



The Family Resources Survey

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

Stephanie Freeth and Philippa Sowman
revised February 2005

Summary

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

Census data were successfully linked to 94.2% of the FRS 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 91.2% of responding and 87.3% of non-responding households.

The non-response rate for households matched with census records was 30.4%, 4.4% non-contacts and 26.0% refusals. In addition, 1.0% of households gave incomplete information and were classified as partially rather than fully responding (68.6%).

Non-contact was higher than average amongst FRS households:

- located in London,
- living in converted or shared houses and flats, or purpose-built flats or maisonettes,
- privately renting their accommodation or renting from a Registered Social Landlord or Housing Association,
- occupying accommodation with fewer than four rooms or
- without a car or van.

In terms of household composition, non-contact was higher than average for households containing:

- one adult,
- one adult employed,
- no dependent children,
- no families (that is households comprising either one person only or a number of adults not in a family group) or
- one or more 'single' persons aged between 35 and 49 years, or one or more 'single' persons aged between 50 and 64 years or aged under 35 years.

Non-contact was also above average where the Household Reference Person was:

- aged between 16-34 years,
- single (never married), or separated (but still legally married) or divorced,
- unemployed, or a student, permanently sick or disabled.

Refusal was more likely to occur in households:

- located in London,
- living in a detached or semi-detached house,
- who owned their home outright or
- who had not moved house in the year before the Census.

Refusal rate was above average for households containing:

- no dependent children,
- three or more adults in employment,
- couples with non-dependent children (and where the younger member of the couple was aged less than 50 years) or
- a 'single' person aged 80 years or more.

Refusal also tended to be higher in households where the Household Reference Person:

- was aged 65 years and over,
- was widowed,
- had qualifications other than a degree, 'A' levels, or 'O' levels and GCSE's,
- was a small employer or own account worker or
- who had not moved house in the twelve months before the Census.

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

Non-contact was significantly associated with:

- area type,
- the type of building occupied by the household,
- the number of people in the household,
- the age of the Household Reference Person or
- the country of birth of the Household Reference Person.

Refusal was significantly related to:

- Government Office Region,
- housing tenure,
- the age of the youngest dependent child,
- the highest qualification of the Household Reference Person or
- the ethnic group of the Household Reference Person.

Total non-response, which included both non-contacts and refusals, was significantly associated with:

- area type,
- housing tenure,
- the age of the youngest dependent child,
- the marital status of the Household Reference Person,
- the highest qualification of the Household Reference Person or
- the economic activity of the Household Reference Person.

1 Background and Methodology

1.1 Background

The Family Resources Survey (FRS), 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 re-weight 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 FRS, 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. Survey sponsors could also request further analysis and the Department for Work and Pensions (DWP) who sponsor the FRS asked ONS to carry out additional analysis on the relationship between non-response and household composition. The results from the study on the FRS are presented in this report.

1.2. The FRS content

The FRS was established to provide facts and figures about living standards in Great Britain and to examine people's relationship, and interaction, with the social security system. The survey was launched in October 1992 to meet the specific information requirements of the DWP. Prior to the conception of the FRS, DWP relied upon other government social surveys, notably the Family Expenditure Survey and the General Household Survey for their information. However, as these surveys had relatively small sample sizes and did not provide sufficient information on many groups of particular interest, it became apparent that a dedicated survey was required to meet DWP's specific needs.

The FRS is a continuous survey based on an annual sample of the residents in private households in Great Britain. The questionnaire, which forms the basis of the survey, is designed to allow core data to be collected on an annual basis and thus permit trends to be monitored and analysed by users. All adults eligible for inclusion in the survey are asked a wide range of questions about their circumstances including: income and benefits; tenure and housing costs; assets and savings; occupation and employment; health and ability to work; pensions and insurance; and childcare and carers.

Information for the FRS is collected by means of an interview carried out with the household reference person and other adult household members. The FRS allows proxy interviews and defines a fully co-operating household as one in which a full interview has been obtained either in person or by proxy with every eligible member of the household. A partially co-operating household is one where information has been obtained for the Household Reference Person's benefit unit but the interviewer has been unable to obtain information from other household benefit units.

1.3 Design of the CNR on the FRS

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 survey. This was carried out in previous years 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 FRS, this involved matching addresses sampled for interview between April and June 2001. In total 10,694 addresses were included for the matching procedure detailed in Section 1.5.

Fieldwork on the FRS is carried out jointly by ONS and the National Centre for Social Research; addresses sampled for interview by both organisations were included in the study.

1.4 Response rate on the FRS

Table 1.1 shows the response rate of the April to June 2001 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.³ The overall response rate for the April to June 2001 sample was 68%; (including 1% classed as partially co-operating). Households who refused to co-operate made up the majority of non-respondents (26% of the total sample). This total includes both refusals to the advance letter and total household refusal to the interviewer. Only 5% of the households in the study were not contacted.

³ The response rate and other percentages shown were calculated using the adjusted eligible households in the denominator and the relevant total number of cases in the second column of Table 1.1 as the numerator. The Annual Technical Report of the 2001/2 FRS (Gatward and Rowland, 2003) contains details on the calculation of the adjusted eligible households.

Table 1.1 FRS sample by response category (April to June 2001)

	Number of cases	Percentage of total
Total responding households	6,546	67.9
Fully co-operating	6,450	66.9
Partially co-operating	96	1.0
Total non-responding households	2,987	31.0
Non-contacts	497	5.2
Refusals	2,490	25.8
Unknown eligibility deemed to be eligible by the process used to calculate the adjusted eligible households	110	1.1
Total adjusted eligible households	9,643	100
Ineligible households or household with unknown eligibility but deemed to be ineligible in the adjustment of eligible households	1,051	
Base (all households)	10,694	

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 10,694 households were included in the FRS sample for matching. Overall, just over 94% of the household in the study (10,072) were successfully matched and census data obtained (Table 1.2). Of the 10,072 records matched, 767 contained

persons or households imputed for the One Number Census¹. In addition, 363 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% of the FRS 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 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 less than 1% of all cases matched so there are only a small number of cases where the match was potentially incorrect.

Table 1.2 Results of the matching exercise for the FRS sample

	Number of cases	Percentage of total
Census record available	10,072	94.2
(of which) Record not altered by the One Number Census	9,305	87.0
Record contained some imputed Persons	316	3.0
Record was wholly imputed	451	4.2
No usual residents	363	3.4
No census record	622	5.8
(of which) Address found but no census form	325	3.0
Address not found	297	2.8
Matched households with usual residents	9,709	90.8
Matched households with usual residents but excluding imputed households	8,942	83.6
Base (all households available for matching)	10,694	100

Table 1.3 shows the effective match rate achieved on the FRS for different categories of household response. The match rate was higher for responding than non-responding households: census data were obtained for 96.6% of responding households compared with 95.3% of non-responding households. This difference was largely due to the lower match rates recorded for households that were not contacted on the survey (91.1%) rather than for refusals (96.1%). Thus it appears that those households that were difficult to contact were also less likely to have been enumerated on the Census or to have filled out a census form.

¹ The One Number Census project integrates the 2001 Census counts with the estimated level of underenumerations 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 FRS response category^a

	Number of households matched	Percentage of total in response category^b
Responding (of which)	6,322	96.6
Fully co-operating	6,230	96.6
Partially co-operating	92	95.8
Non-responding (of which)	2,847	95.3
Non-contacts	453	91.1
Refusals	2,394	96.1
Total eligible households enumerated on the Census as having at least one adult	9,169	
Ineligible and unknown eligibility	734	69.8
Eligible but enumerated on the Census as having no adult	3	
Total matched with usual residents	9,906	92.6
All households	10,694	

a. Excludes households with no usual residents.

b. 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 form 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 census data of FRS sampled addresses to examine non-response on the FRS, households classified as having no usual residents on the Census and those imputed for the One Number Census were not relevant units of analysis. These households were excluded from the analysis. Of the original 10,072 households matched, 363 were excluded because they contained no usual residents and 767 imputed households were excluded leaving 8,580 households. Once usual residency and imputation had been taken into account, another 359 households classified as ineligible on the FRS and three eligible households enumerated on the Census as containing children only were excluded from the analysis leaving 8,580 records.² Table 1.4 shows the number

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.

of households included in the analysis by response category. As a result of excluding the cases that were not relevant to the analysis, the final response and non-response rates for the cases used in the analysis (shown in the final column of Table 1.4) were slightly different from those given in Table 1.1.⁴

Table 1.4 Households included in the analysis by FRS response category

	Number of households included in the analysis	Percentage of total in response category (a)	Percentage of all eligible households included in the analysis (response rate used in the analysis)
Responding (of which)	5,973	91.2	
Fully co-operating	5,887	91.3	68.6
Partially co-operating	86	89.6	1.0
Non-responding (of which)	2,607	87.3	
Non-contacts	378	76.1	4.4
Refusals	2,229	89.5	26.0
Total eligible households used in the analysis	8,580	89.0	100.0
Ineligible	359		
Eligible but enumerated on the Census as having no adults.	3		
Total household with usual residents and not altered by the One Number Census	8,942		

a. The percentages for this column were calculated using the 'Number of Cases' column in Table 1.1.

The main unit of analysis 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. However, since some data for individuals was obtained, it was also possible to investigate the characteristics of adults in responding and non-responding households.

⁴ There are two reasons why the response and non-response rates for the cases used in the analysis (Table 1.4) were slightly different from those given in Table 1.1. The primary reason was because FRS households that were not matched to census records or excluded from the analysis (because the matched record was imputed or contain no usual residents) were excluded from the response rate calculation in Table 1.4. In addition, the number of eligible households in Table 1.4 were not adjusted. Adjustment was not possible because the cases that were not matched to census records were not on the census-linked data set.

Analysis methods

The analysis presented in this report is essentially descriptive, aiming to identify those census characteristics which are significantly associated with household response. Census variables were initially considered singly and Chapter 2 presents tabulations and commentary on how the different elements of non-response vary with different household characteristics. The main aims of the chapter are to identify the subgroups that were most and least likely to be non-responders to the FRS and to identify the bias in the unweighted FRS sample caused by non-response.

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 were therefore explored further by the use of multivariate analysis in Chapter 3. The aim of this 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.

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 will result in the loss of 6.3% of households from the analysis and a loss of 18.1% of non-contacted households in London. Weights can be developed to adjust for the loss of these households but this is beyond the resources currently available to this stage of the project. Excluding the imputed cases without adjustment will reduce the contribution of households in London to the overall results presented in Chapter 2 but it will not affect the ability of this study to identify the household characteristics that are significantly related to non-response. This is because region was included as a control variable in the multivariate analysis presented in Chapter 3.

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 three sections are concerned with household characteristics covering, regional distribution of non-response, household composition, basic descriptors of the accommodation and household, and characteristics of the Household Reference Person. Section 2.6 examines the characteristics of all adults within non-responding households. Section 2.7 discusses the factors related to response on the FRS, and refers to the 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 FRS sample due to household non-response. Finally, section 2.9 shows the results requested by DWP.

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.

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 5.0%. By adopting the conventional 5.0% 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 census characteristic, the next stage was to identify which categories had significantly high or low response rates. This was done by using a test of differences between proportions but only those results which were significant at the 1.0% 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, and also to allow for the effect of the survey's complex sample design.

2.2 Regional distribution of non-responding households

The variation in response by region can be obtained from the analysis of FRS data but the results for the three-month period 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 Government Office Region and area type is based on all the 9,643 eligible households available for matching, including cases with a valid FRS response code but no or only imputed census data. The cases containing no or imputed census data can be included in the regional and area analysis because information on the region and area in which the address was located was available from the survey's sample records. The response details of the households included in the region table are set out in Table 1.1.⁵

The results showed that non-contact and refusal rates varied by region and area type. For example, non-contact was more likely in London (7.0%) than in the South East (3.9%) or South West (3.9%). The refusal rate was also higher in the East Midlands (29.4%) and London (29.1%) than in Wales (21.2%).

In relation to area types, London again had the highest refusal and non-contact rates (29.2% and 7.0% respectively).

Table 2.1

2.3 Household composition of non-responding households

Number of people

FRS non-contact and refusal rates varied with the number of usual residents recorded at the Census. Non-contact was highest for one-person households (8.2%) and declined gradually as the number of household members increased to reach 1.8% for those households consisting of four or more people. 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.

Non-contact was more likely in those households with only one adult (7.5%) compared with households with two or more adults (2.8% or less).⁵ (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 also contain children.)

Households with only one adult in employment had a higher non-contact rate (6.1%) compared with other households (4.1% or less). Although, households with three or more adults in employment were easier to contact compared with households with only one adult in employment, they were more inclined to refuse to take part in the survey (30.0% compared with 23.9% for households with only one working adult).

⁵ To be consistent with the non-response rates presented in Table 1.1, response rate and other percentages on the region and area tables (Table 2.1 and 2.9) were calculated using the adjusted eligible households in the denominator. Section 1.7 explained why it was not possible to use the adjusted eligible households in tables containing census data but the reasons did not apply to Tables 2.1 and 2.9. This was because Tables 2.1 and 2.9 do not contain census data and they were produced using a data set that contained all cases. Unfortunately, that data set could not be used to produce the other tables in this report because it did not contain census data.

⁵ An adult is someone who is not a dependent child. A dependent child is a person aged 0 to 15 years or aged 16 to 18 in full-time education and living with his or her parent(s).

Households with dependent children had lower non-contact rates compared with households with no dependent children.⁶ The non-contact rate for households with two or more dependent children was between 2.2% and this rate rose to 5.1% amongst household with no dependent children. The refusal rate was also higher for those households with no children (27.4%) compared with households with dependent children (between 22.0% and 23.1%). Among households with children, refusal rates were lower for those with pre-school children 19.5% compared with households where the youngest dependent child was aged 10 years and over (25.9%).

Table 2.2

Number of family units

A family is defined on the Census as a married or co-habiting 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 households comprising only non-family members are classed as ‘no family’ households.

Households that did not contain a family had higher non-contact rates (7.8%) than other types of households (2.7%). Refusal rates did not appear to vary with the number of families within the household.

Household type

Table 2.3 gives response rates for three 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 The third classification (Household Type C) is based on a specification provided by the DWP and describes household composition as well as the age of the household members.

‘Household Type A’ highlights the high non-contact rates for one-person households (8.2%) as mentioned earlier in this chapter. Couples and lone parents with one or more dependent children were less likely to refuse (22.7% and 21.1% respectively) compared with couples and lone parents with non-dependent children (31.3% and 31.6% respectively).

‘Household Type B’ shows a similar result. Non-contact was higher for one adult households with no children (8.2% compared with 3.7% or less for other households). With the exception of households containing three or more adults with children, households with no children were more likely to refused the FRS than households with children.

The analysis of ‘Household Type C’ showed that households containing one or more ‘single’ persons aged between 35 and 49 years had the highest non-contact rate (12.5%). (In the context of this analysis, ‘single’ people include those who are never

⁶ A dependent child is a person aged 0 to 15 years or aged 16 to 18 in full-time education and living with his or her parent(s).

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.

married, separated, widowed or divorced and who are not a lone parent or living with somebody as a couple.) The non-contact rate for households containing one or more 'single' persons aged between 50 and 64 years or aged under 35 years was also high (7.8% and 6.0% respectively). Households with the highest refusal rate were those containing a 'single' person aged 80 years or more (36.3%) and households containing a couple with no dependent children and where the younger member of the couple was aged less than 50 years (27.7%). Households containing a lone parent family were the least likely to refuse (20.9%).

Table 2.3

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

Non-response rates were associated with the tenure and accommodation characteristics of households selected to take part in the FRS. Non-contact rates were higher among households living in converted or shared houses and flats (10.3%) or in a purpose-built flat or maisonette (9.5%), compared with those who lived in a house (3.3% or less). Non-contact was also high for households that privately rented their accommodation (9.0%) or those who rented from a Registered Social Landlord or Housing Association (6.7%). Households that refused to take part in the survey were most likely to live in a detached or semi-detached house (27.3% and 26.9% respectively) or to own their homes outright (30.4%).

Households occupying a small number of rooms (between one and three) were more difficult to contact compared with households occupying six or more rooms. Interviewers were unable to contact 9.1% of households living in accommodation containing fewer than four rooms compared with 3.2% for households in accommodation with seven or more rooms.

Non-contact was higher for households which did not own a car (5.9 %) and decreased gradually as the number of cars in the household increased, reaching 3.3% for those with three or more cars. Households who had not moved in the twelve months before the Census were more likely to refuse (26.5%) compared with those that had moved (21.9%).

There was no clear association between either non-contact or refusals and other characteristics of the accommodation such as availability of a bathroom or central heating.

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, country of birth 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 high among households whose HRP was aged between 16 and 34 years old (6.3%) but dropped steadily after this age group to reach 3.0% for HRP aged 65 years and over. Refusals demonstrated the opposite trend, with rates increasing with the age of the HRP, from 22.6% for 16 to 34 year olds to 29.7% for HRP aged 65 and over.

The non-contact rate was highest for households with a single, never married HRP (8.0%), followed closely by households with a separated (but still legally married) or divorced HRP. Non-contact was lowest for households with a married (first or re-marriage) or widowed HRP (2.4% and 3.5% respectively). Households whose HRP was widowed were the most likely to refuse to take part in the survey (31.3% compared with 25.9% or less for other HRPs).

The results show no significant associations between the sex of the HRP and the non-contact or refusal rates of the survey.

Other characteristics of the Household Reference Person

Refusals were highest in households whose HRP had qualifications other than a degree, 'A' levels, or 'O' levels and GCSE's (30.6%). Households most likely to participate in the FRS were households whose HRP had a first or higher degree (19.4%).

Households whose HRP was unemployed (9.9%) or who was a student or was permanently sick or disabled (6.1%) were more difficult to contact compared with those whose HRP was retired (3.0%).

Compared with other households, households where the HRP was a small employer or own account worker, were more likely to refuse the survey (29.1%) compared with households with an HRP was in a managerial or professional occupation (22.7%).

There were no significant associations between the country of birth of HRP and non-contact or refusal but non-contact was higher among households whose HRP was non-White (7.5%) compared with households whose HRP was White (4.2%).

Households whose HRP had not moved in the twelve before the Census were more likely to refuse (26.5%) compared with those whose HRP had not moved (21.4%).

Tables 2.5 and 2.6

2.6 Characteristics of all adults in non-responding households

This section investigates the characteristics of individual adults according to whether their household responded to the FRS or not. This type of analysis is useful in showing the characteristics of the adults who are included in the responding sample and, for those adults not included, whether they have been disproportionately lost to the sample as a result of household non-contact or refusal. Many of the relationships are similar to those for the corresponding characteristics of the Household Reference Person. This is to be expected since the rationale for describing households by the

characteristics of one reference person is that this person is in some way representative of all adults within the household.

The various non-response rates presented for all adults may differ slightly from those for households, because of the effect of differential non-response among households of different size. For example, 26.2% of all adults in the set sample were in refusing households whereas 26.0% of households were classed as refusing. This reflects a higher refusal rate seen for households containing more adults (Table 2.2). The proportion of adults in non-contacted households were, in contrast, slightly smaller than the proportion of households classed as non-contact because of the higher rate of non-contact in single person households.

Individuals who were most likely to be lost to the FRS sample due to non-contact were:

- aged between 16 and 44 years,
- divorced or single (never married),
- born outside the United Kingdom,
- from a Non-white ethnic group,
- not living at the sampled address a year before census night or
- unemployed.

In contrast, people who were most likely to be missed from the sample because of refusal were:

- widowed, single (never married), or married (first marriage),
- qualified with qualifications other than 'O', 'A' levels or first or higher degree, or were not academically qualified,
- born in the United Kingdom,
- of White ethnic origin,
- living at the sampled address a year before census night or
- small employers and own account workers.

Tables 2.7 and 2.8

2.7 Summary and discussion of the factors related to response on the FRS

The previous five sections have shown that non-response on the FRS was related to:

- Government Office Region and area type,
- the size of the household,
- the number of adults in the household and whether they were in employment or not,
- whether the household contains children or not, and if so how many children the household contained,
- household composition,
- housing tenure,
- the type of accommodation occupied by the household,
- the number of rooms (occupied by the household)
- car ownership by the household or
- whether the household had moved in the twelve months before the Census or not.

Non-response was also related to the following characteristics of the HRP:

- age,
- marital status,
- highest qualification,
- economic activity,
- socio-economic classification (NS-SEC) or
- whether the HRP had moved in the twelve months before the Census or not.

These results are similar to 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-contact and refusals, Groves and Couper have also outlined how the factors influence response. The findings of Groves and Couper are summarised here to provide insight into the source 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. These factors may help to explain the high non-contact rates achieved in London. It may also have influenced the high non-contact rates for households that did not own a car. The lack of private transport may require members of these households to spend less time at home and even more time on travel as they have to rely on less convenient public transport for work, domestic and social activities.

Non-contact rates were highest in FRS sampled households containing one person. 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 FRS sample who lived in part of a converted or shared house, or in a purpose-built block of flats 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. The FRS results also highlighted this, with households with pre-school children having the lowest non-contact rates.

There is also some suggestion in the literature that minority groups, perhaps because of the tendency to hold multiple jobs, spend more time out of the home. This may explain why Non-White groups had a comparatively high non-contact rate.

Refusals

Compared with Non-metropolitan Districts, London experienced higher refusal rates on the FRS. Groves and Couper have also 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 larger 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 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.

The literature also highlights the relatively high rates of refusal 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 in other age groups because of their lower employment rates and, at advanced ages, reduce mobility, hence their relatively low non-contact rate. However, health problems prevent some from taking part in surveys.

2.8 Bias in the FRS 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 which results from non-response on the FRS. Tables 2.9 to 2.16 compare the frequency distributions for census characteristics of households that have fully responded to the FRS and for all households in the FRS-census matched sample. The results of the chi-square tests presented in sections 2.2 and 2.6 are shown again on these tables to indicate the census characteristics that were related to non-response on the FRS.

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 FRS-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, most 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 over) and tend to have significantly low response rates.

Sub-groups in the FRS sample with correction factors of 1.10 and above include households:

- who had three or more adults in employment (1.11)
- containing a couple with non-dependent children only (1.14), or a lone parent with non-dependent children only (1.11).
- with one or more 'single' people aged 35 to 49 years (1.16).

Tables 2.9 – 2.16

2.9 Household composition, economic activity and response

One of the key estimates of the FRS is the proportion of dependent children in workless working age households. (Workless working age households are households containing at least one person of working age and where no one in the household is in employment, as defined by the International Labour Organisation). Providing an accurate estimate of the proportion of children in workless working age households requires an accurate estimate of the households containing dependent children and the economic activity of the adults in these households. DWP are interested in assessing the quality of these FRS estimates and asked ONS to carry out additional investigations into this topic.

About one in five (17.6 %) of the working age households in the matched sample contained no adult in employment. Households containing a couple with a dependent child or children or a couple the younger of whom was aged less than 50 years were the least likely to contain no working adults. In contrast, households with at least one 'single' person aged between 50 and 64 years or containing a lone parent family were the most likely to be workless (44.0% and 46.4% respectively).

The proportion of workless working age households in the fully co-operating FRS sample was 17.9%. Again, households containing a couple with a dependent child or children or a couple the younger of whom was aged less than 50 years were the least likely to contain no working adults. Households with a lone parent or households with at least one 'single' person aged between 50 and 64 years were the groups in the fully co-operating sample that were the most likely to contain no working adults (44.5% and 45.9% respectively).

Worklessness amongst households containing a lone parent family was higher in the FRS partially co-operating, non-contact and refusing sample (51.8%) than in the FRS

fully-co-operating sample (44.5%). However, the difference between the estimates was not statistically significant. (The difference between the estimates for parents with a dependent child or children was also not statistically significant.) The results indicate that although the estimate of worklessness amongst lone parent families was lower in the FRS fully co-operating sample than in the partial and non-responding sample, the difference could have occurred by chance. In addition, non-response bias for the estimate was likely to be low. Correction factor similar to the ones presented in the last section was calculated to measure the adjustment that would need to be applied to the fully-responding sample in order to achieve the same distribution as for the total set sample. The correction factor for workless households containing a lone parent family was 1.04.

Tables 2.17 and 2.18

The investigations presented in this section was the first attempt by the CNR to explore the relationship between household composition, economic activity and response. The results presented here are based on census data on unemployment which are subject to measurement error. They should therefore be treated as results from a “work in progress” and be interpreted in conjunction with the findings from other studies on this topic. The joint DWP/ONS working group set up to compare the estimates of the proportion of children in workless households derived from the FRS and from the Labour Force Survey has already carried out a number of investigations into this area. The CNR project team is aware of the work carried out by the joint working group and has met DWP to discuss the findings, compare them with results from other sources and to consider future lines of enquiries. DWP and ONS are planning to use the CNR and other data to carry out further investigations into these issues.

2.10 Summary

This chapter 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 result from one basic relationship or factor manifesting itself 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 FRS. 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. The analysis does, however, have obvious limitations since each characteristic is considered individually even though some are clearly related. As a result, some of the associations may be one 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 analysis 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 earlier 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 the 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 rates. The results of these models are reported briefly.

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

⁸ The logistic regression analyses were 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 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 were mostly 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 (Tables 3.1 to 3.3). The tables indicate the variables that had significant independent effects on non-response and, for each variable, the categories of households that had estimated parameters which were significantly higher than for the reference category. 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 of 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.

3.3 Household non-contact

Household characteristics

The logistic regression identified that non-contact was significantly associated with:

- area type,
- housing tenure or
- the type of building occupied by the household.

The following characteristics no longer had a significant relationship with non-contact once area, tenure and accommodation type had been taken into account: Government Office Region, the number of rooms occupied by the household, the length of residence of the household at the address and the number of households cars or vans.

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.

Household composition

The model describing the relationship between household composition and non-contact identified three characteristics as significant influences on non-contact. These were:

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

When the effect of the variables above had been taken into account the following characteristics were no longer significant factors – the number of adults, dependent children and families in the household and the age of the youngest dependent child.

Characteristics of the Household Reference Person

Logistic regression identified the following characteristics of the Household Reference Person (HRP) as having a significant effect on non-contact:

- marital status or
- ethnic group.

Once the model had taken account of the effect of the variables above, the age, economic activity and socio-economic classification of the HRP were no longer significant influences on non-contact.

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 only variables added in this case were the age, country of birth and economic activity of the HRP. 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 the non-contact rate 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 analysis showed that non-contact was more likely to occur in households:

- located in the Metropolitan Districts.
- living in a purpose-built flat or in a converted or shared house.
- containing between one and three people.
- whose HRP was aged under 55 years or who was born outside the United Kingdom.

Table 3.1

3.4 Household refusal

Household characteristics

The household characteristics which were significantly associated with refusal included:

- Government Office Region,
- housing tenure or
- length of residence at the address.

Once the effect of these factors had been taken into account, area type and the type of building occupied by the household ceased to have an influence on refusal.

Household composition

The logistic regression model identified two characteristics as significantly associated with refusal. These were:

- the age of the youngest dependent child or
- the number of adults in employment.

Once the effect of the variables above had been allowed for, the number of people, adults and dependent children in the household and whether the household contained a lone parent family, a married couple, a cohabiting couple or other people were no longer significantly associated with refusal.

Characteristics of the Household Reference Person

Logistic regression identified the following characteristics of the HRP as having an independent effect on refusal:

- highest qualifications or
- length of residence at the address.

In contrast, the age, marital status, ethnic group, economic activity and socio-economic classification of the HRP were no longer significant variables once the model had allowed for the effect of qualifications and length of residence.

Predictors of refusal

The results of the final stage of the analysis of household refusal combining all categories of variables, are shown in Table 3.2. The additional variables included in the analysis were area type and age, ethnic origin and economic activity of the HRP or whether the household contained a lone parent family, a married couple, a cohabiting couple or other people. The model indicated that refusal was more likely to take place in households:

- located in the North West, Midlands, East of England, South East and London.
- Who owned their accommodation outright or who had a mortgage.
- Who had no dependent children or whose dependent children were aged five years and over.
- Whose HRP did not have degree level qualifications or who was White.

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 FRS 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 model that described the relationship between household characteristics and total non-response showed that total non-response was significantly associated with:

- Government Office Region,
- area type,
- housing tenure or
- number of rooms.

The type of building and the length of residence at the address were no longer linked to total non-response once the model had taken account of the effect of the variables above.

Household composition

The following variables were identified by the logistic regression model as being significantly associated with total non-response:

- number of people in the household,
- number of dependent children or
- age of the youngest dependent child.

Once the effect of these variables had been allowed for, the number of adults, employed adults and families in the household and whether the household contained a lone parent family, married couple, cohabiting couple, or other people were no longer significantly associated with total non-response.

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 or
- highest qualifications.

The ethnic origin and length of residence of the HRP at the address were no longer significantly related to total non-response once the model had allowed for the effect of marital status and qualifications.

Predictors of total non-response

The results of the logistic regression combining the three categories of variables are shown in Table 3.3. The additional variables included in the analysis were the number of adults in the households and economic activity of the HRP.

The logistic regression model indicated that total non-response was more likely to occur in households:

- located in London.
- Who owned their accommodation outright or who had a mortgage.
- Who had no dependent children or whose dependent children were aged five years and over.
- Whose HRP was single or did not have degree level qualifications or who was self-employed.

Table 3.3

Table 3.4 summarises the results for the final models for the various elements of FRS non-response. As expected, there was close correspondence between the characteristics associated with refusal and total non-response and three of the six variables with significant effect on total non-response were also included in the model for refusals.

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

4.1 Introduction

The previous chapters have identified a number of household characteristics that were strongly associated with non-response on the FRS. This chapter examines whether the household characteristics that were significantly associated with non-response were also related to key FRS estimates of income. 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: area type, housing tenure, age of the youngest dependent child, and three characteristics of the Household Reference Person (marital status, highest qualifications and economic activity).

4.2 The relationship between FRS income estimates and selected census household characteristics

Tables 4.1 and 4.4 illustrate, for fully co-operating FRS households only, the associations between selected measures of household income as recorded on the FRS and the household characteristics most strongly related to total non-response.

The household income estimates selected for analysis are strongly associated with the census characteristics shown to be related to total non-response. In addition, the pattern of the relationships was also broadly similar. For example, London and non-Metropolitan Districts had a significantly higher share of households in the highest income category and the various income estimates tended to increase with the qualification of the HRP. Households where the HRP had a higher income were most likely to have a mortgage while lower income households tended to rent from a Local Authority, or a Registered Social Landlord or Housing Association. Income was highest for those households whose HRP was employed or self-employed.

The relationship between income and the marital status of the HRP was more complex but clear relationships can be observed. There was a tendency for unequivalised total income before housing costs and equivalised total income after housing costs to be highest for households with a married HRP. In contrast, unequivalised net disposal income and equivalised total income before housing costs tended to be highest for households whose HRP was single or married.

The results have demonstrated that the key income estimates of the FRS are associated with the household characteristics that are strongly associated with total non-response. Thus 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 survey estimates of income. However, the results presented in this chapter should be interpreted and applied in association with the findings from Chapter 2 which indicate that non-response bias was relatively small on the FRS. It would be expected that more complex categories, derived from the cross-classification of two or more variables, might reveal larger bias. However, the results for the complex categories may be unstable because the cross classifications of variables required to create them may result in categories based on small cell sizes. Further work is therefore required to assess if the bias detected merits adjustment and the type of adjustment required.

References

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

Gatward, R. and Rowland, M. (2003) *Family Resources Survey Annual Technical Report: 2001/2002*. London, ONS.

Groves, R.M. and Couper, M. P. (1998) *Nonresponse in household interview surveys*. New York: Wiley.