The Scottish Household Survey

Report of the 2001 Census-linked study of survey non-response

Stephanie Freeth and Janet Sparks
August 2004

Summary

This report presents information on the characteristics of non-respondents to the Scottish Household Survey (SHS) using linked data from the 2001 Census and the SHS sample from April to June 2001.

Census data were successfully linked to 90.2% of the records included in the study. However, due to the need to exclude households containing no usual residents and households imputed for the One Number Census, useable census data were obtained for 88.2% of responding and 84.7 % of non-responding households.

The non-response rate for households matched with census records was 28.3% (13.3% non-contacts and 15.0% refusals).

SHS households that were difficult to contact tended to be:

- single adult households;
- households with one adult in employment;
- households consisting of people who did not form a family;
- renting from private landlords, Registered Social Landlords or Housing Associations;
- living in purpose built flats or maisonettes;
- occupying one to three rooms only or
- without a car or van.

The Household Reference Person of difficult to contact households tended to be:

- young,
- single, separated/divorced or
- unemployed.

Households who refused to take part in the SHS were most likely to:

- have no working adult or
- be living at the same address as a year ago.

The Household Reference Person of refusing households tended to be:

- older,
- widowed,
- not academically qualified or
- retired.

Logistic Regression analysis was used to identify the characteristics most strongly associated with each of the elements of non-response.

The results showed that non-contact was more likely to occur in households:

- living in a purpose-built flat;
- containing one person only;
- containing no dependent children or one dependent child or
- whose HRP was single (never married) or aged under 65 years.

The non-contact model also showed that the Local Authority area also had a significant associated with household non-contact.

Refusal was significantly more likely to take place in households:

- located in Aberdeen City, Argyll and Bute, Dundee City, East Dunbartonshire, East Lothian, Falkirk, North Lanarkshire, Shetland and Stirling or
- whose HRP was single, separated, divorced or widowed or who had no educational qualifications.

The logistic regression model for total non-response, which included both non-contact and refusals, shows that interviewers were more likely to encounter non-response in households:

- living in Aberdeenshire, Clackmannanshire, Dumfries and Galloway, Dundee City, East Lothian, Falkirk, Fife, North Lanarkshire, South Ayrshire or Stirling;
- living in a purpose-built flat;
- which does not contain a married or cohabiting couple or
- where the HRP had no educational qualifications.

In addition, the number of adults in employment also had a significant associated with total non-response.

Further examination of the types of household most likely *not* to respond on the SHS indicated that the household characteristics that were significantly associated with total non-response were also related to a number of key SHS measures. This indicates that non-response bias may also affect key SHS estimates and it may be desirable to re-weight the data to adjust for these effects.

1 Background and Methodology

1.1 Background

The Scottish Household Survey (SHS), like all voluntary surveys, suffers from some non-response among sampled units. Although there is interest in absolute levels of survey response, a greater concern is that differential response may result in some groups being under-represented in the achieved sample. If non-response bias occurs in respect of any sub-group, survey estimates based on the achieved sample may be inaccurate and, if the bias is likely to be related to key survey estimates, it may be desirable to reweight the data to compensate for these effects.

There are obvious difficulties in collecting information about survey non-respondents since, by definition, most of them are reluctant to co-operate with interviewers. Over recent decades a substantial amount of information about non-respondents to continuous surveys has been derived from studies linked to the 1971, 1981 and 1991 Censuses (Foster 1998). These census-linked studies of survey non-response (CNR), which involved finding the census information for households in the survey sample, have provided a rich source of data on non-responding households.

The 2001 Census has offered the chance for the Office for National Statistics (ONS) to carry out another CNR. Twelve surveys, including the SHS, have been included in the study. As before, the main aim of the 2001 CNR has been to compare the census characteristics of different categories of responding and non-responding households and to identify variables that are independently associated with non-response. The results from the study are presented in this report.

1.2. The SHS content

The Scottish Household Survey is a major cross-sectional survey which began in 1998. It is commissioned by the Scottish Executive to provide reliable and up-to-date information on the composition, characteristics and behaviour of Scottish households at national and sub-national level. The key objectives of the surveys are:

- To provide household and individual information previously unavailable in Scotland. The information is designed to support the work of the Scottish Executive Development Department's transport and social justice policy areas and the work of the Scottish Parliament.
- To permit disaggregation of such information both geographically and in terms of population sub-groups, for example, families with children or the elderly.
- To allow the relationships between social variables within households to be examined. The will support cross-departmental and inter-departmental policies.

The SHS aims to obtain a sample of 31,000 interviews every two years. The sample covers the whole of mainland Scotland and the Islands and each quarter's sample is designed to produce nationally representative results for Scottish households and the adult population, aged 16 and over, in private households. To ensure that the sample in each local authority area does not fall below a pre-determined accuracy threshold, some of the smaller local authorities have been over-sampled while some of the larger

ones have been under-sampled. Weights are applied in the analysis to adjust for unequal probabilities of selection and to counteract the effect of non-response bias.

Fieldwork for the SHS is spread evenly across the 24 months covered in each two-year sweep. Each sampled address is allocated to a data period/month for interview. At the end of the data period, any outstanding cases can remain in the field and the interviews achieved at these outstanding addresses will be carried into the data file relating a later data period. Details on the carrying over of outstanding cases are described in the Technical Report of the SHS (Hope, 2002).

Interviewing was carried out using Computer Aided Personal Interviewing. The interview is divided into two parts. The first part collects information about the composition and characteristics of the household from the highest income householder or their spouse/partner. The second part focuses mainly on the attitudes and experiences of an adult member of the household. This adult household member is randomly selected from the adults resident at the household. Fieldwork on the 2001 and the current SHS is carried out by a consortium formed by NFO System Three Social Research and MORI Scotland.

Although the sample is based on a two-year sweep, key results are made available on a more regular basis and a detailed report entitled *Scotland's People* is published every year (Scottish Executive, 2002)

1.3 Design of the CNR on the SHS

The main purpose of the CNR was to compare the characteristics of non-responding and responding households, so the sample had to include a sufficient sample from both groups. It was also desirable that the cases should have been sampled as close as possible to the date of the Census (29 April 2001), in order to minimise the number of cases in which the occupants had moved between the Census and the survey interview. This was carried out in previous CNR studies by including addresses selected for survey interviews in the months on either side of census night. However, matching addresses selected for interview before April 2001 was not feasible for the 2001 CNR due to data compatibility problems linked to the adoption, in April 2001, of the National Statistics Socio-economic Classification (NS-SEC) and the modification of a number of classificatory questions used on government surveys in Britain. The 2001 CNR therefore had to match census and survey data for addresses selected for interview from April 2001. In the case of the SHS, this involved matching addresses sampled for interview in between April and June 2001. In total, 6,275 addresses were included for the matching procedure detailed in Section 1.5.

1.4 Response rates on the April to June SHS

Table 1.1 shows the response rates of the April to June 2001 SHS sample. It should be noted that the figures used here refer to final response rates of acceptable data, after coding and editing, rather than the number of co-operating households. The overall response rate for the April to June 2001 SHS sample was 68%. (In 64% of cases, interviews were successfully obtained from both the highest income householder and

the randomly selected adult but in 4% of households, the randomly selected adult was either not available or not willing to be interviewed.) Households who refused to co-operate made up 14.5% of the total sample and interviewers were unable to contact another 14.5% of households. In addition, 2.9% of the sample was still in the field at the beginning of 2002; fieldwork for this 2.9% was carried into 2002.

Table 1.1 SHS sample by response category (April and June 2001)

	Number of Cases	Percentage of Total
Fully co-operating (1)	3,666	64.0%
Partially co-operating (2)	232	4.0%
Total responding households	3,898	68.0%
Non-contacts	832	14.5%
Refusals	833	14.5%
Total non-responding households	1,665	29.0%
Still in field	166	2.9%
Total eligible households	5,729	100.0%
Ineligible households (deadwood)	546	
Base (all households)	6,275	

(1) Successful Highest Income Householder interview and Random Adult interview.
(2) Highest Income Householder interview only.

1.5 The census matching process

The matching process was carried out using the Ordnance Survey Address Point Reference (OSAPR) and other identifying information. In the majority of cases, where an address is occupied by a single household, finding the matching OSAPR was sufficient to identify the matching census household. In 2001, as in earlier studies, the link between survey cases and census forms was on the basis of address and did not take into account whether the address was occupied by the same people at the Census as at the time of the survey. For households in multi-occupied addresses, survey data on sex and age or date of birth of household members were used to enable the matching census unit to be identified. Where the information given did not conclusively identify which census household within a multi-occupied address was the matching unit, ONS randomly selected one household within the matched address.

1.6 Results of census matching

A total of 6,275 households were included in the SHS sample for matching. Overall, 90.2% of the households in the study (5,659) were successfully matched and census data obtained (Table 1.2). Of the 5,659 records matched, 305 contained persons or

households imputed for the One Number Census.[1] In addition, 270 households contained no usual residents because they were occupied entirely by visitors or as a second home. (As discussed in more detail later, imputed households and households with no usual residents were excluded from the analysis).

For 3.6% of the SHS sample, the address was enumerated but census household data were not available. This arose, for example, where the unit was vacant at the time of the Census or the occupants had not complied with the Census. In a further 6.3% of cases, the household address was not traced or was found not to have been enumerated.

The success of the matching process should not be judged simply in numerical terms. It is also important to assess the extent to which the correct census household was identified. Random selection of a household at a multi-occupied address was used in 2.2% of all cases matched so there is a small number of cases where the match was potentially incorrect.

Table 1.3 shows the effective match rate achieved on the SHS for different categories of household response. The match rate was higher for responding than non-responding households: census data were obtained for 92.3% of responding households compared with 90.8% of non-responding households. This difference was due to the lower match rates recorded for households that were not contacted on the survey (87.4%) rather than for refusals (94.1%). The match rate was also lower for addresses which were still in the field (84.9%). Thus it appears that those households which were difficult to contact were also less likely to have been enumerated on the Census or less likely to have filled out a census form.

Table 1.2 Results of the matching exercise for the SHS sample

	Number of cases	**Percentage of total**
Census record available	5,659	90.2%
(of which) Record not altered by the One Number Census	5,354	85.3%
Record contained some imputed persons	160	2.5%
Record was wholly imputed	145	2.3%
No usual residents	270	4.3%
No census record	616	9.8%
(of which) Address found but no census form	223	3.6%
Address not found	393	6.3%
Matched households with usual residence	5,389	85.9%
Matched households with usual residence but excluding imputed households	5,084	81.0%
Base (all households available for matching)	6,275	100.0%

[1] The One Number Census project integrates the 2001 Census counts with the estimated level of underenumeration in the Census. The project adjusts the Census database for the estimated undercount so that all statistics add to 'One Number' – the national estimate of the population.

Table 1.3 Matched households* by SHS response category

	Number of households matched	Percentage of total in response category (3)
Responding (of which)	3,596	92.3%
Fully co-operating (1)	3,380	92.2%
Partially co-operating (2)	216	93.1%
Non-responding (of which)	1,511	90.8%
Non-contacts	727	87.4%
Refusals	784	94.1%
Still in field	141	84.9%
Total eligible households enumerated on the Census	5,248	91.6%
Ineligible (deadwood)	141	25.8%
Total matched with usual residents	5,389	85.9%
All households	6,275	

* Excludes households with no usual residents.
(1) Successful Highest Income Householder interview and Random Adult interview.
(2) Highest Income Householder interview only.
(3) The percentages for this column were calculated using the 'Number of Cases' column in Table 1.1.

1.7 The analysis

Data available for analysis
The 2001 Census contained 36 questions, which were either asked of households or of individual household members. The Census data used in the CNR were fully edited and included derived variables, data imputed as part of the One Number Census and households with no usual residents. As the purpose of the analysis was to use the census data of SHS sampled addresses to examine non-response on the SHS, households classified as having no usual residents on the Census, and those imputed for the One Number were not relevant units of analysis. These households were excluded from the analysis. Of the original 5,659 households matched, 270 cases were excluded because they contained no usual residents and 305 imputed households were also excluded leaving 5,084 households. Once usual residency and imputation had been taken into account, another 106 households classified as ineligible (deadwood)

on the SHS were excluded from the analysis leaving 4,978 records.[2] Table 1.4 shows the number of households included in the analysis by response category.

Table 1.4 Households included in the analysis by SHS response category

	Number of households included in the analysis	**Percentage of total in response category** (3)	**Percentage of all eligible households included in the analysis (response rate used in the analysis)**
Responding (of which)	3,439	88.2%	69.1%
Fully co-operating (1)	3,238	88.3%	65.0%
Partially co-operating (2)	201	86.6	4.0%
Non-responding (of which)	1,411	84.7%	28.3%
Non-contacts	663	79.7%	13.3%
Refusals	748	89.8%	15.0%
Still in field	128	77.1%	2.6%
Total eligible households used in the analysis	4,978	86.9%	100.0%
Ineligible (deadwood)	106	19.4%	
Total household with usual residents and not altered by the One Number Census	5,084		

(1) Successful Highest Income Householder interview and Random Adult interview.
(2) Highest Income Householder interview only.
(3) The percentages for this column were calculated using the 'Number of Cases' column in Table 1.1.

Analysis methods

The analysis presented in this report is essentially descriptive, aiming to identify those census characteristics which are significantly associated with household response. The main unit of analysis used in the study was the household since this is the level at which response and non-response are measured and also that at which the link between census and survey data was defined.

In order to ensure that SHS estimates for any Local Authority area would not fall below a pre-determined accuracy threshold, the SHS oversamples in some Local Authority areas (Hope, 2002). The unequal probability of selection was taken into account in the CNR analysis by applying a scaled sample weight to the matched data. Table 1.5 shows the effect of the sample weight on the number of household included in the analysis.

[2] Fewer ineligible households were excluded from the analysis than ineligible cases presented in Table 1.3. This is because some ineligible households also contained no usual residents or imputed households and were therefore excluded in the first round of exclusion.

Table 1.5 Weighted number of households included in the analysis by SHS response category

	Weighed number of households included in the analysis	Weighted percentage of all eligible households included in the analysis (response rate quoted in most of the tables)
Responding (of which)	3,428	68.9%
Interview obtained from Highest Income Householder and the random adult	3,221	64.7%
Interview obtained from Highest Income Householder but no random adult interview	207	4.2%
Non-responding (of which)	1423	28.6%
Non-contacts	677	13.6%
Refusals	746	15.0%
Still in field in June 2001	125	2.5%
Total eligible households used in the analysis	4,976	100.0%

Census variables were initially considered singly and Chapter 2 presents tabulations and commentary on how the different elements of non-response varied with different household characteristics. The main aims of the chapter are to identify the sub-groups that were most and least likely to be non-responders to the SHS and to identify the bias in the unweighted SHS sample caused by non-response. The chapter also presents summary descriptions of non-responders.

Since there is likely to be a considerable degree of overlap between some of the variables used, an approach which considers each variable individually has obvious limitations. The associations between census characteristics and non-response have been therefore explored further by the use of multivariate analysis in Chapter 3. The aim of that part of the analysis was to identify which census characteristics, of those identified by the univariate analysis, were most strongly associated with total non-response, refusal and non-contact. The association between the factors related to total non-response and key SHS estimates is explored in Chapter 4 to assess if non-response bias, which is expressed in terms of demographic and socio-economic characteristics in Chapters 2 and 3, may also affect key SHS estimates.

It was pointed out earlier that households imputed for the One Number Census were excluded from the analysis. Exploratory investigations have shown that excluding the

imputed cases resulted in the loss of 5.4% households from the analysis and a loss of 9.9% of households in the City of Edinburgh, 10.7% of households in West Dunbartonshire , 11.3% of households in Glasgow City and 11.7% of households in East Renfrewshire. Weights can be developed to adjust for the loss of these households but that was beyond the resources currently available to this stage of the project. Excluding the imputed cases without adjustment have reduced the contribution households in the areas mentioned above made to the overall results presented in Chapter 2 but it has not affected the ability of this study to identify the regional and household characteristics that are significantly related to non-response. This is because region was included as a control variable in the multivariate analysis in Chapter 3 and variations in non-response by region was examined using all the eligible households available for matching. The regional analysis is described in more detail in Chapter 2.

2 Characteristics of non-responding households and persons

2.1 Introduction

This chapter describes how the different elements of non-response are associated with the census characteristics of the household. At this stage census characteristics are considered singly, but the analysis is developed further using multivariate techniques in Chapter 3. The first four sections describe the variation in non-response by Local Authority area, household composition, basic descriptors of the accommodation and characteristics of the Household Reference Person. Section 2.7 provides summary descriptions of non-responding households. It also refers to the research literature to provide insight into why non-response was more likely to occur in certain situations. Section 2.8 examines the implications of non-response by identifying the bias in the responding SHS sample due to household non-response.

Tables presenting the results follow a standard layout in which the rows identify different types of household, as defined by census variables, and the columns show different response categories. In order to meet disclosure control rules for census data laid down by ONS and the General Register Office for Scotland, a few of the categories in the tables have to be combined. Response and non-response rates presented in the tables are based on the total number of households in each category, as given in the final column, and they are additive. Thus the responding, non-contact, and refusal rates sum to 100%. A few minor discrepancies may occur because of rounding.

Except for the results shown in Tables 2.1 and 2.9 and reported in Section 2.2, scaled sample weights were applied to the data in most cases in order to take account of the over-sampling of households in some Local Authority areas. Most of the tables accompanying this chapter therefore show weighted results calculated from the totals shown in Table 1.5.

The chi-square test was used to identify significant associations between the various response and non-response rates and each of the census characteristics shown in the tables. An asterisk indicates that the probability of obtaining such a high chi-square statistic by chance was less than 5%. By adopting the conventional 5% significance level for statistical tests, we therefore conclude that there was a significant association between response and the census characteristic in the tables marked with an asterisk.

Having identified where there was a significant association between response and a census characteristic, the next stage was to identify which categories had significantly high or low response rates. This was done using a test of differences between proportions but only those results which were significant at the 1% level are discussed in the commentary. This more stringent level of testing was adopted primarily to allow for the number of related tests being made, which would result in some spuriously significant differences.

2.2 Regional distribution of non-responding households

The variation in response by region can be obtained from the analysis of SHS data but the results for the months included in the study are shown here for completeness. In order to avoid the regional distortion described in Section 1.7, the table showing non-response rate by Local Authority area is based on all the 6,275 eligible households available for matching, including cases with a valid SHS response code but no or only imputed census data. The cases containing only imputed census data or no data at all can be included in the Local Authority level analysis because address information was available from the survey's sample records. The response details of the households included in the Local Authority level table are set out in Table 1.1. **Table 2.1**

2.3 Household composition of non-responding households

Number of people

SHS non-contact rates varied with the number of usual residents recorded at the Census. Non-contacts were highest in one-person households at 18.7% compared with 12.0% or lower for larger households. There were no significant differences in refusal rates. As the number of people in the household is a combination of the number of adults and children, these observed differences are probably best explained by examining the variables separately.

Single-adult households were the most likely group to be missed because of non-contact, their non-contact rate was 18.6% compared with 10.9%, 8.3% and 8.2 % for households with two, three and four or more adults respectively. [6]

Non-contact was highest in households with only one adult employed (18.6%) while households with three or more adults employed were most likely to have someone at home when an interviewer called (7.5%). Households with no working adults were most likely to refuse, 17.1% compared with 13.3% of households with one working adult and 13.1% of households with two working adults.

The presence of two dependent children increased the chances of making contact with the household; 9.4% of such households were non-contacts compared with 14.3% of households with no dependent children. The non-contact rate for households with one dependent child was similar to that of households without dependent children (14.4%). The age of the youngest dependent child did not effect non-contact rates and the number and age of dependent children had no significant effect on refusal rates.

Table 2.2

Number of family units

A family is defined on the Census as a married or cohabiting couple either on their own or with their never-married children (provided these children do not have children of their own) or a lone parent with their never married children. Individuals who cannot be allocated to a family are classed as non-family members and

[6] An adult is someone who is not a dependent child. A dependent child is a person aged 0 to 15 years in a household (whether or not in a family) or aged 16 to 18 in full-time education and living in a family with his or her parent(s). Households with only one adult are not the same as the one-person households referred to in the previous paragraph because some one-adult households may also contain children.

households comprising only non-family members are classed as 'no family' households.

The non-contact rate was higher for households which did not contain a family (18.5%) compared with those containing families (10.8%). There was no significant difference in refusal rates.

Household type
Table 2.3 gives response rates for two classifications of household type. The first classification (Household Type A) uses information about the family units in a household, separating lone parents from couples with children, and distinguishing between those with dependent and non-dependent children. The second classification (Household Type B) is simply a combination of the number of adults and children in the household.[7]

'Household Type A' highlights single person households (18.8%) and lone parents with dependent children (18.2%) as having relatively high non-contact rates.

'Household Type B' shows that the non-contact rate was highest for single adult households. The presence of children in the household made no significant difference to non-contact rates and there was no difference in the response rates of households containing two adults and those containing three adults.

There were no significant differences in refusal rates for either Household Type.

Table 2.3

2.4 Housing tenure and characteristics of the accommodation of non-responding households.

Non-contact rates were associated with the type of accommodation occupied by the households selected to take part in the SHS. However there were no significant differences in refusals rates by type of accommodation.

Non-contact rates were lowest for households living in homes owned outright (10.1%) and highest for households in accommodation rented from a private landlord (19.2%). With a non-contact rate of 17.3%, households who rented from Registered Social Landlords or Housing Associations were also difficult to contact.

Non-contact rates were highest among households living in purpose-built flats or maisonettes (19.5 %) compared with households who lived in a house (11.8% or less). The type of house (detached, semi-detached or terraced) made no significant difference to the non-contact rate.

[7] Household Type B uses a different definition of an adult and a child. An adult in Household Type B is a person aged 16 years and over and a child is a person aged under 16 years.

Households occupying between one and three rooms had the highest rates of non-contact (20.4%) while households in accommodation containing seven or more rooms had the lowest non-contact rate (9.0%).

In contrast, the presence of central heating in some or all rooms made no significant difference to non-contact or refusal rates.

There is a relationship between the number of cars and vans owned by a household and non-contact. Households with no car or van had a higher non-contact (15.4%) compared with households with two cars or vans (10.2%). There were no significant differences in refusal rates by number of cars.

Households that had not moved in the year before Census night had higher refusal rates (15.4% compared with10.9% for moving households). **Table 2.4**

2.5 Characteristics of the Household Reference Person of non-responding households

Tables 2.5 and 2.6 show household response and non-response rates by selected census characteristics of the Household Reference Person (HRP). This person was identified by adapting the census definition of the HRP. In order to satisfy census disclosure control rules, certain categories in the tables describing the HRP's age, marital status, qualifications and ethnicity were combined.

Age, sex and marital status

Age and marital status of the HRP are useful indicators of both non-contact and refusal. Non-contact was highest among those households whose HRP was aged between 16 and 24 years (25.4%) but dropped steadily in subsequent age group to reach 7.7% for HRP aged 75 years and over. Refusals tended to increase gradually with the age of the HRP, from 9.7% for households whose HRP was aged between 16 and 24 years to 18.9% for households whose HRP was aged 75 years and over. The combined effect of differing levels of non-contact and refusal by age is that the households with the oldest HRPs (aged 75 and over) were more likely to give a full interview (70.2%) than the households with the youngest HRPs (aged less than 25 years) - 59.5%.

Non-contact rate was higher in households where the HRP was single (never married) (23.2%) or separated (but still legally married) / divorced (at 17.9%) than in households where the HRP was married (9.2%). Households whose HRP was widowed were the most likely to refuse to take part in the survey (19.3%) and households with a single (never married) HRP were the least likely to refuse (11.0%).

There were no significant differences in non-contact and refusal rates by sex.

Other characteristics of the Household Reference Person

Refusals were highest for HRP with no academic qualifications (18.3% compared with 13.0% for HRP with O, A, GCSE's and A levels and 10.8% for HRP with a degree or higher qualification). Non-contact did not vary significantly with qualification level.

The country of birth and the ethnic origin of the HRP had no significant effect on non-contact or refusals rates.

The Census gives only limited information on an individual's length of residence at the census address but it does distinguish between those who were living at the same address one year ago and those who had moved in the previous year. Refusals were more likely from households whose HRP lived at the same address as the previous year than from households whose HRP had moved(15.7% compared with 10.6%). There was no significant difference in non-contact rates.

Households with an HRP who was retired were the least likely to feature among the non-contacts (8.6%) while non-contact was highest among households whose HRP were unemployed (22.5%). Refusals were more common in those households where the HRP was retired (17.2%) than from households where the HRP was an employee (12.9%).

The only significant differences in non-contact and refusal rates by the socio-economic classification as measured by the NS - SEC group of the HRP were for HRPs who were full-time students, never-worked or long-term unemployed. These households were less likely to non-contacts (11.2% compared with 13.2% or more for other households) and more likely to refuse (17.9% compared with 15.5% or less for other households). **Tables 2.5 and 2.6**

2.7 Summary descriptions and discussion of the characteristics of non-responders on the SHS

The previous sections have shown that households that were difficult to contact tended to be:

- single adult households;
- households with one adult in employment;
- households that did not include a family;
- renting from private landlords, Registered Social Landlords or Housing Associations;
- living in purpose built flats or maisonettes;
- occupying one to three rooms only or
- without a car or van.

The Household Reference Person of difficult to contact households tended to be:

- young,
- single, separated/divorced or
- unemployed.

Households who refused to take part in the SHS were most likely to:

- have no working adult or
- be living at the same address as a year ago.

The Household Reference Person of refusing households tended to be:

- older,
- widowed,
- not academically qualified or
- retired.

These results are in agreement with those cited in the research and other literature on non-response, compiled by Groves and Couper (1998). In addition to highlighting the factors associated with non-response, Groves and Couper have also outlined how the factors may have influenced response. The findings of Groves and Couper are summarised here to provide insight into the sources of non-response.

Non-contact

Extensive research in the USA cited by Groves and Couper suggests that residents of large cities who work outside the home may spend more time travelling to their place of employment than residents of smaller urban areas. Further, they may require more time away from the home for grocery purchases, shopping, and other activities. They may also have more entertainment options that take them away from home. Groves and Couper also suggest that these factors may influence the non-contact rate for households that do not own a car, a suggestion that seems to be reflected in the results for the SHS.

Non-contact rates were highest in SHS sampled households containing one person only. Groves and Couper point out that 'If one person lives alone in a housing unit, contact is completely dependent on when he or she is at home… If the at-home times of members of the same household were completely independent of one another, then the larger the number of persons in the household, the larger the probability that at any one call someone would be contacted'. They also assert that as the number of single person households is increasing, it is getting more difficult for interviewers to contact these households, especially as more time is being spent outside the home.

Those households in the SHS sample who lived in a purpose-built flat or maisonette were also hard to contact. These properties may have security measures, which affect the ability of the interviewer to contact the household. Groves and Couper observe that people from all levels of society now live in flats and other divided accommodation. Many apartment buildings in cities have locked central entrances, doormen or security guards, or intercom systems which can impede the interviewer's chance of contact with the sampled household. Accommodation in suburban and rural areas may have trespass signs, or warning signs about dogs or locked gates, which hinder access by visitors.

Groves and Couper note that households with young children may be easier to contact because they are more likely to have an adult carer at home than households without young children. This study has found households with 2 dependent children having lower non-contact rates, however contact rates for families with one dependent child were similar to those for families with no dependent children.

Refusals
Groves and Couper have found differences in non-response between cities and rural areas in the USA. They suggest that this effect may be a function of inherent features of life in large urban areas – the faster pace, the frequency of fleeting single-purpose contacts with strangers, and the looser ties of community in such areas.

Groves and Couper have also suggested that community ties may also be linked to lower co-operation rates from single-person households and higher co-operation from households with children. Single person households tend to be less socially integrated into the local community and therefore may feel less obligated to take part in a survey. In contrast, households with children are more likely to be integrated into the community through schools and other social networks. However the results presented here do not appear to correspond with this aspect of Groves and Couper's assertions.

The literature also highlights the higher rate of refusals among the elderly, similar to the findings reported here. Groves and Couper point out that elderly people are more likely to be at home compared with people of other age groups, because of their lower employment rates and, at advanced ages, reduced mobility. However, health problems prevent some from taking part in surveys.

2.8 Bias in the SHS responding sample

Having looked in some detail at variation in different types of non-response according to the characteristics of households, we now use the same data to illustrate the bias that results from non-response on the SHS. Tables 2.10 to 2.14 compare the frequency distributions for census characteristics of households that have fully or partially responded to the SHS and for all households in the SHS-census matched sample. The results of the chi-square tests presented in sections 2.2 to 2.6 are shown again on Tables 2.10 to 2.14 to indicate the census characteristics that were related to non-response on the SHS.

In order to give an indication of the direction and size of non-response bias, a correction factor was calculated for each category by dividing the percentage of such households in the SHS-census matched sample by the corresponding percentage in the responding sample. The correction factors show the adjustment, or 'weight', that would need to be applied to the responding sample in order to achieve the same distribution as for the total set sample. However, the 'weights' are used here to illustrate the effect of non-response. It is unlikely that factors based on single variables would actually be used in weighting to adjust for non-response. The more a correction factor departs from 1.0, the greater the effect of non-response. Categories which are over-represented in the responding sample have factors of less than 1.0.

Most of the correction factors for categories defined by single variables were in the range 0.95 to 1.05, indicating that the bias was relatively small. It would be expected that more complex categories, derived from the cross-classification of two or more simple variables, might yield more extreme correction factors. However, it is increasingly likely as variables are combined that small cell sizes may give unstable results.

In assessing bias, emphasis is usually placed on sub-groups which are substantially under-represented in the responding sample. These groups have high correction factors (1.10 or above) and tend to have significantly low response rates. Sub-groups in the SHS sample with correction factors of 1.10 and above included households:

- living in Dumfries and Galloway (1.10), Dundee City (1.17), East Lothian (1.19), Falkirk (1.12), Fife (1.10) and Stirling (1.17);
- in accommodation with between one and three rooms (1.11);
- in accommodation with no central heating (1.11);
- where the HRP was separated or divorced (1.10) and
- where the HRP was unemployed (1.14)

Correction factors of between 1.05 and 1.09 were more common. They were seen for households living in Clackmannanshire, one person households (1.09), households with one adult in employment (1.05), households which did not include people who formed a family (1.09), households living in a purpose built flat or maisonette (1.09) or in a converted/shared house (1/05), households where the HRP was aged between 16 and 24 years (1.09), single (never married) (1.08), had no academic qualifications (1.05), who was looking after the home or family (1.09), or a student or permanently sick or disabled (1.09) and households where the HRP was working for a small employer or on their own account (1.05).

The only sub-group that were considerably over-represented in the responding sample (indicated by having a correction factors of less than 0.9) was households in the Western Isles. **Tables 2.10 to 2.14**

This chapter has examined the relationship between household characteristics and non-response. It has revealed an association between non-response and a wide range of household variables. As mentioned earlier there is likely to be a considerable degree of overlap between some of the variables investigated. A number of the associations described in the text may be one basic relationship or factor shown in alternative ways. Further analysis, using multivariate techniques, have been carried out to allow for the effect of the overlap between variables. This will enable us to identify the factors that were the most strongly associated with non-response on the SHS. The results of these analyses are presented in the next chapter.

3 Multivariate analysis of non-responding households

3.1 Introduction

Chapter 2 provided a substantial amount of information on how non-response rates differ according to the census characteristics of the household. Yet, as indicated earlier, the analysis of individual characteristics has obvious limitations since some characteristics are clearly related. As a result, some of the associations may be the same basic relationship shown in alternative ways.

This chapter presents the results of further analysis to identify which characteristics are most strongly associated with the two elements of non-response and with total non-response. The results are based on logistic regression which can be used to predict the probability of an event occurring, such as non-response versus response, from a set of independent variables.[8] The technique identifies which of a set of independent variables are most strongly associated with the binary dependent variable, i.e. response versus non-response.

3.2 Analysis methods

The analysis presented in Chapter 2 has identified a large number of associations between household characteristics and non-response. The associations can be divided into three types:

- association between housing characteristics and response
- relationship between household composition and response
- association between the characteristics of the Household Reference Person (HRP) and response.

Since the variables in each of the categories of association identified in Chapter 2 are likely to be more closely related to each other than to the variables in the other categories, a separate logistic regression analysis was first run for each of these categories to identify the variables that were most strongly associated with non-response. This approach enabled us to deal with correlation between the variables in each category and to reduce the likelihood of developing models that were over-complicated and difficult to interpret. Only variables found to be significantly associated in the univariate analysis were entered in the model. Categories of a variable associated with similar levels of non-response and which could logically be combined, were grouped in order to simplify the model while still discriminating between groups with high and low response. The results of these models are reported briefly.

For characteristics relating to the composition of the household, the original tabulations included two classifications of household type. These, although shown as

[8] The logistic regression analyses was run in SPSS using forward stepwise variable selection. At each step the procedure enters the variable with the smallest significant level for the score statistic, if less than 0.05, or removes the variable with the highest significance level for the WALD statistic, if greater than 0.1. The process stops when a previously considered model is encountered or no variables meet the entry or removal criteria. A further manual check was made using the goodness of fit statistics for the model, to ensure that each additional variable resulted in a significant improvement in the log-likelihood statistic.

separate single variables in Table 2.3, were themselves actually relatively complex derivations based on a number of separate items of information about household members, and their inclusion helped to give a more detailed picture of how non-response varied with household composition. For the multivariate analysis it was preferable to go back to the simple variables from which they were derived. This enabled us to develop a simple model of associations that also allowed easier comparison between the associations for different rates and different surveys.

The results for the three categories of variables were then combined in a single logistic regression model which included the main characteristics identified at each of the previous stages together with a small number of separately-defined key variables. The additional variables relate mostly to factors other researchers have found to be associated with non-response. They were included to allow for the possibility that different effects might be significant when the three categories of characteristics were combined. The results for these models are tabulated to indicate the variables that had significant independent effects on non-response and the categories of households that had estimated parameters which were significantly higher than for the reference category.[8] The reference category was defined as that with the lowest non-response rate.

For consistency with the results described in Chapter 2, the logistic regression analyses were run using the non-response rates calculated as percentages of the total matched sample of households. This method of calculating the rates is consistent with the usual practice in describing the outcome for a survey sample and is convenient since it gives additive rates which sum to 100 percent.[9] By using additive rates in the models we are able to build on the results shown in Chapter 2 and to identify the main ways in which different types of non-responding households differed from all households. The logistic regression models therefore give the probability of a household falling into a particular response/non-response category.

All the logistic regression models were developed using a scaled sample weight to take account of the over-sampling of households in some Local Authority areas.

3.3 Household non-contact

Household characteristics
The logistic regression for the household characteristics variables showed that non-contact was significantly associated with:

- the Local Authority area,
- housing tenure
- the type of building occupied by the household and
- the number of rooms occupied by the household.

[8] A variable is deemed to have a significant independent effect if the significant level of the WALD statistic is less than 0.05.

[9] An alternative approach, which allows for the sequential nature of the response process, is to use multiplicative rates. For multiplicative rates, the denominator is the number of households which are still in the set sample at that stage. Hence the refusal rate would be based on the total number of contacted households and the partial response rate on the number of co-operating households. Non-contact and total non-response rates are the same using either approach.

Once the effect of these variables had been allowed for, the number of cars in the household was no longer significantly related to non-contact.

Household composition
The household composition model identified two significant characteristics:

- the number of employed adults in the household and
- whether the household contained a lone parent family, a married couple, a cohabiting couple, or other people.

When the effect of the factors above had been taken into account the number of people, adults, dependent children or families in the household were no longer significantly associated with non-contact.

Characteristics of the Household Reference Person
The logistic regression model developed to describe the HRP attributes that were related to non-contact showed that non-contact was associated with:

- the age and
- marital status of the HRP.

The economic activity of the HRP was not a significant factor once the model had allowed for the effect of the other variables.

Predictors of non-contact
For the final stage of the analysis of non-contacted households, the variables identified by the three separate analyses were combined in one logistic regression model. As mentioned earlier, a number of other key variables were also included in the logistic regression to allow for the possibility that different effects might be significant when the three categories of characteristics were combined. The variables added in this case were the number of cars, people, adults and dependent children in the household. Table 3.1 gives details of the results of the analysis and the categories used. An asterisk beside the variable name in the table indicates that the characteristic was significantly associated with non-contact after allowing for the effects of the other variables included in the model. Categories where the estimated parameter was significantly greater than for the reference category (defined as that with the lowest probability of non-contact and listed first) are marked (+).

The results showed that non-contact was more likely to occur in households:

- living in a purpose-built flat;
- containing one person only;
- containing no dependent children or one dependent child or
- whose HRP was single (never married) or aged under 65 years.

The model also showed that the Local Authority area also had a significant associated with household non-contact but for none of the individual authorities was the non-contact rate significantly higher than the measure for the reference category.

Table 3.1

3.4 Household refusal

Household characteristics
The logistic regression model confirmed the results in Chapter and showed that both Local Authority area and whether the household lived at the address a year before the Census were significantly associated with household level refusal.

Household composition
The results presented in Chapter 2 showed that the number of employed adults in the household was the only household composition variable that was significantly associated with refusal. Consequently we did not attempt to develop a model for the relationship between household composition attributes and household level refusal.

Characteristics of the Household Reference Person
The HRP attributes highlighted as being significantly associated with refusal were marital status and highest qualification. Once the effect of marital status qualifications had been taken into account, age, economic activity and the length of time the HRP had lived at the address were no longer significant influences on household level refusal.

Predictors of refusal
The results of the final stage of the analysis of household refusal combined all categories of variables. A number of additional variables were again included to cater for the possibility that different effects may be significant when the three categories of variables were examined together. The additional variables were the number of people, adults and dependent children in the household and the age of the HRP. The model indicated that the likelihood of refusal was less likely to take place in households:

- located in Aberdeen City, Argyll and Bute, Dundee City, East Dunbartonshire, East Lothian, Falkirk, Fife, North Lanarkshire, Shetland and Stirling or
- whose HRP was single, separated, divorced or widowed or who had no educational qualifications. **Table 3.2**

3.5 Total non-response

In the final stage of the logistic regression, the separate results for non-contact and refusal were brought together by developing logistic regression models for total non-response. Since refusals account for the major part of SHS non-response, the characteristics associated with total non-response would be expected to be similar to those associated with refusal. However, since some groups with high refusal rates had low non-contact rates, some of the effects identified when non-contact and refusal were examined separately may cancel out in the model for total non-response.

Household characteristics
The logistic regression analysis for these variables showed that total non-response was significantly associated with:

- Local Authority area,

- the type of accommodation occupied by the household and
- the number of rooms occupied by the household.

Once the model had allowed for the effect of the variables above, the characteristics that were no longer significantly associated with non-response included housing tenure, the number of cars in the household and the length of time the household had been at the address.

Household composition
Two household composition variables were identified as being significantly associated with total non-response. They were:
- the number of employed adults and
- whether the household contained a lone parent family, a married couple, a cohabiting couple, or other people.

The following characteristics ceased to be significant influences on total non-response when the effect of household type and number of employed adults had been allowed for: the number of people, adults, dependent children and families in the household.

Characteristics of the Household Reference Person
The logistic regression model developed to explore the relationship between total non-response and the characteristics of the HRP identified the following as variables that were significantly associated with total non-response:
- marital status
- highest qualification.

The other HRP characteristics that were no longer significantly related to total non-response once the effect of marital status and qualification had been allowed for were age and economic activity.

Predictors of total non-response
Again a number of additional variables were included to allow for the possibility that different effects may be significant when the three categories of variables were examined together. The additional variables were the number of people, adults and dependent children in the household and the age of the HRP. The results of the logistic regression model combining the three categories of variables show that interviewers were more likely to encounter non-response in households:
- living in Clackmannanshire, Dumfries and Galloway, Dundee City, East Lothian, Falkirk, Fife, North Lanarkshire, South Ayrshire or West Dunbartonshire;
- living in a purpose-built flat;
- which did not contain a married or cohabiting couple or
- where the HRP had no educational qualifications.

In addition, the number of adults in employment also had a significant associated with total non-response but for none of the individual categories was the total non-response rate significantly higher than the measure for the reference category. **Table 3.3**

Table 3.4 summarises the results for the final models for the various elements of SHS non-response.

4 Association between key SHS measures and correlates of non-response

4.1 Introduction

The previous chapters have identified a number of household characteristics that were associated with various aspects of non-response on the SHS. This chapter examines whether the household characteristics that were significantly associated with non-response were also related to a number of selected key estimates based on the SHS. If the census characteristics associated with non-response are also associated with the main survey estimates, then it is possible that non-response may affect the survey estimates. If this is the case, the effects of differential non-response may be accounted for, in part at least, by applying weights to correct the bias in these household characteristics.

The analysis presented in Chapters 2 and 3 examined in detail the characteristics associated with each of the separate components of non-response and with non-response as a whole. When investigating methods of weighting survey data, the main interest is clearly in the extent of bias in the achieved sample and hence in the factors associated with total non-response rather than the separate components of non-response. The analyses described in this chapter therefore focused on the effects of the factors that were significantly associated with total non-response, namely: Local Authority area, the type of building occupied by the household, the number of adults in employment, household type and the highest qualification of the Household Reference Person (HRP). The key SHS measures were selected from a list supplied by the Scottish Executive. They included: whether any household member needed regular help or care, had a full driving licence or access to the internet for non-business use.

As with most of the results presented in this report, the results relating to the association between key SHS estimates and the factors linked to non-response were produced using weighted data.

4.2 The relationship between key SHS estimates and selected census household characteristics

The results have demonstrated that some estimates from the SHS are associated with many of the household characteristics linked with non-response. The pattern of the relationships examined were also very similar. For example:

- Compared with households who lived in a house, households who lived in a purpose built flat were less likely to have access to the internet for non-business use or to have a household member who held a full driving licence.
- Compared with other households, households containing two or more employed adults were more likely to have access to the internet or to have a household member with a full driving licence.
- Households containing a lone parent family were less likely to have a household member with a full driving license or to have access to the internet.

- Households whose HRP did not have any education qualifications were more likely to have a member who needed regular help or care and less likely to have access to the internet or to have a member who had a full driving licence.

The results presented in this chapter have reveal associations between the key factors related to non-response and a number of key SHS key variables. This indicates that it is likely that non-response bias, which has so far been expressed in terms of demographic and socio-economic characteristics of the household, may also affect key SHS estimates. It may be desirable to re-weight the data to adjust for these effects. However, more than one strategy can be used to re-weight the data and further work is required to evaluate the various types of non-response adjustments that can be applied to the SHS.

5. Bibliography

Foster, K. (1998) Evaluating non-response on household surveys. *GSS Methodology Series* No. 8, Government Statistical Service, London.

Groves RM, and Couper MP (1998) *Nonresponse in Household Interview Surveys,* John Wiley & Sons Inc, New York.

Hope, S. (2002) *Scotland's People: Results from the 2001 Scottish Household Survey, Volume 6: Technical Report,* Scottish Executive, Edinburgh.

Scottish Executive (2002) *Scotland's People: Results from the 2001 Scottish Household Survey, Volume 5: Annual Report,* Scottish Executive, Edinburgh.