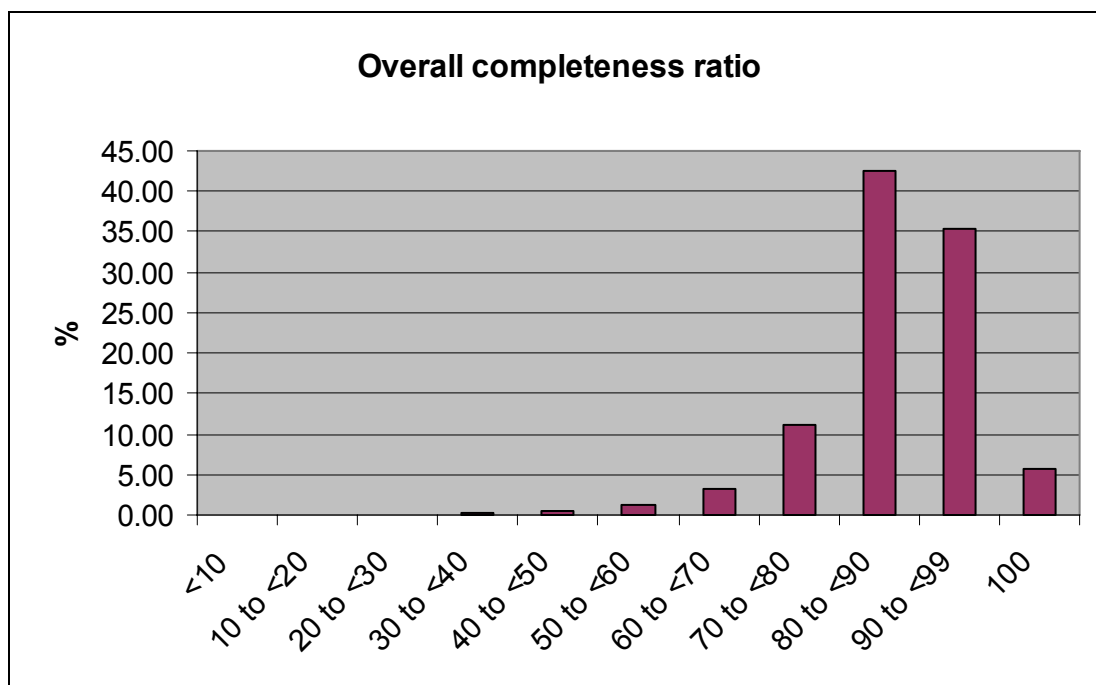
Strictly speaking this is a completeness ratio for households rather than a form completeness ratio, as people who filled in Individual forms are linked into the relevant household even though they are included on a different form. Households with no associated person data have been excluded.

The mean was 86.8% and the distribution is shown in Table 26 and Figure 24:

Table 26 - Overall Completeness Ratio

Completeness ratio (%)	Frequency	% of total	Cumulative %
<10	1	0.00	0.00
10 to <20	4	0.02	0.02
20 to <30	10	0.04	0.06
30 to <40	33	0.14	0.20
40 to <50	113	0.48	0.68
50 to <60	298	1.27	1.95
60 to <70	765	3.25	5.20
70 to <80	2616	11.11	16.31
80 to <90	10012	42.52	58.82
90 to <99	8340	35.42	94.24
100	1356	5.76	100.00
Total	23548	100.00	

Figure 24 - Overall Completeness Ratio



Since this measure combines the household questions and questions for one or more people, it is inevitable that the proportion completing the entire form is

smaller than for either of the other measures - in fact, at under 6%, it is less than half the proportion of individuals completing all the required questions. However, overall the picture is broadly similar to that at individual level, with 83.7% of households completing over 80% of the form, and only 0.7% less than half of it.

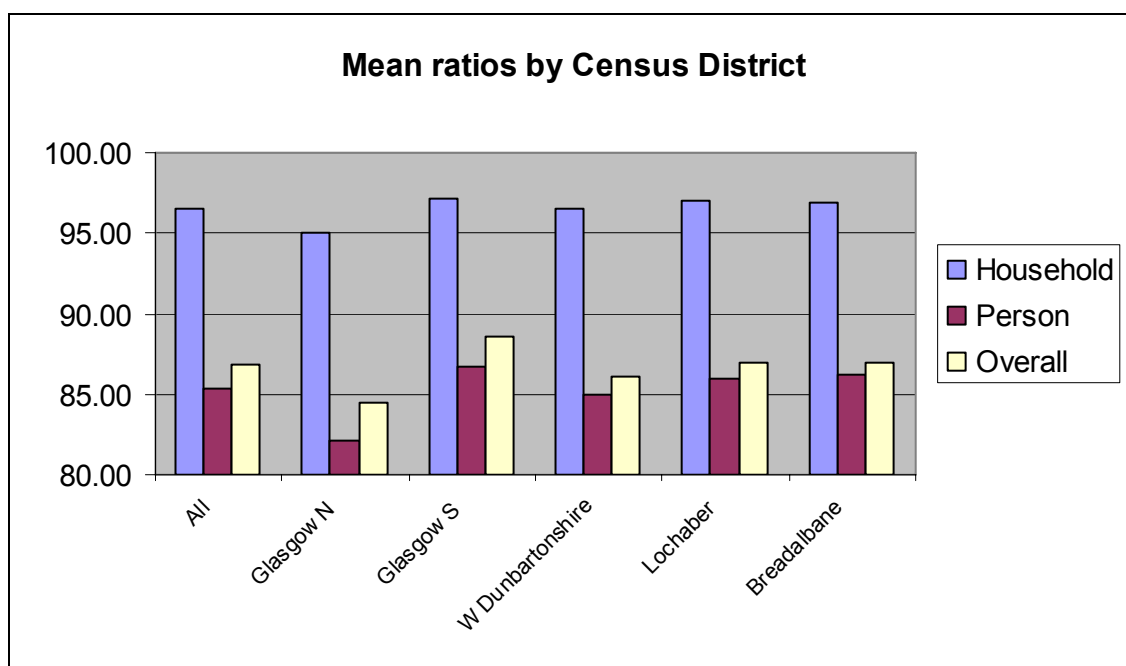
5.7.4 Breakdown by Census District of Completeness Ratios

Table 27 and Figure 25 show the mean value of each of the three ratios, for each of the five Census districts:

Table 27 - Completeness Ratios by Census District

Census District	Mean household ratio (%)	Mean person ratio (%)	Mean overall ratio (%)
01 - North Glasgow	94.97	82.16	84.51
02 - South Glasgow	97.16	86.77	88.59
03 - West Dunbartonshire	96.53	85.00	86.11
04- Lochaber	96.97	86.02	86.93
05- Breadalbane	96.87	86.23	87.02
All districts	96.57	85.39	86.78

Figure 25 - Mean Ratios by Census District



North Glasgow stands out as particularly poor compared to the other four districts. In all three cases it is over 1% below the mean (over 3.2% in the case of the person questions), whereas no other district is more than 1% either side of the mean on any of the three measures. Indeed, on every measure except the overall ratio, it is further from the next poorest district than any other two districts are from each other. The questions with the lowest item response in North

Glasgow were generally the same as those for the Test as a whole, suggesting that the problem in this area is not due to issues or difficulties with any particular question but rather with the form as a whole.

By contrast, South Glasgow is consistently the best-performing district, reaching 1.81% above the mean for the overall ratio. West Dunbartonshire is the second poorest, while Lochaber and Breadalbane vary between second and third place but with no more than 0.21% between them on any measure.

5.7.5 Breakdown of Completeness by Income and Enumeration Method

Table 28 shows the mean ratios for households where the income question was or was not asked.

Table 28 - Completeness Ratios by Income Question

Income question asked?	Mean household ratio (%)	Mean person ratio (%)	Mean overall ratio (%)
No	96.62	85.47	86.80
Yes	96.52	85.35	86.76
All	96.57	85.39	86.78

It appears that the presence of an income question makes little or no difference to whether or not the rest of the form is completed.

Table 29 gives the mean ratios for the two different enumeration methods.

Table 29 - Completeness Ratios by Enumeration Method

Enumeration method	Mean household ratio (%)	Mean person ratio (%)	Mean overall ratio (%)
Hand delivery	96.36	85.09	86.55
Post out	96.80	85.75	87.03
All	96.57	85.39	86.78

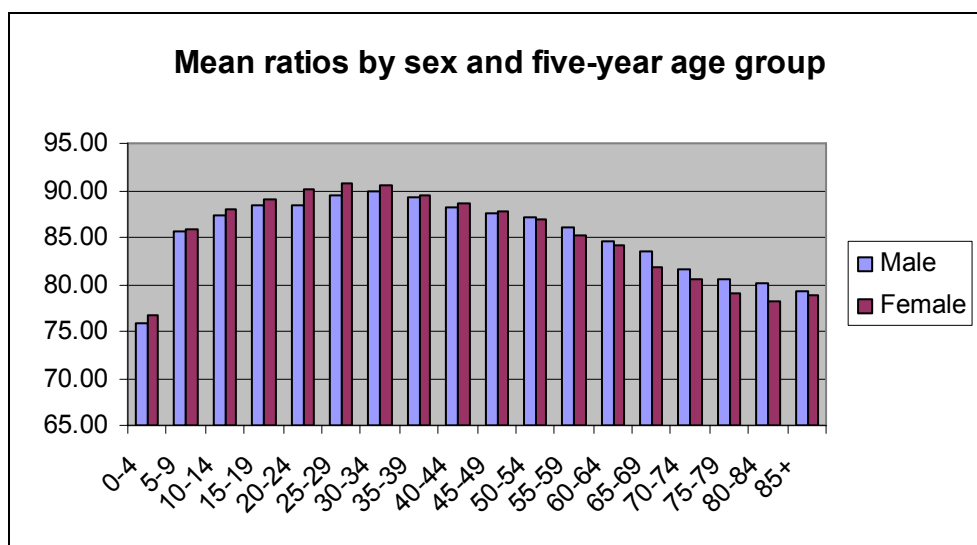
The difference is slightly larger than for the income question, although still small (well under 1% in every case). Forms that are posted out have a slightly higher mean completeness ratio, possibly because there is no direct contact with an enumerator so people are less reticent to give personal details. However this small effect is far outweighed by the clear improvement in form return rates for hand-delivered forms (see section 5.1).

5.7.6 Breakdown of Person Questions Completeness by Age and Sex

Table 30 and Figure 26 below give the mean ratios for the person section by sex and five-year age group.

Table 30 - Completeness Ratios by Five-year Age Group and Sex

Age group	Person completeness ratio (%)	
	Male	Female
0-4	75.75	76.70
5-9	85.56	85.82
10-14	87.33	87.95
15-19	88.30	89.12
20-24	88.46	90.01
25-29	89.47	90.75
30-34	89.94	90.63
35-39	89.23	89.49
40-44	88.17	88.64
45-49	87.51	87.80
50-54	87.09	86.90
55-59	85.97	85.23
60-64	84.61	84.12
65-69	83.47	81.86
70-74	81.51	80.62
75-79	80.53	78.96
80-84	80.09	78.23
85+	79.16	78.74
All	85.89	85.56

Figure 26 - Mean Ratios by Sex and Five-year Age Group

The form was particularly poorly completed for babies and small children. Apart from that, the graph shows a smooth increase in completion rates up to a peak in the late 20s and early 30s, then tailing off with a levelling-out in the 70s and 80s. Females completed a higher proportion of questions until the 50s, after which it is the males that have the higher completion rate. The largest differences between the sexes are in the 20-24 year age range, where females have a completion rate

1.55% higher than males, and the 80-84 year range where the male completion rate is 1.86% higher than for females. No particular questions stand out as explaining the differences between the sexes.

The particularly poor form completeness ratio for children under 5 seems to be largely due to parents omitting questions that are perceived not to apply to children that age. Apart from the language questions (which were poorly completed by all groups) and questions relating to current or former employment (which can only have been answered by mistake for children this age), the questions with the lowest item response rate were method of travel to work or study (52.3%), activity last week (66.4%), "do you feel safe going out in the evening in your area?" (68.2%) and "how many children have you given birth to?" (69.1%). Three of these could reasonably have been considered not applicable to small children, while the question on going out in the evening asks for an opinion, which small children are not likely to be able to provide, and in any case they would not normally be allowed out on their own at any time.

The over-70s also had a comparatively poor ratio, although significantly higher than for the under-5s and, in this case, the culmination of a general trend towards poorer completion among older people. As always, the language questions were particularly poorly completed. The questions asking for details of a second residence also had a poor item response, but these applied to a very small proportion of this age range, as did the question on termtime address. Apart from these, the question with the lowest response was the travel to work or study question (53.2%) followed by "have you ever worked?" (58.6%), hours worked (67.0%), occupation (71.8%) and supervisor status (76.4%). It is likely that people of this age group, few of whom would be currently working, either assumed that the questions did not apply to retired people or else were unable to recall the relevant details of their last job.

5.8 Ethnic group

The frequencies for the responses to the ethnic group question can be found in Annex C.

Gypsy/Travellers were a group of special interest in the Test. In total only 21 people identified themselves as in this group, of which only 14 were in the Highland areas of Lochaber and Breadalbane where they were expected to be concentrated. However this does not include multi-tick responses - it is possible that some people identified themselves as both a Gypsy/Traveller and a member of one or more other ethnic groups. Other people may also have attempted to identify themselves in more than one category, and work is needed to quantify this.

The table in Annex C does not distinguish between the various write-in responses in categories such as "European - Other" or "Asian - Other". These have not been

captured, and so it will be necessary to view the images of forms that have something written in these categories.

Other recommended work on this topic is:

- on interaction of national identity and ethnic group questions
- to match these Test results at individual level with 2001 Census forms to identify changes in self-perception
- further investigation of numbers of Gypsy/Travellers

Recommendation 15

It is recommended that more in-depth analysis be carried out on the ethnic group question, in conjunction with the Scottish Executive.

The area profiles at section 5.15 currently lack any information on the ethnic makeup of the districts. Including this information would improve understanding of response levels.

Recommendation 16

It is recommended that a profile is created of the ethnic makeup of each Census District involved in the Test.

5.9 Marital and Civil Partnership Status

For the first time in a Census-type exercise in the UK, the marital status question incorporated the concept of a same-sex civil partnership, which became law in Scotland only four months before Census Test day. For each category that applied to marriage, it was necessary to provide an equivalent category for civil partnership: for instance, the category “divorced” was paralleled by “dissolved from a civil partnership”.

Table 31 reproduces from Annex C the frequency counts for this question.

Table 31 - Frequency Counts for Marital/Civil Partnership Status

Response	Frequency	%
Single	16786	36.1%
Married (first marriage)	16696	35.9%
Re-married	2114	4.5%
Separated	1282	2.8%
Divorced	2998	6.4%
Widowed	4298	9.2%
In a same-sex civil partnership	69	0.1%
Second or subsequent civil partnership	203	0.4%
Separated from a civil partner	29	0.1%
Dissolved from a civil partnership	67	0.1%
Surviving civil partner	137	0.3%
Missing	1511	3.2%
Error	304	0.7%

Those claiming to be, or to have been, in a civil partnership account for only about 1% of the total. Nevertheless, they are much higher than might be expected (only 259 partnerships were registered in the whole of Scotland in the first quarter of 2006) and show an unexpected pattern, with three times as many people in their second or subsequent civil partnership than their first, despite the recent introduction of the law - in fact no second civil partnerships had been registered anywhere in the UK at the time of the Test. Many people evidently misunderstood the concept of a civil partnership, perhaps thinking that it referred to heterosexual couples who lived together but were not married, or those who were married in a civil, rather than religious, ceremony. Further work is needed on how to avoid such misunderstandings.

Recommendation 17

It is recommended that further work be carried out on the marital and civil partnership status question, to determine whether the question was understood correctly and how it can be improved.

5.10 Second Residency

1729 individual respondents (3.7% of the total) stated at question 8 that they had at least one other residence besides the one at which they were enumerated. This excludes those that stated that they had no other fixed address, but includes all others who ticked one of the “yes” boxes of this multi-tick question, even if they also ticked “no”. For a more detailed breakdown, see Annex C.

Table 32 shows how well question 9 (“what is your other address?”) was filled in for these 1729 people. Where something has been written in the address text boxes, no attempt has been made to establish its validity. Postcodes, however, have been validated against Royal Mail’s Postal Address File (PAF).

Table 32 - Number of Postcodes Given for Second Residence

Census District	01	02	03	04	05	Total	Percentage
Valid postcode	33	205	82	202	315	837	48.4%
Partial postcode	10	36	11	12	14	83	4.8%
Invalid postcode	1	11	7	2	3	24	1.4%
Address but no postcode	34	131	98	136	120	519	30.0%
No address information	49	69	63	37	48	266	15.4%
Total	127	452	261	389	500	1729	100.0%

Further work on this topic is recommended by ONS's report (Annex A), as stated in Recommendation 18:

Recommendation 18

It is recommended that analysis is carried out on the format respondents used to write in their second address, and whether the space provided was sufficient.

5.11 Visitors

The number of visitors in a household could be measured in two ways - by the answer to V1 which asked "How many visitors stayed here on Census Test night (22/23 April 2006)?" and by the number of visitors who completed visitors information in V2. A maximum of 4 visitors could be recorded in V2 and households were instructed to only complete details for the first 4 visitors.

The total number of visitors recorded in V1 is 802. Table 33 gives the number of visitors in each Census District from the answers given in V1.

Table 33 - Number of visitors recorded on V1

Number of visitors recorded on V1						
Census District	01	02	03	04	05	Total
Number of Visitors	116	187	107	144	248	802
Percentage	14%	23%	13%	18%	31%	100%

2.4% (563 of 23,770) of households recorded visitors. Table 34 gives the number of households in each Census District with visitors recorded on V2.

Table 34 - Number of households with visitors recorded on V2

Number of households with visitors recorded on V2						
Census District	01	02	03	04	05	Total
Number of Households with Visitors	75	161	101	111	115	563
Percentage	13%	29%	18%	20%	20%	100%

The number of visitors per household for all Census Test areas according to V1 is show in Table 35. Households that omitted the question are considered to have recorded no visitors in V1. Note that the household that recorded 46 visitors was in fact a hotel, although the household questions were apparently filled in only for the part of the building occupied by the owner's family. No details were given at V2 for any of the visitors. It is likely that some of the other households with a large number of visitors were also hotels or guest houses.

Table 35 - Number of Visitors per Household (V1)

Number of Visitors per Household (V1)		
	Frequency	Cumulative Frequency
1	289	289
2	90	379
3	37	416
4	24	440
5	2	442
6	1	443
7	1	444
8	3	447
9	1	448
10	1	449
14	1	450
46	1	451

Table 36 shows the number of visitors recorded in V2. There was only space for four visitors' details on the forms, so this is the maximum number that could be recorded at this point.

Table 36 - Number of Visitors per Household (V2)

Number of Visitors per Household (V2)		
	Frequency	Cumulative Frequency
1	359	359
2	119	478
3	45	523
4	40	563

Even allowing for the space limitation in V2, it is clear that a number of households (at least 111) have included details of visitors at V2 without filling in V1. Table 37 cross-tabulates the number of visitors recorded in V1 and V2.

Table 37 - Comparison of V1 and V2

Comparison of V1 and V2						
V1	Number of visitors recorded in V2					Total
	0	1	2	3	4	
missing	16870	76	32	12	11	17001
0	6293	20	3	0	2	6318
1	27	259	3	0	0	289
2	7	3	79	0	1	90
3	4	1	1	30	1	37
4+	7	0	1	3	24	35
Total	23208	359	119	45	39	23770

This confirms that there are more occasions (161) where more visitors are recorded in V2 than vice versa (54). This suggests that a question asking for basic details of visitors is more accurate than one asking for a simple count of visitors.

Table 38 gives an indication of how well usual address information was provided for visitors. The "valid postcode" category includes all those who have ticked the box to say that they have the same address as another visitor. No attempt has been made to determine whether this other visitor has provided a postcode. Visitors who have ticked "no usual address" but have also provided a postcode are included under "postcode given".

Table 38 - Number of Visitors Giving Postcode

Census District	01	02	03	04	05	Total	Percentage
Valid postcode	98	146	82	139	140	605	67.8%
Partial postcode	7	9	14	11	10	51	5.7%
Invalid postcode	0	1	1	3	4	9	1.0%
No usual address	8	8	6	1	3	26	2.9%
Postcode not given	28	55	47	36	35	201	22.5%
Total	141	219	150	190	192	892	100.0%

735 of the 892 visitors had written some text in the usual address space on the form. Of other 157, 2 gave a valid postcode, while 107 stated that they had the same address as another visitor and 21 that they had no usual address. 27 visitors (3.0%) had no address information at all. These data are reasonably encouraging, as it should be possible to obtain a postcode from a written address as long as at least a street and town are given, although this may be labour-intensive depending on the quality of the address information. The address details may in any case be sufficient for matching to the form of the visitor's home address.

Recommendation 19

It is recommended that research be carried out into the quality of address data provided by those visitors who did not provide a valid postcode.

5.12 Respondents' Views on the Questions

Table 39 gives the frequency of each response on the evaluation page at the back of the Household form. Respondents were asked to tick a box if they were unhappy with any particular question on the form. There was also a text box, covering approximately one third of the page, for the respondent to provide any comments. The table shows the number expressing dissatisfaction with each question, along with the number who wrote something in the comments box. The percentage of households responding is also shown; the denominator is smaller for the income question as this was only asked of approximately half the households.

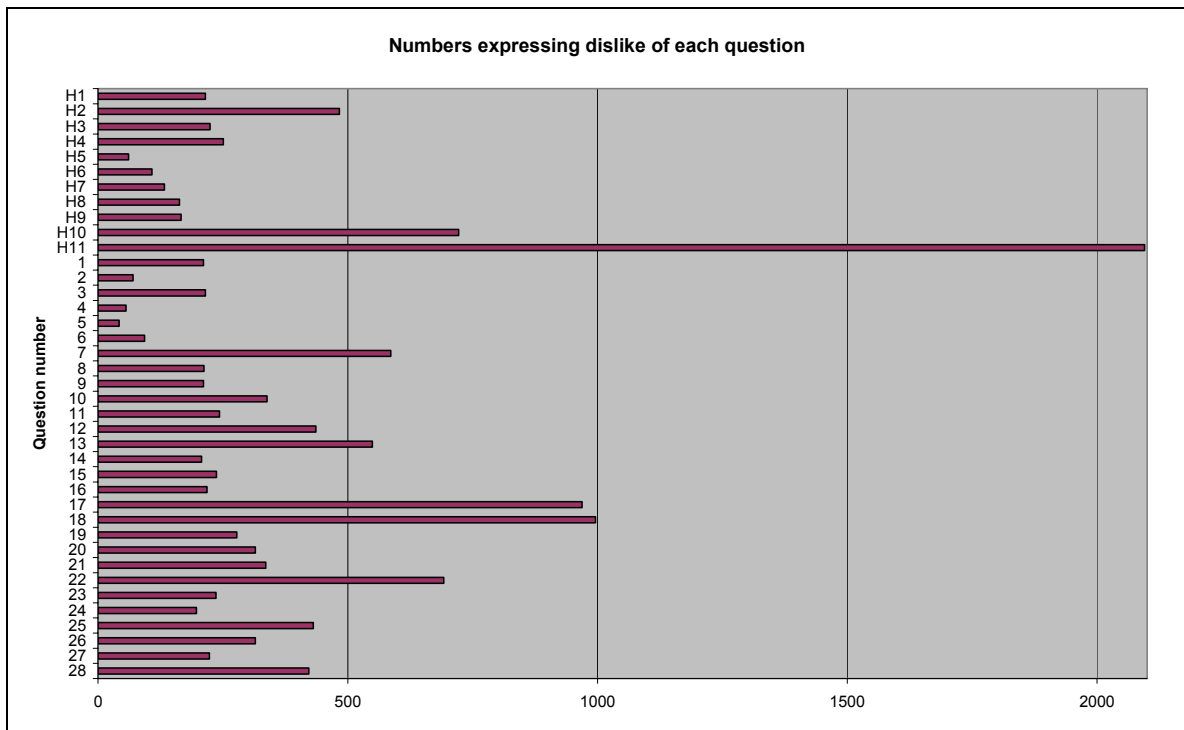
Table 39 - Frequency of responses to "Your Views" page

Question	Total	%
H1 Type of Accommodation	215	0.9%
H2 Repairs or adaptation	483	2.0%
H3 Type of central Heating	224	0.9%
H4 Number of Rooms	251	1.1%
H5 Self-contained	61	0.3%
H6 Tenure	108	0.5%
H7 Landlord	133	0.6%
H8 Croft	163	0.7%
H9 Garden	166	0.7%
H10 Eat Together	722	3.0%
H11 Income	2095	16.9%
1 Name	211	0.9%
2 Sex	70	0.3%
3 Date of Birth	215	0.9%
4 Student	56	0.2%
5 Term Address	42	0.2%
6 Country of Birth	93	0.4%
7 Marital or civil partnership status	586	2.5%
8 Other Address status	212	0.9%
9 Other Address	211	0.9%
10 Nights at other Address	338	1.4%
11 Weeks at other Address	243	1.0%
12 General Health	436	1.8%
13 Health conditions	549	2.3%
14 Number of Children	207	0.9%
15 Safety going out	237	1.0%
16 Experience of discrimination	218	0.9%
17 Current Religion	969	4.1%
18 Religion of Upbringing	996	4.2%
19 National Identity	278	1.2%
20 Ethnic Group	315	1.3%
21 Employment Status	336	1.4%
22 Ever worked	692	2.9%
23 Main Job	236	1.0%
24 Supervisor	197	0.8%
25 Work hours	431	1.8%
26 Main Job	315	1.3%
27 Mode of Travel to work	223	0.9%
28 Language	422	1.8%
Comments	3594	15.1%

By far the highest proportion of respondents, 16.9%, said that they were unhappy with question H11, “What is your household’s total income from all sources?” This is despite the fact that this question only appeared on approximately half of the household forms. The next highest levels of dissatisfaction were with the two religion questions, with 4.2% expressing a dislike of question 18 which asked for the respondent’s religion, denomination or body of faith of upbringing, and 4.1% for question 17 asking for current religion or faith. The highest level of unhappiness for the household questions, besides the income question, was 3.0% for question H10, “How many times last week did your household sit down at a table and eat a meal together?”

Figure 27 shows the total for each question in graphical form (refer to Table 39 for a key linking question number to subject).

Figure 27 - Number expressing dislike of each question



Annex D gives a cross-tab of those expressing dissatisfaction with each question against their response to that question, including non-response.

Recommendation 20

It is recommended that analysis be carried out of cross-tabs of views with response to significant questions, particularly income. The written comments on the views page should also be investigated.

5.13 Placeholder Forms

Parts A-C of the Placeholder form gave the enumerator an opportunity to give a reason why no Household or CE form was returned from the address. Although the three parts were captured separately, the instructions required the enumerator to only complete one part. Forms where more than one were completed are therefore considered to be multi-tick responses.

The frequencies were as follows:

Table 40 - Reason No Form Received

Reason	Frequency	Percentage
Derelict	733	2.55%
Demolished	14	0.05%
Duplicate address	263	0.91%
Doesn't exist	394	1.37%
Non-residential	254	0.88%
Not completed	917	3.19%
Line number used in error	85	0.30%
No form expected - other	365	1.27%
Second residence, holiday accommodation	762	2.65%
Vacant household space	978	3.40%
New build but not yet occupied	45	0.16%
Vacant communal establishment	23	0.08%
Absent household	1138	3.96%
Household refusal	4438	15.43%
No contact but household present	16495	57.37%
Communal establishment refusal	48	0.17%
Absent communal establishment	9	0.03%
Multi-tick	1155	4.02%
Missing	638	2.22%
Total	28754	100.00%

Note that the total includes Placeholder forms where a Household form was also subsequently received from the same address (see section 4.3). No detailed work has yet been carried out to determine whether or not use of Placeholders made a positive contribution to the aims of the Test.

The large number of occurrences of “no contact but household present” illustrates the increasing difficulty in making contact with residents, for instance due to the use of entryphones and the reduction in the number of people at home during the day. This will be one of the major challenges of the 2011 Census.

Recommendation 21

It is recommended that the design and content of the Placeholder form be re-evaluated to determine its usefulness as part of the GROS enumeration strategy.

5.14 Communal Establishments

A total of 56 CE forms were returned. Table 11 in section 4.4.4 gives the breakdown by area. Table 41 shows the breakdown by establishment type.

Table 41 - Number of Communal Establishments by Type

Establishment type	Frequency
General Hospital	3
Other Hospital	1
Nursing Home	3
Residential Care Home	17
Defence Establishment (including ships)	1
Educational Establishment (including Halls of Residence)	2
Hotel, Boarding House, Guest House	9
Hostels (including Youth Hostels, Hostels for Homeless)	4
Other	9
Missing	4
Error	3
Total	56

The number of people counted in communal establishments by area is shown in Table 42.

Table 42 - Number of Communal Establishment Residents by Area

Census District	Frequency
North Glasgow	31
South Glasgow	286
West Dunbartonshire	145
Lochaber	199
Breadalbane	298
Total	959

Further investigations are needed into communal establishment data.

Recommendation 22

It is recommended that further work be carried out on communal establishments, including whether particular types of establishment were less likely to respond and whether all those present were enumerated.

5.15 Test Area Profiles

This section gives background information for each of the five Test areas, derived both from the Test itself and from the Scottish Index of Multiple Deprivation. This information is useful in interpreting the Test results, for instance considering what effect deprivation has on response rates.

5.15.1 Deprivation

This section explores the profile of the five Test areas in terms of deprivation levels, and also looks at how levels of response to the Test varied by deprivation level.

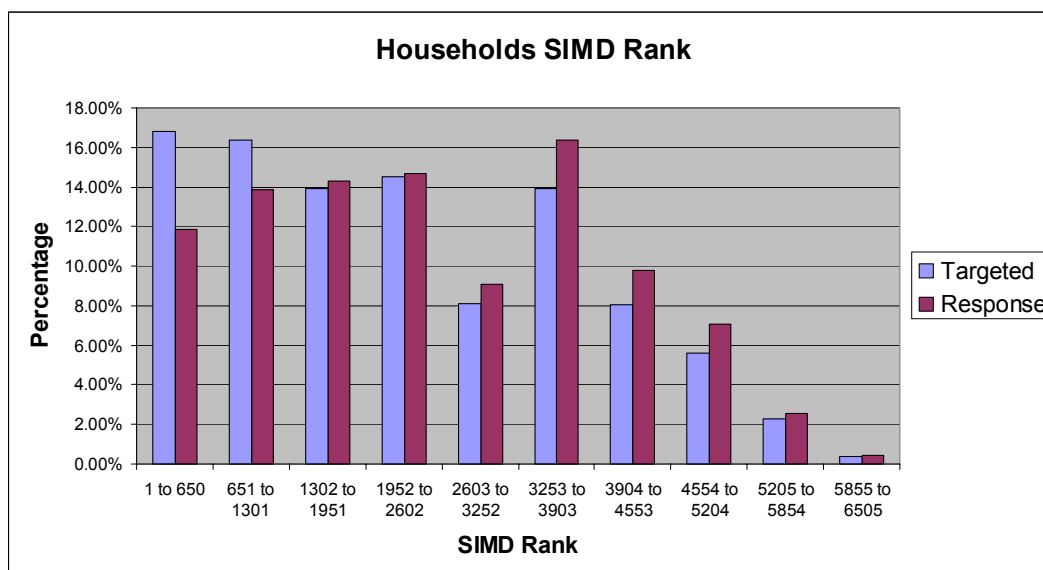
The Scottish Index of Multiple Deprivation 2006 (SIMD) was used to characterise the 5 Census Districts (CDs). The SIMD identifies small area concentrations of multiple deprivation across Scotland in a consistent way. The small area geography divides Scotland up into 6505 “data zones”, with a median population size of 769. These are ranked from 1 (most deprived) to 6505 (least deprived) using 37 indicators of deprivation across seven categories: current income, employment, health, education, geographic access to services, housing and crime. In total, 153 data zones fell wholly or partially in the Test areas.

To characterise the Census Test areas each household was allocated the SIMD rank corresponding to the data zone in which it was located - the lower the rank, the more deprived the area. The histogram below shows the distribution of SIMD rank for households grouped by decile for all areas. The distribution for both household targeted in the Test and the households that responded to the Test are shown. A household that has responded to the Test is a household that has returned a completed form.

It is important to note that the percentage of households in each decile is relative and so a decile with a higher percentage of responded households than targeted households does not represent a higher number of households having responded.

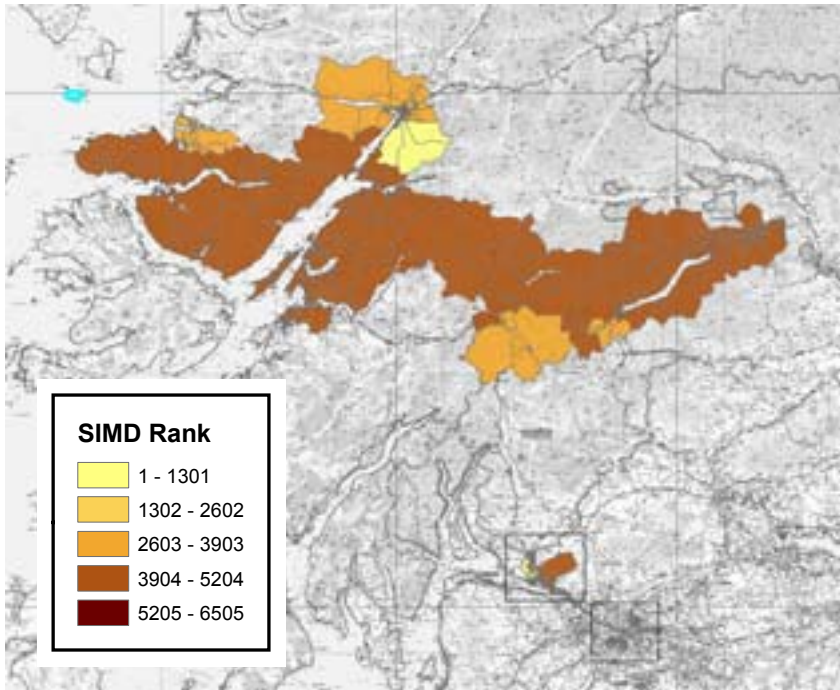
The Census Test enumerated households in all SIMD rank deciles, but the sample was weighted to intentionally target the more deprived areas. As is shown in Figure 28, over 30% of targeted households come from the lower two deciles, while only 5% of households came from the upper two deciles. The distribution of responses shows that a lower proportion of the lower two deciles responded to the Census Test, with 25% of households that responded coming from the lower 2 deciles compared with 30% targeted. Thus response rates were lower in the more deprived areas. This pattern of differential response is not unexpected.

Figure 28 - SIMD Rank of Households



The maps below shows the areas enumerated in the Census Test and their corresponding SIMD rank quintile. Quintiles were used for the maps because they define the differences more clearly. The most deprived areas are represented by the light yellow colour for SIMD rank in the range 1 to 1301. The most affluent areas are represented by a dark red and are in the range 5205 to 6505.

The first map shows the SIMD rank for Lochaber and Breadalbane CDs, with the location of West Dunbartonshire and Glasgow highlighted. The second map shows the location and SIMD rank of North and South Glasgow Census Districts and the third map shows the SIMD rank and location of West Dunbartonshire Census District.

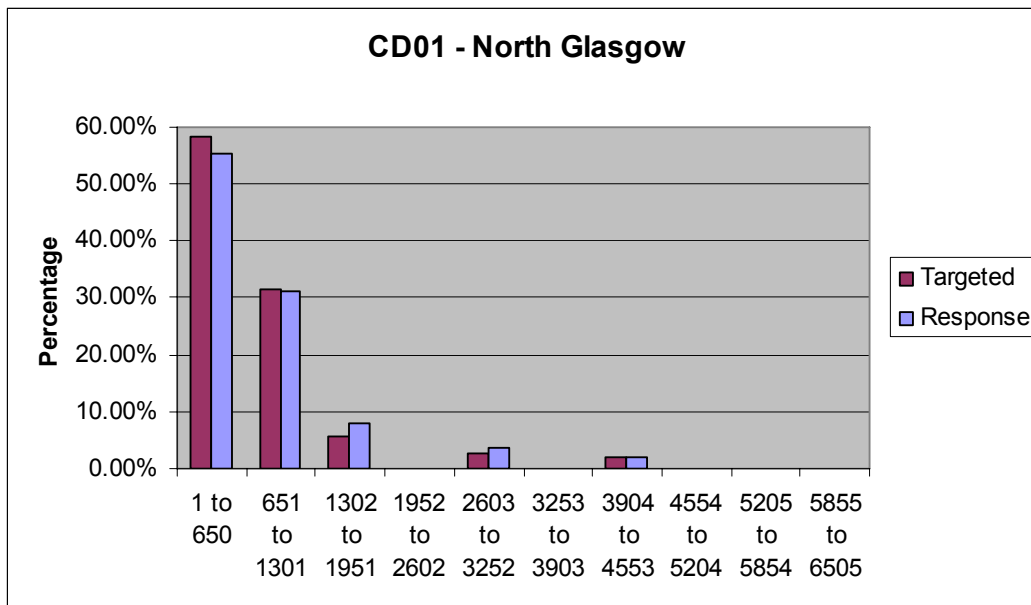


Lochaber and Breadalbane
with West Dunbartonshire and North and South Glasgow highlighted

Census District 01 – North Glasgow

The histogram for North Glasgow shows that this area has the highest proportion of respondents in deprived areas with over 90% in the lower three SIMD deciles. It can also be seen that a slightly lower proportion of households in the lower two deciles responded than were targeted.

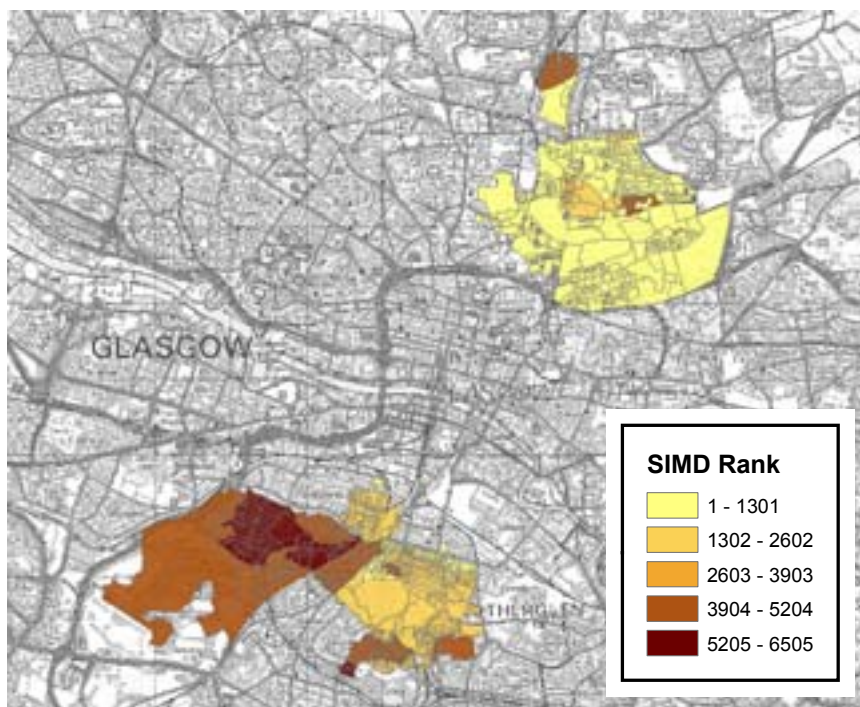
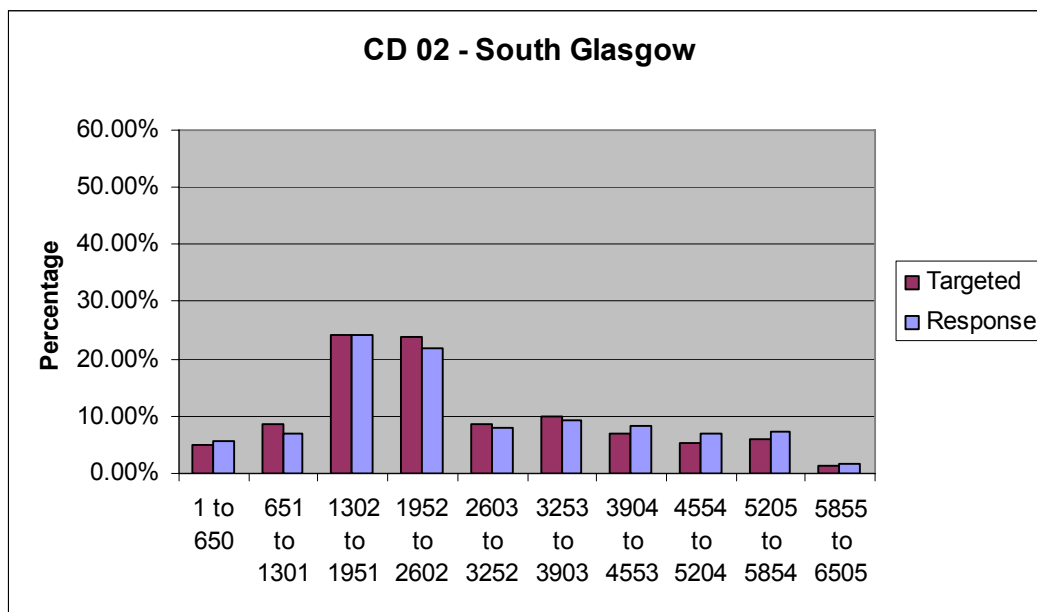
Figure 29 - SIMD Rank of Households in North Glasgow



Census District 02 – South Glasgow

The households in South Glasgow cover the full range of SIMD rank deciles and is the only area to have households in every decile. The highest proportion of households came from the second and third decile. Response is shown to be marginally better in the upper deciles.

Figure 30 - SIMD Rank of Households in South Glasgow

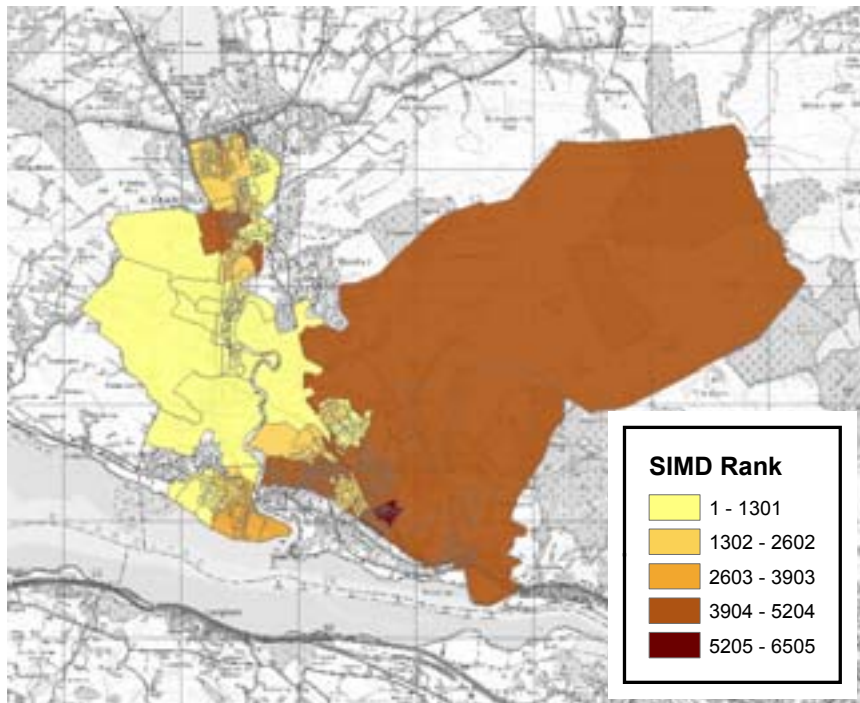
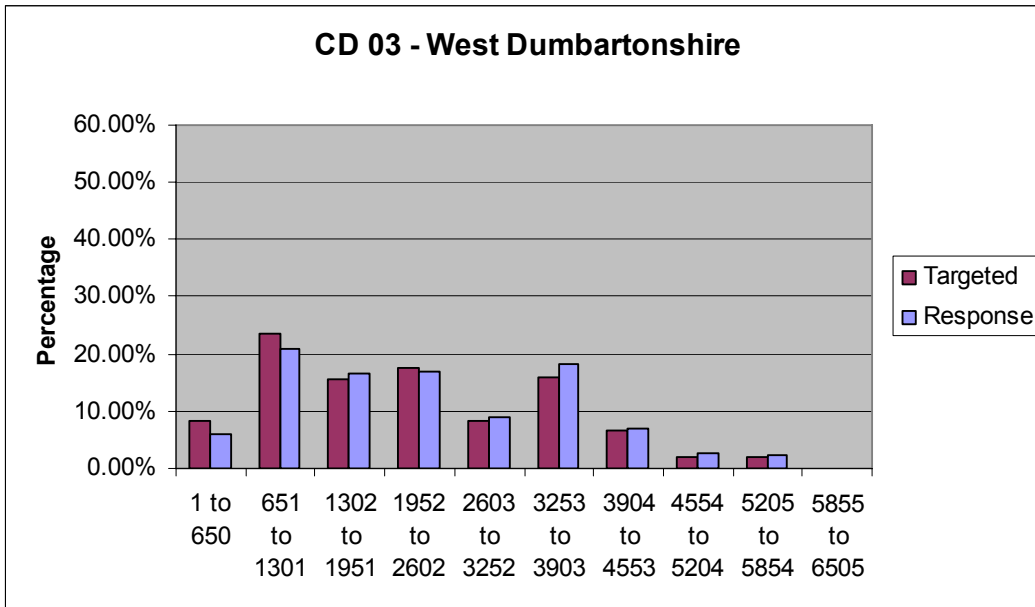


North and South Glasgow

Census District 03 – West Dunbartonshire

West Dunbartonshire contains households in all SIMD rank deciles except the upper decile. Over 75% of responses came from households in areas with SIMD ranks in the range 651 to 3903. The proportion of households that responded from the lower two deciles was lower than the proportion targeted.

Figure 31 - SIMD Rank of Household in West Dunbartonshire



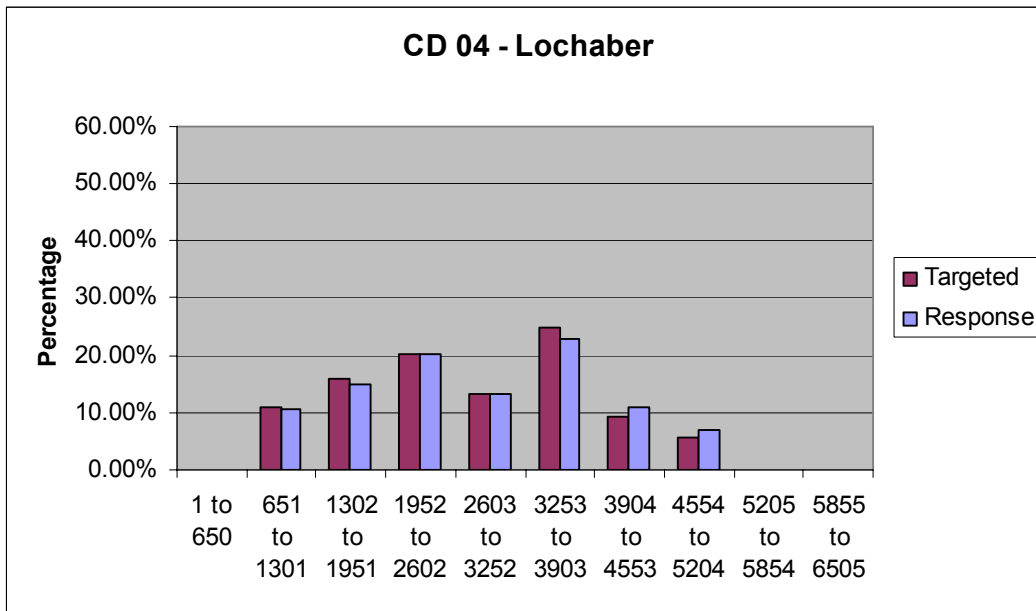
West Dunbartonshire

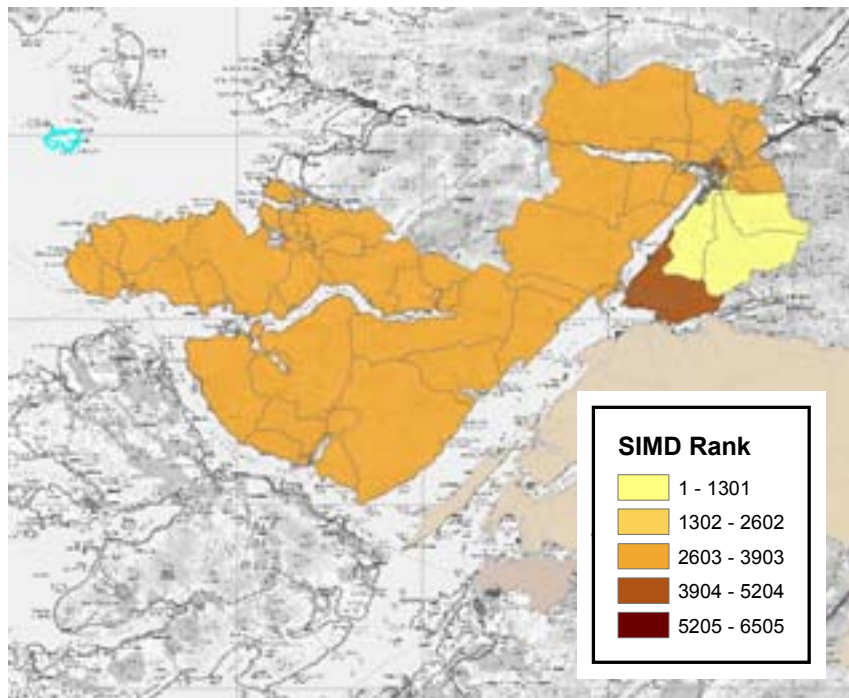
Census District 04 – Lochaber

Both of the rural Census Districts lacked households at the extremes of the deprivation range. Lochaber didn't have any households in the most deprived decile or two most affluent deciles. The best responses came from households in the SIMD rank range 3904 to 5204 in that the proportion of households that responded in this range was higher than the proportion targeted.

All the households in the most deprived quintile were in the major town in this area, Fort William, and surrounding area.

Figure 32 - SIMD Rank of Households in Lochaber



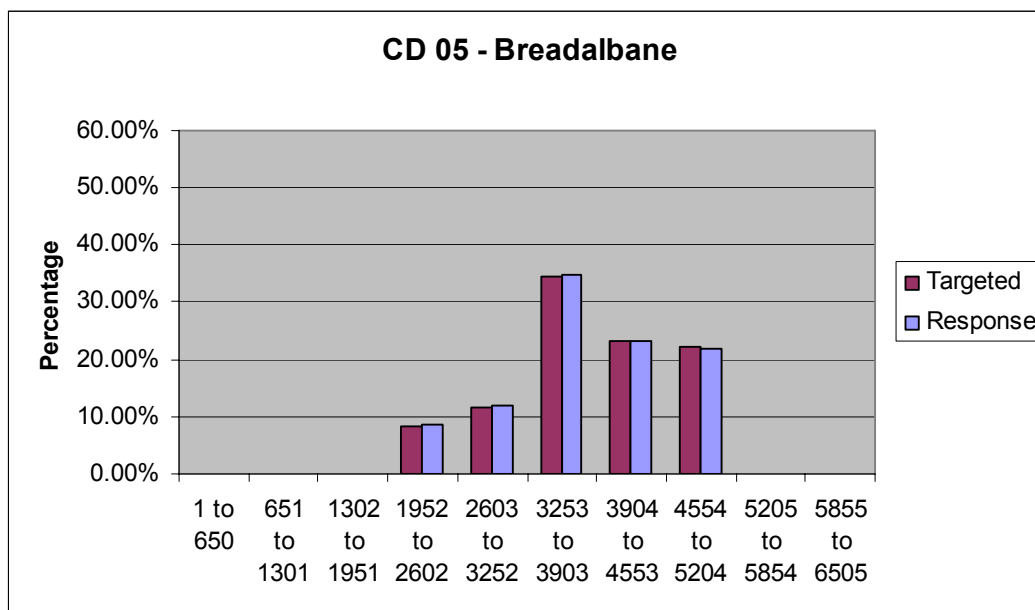


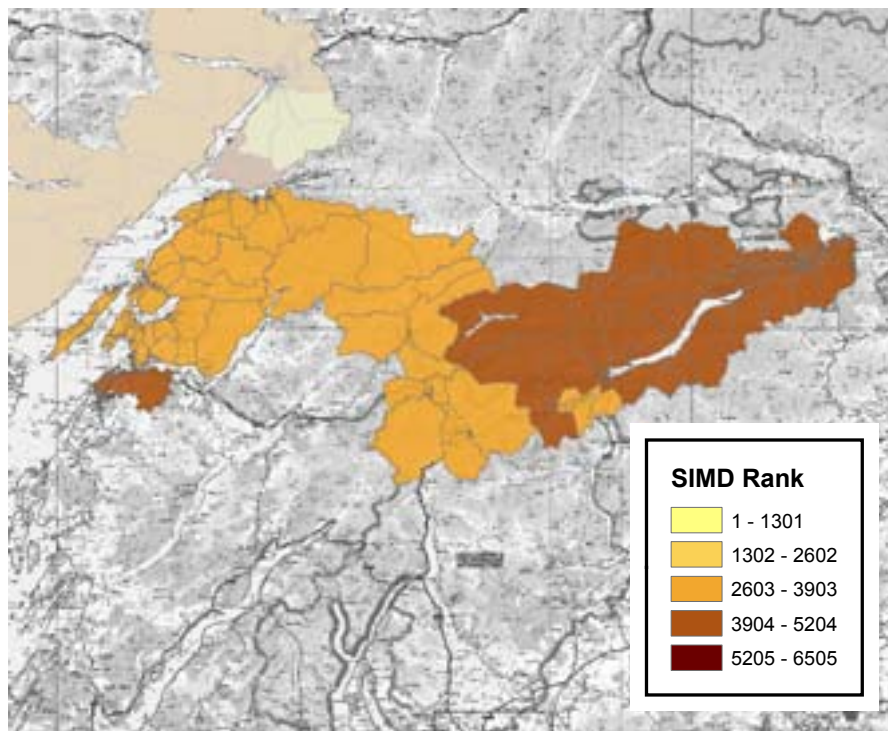
Lochaber

Census District 05 – Breadalbane

Breadalbane did not have any households in the lowest three deciles or the upper two deciles. There was no large difference between the proportion of households that responded and the proportions targeted in each SIMD rank decile.

Figure 33 - SIMD Rank of Households in Breadalbane



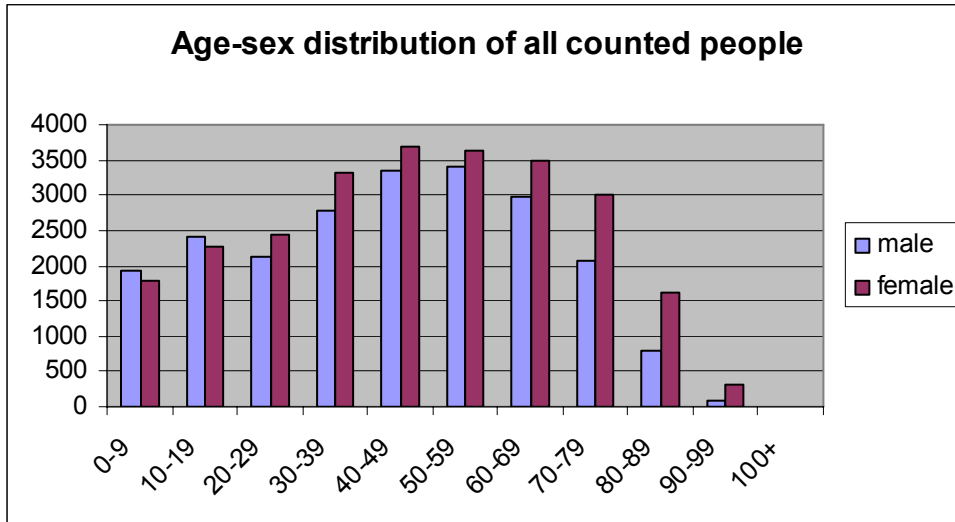


Breadalbane

5.15.2 Age and Sex of all respondents

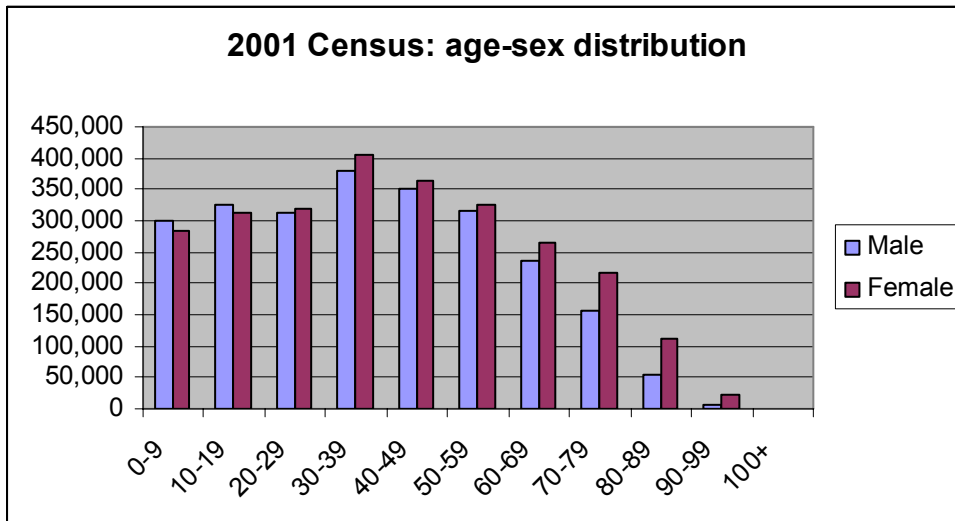
Below is the age distribution by sex for everyone who responded to the Census Test and gave their age and sex. The distribution displays a fairly typical pattern for a large population with the middle age ranges, 30 to 69, representing the highest frequencies. There were more males than females counted in the age range 0 to 19 but above that there were more females than males in every age range. As expected, the ratio of females to males increased across the age groups with more than twice as many females in the age range 80-89 and 90 to 99. The number of people counted in the 20 to 29 age group was low compared with the 10 to 19 and 30 to 39 age groups. This effect is particularly notable for males.

Figure 34 - Age-Sex Distribution of All People



For comparison, the age-sex distribution from the 2001 Census is shown below.

Figure 35 - Age-Sex Distribution of All People in 2001 Census

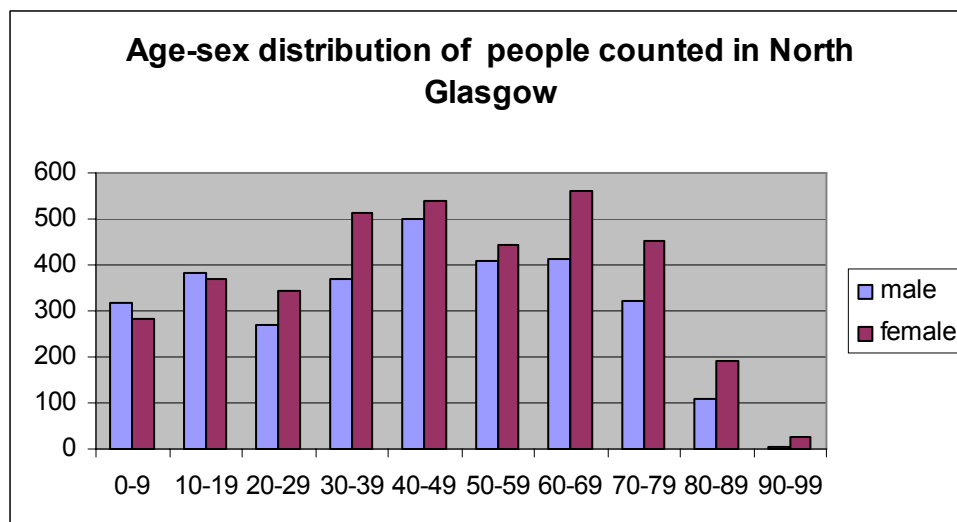


The two distributions are not strictly comparable, as the 2001 Census was compulsory and covered the whole of Scotland. However, the pattern is broadly similar, with the most obvious difference being that the population in 2001 peaks in the 30-39 age group rather than 40-49 as in 2006. This may indicate that people in their 30s are less likely to respond to a voluntary survey, but no definite conclusions can be drawn due to the differences noted above.

The following sections present an analysis of the 2006 data by Census District.

Census District 01 – North Glasgow

Figure 36 - Age-Sex distribution for North Glasgow

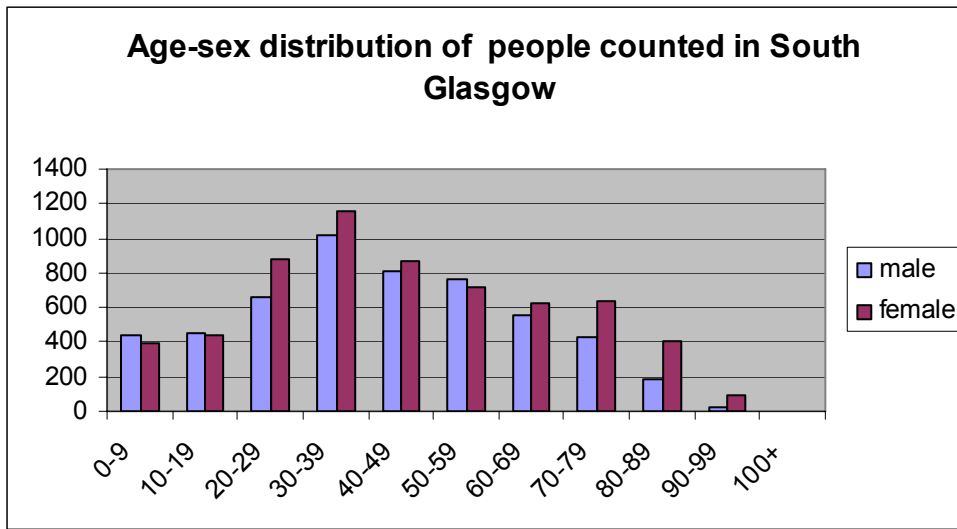


North Glasgow displays a fairly even distribution across the age groups representing a relatively high number of children and young adults compared to the population for all Census Test areas. Again the age group 20 to 29 contains a relatively small number of people compared to adjacent age groups, particularly for males.

It is interesting to observe the relatively high number of females to males in all age ranges from 20 years and over. The 60 to 69 age group has a particularly high number of females relative to other groups. This is almost certainly because men were less diligent in taking part in the Test than because fewer men were actually present.

Census District 02 - South Glasgow

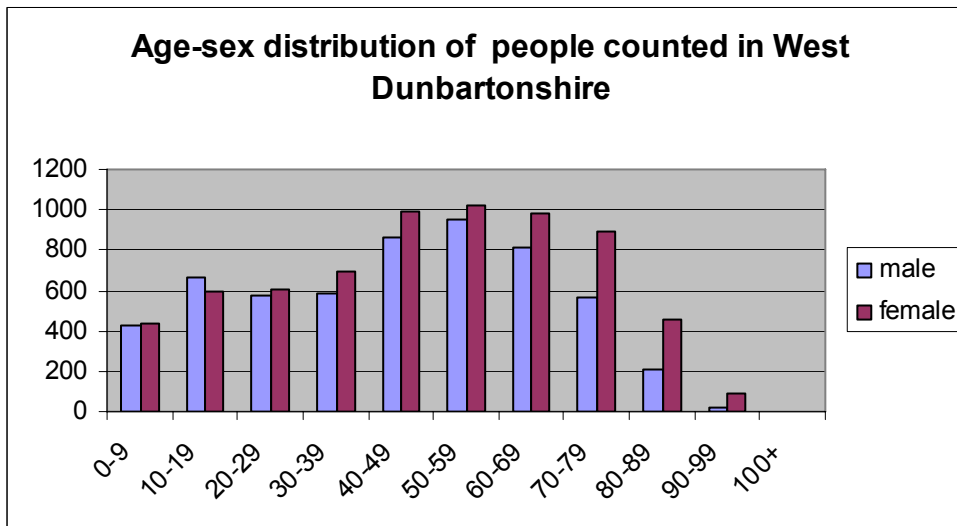
Figure 37 - Age-Sex Distribution for South Glasgow



The age distribution in South Glasgow is slightly skewed towards the younger age groups. This distribution has the lowest mode of all the Census Districts at 30 to 39 years of age. It is also the only Census District to have more people in the 20 to 29 age group than the 10 to 19 age group.

Census District 03 – West Dunbartonshire

Figure 38 - Age-Sex Distribution for West Dunbartonshire



The low number of young people observed in the whole Census Test population is evident in West Dunbartonshire across two age groups, 20 to 29 and 30 to 39. Other than this, the distribution for West Dunbartonshire is similar to the population distribution for all the Census Test Areas taken together.

Census District 04 and 05 – Lochaber and Breadalbane

Figure 39 - Age-Sex Distribution for Lochaber

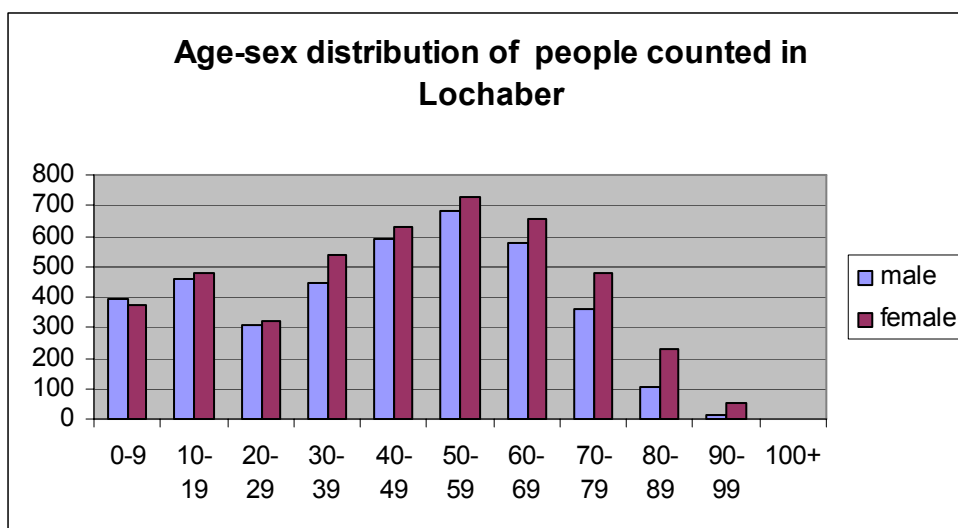
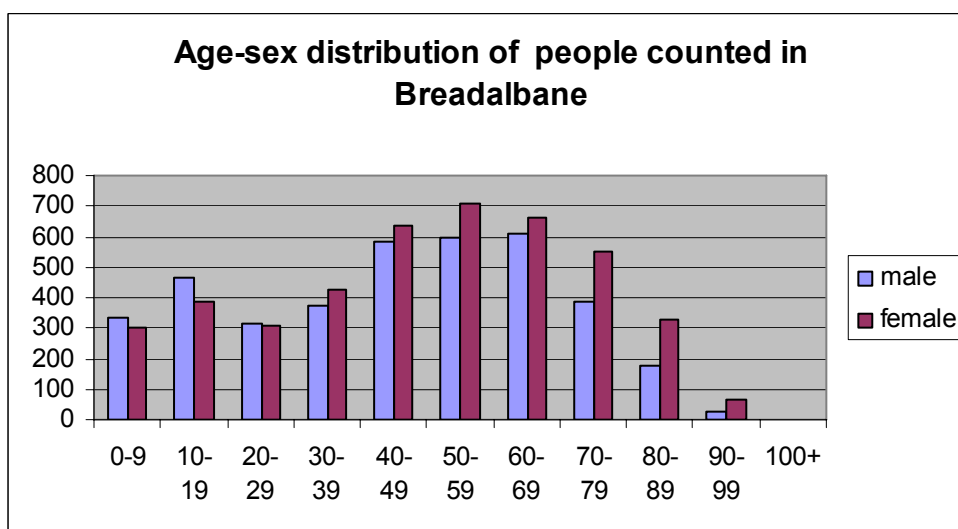


Figure 40 - Age-Sex Distribution for Breadalbane



The Lochaber and Breadalbane sex-age distributions are very similar. Lochaber and Breadalbane both have a relatively low numbers of people in the 20 to 29 and 30 to 39 age ranges compared to other age ranges.

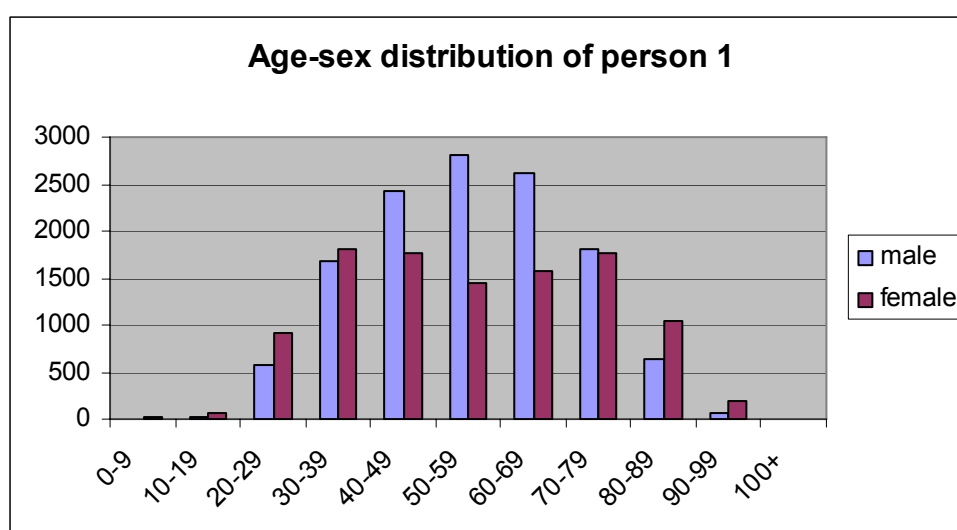
5.15.3 Person One on the Household Questionnaire

The age and sex profile of Person One in the Household form was analysed since this is usually the person filling out the form. The instructions on the form were to “start with the householder or joint householders”, although no definition

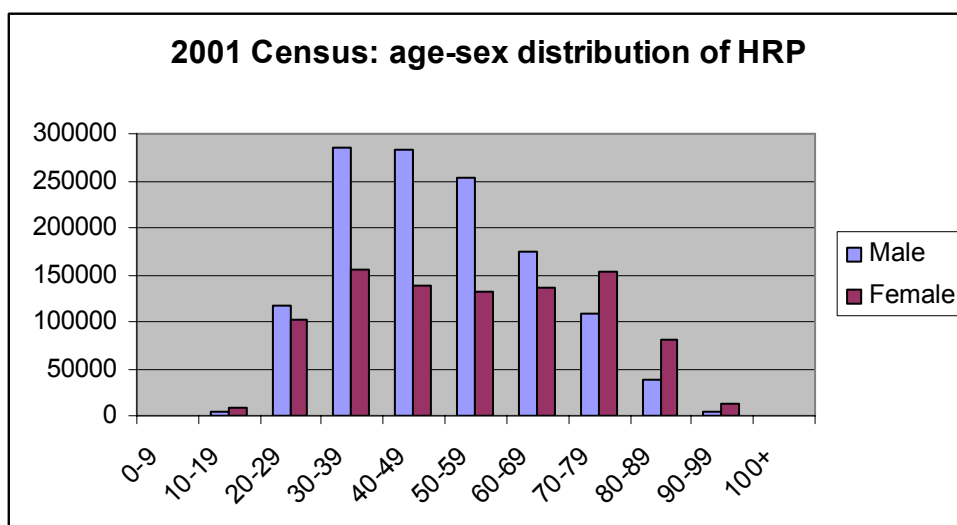
of “householder” was given. The following distributions are the age and sex distributions of Person One on the Household questionnaire.

The chart below shows the distribution of Person One in all Census Test areas. A high proportion of the Person Ones aged 40 to 69 are males suggesting that in this age range males are more likely to be considered the “head of the household”. As age increases above 70, females are more likely to be Person One on the questionnaire, probably because there are increasing numbers of females in the higher age groups. We can also see that there are a number of people, particularly females, who have been entered as Person One in the age range 0 to 19. This might suggest that the assumption that Person One is the form filler is untrue or it could indicate households where the children have a better level of written English than their parents.

Figure 41 - Age-Sex Distribution for Person One



Statistics were not published on Person One in 2001. The nearest equivalent in the published data is Household Reference Person (HRP) which is based on economic activity and age, as well as order on the form. Figure 42 shows the distribution of HRP in the 2001 Census.

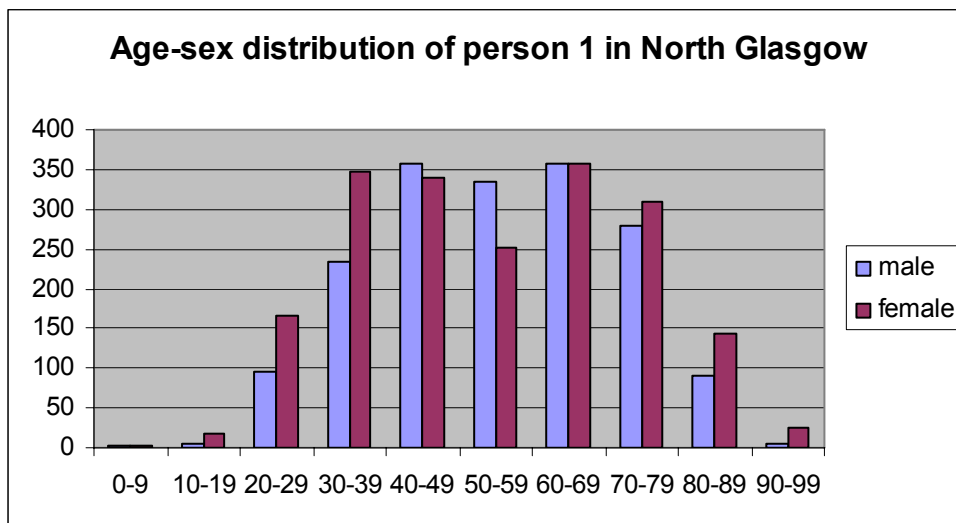
Figure 42 - Age-Sex Distribution for Household Reference Person in 2001 Census

As they are not measuring the same thing, the 2001 and 2006 distributions are even less comparable than for the overall distribution. The fact that the main breadwinner of the household still tends to be male means that males are much more likely than females to be the HRP in the 30-59 age range. On the other hand, the difference is narrower in the 60-69 range, probably because many people of both sexes are no longer economically active at this age. Given the differences in definition, it is difficult to draw any conclusion about the types of people counted in 2006.

Census District 01 – North Glasgow

Person One in North Glasgow was less likely to be female in the 40 to 69 age range, reflecting the higher proportion of total females counted in North Glasgow. The largest difference between the sexes was observed in the 30 to 39 age group where many more females were Person One on the form and the 50 to 59 age group where more males were recorded as Person One.

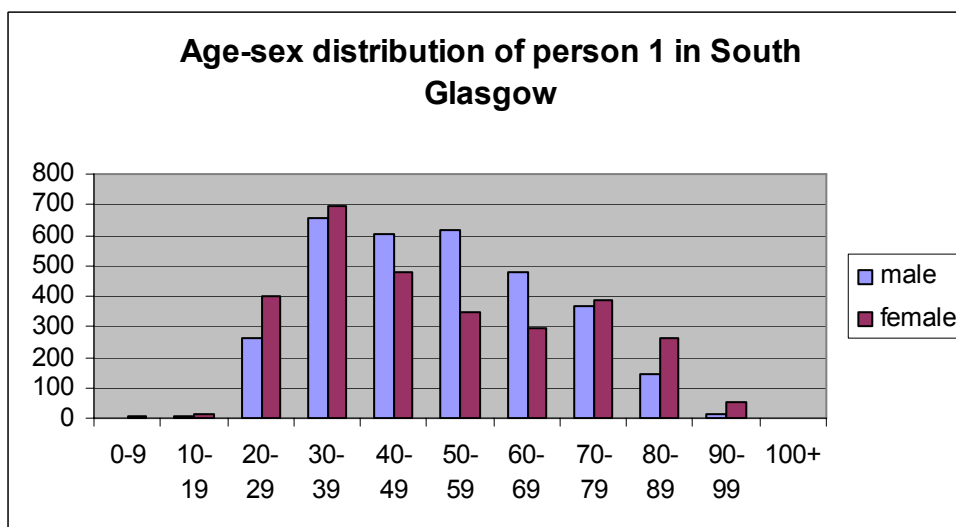
Figure 43 - Age Sex Distribution for Person One North Glasgow



Census District 02 – South Glasgow

The age-sex distribution for Person One in South Glasgow has the lowest mode of all the Census Districts at 30 to 39 years. The low mode age reflects the overall age distribution for that area.

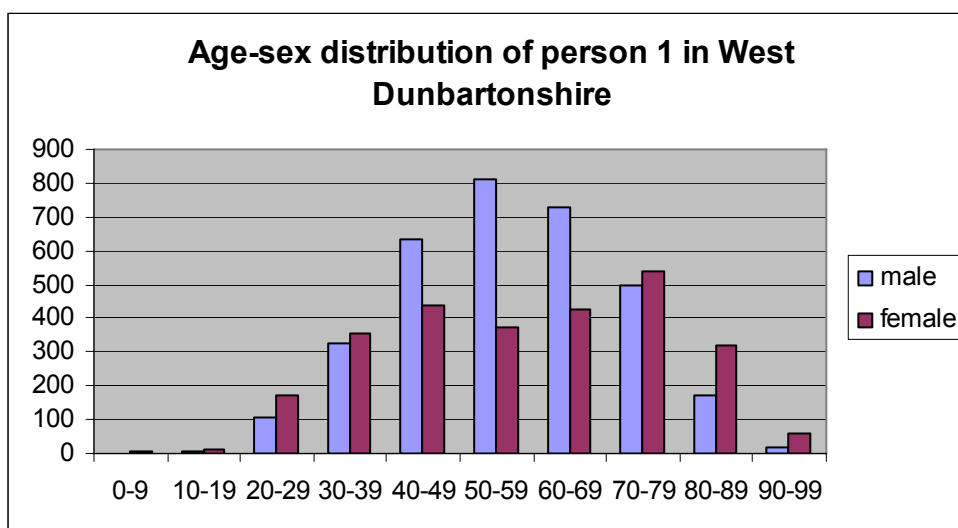
Figure 44 - Age Sex Distribution for Person One South Glasgow



Census District 03 – West Dunbartonshire

The age-sex distribution for West Dunbartonshire is similar to the distribution for all Census Test areas taken together. A large proportion of the people recorded as Person One in the age range 40 to 69 are males. In the 50 to 59 age group over half the Person One responses were male.

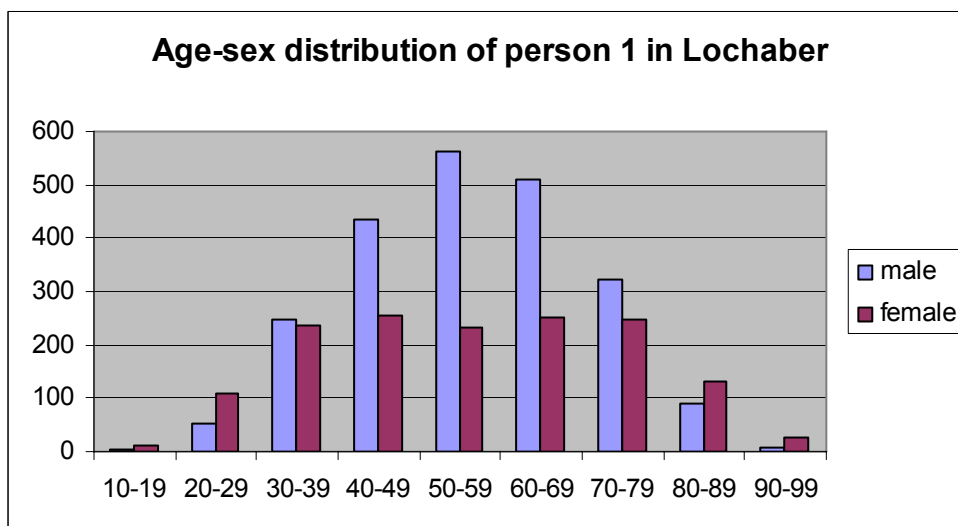
Figure 45 - Age Sex Distribution for Person One West Dunbartonshire



Census District 04 – Lochaber

The predominance of males as Person One in the 40 to 69 age range in Lochaber is very pronounced. Twice as many males were recorded as Person One in the age groups 50 to 59 and 60 to 69 in Lochaber. Lochaber is the only Census District not to have recorded people in the 0 to 9 age range and the over 100 age range as Person One.

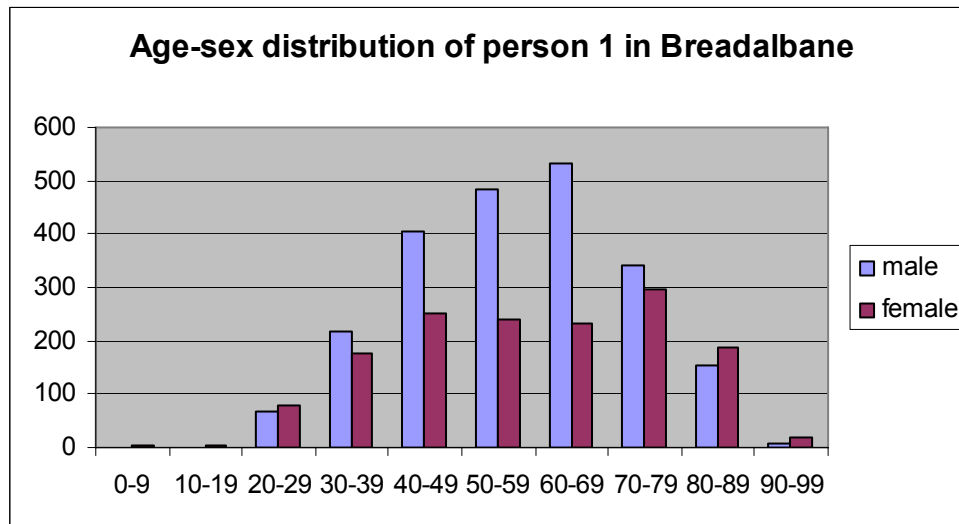
Figure 46 - Age Sex Distribution for Person One Lochaber



Census District 05 – Breadalbane

Breadalbane has the highest mode age of Person One of all the Census Districts at 60 to 69. Like Lochaber, Breadalbane has one of the highest proportion of males in the 40 to 69 age group.

Figure 47 - Age Sex Distribution for Person One Breadalbane



5.15.4 Further Work

It would be informative to also have an ethnicity profile for each Census Test District.

6. Conclusions and Recommendations

6.1 Conclusions

As has been stated, it is difficult to draw definitive conclusions because of the design of the Test (purposive rather than random sampling) and the fact that it was a voluntary exercise, which makes it unlike the 2011 Census which will be compulsory. All the conclusions given below are therefore subject to the proviso that results may not be generalisable to a full Census.

6.1.1 Enumeration Method

Most Test areas showed an improved response rate when the questionnaire was delivered by hand rather than sent through the post. However, work on the item response to the income question shows that people are less likely to answer this particular question if the form is posted rather than hand delivered. Work is yet to be carried out to determine whether this is an issue unique to the income question or whether other questions show a similar pattern.

6.1.2 Income Question

In every Test area, households that received a form containing the income question were more likely to respond to the Test than those whose form did not include this question. It was also found that the presence or absence of the income question made no appreciable difference to how well the rest of the form was completed.

Of those households that received and returned a form with this question, approximately 1 in 8 left the income question blank. This is nearly double the non-response rate of the next poorest household question, although far better than some of the person questions. However there is a concern that this non-response may occur across different sections of society and therefore skew the results, reducing their value.

16.9% of those returning a form with the income question went to the trouble of ticking a box to express their unhappiness with the question. This is more than four times as many as for the next most unpopular question. In a real Census it is important to take into account the reaction of the general public, which may reduce response levels to the Census as a whole.

6.1.3 Item Response

Apart from the income question, whose relatively low item response rate was presumably due to its controversial nature, the questions that were poorly completed seem to be suffering from poor form design. For instance, the question "have you ever worked?" was asked of all people, regardless of their age and whether or not they were currently working. The language question required

respondents to tick a box for each language even if they had no ability in that language - most did not do so.

6.1.4 Form Completeness

Babies and pre-school children, or rather those filling in the form on their behalf, completed the lowest proportion of the form on average. Again this appears to be largely due to poor form design, with questions being asked of small children that do not really apply to them (e.g. about work and fertility). A similar problem exists in the oldest age groups, where respondents often did not give details of jobs that in some cases ended many years ago.

6.1.5 Civil Partnerships

The number of people claiming to be, or have been, in a same-sex civil partnership was far higher than would be expected at such an early stage of the new legislation, and the pattern of current and former partnerships was very unlikely. It seems that people were confused between a same-sex civil partnership and a heterosexual relationship.

The marital status/civil partnership question had a higher level of respondent dissatisfaction than might be expected (see section 6.1.7), possibly due to the presence for the first time of the civil partnership categories.

6.1.6 Second Residency and Visitors

These were both new items designed to help improve coverage for 2011. They are related because both ask for details of another address where the person may be considered resident. The proportion of visitors giving a valid postcode for their usual residence (67.8%) was higher than the proportion of those with a second residence who gave a valid postcode for that (48.4%). However it is of course not possible to tell how many visitors, or people with a second residence, were recorded at all.

In a full Census it may be possible to match these postcodes against the main Census database to estimate overcount in the case of second residences (as people may have been counted at both addresses) and undercount in the case of visitors (as they may have been missed at their home address). This was not done in the Test, as the alternative address was likely to be outside the Test areas. The proportion of valid postcodes, especially for second residence, suggest that this exercise may not be worth the resources that would be required, especially as a large proportion of these addresses were outside Scotland so that this exercise could not be completed until all UK Census data had been processed.

6.1.7 Controversial Questions

As mentioned in section 6.1.2, the income question had by far the highest dissatisfaction rate based on responses to the "Your Views" section on the back of the questionnaire. Other questions with a relatively high level of unhappiness were the questions on religion, whether or not the person has ever worked, marital or civil partnership status and how often the household eats together. It has already been noted (section 6.1.3) that the design of the form caused problems with the question on previous jobs. Religion is always likely to be a controversial subject (and was asked about in the 2001 Census on a voluntary basis), and the dissatisfaction with the marital status question is probably down to the changes made to incorporate civil partnerships. The question on eating together was being tried out for the first time, having been suggested by school pupils in one of the Test areas, and households may have felt that it was too intrusive for a Census. However with the exception of the question on previous work, none of these was among the questions with the lowest item response.

6.2 List of Recommendations

Recommendation 1

For future work it is important that data on the number of forms actually delivered should be readily available to analysts.

Recommendation 2

It is recommended that more time and effort is spent on providing an output specification that accurately represents the format of the data as delivered.

Recommendation 3

It is recommended that the accuracy of answers given by enumerators on the Placeholder forms is assessed by comparing them to the corresponding Household Form.

Recommendation 4

It would be beneficial if a data quality report or incident log were passed over from the data capture team to the statisticians to highlight the reasons for apparent discrepancies in the data. It would also be helpful to receive an incident log from the fieldwork team detailing major issues occurring in the field.

Recommendation 5

Deleted responses should be identified at data capture so that they are not incorrectly identified as images, leading to an invalid multi-tick response. It will not be practical to check images manually in the Census itself.

Recommendation 6

It is recommended that research is carried out into the form reconciliation system that was used in the Test to identify a better system for 2011.

Recommendation 7

It is recommended that the marital/civil partnership status question comes before any filter questions on the questionnaire. A “2 of 4” rule can then be used, similar to that in 2001, to define a valid response, with any 2 questions out of name, sex, date of birth and marital/civil partnership status needing to be answered.

Recommendation 8

It is recommended that work is carried out to investigate the usefulness of Placeholder forms as compared to other options.

Recommendation 9

It is recommended that communal establishments are more clearly defined by HQ so that they can be easily distinguished by enumerators, and clearer instructions given to aid the collection of data from such establishments.

Recommendation 10

It is recommended that item response rates for other questions, besides income, are investigated to determine whether they are affected by the enumeration method.

Recommendation 11

It is recommended that more work be done looking at the quality and usefulness of the response to the income question. It is also recommended that we work closely on this with NISRA and ONS, who are testing a person-level income question in 2007.

Recommendation 12

It is recommended that the images of all multi-tick responses are examined to determine which were genuine multi-ticks and which were corrections. This should be done for as many questions as possible, but especially for those where multi-ticking is thought to be a particular issue.

Recommendation 13

It is recommended that investigative work is carried out into question 22, “have you ever worked?”, to determine reasons for the high item non-response and how this information could be better obtained.

Recommendation 14

It is recommended that non-response by age to the question “have you ever worked?” is investigated.

Recommendation 15

It is recommended that more in-depth analysis be carried out on the ethnic group question, in conjunction with the Scottish Executive.

Recommendation 16

It is recommended that a profile is created of the ethnic makeup of each Census District involved in the Test.

Recommendation 17

It is recommended that further work be carried out on the marital and civil partnership status question, to determine whether the question was understood correctly and how it can be improved.

Recommendation 18

It is recommended that analysis is carried out on the format respondents used to write in their second address, and whether the space provided was sufficient.

Recommendation 19

It is recommended that research be carried out into the quality of address data provided by those visitors who did not provide a valid postcode.

Recommendation 20

It is recommended that analysis be carried out of cross-tabs of views with response to significant questions, particularly income. The written comments on the views page should also be investigated.

Recommendation 21

It is recommended that the design and content of the Placeholder form be re-evaluated to determine its usefulness as part of the GROS enumeration strategy.

Recommendation 22

It is recommended that further work be carried out on communal establishments, including whether particular types of establishment were less likely to respond and whether all those present were enumerated.

Annex A -The Scottish 2006 Census Test: How well did respondent complete the questionnaire?

A Report Produced by

**Data Collection Methodology (Census)
Office for National Statistics**

November 2006

(See separate document)

Annex B - 2006 Census Test Follow Up Survey**A Report Produced by****Statistical Methodology and Geography
GROS****January 2007**

(See separate document)

Annex C - Frequency counts for Test questions

(See separate document)

Annex D - Cross-tabs of Respondents' Views with Responses to Questions

(To be added)