

# **2003-4 British Crime Survey (England and Wales)**

## **Technical Report Volume I**

Keith Bolling, Catherine Grant, Patten Smith and Matthew Brown.

Prepared for the Measuring and Analysing Crime Programme  
Research, Development and Statistics Directorate  
Home Office

December 2004

## TABLE OF CONTENTS

|  |    |
|--|----|
| 1. Background.....   | 1  |
| 1.1 Introduction to the British Crime Survey .....                                 | 1  |
| 1.2 Structure of the technical report.....   | 2  |
| 2. Sample design.....  | 4  |
| 2.1 Introduction.....  | 4  |
| 2.2 Sample size and structure.....   | 4  |
| 2.3 Sample frame.....  | 5  |
| 2.4 Stratification .....   | 5  |
| 2.5 Clustering.....  | 5  |
| 2.6 Rotation of sectors.....   | 6  |
| 2.7 Procedures for selecting the sample.....                                       | 6  |
| 2.8 Non-white boost sample.....  | 8  |
| 2.9 Evaluation of focused enumeration.....   | 10 |
| 2.10 Youth Boost sample.....   | 14 |
| 3. Questionnaire Content and Development.....                                      | 15 |
| 3.1 Structure and coverage of the questionnaire.....                               | 15 |
| 3.1.1 Household Grid and Main questionnaire.....                                   | 19 |
| 3.1.2 Screener questionnaire.....  | 19 |
| 3.1.3 Victim Forms.....  | 20 |
| 3.1.4 Module on mobile phone theft.....  | 23 |
| 3.1.5 Module on performance of the Criminal Justice System and witnessing<br>crime | 24 |
| 3.1.6 Follow-up modules.....   | 24 |
| 3.1.7 Technology crimes.....   | 25 |
| 3.1.8 Anti-social behaviour module.....  | 25 |
| 3.1.9 Demographics module.....   | 26 |
| 3.1.10 Self – completion modules .....   | 27 |
| 3.2 Life event calendar.....   | 27 |
| 3.3 Questionnaire development.....   | 28 |
| 3.3.1 Main stage piloting .....  | 28 |
| 3.4 Final questionnaire and revisions .....  | 29 |
| 3.5 Allocation of sample within CAPI.....  | 30 |
| 3.6 Features of Quancept used in the BCS.....                                      | 30 |
| 3.6.1 Don't Know and Refusal Keys.....   | 30 |
| 3.6.2 Different question types.....  | 31 |
| 3.6.3 Logic and consistency checks .....   | 31 |
| 3.6.4 Date calculation and text substitution.....                                  | 31 |
| 4. Fieldwork.....  | 33 |
| 4.1 Briefing of interviewers .....   | 33 |
| 4.2 Supervision and quality control .....  | 34 |
| 4.3 Fieldwork dates and fieldwork management.....                                  | 34 |
| 4.4 Fieldwork procedures and documents.....  | 35 |
| 4.4.1 Advance letter and leaflet.....  | 35 |
| 4.4.2 Address Contact Sheets (ACS) .....   | 36 |
| 4.5 Presence of others during the interview.....                                   | 37 |
| 4.5.1 Presence of others during the screener interview .....                       | 37 |
| 4.5.2 Presence of others during the self-completion.....                           | 38 |
| 4.6 Length of interview.....   | 40 |

|       |  |    |
|-------|--|----|
| 4.6.1 | Introduction .....   | 40 |
| 4.6.2 | Overall length of interview.....   | 41 |
| 4.6.3 | Average time for different sections of the interview.....                | 42 |
| 4.6.4 | Length of victim forms.....  | 42 |
| 4.6.5 | Length of follow-up modules .....  | 43 |
| 4.7   | Response rate and reasons for non-response: core sample .....            | 44 |
| 4.7.1 | Overall core response rates .....  | 44 |
| 4.7.2 | Core response rates by Government Office Region.....                     | 45 |
| 4.7.3 | Core response rates by type of property and type of area .....           | 46 |
| 4.8   | Response rates and reasons for non-response: Non-white boost sample..... | 48 |
| 4.9   | Response rates and reasons for non response: Youth boost sample .....    | 50 |
| 4.10  | Response to the self-completion questionnaire.....                       | 51 |
| 4.11  | Full and Partial Interviews .....  | 54 |
| 5.    | Data Processing .....  | 55 |
| 5.1   | Offence coding.....  | 55 |
| 5.1.1 | The automatically generated offence code.....                            | 58 |
| 5.1.2 | The coding task .....  | 58 |
| 5.1.3 | Home Office coding.....  | 58 |
| 5.1.4 | Offence coding checks in 2003-4.....                                     | 59 |
| 5.1.5 | Final Offence Code.....  | 60 |
| 5.1.6 | Checks on final offence code.....  | 60 |
| 5.2   | Other coding .....   | 61 |
| 5.3   | Coding of occupation and social class .....                              | 61 |
| 6.    | Data Output.....   | 62 |
| 6.1   | Introduction .....   | 62 |
| 6.2   | Delivery of data output.....   | 62 |
| 6.3   | Content of SPSS data file.....   | 63 |
| 6.4   | Conventions used on SPSS Data Files.....                                 | 63 |
| 6.4.1 | Case identifier.....   | 64 |
| 6.4.2 | Naming conventions.....  | 64 |
| 6.4.3 | Labelling variables.....   | 64 |
| 6.4.4 | Don't Know and Refused values .....                                      | 65 |
| 6.4.5 | Multiple response variables.....   | 65 |
| 7.    | Weighting the Data.....  | 66 |
| 7.1   | Reasons for weighting.....   | 66 |
| 7.2   | Component weights.....   | 66 |
| 7.2.1 | Police Force Area weight ( $w_1$ ).....                                  | 66 |
| 7.2.2 | Inner city weight ( $w_2$ ).....   | 68 |
| 7.2.3 | Dwelling unit weight ( $w_3$ ).....                                      | 68 |
| 7.2.4 | Individual weight ( $w_4$ ).....   | 68 |
| 7.2.5 | Series weight (numinc).....  | 68 |
| 7.3   | Core sample weights.....   | 68 |
| 7.4   | Non-white sample weights.....  | 69 |
| 7.5   | Youth weights.....   | 71 |
| 7.6   | Calibration Weights .....  | 71 |
| 8.    | Comparing key survey variables with the population.....                  | 72 |

## TABLE OF FIGURES

|            |  |    |
|------------|--|----|
| Table 2.1  | Identification of eligible addresses and yield of interviews from 2003-4 BCS focused enumeration procedures.....   | 10 |
| Table 2.2  | Comparison of core sample and focused enumeration sample from the 2001 British Crime Survey.....   | 12 |
| Table 2.3  | Comparison between non-white respondents identified in the core sample and those identified in the focussed enumeration sample (2001 British Crime Survey) ..... | 13 |
| Fig. 3.1   | Flow diagram of the 2003-4 BCS Questionnaire.....  | 16 |
| Table 3.1  | Number of respondents who completed Victim Forms.....  | 22 |
| Table 3.2  | Allocation of interviews to follow-up modules.....   | 30 |
| Table 4.1  | Whether anyone else was present or not during the screener questionnaire .....   | 38 |
| Table 4.2  | Whether anyone else was present or not during the self-completion questionnaire .....  | 39 |
| Table 4.3  | Amount of assistance given by interviewer with self-completion questionnaire .....   | 40 |
| Table 4.4  | Length of interview by number of Victim Forms (Core and non-white boost sample) .....  | 41 |
| Table 4.5  | Average time for each module of the 2003-4 questionnaire (Core and non-white boost sample) .....   | 42 |
| Table 4.6  | Average time taken for each Victim Form (Core and non-white boost sample) .....  | 43 |
| Table 4.7  | Average length of each follow up module (Core and non-white boost sample) .....  | 43 |
| Table 4.8  | 2003-4 Core sample response rate and non-response breakdown ..   | 45 |
| Table 4.9  | 2003-4 Core sample response rate by Government Office Region ..  | 46 |
| Table 4.10 | 2003-4 Core sample response rate by type of property .....   | 47 |
| Table 4.11 | 2003-4 Core sample response rate by area characteristics.....  | 48 |
| Table 4.12 | 2003-4 Non-white boost sample screening outcomes, response rate and non-response breakdown.....  | 49 |
| Table 4.13 | Proportion of Non-white boost sample eligible addresses and achieved interviews by Police Force Area .....   | 50 |
| Table 4.14 | 2003-4 Youth boost sample screening outcomes and response rate ..  | 51 |
| Table 4.15 | Response to self-completion questionnaire by type of sample .....  | 52 |
| Table 4.16 | Response to the self-completion questionnaire by demographic characteristics – combined core and non-white boost sample.....                                     | 53 |
| Table 4.17 | Reasons for refusing self-completion questionnaire or for completion by interviewer –combined core and non-white boost sample .....                              | 53 |
| Fig. 5.1   | British Crime Survey Offence Coding Flowchart .....  | 56 |
| Table 7.1  | Total number of addresses in different strata over 2003-4 .....  | 70 |
| Table 8.1  | Distribution of the 2003-4 BCS achieved sample by Government Office Region compared with the population.....   | 72 |
| Table 8.2  | Age and sex distribution of 2003-4 BCS achieved sample compared with the population.....   | 73 |
| Table 8.3  | Other characteristics of 2003-4 BCS achieved sample compared with the population .....   | 74 |

# 1. Background

## 1.1 Introduction to the British Crime Survey

The British Crime Survey (BCS) is now a well-established study, one of the largest social research surveys conducted in England and Wales. The 2003-4 BCS was the third year the survey has been conducted using the new design parameters introduced in 2001.

The survey is carried out for the Home Office, and is managed by a team of researchers in the Research, Statistics and Development Directorate. They develop each survey in collaboration with an external research organisation. In 2003-4, *BMRB Social Research* was the sole contractor for the survey.

The BCS is primarily a **victimisation** survey, in which respondents are asked about the experiences of **property crimes** of the household (e.g. burglary) and **personal crimes** (e.g. theft from a person) which they themselves have experienced. Following the move to continuous interviewing in 2001 the reference period for all interviews now relates to the last 12 months before the date of interview. Although there have been changes to the design of the survey over time, the wording of the questions that are asked to elicit victimisation experiences, have been held constant throughout the life of the BCS.

Members of the public are asked directly about their experience of crime, irrespective of whether they reported these incidents to the police. As such the BCS provides a record of peoples' experiences of crime which is unaffected by variations in reporting behaviour of victims or variations in police practices of recording crime. The BCS and police recorded figures should be seen as a complementary series, which together provide a better picture of crime than could be obtained from either series alone.

The scope of the BCS goes well beyond the counting of criminal incidents, although it is for this estimate that it has become established as a definitive source of information. In order to classify incidents, the BCS collects extensive information about the victims of crime, the circumstances in which incidents occur and the behaviour of offenders in committing crimes. In this way, the survey provides information to inform crime reduction measures and to gauge their effectiveness.

As well as providing estimates of victimisation, the BCS has been used to collect high-quality information on a range of other crime-related topics, which are designed to inform the Home Office's other performance targets. Some of these topics have been well-established aspects of the BCS, providing time-series data on matters such as contacts between the public and the police, attitudes towards aspects of the criminal justice system and exposure to illegal drugs.

The data arising from the BCS are mainly reported by the Home Office's Strategic Data Flows Programme. These reports now include:

- A full statistical bulletin based on BCS interviews in the financial year, which is published in the summer following the end of each financial year. The latest of these

reports, which also provides detailed information on police recorded crime figures, appeared in July 2004, and can be found at<sup>1</sup>

<http://www.homeoffice.gov.uk/rds/pdfs04/hosb1004.pdf>

- A supplementary bulletin covering topics such as confidence in the Criminal Justice System, worry about crime, and Anti Social Behaviour. The most recent bulletin for the period 2002-3 can be found at <http://www.homeoffice.gov.uk/rds/pdfs2/hosb0204.pdf>
- Shorter statistical updates produced on a quarterly basis, focusing specifically on victimisation rates and trend patterns.
- Longer subject-specific reports in the Home Office Research Series (HORS). One example of a recently published report based on the 2001 BCS survey is in relation to domestic violence, sexual assault and staking. This can be found at <http://www.homeoffice.gov.uk/rds/pdfs04/hors276.pdf>
- Brief Research Papers dealing with specific topics. Some of these reports are only available online. One example of a recently published online report based on the 2003-4 BCS survey is in relation to Anti Social Behaviour. This can be found at <http://www.homeoffice.gov.uk/rds/pdfs04/rdsolr4904.pdf>

The above references are intended only to illustrate the types of reports and findings that are produced from the BCS. For more details on all RDS publications associated with the BCS see <http://www.homeoffice.gov.uk/rds/bcs1.html>

## 1.2 Structure of the technical report

This report documents the technical aspects of the 2003-4 British Crime Survey carried out in England and Wales. The analysis in this report relates to the total sample that was issued in the financial year 2003-4, irrespective of when interviews actually took place. The distinction between issued sample and achieved sample is explained in more detail in section 4.3 of the report.

The sample design is set out in Chapter 2, showing how the Home Office's requirements were translated into a detailed specification. Data collection is the major task for the organisation commissioned to conduct the BCS and forms the central part of this report. Chapter 3 covers the content and development of the questionnaire, while Chapter 4 examines the fieldwork. Chapter 5 and 6 give details of the tasks that are involved in preparing the data for analysis, including the coding and offence classification and the preparation of the BCS data files. Chapter 7 outlines the weighting required for analysis of the data. Chapter 8 provides the results of some checks on the profile of the BCS achieved sample against estimates for the population that the BCS aims to represent.

Although it documents the way in which the BCS was conducted, this report does not explain how to analyse the data set. It is worth emphasising that the BCS is a complex study with data organised at different levels (households, individuals, and incidents) and it has numerous sub-samples that were asked specific questions. Accordingly considerable effort and expertise is required to analyse the data and to interpret it in a valid manner. Some of the analysis routines that play a key role in the published estimates are implemented after the data have been handed over to the Home Office, and are not

---

<sup>1</sup> Dodd T, Nicholas S, Povey D and Walker A (Editors) Crime in England and Wales 2003/2004 Home Office Statistical Bulletin 10/04

documented in this report. Data files from the BCS are deposited in SPSS format with the Data Archive at the University of Essex.

Considerable emphasis is given in the course of conducting the interview to assure respondents that the information they provide will be held in confidence. For this reason, the data set does not identify the location of the sampled areas (postcode sectors) and this information is not released to the Home Office by the survey organisation.

The Home Office produces a set of training notes for those interested in analysing BCS data which contain further detail on the content and structure of the data and guidance on analysis.

## 2. Sample design

### 2.1 Introduction

Following a methodological review<sup>2</sup>, significant changes were made to the design of the British Crime Survey in 2001. The two main changes introduced at this time were an increase in the sample size and a move to continuous fieldwork. The move to continuous data collection was also associated with a change in the survey reference period, with respondents being asked about events that had taken place in the last full 12 months from the date of interview. Prior to this respondents had been asked about events that had taken place in the last calendar year.

As part of these design changes a ‘spliced design’ was carried out during the first 6 months of 2001 in order to assess the impact on victimisation rates of the change in reference period. This design involved randomly allocating all issued addresses during the first 6 months of 2001 either to survey A, which used the old reference period, or to survey B which used the new reference period<sup>3</sup>.

Subsequent analysis undertaken by Peter Lynn (Professor of Survey Methodology in the Institute of Social and Economic Research at the University of Essex) indicated that the change to the reference period may, for some household offences, have improved recall of offences (i.e. may have increased the household crime count). However, for crimes against the person there appeared to be no difference in recall of offences, possibly because victims tend to remember such crimes more easily.

The design of the 2003-4 survey was largely the same as that introduced for 2001 and the rest of this chapter briefly outlines the main elements of this design.

### 2.2 Sample size and structure

The 2003-4 survey was designed to be representative of two linked populations:

- households in England and Wales living in private residential accommodation; and
- adults aged 16 and over living in such households.

The survey did not cover the population resident in institutions such as halls of residence, those in residential care, those in prison, or members of the armed forces. Nor did it cover the experiences of those aged under 16.

The 2003-4 survey was also designed to be representative over time, with each quarter of issued sample being fully representative.

Over the whole year the aim was to achieve approximately 37,000 interviews as part of the core sample. Additionally, the survey aimed to achieve a boost of 3,000 interviews with individuals from non-white groups identified through focused enumeration (see

---

<sup>2</sup> Lynne P and Elliot D (2000) *The British Crime Survey : A Review of Methodology* (London : National Centre)

<sup>3</sup> More details of the spliced design can be found in Bolling, K *et. al.* (2002) *2001 British Crime Survey (England and Wales) Technical Report* (London : BMRB)

section 2.8), and a boost of 2,000 interviews with 16 to 24 year olds identified through screening at core addresses (see section 2.9).

In addition to representing the above populations, the 2003-4 survey was also designed to ensure that:

- at least 600-700 core sample interviews were conducted in each Police Force Area (PFA); and
- 50% of the primary sampling units (PSUs) used on the survey were ones that had been used on the 2002-3 survey. This was done in order to improve the precision of year on year comparisons.

### **2.3 Sample frame**

The small user Postcode Address File (PAF) was used as the sample frame. This is generally accepted as being the best general population sampling frame in England and Wales. It has the best coverage of both residential addresses and of the private household population of individuals, and what non-coverage it has is less concentrated in particular population sub-groups. Furthermore PAF is structured hierarchically, is available in computerised form and can be linked to Census data thereby permitting considerable control to be exercised over the sampling process.

### **2.4 Stratification**

As mentioned in section 2.2, one of the survey requirements was a design that achieved at least 600-700 core sample interviews per Police Force Area (PFA)<sup>4</sup>. If PSUs had been selected with probability proportional to size then the number of interviews achieved in many smaller Police Force Areas would not have reached the minimum number required. Consequently the survey was designed to over sample in the smallest PFAs in such a way to ensure the minimum number of interviews was achieved. One result of this is that the data have to be weighted to remove the effect of differences in the probability of selection of addresses within different Police Force Areas (see Chapter 7).

As well as stratifying disproportionately by PFA, the sample was stratified by other socio-demographic variables in order to maximise the precision of estimates. The stratifiers used in 2003-4 were the same as on previous surveys, namely population density and the proportion of household heads in non-manual occupations. Further details of how the sample frame was stratified are outlined in section 2.7

### **2.5 Clustering**

Whole postcode sectors were used as the primary sampling units (PSUs) and 32 addresses were issued in each PSU. Although issuing a smaller number of addresses in each PSU would have been possible it was felt that the sample efficiency gains attached to reducing the cluster size would be more than offset by the concomitant cost increases.

---

<sup>4</sup>There are 43 Police Force Areas in England and Wales, with each PFA being roughly the same as a county.

However, in PSUs where the proportion of non-white households exceeded 20% (based on the 1991 Census) only 16 addresses were issued. This was done for practical reasons to try and limit the variation in interviewer assignment sizes which would arise from the focused enumeration procedures used to generate the non-white boost sample (see section 2.8). In order to ensure that overall the address sample was self-weighting, PSUs defined as being of high non-white population were duplicated in the sample frame.

Small sectors (containing fewer than 500 delivery points) were amalgamated with neighbours before sample selection to ensure the sample had a reasonable geographic spread in these areas.

## **2.6 Rotation of sectors**

A requirement of the survey design was that approximately 50% of all PSUs issued on the 2002-3 survey should be rotated forward and issued on the 2003-4 survey. This was done in order to increase the precision of year on year estimates of change. Where PSUs were rotated forward from the previous year a fresh set of addresses is selected. New addresses were always selected with reference to the addresses that had been selected last year to ensure there was no chance of the same address being selected two years in a row.

In total, 908 out of the 1,790 PSUs issued on the 2003-4 survey were rotated forward from the previous year. Wherever possible an effort was made to rotate PSUs forward by exactly 12 months (i.e. if a PSU was issued for the first time in July 2002, the aim was to issue it again in July 2003). In practice, it was not always possible to allocate rotated points to exactly the same fieldwork month but in all cases rotated points were allocated to the same quarter as they had been the previous year.

Because of the requirement to rotate sectors forward, in essence the 2003-4 sample consisted of two different samples, namely the sample of rotated PSUs (“the rotated sample”) and the sample of fresh PSUs selected for the first time in 2003-4 (“the fresh sample”). Since the fresh sample was selected from the universe of PSUs this meant that a small proportion of areas in 2003-4 were selected twice in the same survey year (that is, the same postcode sector was rotated forward from 2002-3 **and** was also selected as part of the fresh sample).

In fact, in 2003-4 118 PSUs were selected in both the rotated and the fresh sample. Where this situation arose the rotated and the fresh PSU were treated entirely separately in terms of each having their own assignment number and, in most cases, being issued at different times of the year. However, the selection of addresses for these ‘duplicated’ PSUs was done as a single exercise to prevent addresses being selected twice in the same year (i.e. 64 addresses were selected from the PAF and then randomly allocated to the two assignments).

## **2.7 Procedures for selecting the sample**

The sample of PSUs for the 2003-4 survey was selected as follows:

1. The first stage in the process was to amalgamate any postcode sector containing fewer than 500 delivery points. Small sectors were amalgamated with neighbouring sectors in such a way to ensure that sector combinations did not cross Police Force Area boundaries.

2. All PSUs were then sorted into Police Force Area<sup>5</sup>. Since postal boundaries and administrative boundaries are not co-terminous, where a postcode sector crossed a PFA boundary it was allocated to the PFA that the majority of the PSU fell into.
3. All PSUs were flagged as being either 'high' or 'low' concentration of non-white households based on the 1991 Census. Areas of high non-white population were defined as those where 20% or more of households were of non-white origin, while all other areas were defined as being of low non-white population. High non-white PSUs were duplicated in the sample frame for reasons outlined in section 2.5.
4. All PFA's were designated as 'large' or 'small' depending upon whether a minimum of 650 interviews could be achieved based on drawing a sample with probability proportional to size. In total, there were 28 'small' PFA's that required over sampling and these were each treated as separate strata. The remaining 14 'large' PFA's were treated as a single stratum.
5. Within each PFA, PSUs were ordered by population density and then divided into three bands each containing approximately an equal number of delivery points (high, medium and low density).
6. Within each density stratum, PSUs were then ordered by proportion of households with a non-manual household head.
7. PSUs were then sampled within each PFA with probability proportional to number of delivery points by the method of random start and fixed interval.

The sample of PSUs for the whole 12-month period was selected at the beginning of the year. They were then systematically allocated to quarter to ensure that a nationally representative stratified sample of sectors was issued in each quarter. Within quarter points were then systematically allocated to month.

It is important to note that although each month's issued sample is broadly spread geographically, the sample is designed only to be representative on a quarterly basis and **not** on a monthly basis.

Once the PSUs had been selected, 32 delivery points were selected by the method of random start and fixed interval. Addresses were ordered by postcode before selection to maximise the geographical spread across the sector.

In the relatively infrequent cases where a PAF address generated more than one household, the interviewers in the field used a random selection method to select one.

Individuals aged 16+ in the selected households (and in eligible non-white households in the non-white boost sample) were listed in alphabetical order of first name and one was selected for interview by a random (Kish grid based) method. No substitutes were permitted. Selection of households where more than one existed and selection of an adult within the household were done by interviewers as part of their doorstep screening procedures.

---

<sup>5</sup> City of London and Metropolitan were treated as a single PFA.

## 2.8 Non-white boost sample

For some analyses, the Home Office provides separate estimates for the non-white population, and for specific groups within this population. Since the number of non-white respondents identified within the core sample is not sufficient to allow for sufficiently robust analyses<sup>6</sup>, a boost sample was incorporated in the design of the survey.

Prior to 2001 the boost sample on the BCS was restricted to black and Asian populations. From 2001, it was decided to adopt the categories used in the 2001 Census and so broaden the population covered. Thus, the groups defined as 'non-white' were:

- Indian
- Pakistani
- Bangladeshi
- Other Asian background
- Caribbean
- African
- Other black background
- Chinese
- White and black Caribbean mixed
- White and black African mixed
- White and Asian mixed
- Any other mixed background where at least part of the mix includes one of the above groups.

The key difficulty in drawing any probability sample of people of non-white origin is identifying them in a cost-effective manner. Prior to 2001, two different methods were used on the BCS to achieve a non-white boost sample. First, issuing a large number of pre-selected addresses in areas that had a high proportion of non-white households (these were known as boost areas). Direct face-to-face screening was then carried out at the pre-selected addresses for non-white households. And second, using a method known as focused enumeration, whereby interviewers screen addresses that are adjacent to the core sample address for eligible respondents.

The 2003-4 survey continued the protocol started in 2001 of using only focused enumeration to boost the sample of non-white respondents. This is felt to be preferable to the alternative method since it produces a sample that is representative of the entire non-white population, rather than just the non-white population living in areas of high concentration. The method is designed to try and ensure that interviewers can screen a large number of addresses as efficiently as possible by using proxy information. It is far more cost effective than direct face-to-face screening, especially in areas of low non-white concentration.

Focused enumeration involves screening addresses by proxy where possible. Interviewers screen groups of pre-identified addresses that are adjacent to the core sample address by contacting a responsible adult at either the core sample address or one of the identified adjacent addresses and asking about the ethnic origins of those living at the adjacent

---

<sup>6</sup> In 2003-4, the core sample contained 2,149 respondents who defined themselves as non-white.

addresses. The method has taken a number of forms since it was devised<sup>7</sup>, and the version used on the BCS is one variant of the basic methodology.

In practice, the method used on the 2003-4 survey was as follows:

- The interviewer identified two adjacent addresses on each side of a core sample address using a strict set of rules.
- Wherever possible, interviewers used the contact at the core sample address to establish whether anyone lived at the two addresses either side who they thought was non-white.
- In cases where this was not possible, either because no contact was made at the core sample address or because the person at the core address refused to give the information or did not know anything about their neighbours, the interviewer approached one of the four identified addresses to try and gain the information.
- The interviewer continued like this until they had information about all four identified addresses. The aim was to obtain information about all four addresses in a single visit, so that no additional visits were needed.
- If a household with eligible respondents was identified, the interviewer sought to carry out an interview. Where more than one eligible respondent existed, the interviewer carried out a selection procedure exactly the same as on the core sample.

Since ethnic origin is a matter of self-definition interviewers were briefed to only use the wording provided, which was *“Is there anyone living at [specific address]...who is black, Asian, Chinese, or from any other non-white group”*. Interviewers were also briefed to accept the responses given and not to try and apply any of their own definitions. If respondents had difficulty in giving a direct ‘Yes’ or ‘No’ answer to the question, interviewers could use a card with the census categories on it to help respondents.

Calculating the number of addresses that need to be screened in order to achieve a sample of 3,000 respondents over the whole year is complex, especially considering that information on non-white population came from the 1991 Census which was 10 years out of date. During the previous two years of the survey the proportion of addresses used for screening had to be adjusted during the fieldwork year to ensure that the overall target was met.

In 2003-4 no adjustment was required during the survey year and so the proportion of addresses used for screening throughout the year was as follows:

- In low density PSUs (i.e. less than 20% non-white population) non-white screening was carried out at 75% of core sample addresses (i.e. 24 out of 32 addresses); and
- In high density PSUs (i.e. 20% or more non-white population) non-white screening was carried out at 100% of core sample addresses (i.e. 16 out of 16 addresses)

---

<sup>7</sup> See, for example, Smith P and Prior G (1997) *The Fourth National Survey of Ethnic Minorities* : Technical Report (London : SCPR)

Screening was controlled through the use of different coloured Address Contact Sheets, which easily identified to interviewers whether screening was required or not at a particular address (see section 4.4).

The number of addresses screened at, the number of respondents identified as eligible and the number of interviews achieved over the whole year are shown in Table 2.1. Over the whole year 4.0% of addresses screened contained an eligible respondent. As with previous years of the survey, focused enumeration seems to result in more non-white respondents being identified at the addresses closest to the core sample address. Thus, there was an identification rate of 4.3% at addresses that were first to the left and right of the core sample address compared to an identification rate of 3.6% at addresses that were second to the left and right of the core sample address.

**Table 2.1 Identification of eligible addresses and yield of interviews from 2003-4 BCS focused enumeration procedures**

|              | Number of addresses Screened | Addresses at which Non-white resident identified |            | Addresses subsequently identified as ineligible |             | Eligible addresses for Non-white boost sample |            | Achieved interviews |             |
|--------------|------------------------------|--|------------|---|-------------|---|------------|---------------------|-------------|
|              |                              | N  | %          | N   | %           | N   | %          | N                   | %           |
| Left 1       | 41,688                       | 2,024  | 4.9        | 205   | 10.1        | 1,819   | 4.4        | 903                 | 49.6        |
| 2            | 41,688                       | 1,740  | 4.2        | 221   | 12.7        | 1,519   | 3.6        | 772                 | 50.8        |
| Right 1      | 41,688                       | 1,975  | 4.7        | 210   | 10.6        | 1,765   | 4.2        | 888                 | 50.3        |
| 2            | 41,688                       | 1,716  | 4.1        | 201   | 11.7        | 1,515   | 3.6        | 755                 | 49.8        |
| <b>Total</b> | <b>166,752</b>               | <b>7,455</b>                                     | <b>4.5</b> | <b>837</b>                                      | <b>11.2</b> | <b>6,618</b>                                  | <b>4.0</b> | <b>3,318</b>        | <b>50.1</b> |

## 2.9 Evaluation of focused enumeration

Over the past few years of the survey it has been apparent that the variant of focussed enumeration used tends not to deliver as many interviews with non-white adults as might be expected from Census data. While this can be partly explained by a lower than anticipated response rate, it is also clear that the identification rate (i.e. the proportion of households where at least one non-white adult is identified) tends to be lower than Census figures would suggest.

Over the last three years of the survey the identification rate has remained reasonably consistent at around 4-5% of all screened households. Even allowing for the design of the survey (which over samples smaller Police Force Areas that, on the whole, tend to have a relatively low non-white population) it was felt that based on the 1991 Census figures, the identification rate should be somewhere around 5-6% of households screened.

In order to investigate this further BMRB undertook an analysis of the 2001 BCS data to investigate the efficacy of the BCS variant of focussed enumeration as a screening method. The results of this analysis are reported here.

The first part of the analysis was to compare the non-white interview "productivity" of adjacent addresses that were screened by focussed enumeration with that of core addresses at which an interview was achieved regardless of the ethnic origin of the

respondent. The hypothesis was that an address containing non-white adults would be about as likely to generate an interview if it were sampled as an address adjacent to a core sample address where it had been subject to focussed enumeration, as if it were sampled as a core sample address. If eligible addresses screened by focussed enumeration prove **less** likely to deliver interviews than equivalent addresses covered as part of the core sample, this might suggest that focussed enumeration is performing less than ideally.

It should be noted that because the two methods of sampling differ in a number of ways it is not possible to treat the two respective measures of response rate and identification rate as *precisely* equivalent.

One especially important procedural difference concerns the definition of non-white addresses. At focussed enumeration addresses, interviewers screened for the presence of one or more non-white adult resident. Any such address was defined as eligible and a non-white adult was randomly selected for interview. However, at each core sample address one adult was selected for interview at random and if this person happened to be a non-white person, the address was defined as eligible. Therefore, ethnically mixed households (i.e. ones containing both white and non-white adults) were treated differently in the two samples. In the focussed enumeration sample, mixed households were always counted as non-white ones, whereas in the core sample, some mixed households (those where the respondent was white) were not counted as such.

Unfortunately, in the 2001 survey at both types of address ethnic origin was recorded for the respondent but not for other household members, meaning that it is impossible to eliminate the impact of this procedural difference through weighting<sup>8</sup>. It is worth noting that because of this difference we would expect to find, other things being equal, more focussed enumeration than main addresses categorised as eligible, which *is the opposite of what we find*.

Despite these differences in definition our view was that it was still a valid exercise to compare the two and that any large differences observed should be treated as interpretable.

For the 2001 BCS data we compared the rate at which core and focussed enumeration sample addresses generated productive interviews after weighting to even out variations in the proportion of addresses issued for focussed enumeration in areas of different non-white concentration<sup>9</sup>. Overall, what we found was that each eligible address generated over twice as many achieved interviews with non-white individuals when it was covered as a core address than when it was covered as an adjacent address (see Table 2.2).

---

<sup>8</sup> Since these data were collected the questionnaire has been amended and information on ethnic origin is now collected for all household members.

<sup>9</sup> The focussed enumeration and core sample addresses used in these comparisons were all taken from the same postcode sectors. Weighting was undertaken to match the sector profiles of the two samples.

**Table 2.2 Comparison of core sample and focused enumeration sample from the 2001 British Crime Survey**

|  | <b>Core sample</b> | <b>Focused enumeration sample</b> |
|--|--------------------|-----------------------------------|
| Number of issued eligible addresses required to generate one non-white interview | 23.6               | 51.3                              |
| Estimated identification rate  | 5.8%               | 4.0%                              |
| Estimated response rate  | 73.2%              | 48.4%                             |

As can be seen from Table 2.2 both the identification rate and response rate were around 50% greater for core sample respondents than for focussed enumeration sample respondents. In other words, the focussed enumeration method being used seemed to fail to identify a substantial number of eligible addresses and was also associated with lower response rate at eligible addresses.

This analyses suggests that focussed enumeration as applied on the BCS is delivering perhaps only half the number of interviews with eligible non-white individuals that might be expected from analysis of the core sample data. The question is whether this matters *in practice*. It could be argued that those who are missed by the method are effectively a random subset of all those who are eligible and that therefore their omission will not bias the survey results. To try and answer this question we can compare the profile of those non-white respondents who were identified through focussed enumeration with those identified in the core sample in the same postcode sectors.

Such comparisons indicate that there were a number of differences between the two samples in 2001. When compared with those identified as belonging to a non-white group in the core sample, focussed enumeration respondents were:

- more likely to be Asian;
- less likely to belong to mixed and "other" ethnic groups;
- more likely to live in terraced houses and less likely to live in flats;
- more likely to live in households with children;
- more likely to live in households with three or more adults;
- more likely to be married;
- more likely to be aged 25-44 and less likely to be aged 55-64; and
- more likely to be male than female.

Some of the significant differences are illustrated in Table 2.3.

**Table 2.3 Comparison between non-white respondents identified in the core sample and those identified in the focussed enumeration sample (2001 British Crime Survey)**

|   | Core sample (%) | Focussed enumeration sample (%) |
|---|-----------------|---------------------------------|
| Asian   | 51.1            | 61.5                            |
| Mixed   | 8.4             | 4.9                             |
| Other ethnic group                            | 12.5            | 4.8                             |
| Female  | 54.0            | 48.0                            |
| Aged 25-44                                    | 47.3            | 51.1                            |
| Aged 45-64                                    | 24.8            | 20.9                            |
| Living in a mid-terrace house                 | 32.9            | 37.3                            |
| Living in flat, room/bedsitter                | 20.0            | 15.3                            |
| No security device visible at address         | 63.5            | 71.1                            |
| Sampled house in very good physical condition | 27.0            | 20.6                            |
| Three or more adults in household             | 42.7            | 49.4                            |
| Respondent married                            | 54.3            | 58.3                            |
| Children in household                         | 48.1            | 54.6                            |
| <i>Unweighted base</i>                        | <i>1971</i>     | <i>3269</i>                     |

These results suggest that the under-coverage associated with focussed enumeration is not random, but rather associated with a range of biases. These biases are, perhaps, those that one might expect:

- less visible minority groups (e.g. mixed ethnic origin) are less well represented;
- larger households, which are likely to be more visible and are associated with better response, are better represented;
- the kind of property in which it is easy to apply focussed enumeration are better represented (property types such as terraced housing may be associated with better response rates too).

The preliminary conclusions from these analyses are that (i) focussed enumeration is associated with under-coverage of eligible non-white households (ii) that this under-coverage is non-random and (iii) that this under-coverage is due to a combination of both misidentification and relatively poor response rates.

As a result of this analysis the way in which focused enumeration was conducted on the BCS was changed slightly to try and improve the identification rate. Two changes were made to how the screening was carried out in the field:

- Interviewers were required to use a show card which lists all the ethnic identity categories, rather than verbally ask the screening question. It was felt that where information was being collected by proxy, having a show card with all the different ethnic categories would make it easier for respondents to give accurate information.
- Interviewers were asked to do the screening process on their first visit to the core sample address by using the different strategies available to them. This was intended to ensure that potential non-white households were identified as early as possible in the fieldwork period, thereby giving interviewers a better chance of obtaining an interview. In practice, many interviewers already did screening in this way and so this change represented a tightening up of the fieldwork rules

## **2.10 Youth Boost sample**

As well as increasing the number of interviews conducted with non-white respondents, the Home Office also wanted a boost of young people (aged 16-24) to be carried out. For the first time the Youth Boost part of the survey was carried out over the whole year as an integrated part of survey. The aim was to achieve a sample of 2,000 respondents aged 16 to 24. The 'youth questionnaire' covered fewer topics than the main questionnaire and consequently average interview length was considerably shorter.

A separate screening exercise was developed to generate sample for this age group. Interviews were only sought with eligible respondents at addresses that had been selected as part of the core sample.

Since youth screening could involve conducting two interviews in a single household, the selection for the core sample always took place first. This was to ensure that ALL adults (aged 16 and over) in the household were included in the main selection process. If the person selected as the core sample respondent at the address was aged 16 to 24, a youth boost interview was not conducted, regardless of whether an interview was achieved with a core sample respondent or not. This was to ensure that no more than one 16 to 24 year old was ever interviewed in the same household.

Interviewers screened for 'youth sample' by asking a responsible adult at the core address whether there was anybody living at the address aged 16-24 years old. If more than one 16-24 year old was identified at the address, the same random selection procedure was applied as with the core sample to identify one person for interview.

Details of the youth screening and response rate for 2003-4 can be found in section 4.9.

## **3. Questionnaire Content and Development**

### **3.1 Structure and coverage of the questionnaire**

The 2003-4 BCS questionnaire comprised 10 main sections as follows:

- Household Grid and Main questionnaire
- Screener questionnaire which identified respondents' experience of victimisation within the reference period
- Victim Forms for any incidents identified at the screeners (up to a maximum of six)
- Module on mobile phone theft
- Module on the performance of the Criminal Justice System
- Follow-up modules (A-D) covering different crime-related topics
- Technology crimes module
- Anti-social behaviour module
- Demographic module
- Self-completion modules on drug use and drinking

The basic structure of the questionnaire is shown in Figure 3.1. This also shows what proportion of the sample was allocated to each different module of the questionnaire (see section 3.1.6). The complete questionnaire is documented in Appendix D of Volume 2.

The fires module and the self completion module on stolen goods were removed from the 2003-4 survey.

A brief description of each section or module of the questionnaire is outlined below.

Figure 3.1 Flow diagram of the 2003-4 BCS Questionnaire

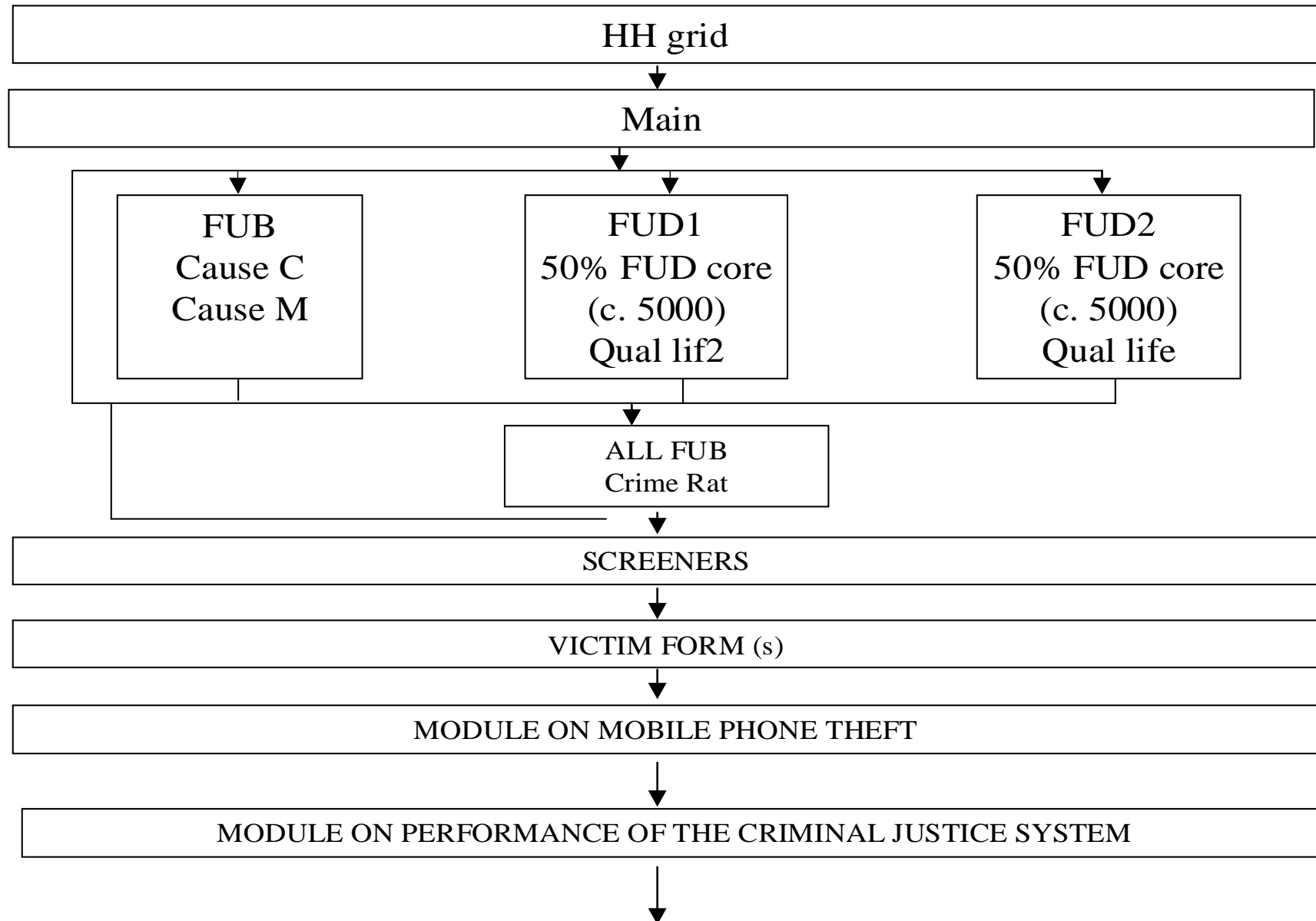
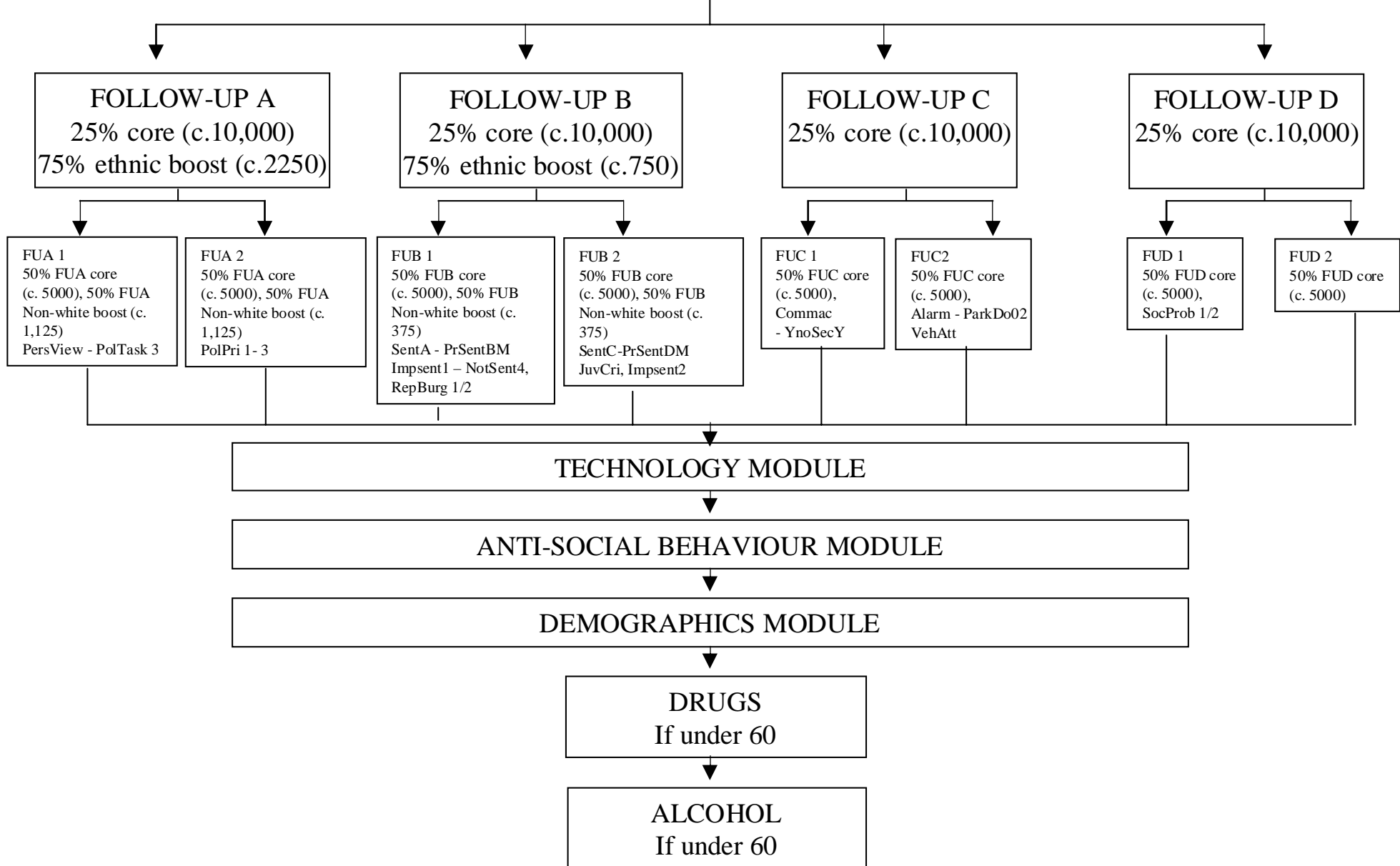


Figure 3.1 (cont.) Flow diagram of the 2003-4 BCS Questionnaire





### 3.1.1 Household Grid and Main questionnaire

Basic socio-demographic details (age, sex, marital status, etc.) were collected in the Household Grid for every adult in the household. Additionally, demographic details of all children under 16 were collected.

The Household Grid was also used to establish the **Household Reference Person**<sup>10</sup>. Household Reference Person (HRP) is the standard classification now used on most government surveys and is based on the following criteria:

- The HRP is the member of the household in whose name the accommodation is owned or rented, or is otherwise responsible for the accommodation. In households with a *sole* householder that person is the HRP.
- In households with *joint* householders the person with the *highest income* is taken as the HRP.
- If both householders have exactly the same income, the *older* is taken as the HRP.

In 2002-3, ethnicity was added to the Household Box and asked of all adults in the household. On previous surveys this had only been asked of the respondent.

The Household Grid was followed by the Main questionnaire, a series of primarily attitudinal questions, which were mostly asked of all respondents. Many of these questions relate to respondents' own local area, such as how long respondents had lived in their local area, what they thought were the main problems in their area, how safe they felt, and how often they went out.

A few of the questions in the Main questionnaire were only asked of a random sub-sample of respondents (e.g. main causes of crime, how crime/fear of crime affects quality of life, and how crime rates have changed over the past two years).

### 3.1.2 Screener questionnaire

Following the Main questionnaire, all respondents were asked whether they had experienced certain types of crimes or incidents within a specified reference period, namely the last 12 months from the date of interview. To try and encourage respondents to recall events accurately, a life event calendar was given to all respondents to act as a visual prompt when answering the screener questions (see section 3.2).

Depending upon individual circumstances a maximum of 25 screener questions were asked. These can be grouped into four main categories:

- All respondents living in households with a vehicle or bicycle during the reference period were asked about experience of vehicle-related crimes (e.g. theft of vehicle, theft from vehicle, damage to vehicle, bicycle theft)
- All respondents who had moved in the reference period were asked about experience of property-related crimes in their **previous** residence(s) (e.g. whether anything was stolen, whether the property was broken into, whether any property was damaged)

---

<sup>10</sup> Prior to 2001 all previous surveys collected details of the Head of Household.

- All respondents were asked about experience of property-related crimes in their **current** residence.
- All respondents were asked about experience of personal crimes (e.g. whether any personal property was stolen, whether any personal property was damaged, whether they had been a victim of force or violence or threats).

The wording of the screener questions has been kept consistent since the BCS began to ensure comparability. They are designed to ensure that all incidents of crime within the scope of the BCS, including relatively minor ones, are mentioned. The screener questions deliberately avoid using terms such as ‘burglary’, ‘robbery’, or ‘assault’, which have precise definitions that many respondents might not be expected to know.

The questions are also designed to ensure that the respondent does not mention the same incident more than once. At the end of the screener questions, the interviewer is shown a list of all incidents recorded and is asked to check with the respondent that all incidents have been recorded and nothing has been counted twice. If this is not the case, the respondent has an opportunity to correct the information before proceeding.

Within the screener questions a crucial distinction exists between **household incidents** and **personal incidents**.

All vehicle-related and property-related crimes are considered to be household incidents, and respondents are asked about whether **anyone** currently resident in the household has experienced an incident within the reference period. A typical example of a household incident is criminal damage to a car. It is assumed that the respondent will be able to recall these incidents and provide information even in cases where he/she was not the owner or user of the car.

Personal incidents refer to all crimes against the individual and only relate to things that have happened to the respondent **personally**, but not to other people in the household. An example of a personal incident would be a personal assault. An assault against other household members would not be recorded, unless the respondent was also assaulted in the course of the incident. In such cases, the offence would be coded according to the crime experienced by the respondent (which may not be the same as the experience of another household member).

### 3.1.3 Victim Forms

All incidents identified at the screener questions are followed through in more detail in the Victim Form. Incidents are covered in a specific priority order, which has been kept consistent since the start of the BCS.

#### **3.1.3.1 Identification and ordering of incidents for Victim Forms**

In 2003-4, 73% of respondents did not report any victimisation over the reference period, meaning that no Victim Forms had to be completed. This is a slightly higher proportion of respondents compared with the 2002-3 survey, when 71% of respondents did not report any victimisation.

Where a respondent had experienced one or more incidents in the reference period, the computer programme automatically identified the order in which the Victim Forms were asked.

This meant that the interviewer had no discretion about the selection or order of Victim Forms<sup>11</sup>.

If six or fewer incidents were identified at the screener questions then a Victim Form was completed for all of the incidents reported. The priority ordering used by the computer was as follows:

- According to the **type** of crime. Victim Forms were asked in reverse order to the screener questions. Broadly speaking this means that all personal incidents were asked before property-related incidents, which were asked before vehicle-related incidents.
- **Chronologically** within each type of crime. If a respondent reported more than one incident of the same type of crime, Victim Forms were asked about the most recent incident first and worked backwards chronologically.
- The first three Victim Forms were **long forms**, which contain all the detailed questions relating to each incident. The second three Victim Forms were **short forms**, a cut down version of the questions that are much quicker to complete.

If the respondent had experienced more than six incidents in the reference period, only six Victim Forms were asked using the above priority ordering. The priority ordering means that the survey does not collect details or only collects limited details (through the short Victim Form) for the crimes or incidents that tend to be more common (e.g. criminal damage to vehicles).

In 2003-4, a total of 16,667 Victim Forms were completed on the core and non-white boost sample and 27% of respondents reported at least one incident (see Table 3.1). Respondents in the non-white boost sample were more likely than those in the core sample to report a crime (33% and 27% respectively).

Among victims, about two-thirds of respondents (69%) had experienced one crime in the reference period and so had completed only 1 Victim Form. Only about one in five (5%) respondents who had been the victim of crime completed 4 or more Victim Forms.

Table 3.1 shows that just under one in five (19%) of all respondents completed 1 Victim Form, while only 1% of all respondents completed 4 or more Victim Forms.

---

<sup>11</sup> In the case of the incidents of sexual victimisation or domestic violence, the interviewer had an option to suspend the Victim Form, as this might embarrass or endanger the respondent in some situations. The interviewer would then attempt to arrange a revisit at a time that would be more convenient (in particular when other household members would not be present).

**Table 3.1** Number of respondents who completed Victim Forms

|                                   | Core sample |      | Non-white boost sample |      | Total  |      |
|-----------------------------------|-------------|------|------------------------|------|--------|------|
|                                   | N           | %    | N                      | %    | N      | %    |
| None (Non Victim)                 | 27,241      | 73.2 | 2,214                  | 66.7 | 29,455 | 72.7 |
| One or more (Victim)              | 9,972       | 26.8 | 1,104                  | 33.3 | 11,076 | 27.3 |
| Number of Victim Forms completed: |             |      |                        |      |        |      |
| 1                                 | 6,863       | 18.4 | 746                    | 22.5 | 7,609  | 18.8 |
| 2                                 | 2,002       | 5.4  | 223                    | 6.7  | 2,225  | 5.5  |
| 3                                 | 656         | 1.8  | 80                     | 2.4  | 736    | 1.8  |
| 4                                 | 229         | 0.6  | 35                     | 1.1  | 264    | 0.7  |
| 5                                 | 99          | 0.3  | 9                      | 0.3  | 108    | 0.3  |
| 6                                 | 123         | 0.3  | 11                     | 0.3  | 134    | 0.3  |

**3.1.3.2 Series of incidents**

Most incidents reported represent one-off crimes or **single incidents**. However, in a minority of cases a respondent may have been victimised a number of times in succession. At each screener question where a respondent reported an incident, they were asked how many incidents of the given type had occurred during the reference period. If more than one incident had been reported, the respondent was asked whether they thought that these incidents represented a ‘series’ or not. A **series** was defined as “*the same thing, done under the same circumstances and probably by the same people*”. Where this was the case, only one Victim Form was completed in relation to the **most recent** incident in the series. In 2003-4, 83% of all Victim Forms related to single incidents and 17% related to a series of incidents.

There are two practical advantages to this approach of only asking about the most recent incident where a series of similar incidents has occurred. First, since many (although not all) incidents classified as a series tend to be petty or minor incidents (e.g. vandalism) it avoids the need to ask the same questions to a respondent several times over. Secondly, it avoids ‘using up’ the limit of six Victim Forms on incidents which tend to be less serious.

In the rare cases where a respondent has experienced a mixture of single incidents and a series of incidents the Quanquest interview program has a complex routine which handles the sequence of individual and series incidents and allows the priority ordering of the Victim Forms to be decided.

In terms of estimating the victimisation rates, series incidents receive a weight corresponding to the number of incidents up to a maximum of five (see section 7).

**3.1.3.3 Content of Victim Forms**

The Victim Form is the key to the estimate of victimisation and collects three vital bits of information:

- The exact month(s) in which the incident or series of incidents occurred. In a few cases, respondents may have reported an incident, which later turned out to have been outside the reference period. In such cases, the Victim Form was simply by-

passed by the computer. If respondents were unsure about the exact month in which something happened, they were asked to narrow it down to a specific quarter. For incidents that were part of a series, respondents were asked how many incidents occurred in each quarter and the month in which the most recent incident had occurred.

In the questionnaire program dates were automatically calculated based on the date of interview and appropriate text substitution was used to ensure that the questions always referred to the correct reference period. Because the 12 month reference period changed throughout the fieldwork year, this meant that some date-related questions in the Victim Form had different text each month to reflect this changing reference period. Details of these questions and the appropriate reference periods used for each month of the 2003-4 survey can be found in Appendix F of Volume 2.

- An open-ended description of the incident where the respondent describes exactly what happened in their own words. The open-ended description is vital to the accurate coding of offences that takes place back in the office. Short, ambiguous or inconsistent descriptions can often make offence coding difficult. At the end of each Victim Form, the original open-ended description that the interviewer had entered at the start of the Victim Form is re-capped, along with the answers to some of the key pre-coded questions. By presenting this information on a single screen, interviewers have the chance to confirm with respondents that the information was correct and consistent. If the respondent and/or interviewer wish to add or clarify any information they then have the opportunity to do this.
- A series of key questions used to establish important characteristics about the incident. Examples of the sort of information collected includes where and when the incident took place; whether there was a racial element to the incident; whether anything was stolen or damaged and, if so, what; the costs of things stolen or damaged; whether force or violence was used and, if so, the nature of the force used and any injuries sustained; and whether the police were informed or not.

The questions within the Victim Form have remained largely unchanged from previous years of the survey.

### **3.1.4 Module on mobile phone theft**

To reflect this new area of crime concern, a module on mobile phone theft was first included on the survey from July 2001. Although mobile phones stolen from the respondent should be identified in the Victim Form, thefts from other members of the household (including children) are not covered. Consequently, in this module all respondents were asked who in the household, if anyone, used a mobile phone, whether anyone in the household had had a mobile phone stolen in the last 12 months and, if so, details of the incident.

Data from these questions should be analysed using the household weights to generate victimisation rates per household. These are discussed in greater detail in Chapter 7. If one wanted to generate figures for victimisation rates per head of population, this can be done but involves generating a victimisation rate within a household (i.e. average number of mobile phone thefts per person in the household, including children) and generating sample weights equal to the household weights multiplied by the number of

persons in the household (rather than the number of adults) For victimisation rates for age groups the same process of averaging and generating new sample weights would have to be followed, but restricted to the age group of interest.

### **3.1.5 Module on performance of the Criminal Justice System and witnessing crime**

This short module asked all respondents about their confidence in the Criminal Justice System, and how good a job they thought the different elements of the Criminal Justice System (the courts, the Crown Prosecution Service, the police, the prison service, the probation service) are doing.

From October 2003 a new set of questions relating to witnessing crime were asked of all respondents. These had previously been asked only of respondents on follow-up C. In addition to asking respondents whether they had witnessed particular crimes or incidents in the last 5 years and the last 12 months respondents who had witnessed anything were also asked whether they had reported this to the police.

### **3.1.6 Follow-up modules**

Respondents were randomly allocated to one of four follow-up modules (see section 3.5. for how this was done). Core sample respondents were allocated equally to each follow-up, meaning that approximately 9,250 respondents were asked each module. Respondents who were part of the non-white boost sample were randomly allocated to only modules A or B, in a ratio of 3:1 (i.e. approximately 2,250 respondents were asked Follow-up A and approximately 750 respondents were asked Follow-up B).

#### ***3.1.6.1 Follow-up A: Contact with and attitudes towards the police***

In 2003-4, Follow-up A included questions that had been asked in previous years, as well as a number of new questions. Some of the topics covered in this module included:

- whether respondents knew anyone in the police or had any contact with the police
- whether they had been stopped by the police either in a vehicle or on foot
- if so, the reason for this and the nature of the contact
- attitudes to the police
- how effective respondents thought the police were at different things
- how different things (e.g. police on patrol, CCTV, Neighbourhood Watch, etc) influenced respondents feelings of personal safety; and
- who respondents thought the police should work with to tackle crime and make people feel safer

#### ***3.1.6.2 Follow-up B: Contact with and attitudes towards the Criminal Justice System***

Again, this module included questions that had been asked in previous years as well as new questions. Topics covered in this module included:

- priorities of the Criminal Justice System

- where people get information about the Criminal Justice System
- knowledge of sentencing practices
- what people felt the purpose of sentencing should be
- attitudes to sentencing policy, including what respondents thought sentences should be for particular crimes and what they thought they actually were.
- recent contact with different parts of the Criminal Justice System

### **3.1.6.3        *Follow-up C: Crime prevention and security and witness intimidation***

Topics covered in this module included:

- home security measures and reasons for lack of home security measures
- vehicle security measures and reasons for lack of such measures
- attitudes to vehicle security
- whether people were harassed or intimidated as a result of witnessing crimes

### **3.1.6.4        *Follow-up D: Ad hoc crime***

This follow-up module was broadly similar to the 2002-3 survey and contained quite a wide variety of questions. These included:

- concerns about crime and fear of crime
- attitudes to the local community and social cohesion
- involvement in community activity/ voluntary activities
- experience of anti-social behaviour
- vicarious victimisation

Questions on attitudes to the local community and social cohesion were also asked of all respondents on the non-white boost sample.

### **3.1.7    Technology crimes**

In 2002-3 a new module was added to the survey to focus on the area of technology crime. This module remained on the 2003-4 survey. The main areas covered by this module included:

- use of credit cards and worry or experience of credit card fraud
- attitudes and concerns about shopping on the Internet
- experiences of computer viruses, hacking, and accessing offensive or inappropriate material on the Internet.
- knowledge of measures to prevent these types of incidents

### **3.1.8    Anti-social behaviour module**

The major addition to the 2003-4 survey was the introduction of a new module on Anti-social behaviour. Questions on ‘problem’ behaviours in local areas have been on the

survey for many years but in 2003-4 the list of problems was expanded to include issues such as speeding traffic, fireworks, and uncontrolled dogs.

Five particular problems were identified as key areas, namely:

- Noisy neighbours or loud parties
- Teenagers hanging around on the streets
- Vandalism and graffiti
- People using or dealing drugs
- People being drunk or rowdy in public places

Respondents who perceived any of these issues as being a 'very' or 'fairly' big problem in their area were asked a series of follow up questions related to the problem. The structure of the module was similar to that of the Victim Forms, with respondents being asked about up to a maximum of three problems. As a result of feedback from interviewers concerning the length of the module the number of problems that a respondent was asked about was reduced to a maximum of two problems from October 2003.

The follow-up questions on each strand of anti-social behaviour included:

- The precise nature of the problem
- How respondents had got the impression that the problem existed
- Where in their local area the problem existed
- Who was responsible for the problem
- How the problem affected their quality of life
- The main causes of the problem
- Who was thought to be responsible for stopping or reducing the problem
- Whether respondents had changed their behaviour in any way because of the problem

In addition to questions about the five specific problems the module also consisted of a short module of questions about crime and disorder on public transport and another module on crime and disorder in town centres.

### 3.1.9 Demographics module

This section collected additional information on the respondent and the Household Reference Person (where this was not the same as the respondent). Questions included:

- general health, including smoking and drinking behaviour
- employment details<sup>12</sup>
- educational attainment and qualifications
- nationality, country of birth and religion (of respondent and HRP)
- housing tenure
- newspaper readership

---

<sup>12</sup> Where the respondent was not the Household Reference person occupation details were also collected about the HRP

- household income

### 3.1.10 Self – completion modules

The self-completion modules were only asked of respondents aged under 60 years of age. They were all presented as computer assisted self-completion (CASI) modules to ensure respondent confidentiality in answering these questions. The respondent was asked to follow the instructions on the screen of the laptop and enter their answers appropriately. Practice questions were included before the start of the self-completion module to give the interviewer an opportunity to show the respondent the different functions of the computer. If the respondent was unable or unwilling to complete the modules using the computer the interviewer could administer the self-completion.

Interviewer assistance and the presence of others while completing these modules was recorded by the interviewer (see Chapter 4).

#### 3.1.10.1 *Drug usage*

The module on drug use remained virtually unchanged from the previous survey. A total of 17 illicit drugs were asked about. Questions included:

- whether ever taken illegal drugs
- whether taken illegal drugs in last 12 months
- whether taken illegal drugs in last month
- age first started taking drugs
- frequency of drug use (16-24 year olds only)

#### 3.1.10.2 *Drinking*

In 2003-4, only a few questions on drinking were asked of all 16-30 year olds as part of the self-completion module. This module covered:

- frequency and volume of alcohol drunk in the last 12 months
- how often felt drunk in the last 12 months

## 3.2 Life event calendar

Due to the change in the reference period that was introduced to the 2001 survey, the Home Office wished to take the opportunity to try and improve the accuracy with which respondents recalled events.

Whenever respondents are asked to think about events that have happened during a particular time period there is likely to be a certain level of response error. Errors most salient to the BCS include:

- Respondents forgetting about more trivial incidents.
- Respondent not knowing about incidents, especially in relation to household crimes which might not have involved them personally.

- Respondents remembering an incident but placing it wrongly in time either by remembering an incident as happening earlier than it actually did (backwards telescoping) or later than it actually did (forward telescoping).
- Respondents deliberately concealing incidents, such as sexual assault, which they may not wish to reveal.

It was decided to try and address issues of recall by using a **life event calendar** on the survey. Such a calendar works by trying to place events or incidents in some sort of meaningful context for each respondent by building up a picture of events that have happened to them in the last year (e.g. birthdays, anniversaries, holidays, starting a new job, etc.) which are memorable to the respondent. Additionally, national dates such as Christmas, Easter, or Bank Holidays can be put on the calendar as common reference points.

Further details about the thinking behind the life event calendar and its development can be found in the 2001 BCS Technical Report.

In relation to the BCS, the life event calendar can be used for two purposes:

- First, to provide respondents with a visual aid throughout the screener questions; and
- Second, to help respondents who were having difficulty recalling in which particular month an incident may have occurred.

Appendix E in Volume 2 has an example of the calendar used on the 2003-4 survey.

### **3.3 Questionnaire development**

Since most of the questions on the 2003-4 BCS had been included in previous years of the survey, it was decided to concentrate piloting efforts only on new questions. Details of the questions that were piloted for the 2003-4 survey can be found in Appendix J of Volume 2. The main areas covered in the 2003-4 piloting were:

- Questions on reassurance and policing priorities
- Questions about the Criminal Justice System
- Anti-social behaviour questions

#### **3.3.1 Main stage piloting**

Piloting of the main parts of the questionnaire was done using **dynamic piloting**. This is a form of piloting which is regularly used by BMRB to develop questionnaires across many different surveys. Dynamic piloting uses cognitive interviewing techniques to try and understand the thought processes that a respondent uses in answering a survey question. It is designed to see whether the respondent understands the question, or specific words and phrases contained within the question; what sort of information the respondent needs to retrieve in order to answer the question; and what decision processes the respondent uses in coming to an answer.

### **3.3.1.1 Dynamic piloting methodology**

Dynamic piloting involves a small number of highly experienced interviewers working together with researchers in a field setting. The interviewer fulfils his or her conventional role, while the researcher observes the interview at first hand. This enables the observer to identify and note areas of doubt, misunderstanding or incomprehension on the part of the respondent during the interview. By observing the interview as it takes place, the researcher is able to witness not only the verbal communication that takes place, but also any non-verbal reactions of the respondent. At the end of the interview or immediately following questions or sections of particular interest, the researcher probes on specific aspects of the interview. Although each observer had a standard list of probes that are developed prior to the pilot, the method is flexible enough to allow observers to carry out spontaneous probing based on what happens during each individual interview.

A small scale and intensive dynamic pilot of this nature is preferable to a larger scale exercise where the interviewers are unaccompanied. In the latter instance, interviewers tend to report back on how they rather than the respondents found the questionnaire.

With all dynamic pilots the interviewers and researchers are thoroughly briefed before starting the pilot and an informal debrief is held at the end of the pilot.

### **3.3.1.2 Structure of piloting on the BCS**

Dynamic piloting on the BCS was carried out in two rounds during February 2003. This approach allowed the preliminary findings from the first round of piloting to be considered by researchers and the Home Office and amendments made to specific questions where necessary before round two. In total four interviewers and four researchers worked on each round of the pilot and a total of 40 interviews were conducted over the two rounds of piloting. Respondents were pre-recruited by interviewers and all respondents were paid a small incentive. Quotas were set to ensure that at least half of the respondents had experienced some sort of crime in the previous twelve months. To make it easier to find such respondents, all the areas used for the pilot were high crime areas.

Interviews lasted approximately 30 minutes, with a further 20-30 minutes of probing by the observer.

## **3.4 Final questionnaire and revisions**

Following feedback from the piloting and detailed analysis of the timings, further modifications were made to the questionnaire to bring the length of the questionnaire in line with previous surveys.

A paper questionnaire was produced from the Quanquest software that detailed the questions and their routing instructions as specified in the Quanquest code. This was translated into a Word document to provide a more user-friendly questionnaire.

Once all changes had been approved the questionnaire was thoroughly checked by BMRB researchers and Home Office research staff.

### 3.5 Allocation of sample within CAPI

In the 2003-4 survey the unique serial number entered by interviewers into the computer had to be capable of the following:

- to randomly allocate respondents to one of four follow-up modules (and within each module to further allocate respondents into a sub-sample).
- to distinguish between a core sample respondent, a non-white boost sample respondent and a youth boost respondent.

The unique serial number pre-printed on all core Address Contact Sheets and transferred by interviewers into the CAPI consisted of 6 digits. The first 4 digits (1000-9999) represented the area or sample point number and the last 2 digits (01-99) represented the address number. Additionally, the interviewers had to enter a screen number which denoted whether the interview was a core sample interview (screen number 0) or a non-white boost sample interview (screen number 1-6). Various checks were incorporated into the questionnaire to minimise the chances of errors being made by interviewers when entering the serial and screen numbers.

Allocation of respondents to each follow-up module was done on the basis of the address number, using an algorithm based on division of the address number by 8 as shown in Table 3.2. For non-white respondents, who were allocated only between Follow-up modules A and B in the ratio of 3:1, the algorithm was amended to ensure a similar random allocation of respondents.

Since each sample point contained either 16 or 32 addresses the above algorithm ensured that within each sample point a similar number of issued addresses were randomly allocated to each follow-up module.

**Table 3.2 Allocation of interviews to follow-up modules**

| Address Number | Remainder divided by 8 | Follow-up module |                        |
|----------------|------------------------|------------------|------------------------|
|                |                        | Core sample      | Non-white boost sample |
| 01,09,17,etc.  | 1                      | A1               | A1                     |
| 02,10,18,etc.  | 2                      | B1               | B1                     |
| 03,11,19,etc.  | 3                      | C1               | A1                     |
| 04,12,20,etc.  | 4                      | D1               | A1                     |
| 05,13,21,etc.  | 5                      | A2               | A2                     |
| 06,14,22,etc.  | 6                      | B2               | B2                     |
| 07,15,23,etc.  | 7                      | C2               | A2                     |
| 08,16,24,etc.  | 8                      | D2               | A2                     |

### 3.6 Features of Quancept used in the BCS

#### 3.6.1 Don't Know and Refusal Keys

In the Quancept script, Don't Know and Refused are special codes. Rather than entering numeric codes for these options, interviewers enter DK and REF respectively. As with previous years of the survey, almost every question had a Don't Know and

Refused option that the interviewer could use. However, at most questions they were hidden, and so did not appear on the screen as an explicit option. In the paper questionnaire in Appendix D, Don't Know and Refused are only shown if they actually appeared as an option on the screen.

### 3.6.2 Different question types

The vast majority of questions were pre-coded, meaning that a list of answer categories appears on the laptop screen and the interviewers enter the appropriate numeric code. Questions were either single response (i.e. only one code can be entered) or multi-response (i.e. more than one code can be entered). In the latter case, answers are entered separated by spaces. In multi-response questions it is possible to allow a combination of either multi-response or single response options at the same question. In the case of numeric questions, where an actual value is required, the interviewer simply types in the appropriate number.

Many pre-coded questions had an 'Other –specify' option, and if this option was selected by a respondent, the interviewer would simply type in the answer given. In all these questions, the answers were later examined by coders to see if the other answer could be back coded into one of the original pre-coded options (see section 5.2).

In Quancept the standard keys that interviewers use to move forwards and backwards through the questionnaire are *Ctrl + Enter* and *Ctrl + Backspace* respectively. It was felt that these keystroke combinations might be awkward for respondents when completing the self-completion part of the questionnaire. Consequently, a modified version of the software was used for the BCS which allowed respondents to use single keystrokes (F2 for forward, F1 for backward) to complete the self-completion.

### 3.6.3 Logic and consistency checks

A number of logic and consistency checks were built into the Quancept script. These were of two types: hard checks and soft checks. Hard checks are ones where the interviewer is unable to move to the next question until the discrepancy or inconsistency has been resolved. Soft checks are ones where the interviewer is asked to confirm that the information entered at a specific question is correct but is able to pass on to the next question.

A full list of all the logic and consistency checks in the 2003-4 script can be found in Appendix I of Volume 2.

### 3.6.4 Date calculation and text substitution

Text substitution and date calculations were used extensively throughout the questionnaire.

**Text substitution** is where alternative text is used in a question depending upon the series of answers given by a respondent to previous questions. In the paper questionnaire, square brackets are used to denote the existence of text substitution in a question.

Two main types of **date calculations** were used in the questionnaire:

- First, the precise reference period was calculated based on the date of interview and this was then substituted into the text of many questions. In all cases it was decided to calculate the date to the first of the month 12 months previous. Thus, for example, any interviews conducted in July 2002 would use the reference period “*since the first of July 2001*”. This means that in practice the 12 month reference period consisted of the last 12 full calendar months, plus the current month (i.e. slightly more than 12 calendar months). This fact is taken into account when the victimisation rates are being estimated.
- Second, some code frames consisted of particular time periods (e.g. months or quarters) which changed on a month by month basis. With these type of questions the Quancept script was programmed to allow the whole reference period covered by the questionnaire (that is, from April 2002 to June 2004 – a total of 27 months). However, interviewers only saw on screen the sub-set of codes that were appropriate to the correct reference period (i.e. 13 calendar months) for the month they were interviewing in.

Since some questions use these constantly rotating code frames based upon date of interview it is impossible to label these variables in any meaningful way in the SPSS data file. A list of these questions and the appropriate code frames that actually appeared on screen depending upon the month of interview can be found in Appendix F of Volume 2.

## 4. Fieldwork

This chapter documents all aspects of the data collection process, focusing on fieldwork procedures, the management of fieldwork across the survey year, quality control procedures and response rates achieved across the different samples.

### 4.1 Briefing of interviewers

All interviewers working on the 2003-4 survey attended one of two types of briefings during the year:

All interviewers who had not previously done a BCS assignment were required to attend a full day face-to-face briefing before they could work on the survey. These briefings were held throughout 2003-4 as required. In total, 11 full briefings of new interviewers were held and 130 interviewers were briefed. All briefings were attended by researchers and field staff working on the survey, and many were also attended by Home Office researchers.

Each briefing covered the following topics:

- Background to the BCS and how the information is used by the Home Office
- Details about sampling and fieldwork procedures and techniques on how to obtain high response rates
- Explanation of the screening procedures used on the non-white boost sample
- Details of certain questions and topics covered in the interview. The primary focus of the briefing was on how to ensure that information from the screener questions and the Victim Form are collected correctly. Additionally, how to approach the self-completion sections of the interview was covered.

In addition to a face-to-face briefing, before starting a BCS assignment for the first time all interviewers had to watch a video briefing that provided further details about the procedures used for the youth boost sample. Video briefings are recorded on CD and interviewers are able to play them on their laptop computers. Interviewers were also required to read the written Interviewer Instructions and carry out at least two practice interviews based on particular scenarios provided in the Instructions.

Interviewers who had already been briefed on BCS were required to attend a refresher briefing. In total 26 half-day refresher briefings were held in July and August 2003 and 320 interviewers attended these events. Since the survey had only changed slightly between 2002-3 and 2003-4 the objectives of these meetings were different from the initial briefings. Although they did provide an opportunity to tell interviewers about changes to the questionnaire, the primary objectives were to provide feedback to interviewers on how the survey had gone in 2002-3; to highlight areas of field procedures and data collection that could be improved; and to generally re-motivate interviewers, many of whom had been working on the survey for a relatively long time. In addition, the review meetings gave interviewers a chance to provide feedback on the survey and fieldwork procedures.

All refresher briefings were attended by researchers and field staff working on the survey, and most were attended by Home Office researchers. Each briefing covered the following topics:

- Details on the 2002-3 BCS survey results and how these had been reported in the media
- Analysis of 2002-3 response rates and discussion about how response rates could be improved
- Discussion of the questionnaire and particular areas of the questionnaire that interviewers found difficult or problematic. In 2003-4 the main focus of the review meetings was on the new Anti Social Behaviour module that was introduced at the start of the survey year.
- Discussion about field issues and other information about future developments.

## **4.2 Supervision and quality control**

Several methods were used to ensure the quality and validity of the data collection operation.

A proportion of interviewers, particularly those less experienced, were accompanied in the field by supervisors. This included interviewers who were new to random probability, who were accompanied on the first day of a BCS assignment by a supervisor.

A proportion of addresses were re-contacted, to verify that the interviewer had contacted someone at the address, whether or not an interview had resulted. In total, 5,889 respondents were re-contacted to verify that the interviewer had contacted someone, and whether or not an interview had resulted. These back checking procedures were mainly carried out by telephone.

## **4.3 Fieldwork dates and fieldwork management**

During 2003-4 the survey was managed on a monthly basis. Approximately 150 assignments were issued to interviewers at the start of each month, with fieldwork normally starting on the first Monday of the new month.

Interviewers were encouraged to start their assignment as early as possible in the month to minimise the time between respondents receiving the advance letter and an interviewer calling (see section 4.4.1). Interviewers had until the end of the calendar month to cover all the addresses in their assignment.

Once all the issued addresses had been covered the Address Contact Sheets were returned to Head Office and a decision was taken about re-issuing non-productive outcomes. As a general rule all non-productive addresses (non-contacts, refusals, broken appointments, etc.) were reissued unless there was a specific reason not to or it would not have been cost effective (e.g. only one address in an assignment). Re-issued addresses were generally given to a supervisor or senior interviewer. Once the first reissue period had been completed a decision was taken about whether to re-issue addresses that were still non-productive for a second or third time.

In total, 12,468 addresses were re-issued on the core sample, which represented 23% of the original sample. A further 11% of addresses were issued for a second time, and 4%

of addresses were issued for a third time. Of all the addresses re-issued, 28% were converted into productive outcomes. Addresses where the original outcome had been a refusal were less likely to be converted than those that had been a non-contact or some other unproductive outcome (e.g. broken appointment, away, etc.). Overall, the impact of the re-issue process was to increase the response rate on the core sample from 68.0% after the initial issue to the final response rate of 74.1% (see section 4.7.1).

The nature of the re-issue system means that each fieldwork period lasted approximately 3 months. This meant that for sample issued in January, fieldwork was carried out until the end of March, while sample issued in February was in the field until the end of April, and so on. Due to the data delivery requirements there was a non-moveable cut-off point at the end of each quarter by which fieldwork for the preceding quarter had to be finished.

Because of this time lag between addresses being issued and interviews being achieved, the time period covered by the 2003-4 **issued sample** and the time period covered by the 2003-4 **achieved sample** are different. Although sample for the survey was issued between April 2003 and March 2004, the actual fieldwork dates over which interviews were achieved ran from April 2003 to June 2004. At the quarterly level, this means that not all interviews were actually achieved in the quarter of issue. In fact, approximately 80% of interviews were achieved in the same quarter as they were issued, with 20% of interviews falling into the next quarter. Not surprisingly, most of the interviews that fell into the following quarter were those issued in the last month of a quarter (i.e. March, June, September, and December).

Details about how the issued sample and achieved sample relate to the data output can be found in section 6.2.

#### **4.4 Fieldwork procedures and documents**

In most cases an interviewer assignment consisted of 32 issued addresses. In areas where the proportion of the non-white population was greater than 20%, interviewers were only issued with 16 addresses as it was anticipated that they would achieve a higher number of non-white boost interviews. Interviewers were only allowed to conduct interviews at the addresses they were issued with.

##### **4.4.1 Advance letter and leaflet**

All selected addresses were sent a letter from the Home Office in advance of an interviewer calling at the address. For addresses in Wales, a Welsh translation was provided on the reverse of the letter. This explained a little about the survey, why this particular address had been selected and telling the occupiers that an interviewer from BMRB would be calling in the next few weeks. The letter also provided a telephone number and an email address for people to contact to find out more about the survey, to make an appointment for an interviewer to call, or to opt out of the survey. Over the course of the whole year only 669 people, representing less than 2% of addresses issued, opted out of the survey by contacting either BMRB or the Home Office.

In addition to the advance letter for the core sample there was also a similar letter for the non-white boost sample and the youth sample. Since the boost samples were not pre-identified, these letters were given to potential respondents by interviewers once they had identified a potentially eligible household.

Included with the advance letter was a coloured leaflet from the Home Office which provided people with some more details about the survey, including findings from the previous survey. The leaflet also tried to answer some questions that potential respondents might have such as issues relating to confidentiality.

Examples of the advance letters used on the core, non-white and youth boost samples can be found in Appendix A of Volume 2 and a copy of the leaflet can be found in Appendix B of Volume 2.

#### 4.4.2 Address Contact Sheets (ACS)

Interviewers were issued with an Address Contact Sheet (ACS) for each sampled address. This is the key document that allows interviewers to carry out the different tasks that make up the BCS assignment and to record and manage their own calling strategies for each address. In total there are four different types of Address Contact Sheet used on the BCS, all of which are colour coded to avoid confusion. These are:

- **Yellow Address Contact Sheet** – A yellow ACS indicates a pre-identified core sample address where screening at adjacent addresses for non-white respondents is needed. All core sample Contact Sheets have the full address printed on the front page, as well as details of the serial number for that address.
- **White Address Contact Sheet** – A white ACS is identical to the yellow ACS, except that it indicates that no screening of adjacent addresses is needed. As such, it is two pages shorter than the yellow ACS.
- **Blue Address Contact Sheet** – A blue ACS is created by interviewers once they have identified a potentially eligible non-white household. Since these addresses are not pre-identified, the blue ACS is blank and interviewers have to fill in the address and serial number details themselves.
- **Pink Address Contact Sheet** – A pink ACS is created by interviewers once they have identified a potentially eligible youth respondent. As with the blue ACS, the pink ACS is blank and interviewers have to fill in the address and serial number details themselves.

The Address Contact Sheets are crucial documents to the management of the BCS, both at the level of the individual assignment and for the management of the survey overall. The primary functions of the Address Contact Sheet are as follows:

- To allow interviewers to record the days and times that they have called at an address. Additionally, there is space for interviewers to record details or comments that may be useful should the address be re-issued to another interviewer.
- To provide a record of all the outcomes achieved at the address. The ACS allows the outcome at each re-issue stage to be recorded separately, so that there is a complete record of outcomes for each address. Although these outcomes are recorded by interviewers on the ACS, they are also reported electronically to Head Office on a daily basis so that overall progress can be monitored and managed.
- To allow the interviewer to carry out any selection procedures where necessary. Where an interviewer finds more than one dwelling unit at an address they need to carry out a procedure to randomly select one dwelling unit for interview. Similarly, where more than one eligible respondent exists, interviewers need to randomly select

one person for interview. The ACS allows them to carry out this procedure and record the details for future reference or checking.

- To allow the interviewer to carry out the screening process for both the non-white and youth boost samples. The ACS has step by step instructions for interviewers about how to carry out these procedures and also allows them to record the screening outcomes for every address. As with the final response outcomes, all screening outcomes are reported back to Head Office on a daily basis.
- To collect some basic information about the area and the selected address (e.g. type of property, condition of the property, whether it is in a Neighbourhood Watch area, etc.). This information is collected by interviewers based on their own observations and, as such, is highly subjective. Nevertheless, such information does tend to be highly associated with non-response and is also used by the Home Office as an area based disorder measure. This observational data is recorded by interviewers on the back page of the ACS and returned to Head Office where the information is then scanned. The data is then added to the main data files at a later stage.

Examples of both core sample Address Contact Sheets and the non-white and youth Boost Sample ACS can be found in Appendix C of Volume 2.

#### **4.5 Presence of others during the interview**

In briefing interviewers, emphasis was placed on trying to conduct the interview in private. It was felt that this would not only make the interview run more smoothly, it might also prevent some respondents from not mentioning certain incidents, which they might be embarrassed or worried of talking about in front of others.

This was a particular concern for respondents who had experienced domestic violence or sexual assault. Where respondents had experienced such incidents in the last 12 months interviewers had the option of suspending the Victim Form (simply by skipping over it) if they felt it was inappropriate to continue with the questions due to the presence of others in the room. This procedure meant that the interviewer could complete the rest of the interview. During 2003-4 a total of 18 Victim Forms were suspended by interviewers for this reason.

Despite such efforts, it is recognised that in some situations the presence of others can improve the accuracy of the information collected. This is particularly the case in incidents of vehicle crime or property crime, where the respondent may not have been personally present, reported the incident to the police, etc. Moreover, in some cases it will simply not be possible for the interview to be conducted without others present in the room.

##### **4.5.1 Presence of others during the screener interview**

The key point at which the presence of another person could affect the estimate of victimisation is during the initial set of screener questions. Therefore, at the end of these questions, the interviewer recorded whether anyone else was present. Table 4.1 shows whether anyone else was present or not during the initial screener part of the questionnaire, when respondents are giving details about their experiences of crime.

**Table 4.1 Whether anyone else was present or not during the screener questionnaire**

|                                 | <b>Core sample</b> | <b>Non-white boost sample</b> | <b>Youth boost sample</b> | <b>Total</b>  |
|---------------------------------|--------------------|-------------------------------|---------------------------|---------------|
|                                 | <b>%</b>           | <b>%</b>                      | <b>%</b>                  | <b>%</b>      |
| No-one present                  | 71                 | 59                            | 68                        | 70            |
| Children under 16               | 8                  | 19                            | 5                         | 9             |
| Spouse/partner                  | 16                 | 16                            | 3                         | 15            |
| Other adult                     | 8                  | 16                            | 27                        | 9             |
| <i>Base: (All)<sup>13</sup></i> | <i>37,213</i>      | <i>3,318</i>                  | <i>2,294</i>              | <i>42,285</i> |

In seven out of ten interviews (70%), the respondent was interviewed without anyone else other than the interviewer being present. Respondents interviewed as part of the non-white boost sample were more than twice as likely as respondents from the core sample to be interviewed with children under 16 present (19% and 8% respectively) and with other adults present (16% and 8% respectively). Respondents interviewed on the youth boost sample were more likely than other respondents to have been interviewed with some other adult present, possibly a parent

Looking at the figures by sex and age shows that on the core sample male and female respondents were just as likely to have done the interview with no-one else being present (72% of men and 70% of women). However, on the non-white boost sample female respondents were significantly less likely than male respondents to have done the screener questionnaire with no-one else present (64% of men and 55% of women).

Respondents aged 16 to 24 were less likely than average (64%) to have done the screener questionnaire with no-one else present, with more than a quarter (26%) having done the interview with some other adult present. Interestingly, 16 to 24 year olds who were interviewed on the youth boost sample were more likely to have done the interview with no-one else present compared with 16 to 24 year olds interviewed on the core sample (68% and 62% respectively). This may reflect the nature of the methodology and the fact that a youth boost interview is generally done after a core sample interview has already taken place.

#### 4.5.2 Presence of others during the self-completion

For those who did the self-completion, the presence of others during this part of the interview was also recorded. Table 4.2 shows that the proportion of people who did the self-completion with no-one else in the room was similar to those who did the screener questionnaire with no-one else in the room. In over seven out of ten (73%) interviews the self-completion was done with no-one other than the respondent and interviewer present. Respondents from the core sample were more likely than respondents from the non-white boost sample to have completed the self-completion with no-one else present (74% and 65% respectively).

<sup>13</sup> Percentages add up to more than 100% since more than one answer could be coded at this question

**Table 4.2 Whether anyone else was present or not during the self-completion questionnaire**

|  | <b>Core sample</b> | <b>Non-white boost sample</b> | <b>Youth boost sample</b> | <b>Total</b>  |
|--|--------------------|-------------------------------|---------------------------|---------------|
|  | <b>%</b>           | <b>%</b>                      | <b>%</b>                  | <b>%</b>      |
| No-one else  | 74                 | 65                            | 70                        | 73            |
| Children under 16  | 8                  | 15                            | 5                         | 9             |
| Spouse/partner   | 12                 | 12                            | 3                         | 11            |
| Other adult  | 6                  | 12                            | 23                        | 8             |
| <i>Base (all who completed self-completion)<sup>14</sup></i> | <i>23,931</i>      | <i>2,588</i>                  | <i>2,278</i>              | <i>28,797</i> |

Where anyone else was present during the self-completion section, interviewers were briefed to try and ‘arrange’ the room whenever possible so that the respondent had a degree of privacy to do the self-completion. Thus, for example, interviewers might try to ensure that the respondent is sitting with the screen facing a wall or in such a position that no-one else in the room could actually read the computer screen.

Where anyone else was present, the extent to which they were involved in answering questions was noted as was whether the interviewer was involved in the self-completion sections.

Even where someone else was present during the self-completion, it was not common for others to become involved in answering the questions. In 88% of interviews where someone else was present, the respondent completed the self-completion section entirely on their own. In 6% of interviews someone else actually looked at or read the self-completion with the respondent, while in another 6% of cases the respondent discussed the self-completion with other people. Respondents from the non-white boost sample were more likely than average to have had someone else look at or read the self-completion (12%) or to have discussed the self-completion with someone else (10%).

Table 4.3 shows the amount of assistance that interviewers gave to respondents on the self-completion section. This shows that almost one in eight (12%) respondents who did the self-completion asked the interviewer to complete the self-completion for them. Respondents from the non-white boost sample were more likely to choose this option than respondents from the core sample (23% and 12% respectively).

However, the vast majority of respondents (84%) who completed the self-completion did it entirely themselves with no help from the interviewer. Respondents from the non-white boost sample were less likely than those from the core sample to have done it without any assistance (69% and 85% respectively). Respondents from the youth boost sample were the most likely to have done the self-completion themselves and without any assistance (94%).

At the end of the interview, interviewers were asked to record their assessment of whether the respondent had had any difficulties reading (English) during the interview.

<sup>14</sup> Percentages add up to more than 100% since more than one answer could be coded.

Not surprisingly, where those who had asked the interviewer to do the self-completion for them or had needed help with nearly all the questions were more likely to have difficulties reading English than those who had needed little or no assistance. Thus, over half (51%) of respondents who the interviewer assessed as having difficulty with reading English either asked the interviewer to do the self-completion for them or needed significant help with the questions.

**Table 4.3 Amount of assistance given by interviewer with self-completion questionnaire**

|  | Core sample   | Non-white boost sample | Youth boost sample | Total         |
|--|---------------|------------------------|--------------------|---------------|
|  | %             | %                      | %                  | %             |
| All done by respondent   | 85            | 69                     | 94                 | 84            |
| Help given with one or two questions                           | 2             | 4                      | 1                  | 2             |
| Help given with more than one or two questions, less than half | <1            | 1                      | <1                 | <1            |
| Help given with more than half, but not all                    | <1            | <1                     | <1                 | <1            |
| Help given with all/nearly all                                 | 1             | 3                      | 1                  | 1             |
| Completed by interviewer                                       | 12            | 23                     | 4                  | 12            |
| <i>Base (all who completed self-completion)</i>                | <i>23,931</i> | <i>2,588</i>           | <i>2,278</i>       | <i>28,797</i> |

## 4.6 Length of interview

### 4.6.1 Introduction

Timing stamps were placed throughout the questionnaire to allow timing of individual sections. Due to various technical issues associated with CAPI systems, it is not always possible to derive meaningful time stamps from every interview. For example, should an interviewer briefly go back into an interview at a later time to check or amend a response the time stamps can be set to show an apparently very short (2-3 minutes) interview. Similarly, if an interviewer has to temporarily stop or suspend an interview for an hour or so and fails to come out of the questionnaire in the intervening period (simply powering down the computer instead) the time stamps can show an interview of 4-5 hours.

To eliminate the effects of these outlying cases on the calculation of average timings, it was decided to only include interviews where the total length of interview was in the range 15 minutes to 180 minutes. On the 2003-4 survey, 95% of interviews (40,650 out of 42,812) had a valid time within these ranges and are included in the analysis below.

Since the calculation of interview times is based on time stamps generated within the interview (rather than an interviewer estimate), they represent the elapsed time from the first question to the last question. As such they do **not** include the time during which the interviewer is introducing the survey, setting up the laptop, or packing up at the end of the interview.

#### 4.6.2 Overall length of interview

The average (mean) length of the interview during 2003-4 was **47 minutes**. The main influence on length was whether or not the respondent was a victim of crime or not. The average interview length for **non-victims** was **40 minutes** compared to **63 minutes** for **victims**. The average length of core sample interviews was **47 minutes**, while the average length of the non-white boost sample interviews was **48 minutes**. The average length of the youth boost sample interviews was **23 minutes**<sup>15</sup>.

The average length of interview by number of Victim Forms is shown in Table 4.4 below. This shows that the length of interview was strongly related to the number of Victim Forms completed by the respondent, with those completing 4 or more Victim Forms having an average interview length of 94 minutes.

**Table 4.4 Length of interview by number of Victim Forms (Core and non-white boost sample)**

| <b>Number of Victim Forms</b> | <b>Average time (minutes)</b> |
|-------------------------------|-------------------------------|
| <b>Non victims</b>            | <b>40</b>                     |
| <b>All victims</b>            | <b>63</b>                     |
| 1                             | 56                            |
| 2                             | 70                            |
| 3                             | 84                            |
| 4 or more                     | 94                            |
| <b>All respondents</b>        | <b>47</b>                     |

The most common interview length was between 30 and 60 minutes, with 56% of all respondents completing the survey in this time. About a quarter (24%) of respondents completed the survey in under 30 minutes, while 5% of respondents took over 90 minutes.

The average interview length of the 2003-4 survey was about 2 minutes longer than the 2002-3 survey, although of a similar length to the 2001-2 survey. The increase in length was due primarily to the addition of the new module on Anti Social Behaviour.

Since the module was only asked of half the sample, this meant that there was a noticeable difference in average interview length depending upon the half of the sample a respondent was in. Respondents who were asked Follow-up A or Follow-up C plus the Technology Module had an average interview length of 42 minutes compared to an average length of 51 minutes for respondents who were asked Follow-up B or Follow-up D plus the Anti Social Behaviour Module. To try and better equalise questionnaire length across the sample the length of the Anti Social Behaviour Module was reduced slightly after 6 months (see section 3.1.8)

---

<sup>15</sup> Youth boost interviews are NOT included in the rest of the analysis since they would act to lower overall average times since the youth boost interview is a cut down version of the main questionnaire.

### 4.6.3 Average time for different sections of the interview.

The average times for each of the main sections of the questionnaire are shown below in Table 4.5.

**Table 4.5 Average time for each module of the 2003-4 questionnaire (Core and non-white boost sample)**

| Questionnaire module             | Average time (minutes) |
|----------------------------------|------------------------|
| Household Box                    | 5.4                    |
| Main questionnaire               | 6.4                    |
| Screener questions               | 3.1                    |
| Victim forms <sup>16</sup>       | 5.2                    |
| Mobile phones                    | 0.6                    |
| Criminal Justice System          | 4.0                    |
| Follow-up module <sup>17</sup>   | 6.7                    |
| Technology                       | 1.5                    |
| Anti Social Behaviour            | 4.4                    |
| Demographics                     | 5.8                    |
| Drugs and Drinking <sup>18</sup> | 2.5                    |
| End of interview admin.          | 1.3                    |
| <b>Average time</b>              | <b>47</b>              |

### 4.6.4 Length of victim forms

As mentioned above the average length of the questionnaire is affected primarily by the number of Victim Forms completed by a respondent with the average time for non-victims being 40 minutes compared to an average of 63 minutes for victims of crime.

Although the average time taken to complete the Victim Forms was only 5.2 minutes, this time is skewed by the fact that non-victims take no time to complete this section. Therefore, a more meaningful timing is the average time taken to complete each Victim Form by those who completed each one.

Table 4.6 shows that long Victim Forms averaged about 11.6 minutes per form, while short Victim Forms averaged 4.5 minutes per form. Table 4.6 also shows that the time taken to complete the first Victim Form was greater than for forms two or three, suggesting that respondents speed up as they go through each subsequent Victim Form. This pattern has been evident in all previous surveys.

---

<sup>16</sup> This is an average time across all respondents, with non-victims having a time of 0.

<sup>17</sup> This represents an average time across all the follow-up modules.

<sup>18</sup> This is an average time across all respondents. Those who either refused the self-completion or were not eligible (those aged 60 or over) would have a time of 0

**Table 4.6 Average time taken for each Victim Form (Core and non-white boost sample)**

| <b>Victim Form number</b> | <b>Average time (minutes)</b> |
|---------------------------|-------------------------------|
| Victim Form 1             | 13.9                          |
| Victim Form 2             | 11.2                          |
| Victim Form 3             | 9.8                           |
| Victim Form 4             | 4.7                           |
| Victim Form 5             | 4.3                           |
| Victim Form 6             | 4.5                           |

#### 4.6.5 Length of follow-up modules

Table 4.7 below shows the average time taken for each of the follow up modules based only on those who actually answered the module.

**Table 4.7 Average length of each follow up module (Core and non-white boost sample)**

| <b>Follow-up module</b>      | <b>Average time (minutes)</b> |
|------------------------------|-------------------------------|
| Follow-up A                  | 6.1                           |
| Follow-up B                  | 8.6                           |
| Follow-up C                  | 4.9                           |
| Follow-up D                  | 6.6                           |
| Technology Module            | 2.9                           |
| Anti Social Behaviour Module | 10.7                          |

Similarly, the overall timings of the self-completion are masked by the fact that all those who are not eligible for the self-completion (i.e. those aged 60 or over) and those who refuse the self-completion have an average time of zero. Considering only those respondents who actually did the self-completion sections the average time of the Drugs and Drinking module was 3.9 minutes. In fact, 75% of respondents who did the self-completion did it in less than 5 minutes, while around a quarter (23%) took 5-10 minutes. Only 2% of respondents took more than 10 minutes to do the self-completion.

The time of the Anti Social Behaviour Module was directly related to the number of strands of Anti Social Behaviour that respondents perceived to be a problem in their area. For those who perceived no problems in their area the module took, on average, 4 minutes compared to an average of 26 minutes for those who reported 3 or more problems. To try and reduce the burden on individual respondents the maximum number of problems that any one person could be asked about was reduced from three to two in October 2004. This reduced the average length of the Anti Social Behaviour Module by 2 minutes, from 11.7 minutes in the first half of the year to 9.7 minutes in the second half of the year.

## **4.7 Response rate and reasons for non-response: core sample**

### **4.7.1 Overall core response rates**

The full response rate analysis for the 2003-4 issued core sample is shown in Table 4.8. Overall, the response rate and non-response breakdown on the 2003-4 survey was extremely similar to the 2002-3 survey.

Around one in twelve (8%) issued core addresses were identified as not being eligible residential addresses (or deadwood). The most common types of deadwood were empty or vacant residential properties (4.5%) and business or industrial premises (1.1%). The total proportion of addresses that were identified as being deadwood was similar to that identified in the previous two surveys.

At eligible addresses the most common reason for not getting an interview was because of a refusal, which accounted for 15% of all issued eligible addresses. The most common type of refusal was after the person selection had been made, accounting for 9% of the eligible sample compared to 5% of addresses where all information was refused or someone had contacted Head Office to refuse.

Non-contact formed a much smaller proportion of unproductive addresses, accounting for 5% of all issued eligible addresses. The vast majority of such cases represented situations where no contact had been made with anyone at the address despite repeated calls over a lengthy period of time. It is possible that some of these addresses were actually empty or vacant and so should have been coded as deadwood. However, the impact that this would have on the overall response rate is likely to be minimal

Finally, a further 6% of issued eligible addresses were categorised as unproductive for other reasons including broken appointments, people who were ill or away during the period of the survey, and people who had inadequate English to complete the survey.

**Table 4.8 2003-4 Core sample response rate and non-response breakdown**

|  | N             | %           |
|--|---------------|-------------|
| <b>Total addresses issued</b>            | 54,736        | 100         |
| Addresses not traced                     | 534           | 1.0         |
| Not built/ does not exist                | 62            | 0.1         |
| Derelict/ demolished                     | 290           | 0.5         |
| Empty/vacant                             | 2,437         | 4.5         |
| Second home/not main residence           | 375           | 0.7         |
| Business/ industrial                     | 587           | 1.1         |
| Institution/communal establishment       | 109           | 0.2         |
| Other deadwood                           | 139           | 0.3         |
|  |               |             |
| <b>Total ineligible addresses</b>        | <b>4,533</b>  | <b>8.3</b>  |
|  |               |             |
| <b>Total eligible addresses issued</b>   | 50,203        | 100         |
|  |               |             |
| No contact made with household           | 2,450         | 4.9         |
| No contact made with selected respondent | 276           | 0.5         |
| <b>Total non contact</b>                 | <b>2,726</b>  | <b>5.4</b>  |
|  |               |             |
| Office refusal                           | 669           | 1.3         |
| Refused all information                  | 1,700         | 3.4         |
| Personal refusal                         | 4,409         | 8.8         |
| Proxy refusal                            | 628           | 1.3         |
| <b>Total refusal</b>                     | <b>7,406</b>  | <b>14.8</b> |
|  |               |             |
| Broken appointment                       | 741           | 1.5         |
| Temporarily ill/incapacitated            | 257           | 0.5         |
| Physically or mentally unable            | 484           | 1.0         |
| Away/ in hospital                        | 632           | 1.3         |
| Inadequate English                       | 189           | 0.4         |
| Other unsuccessful                       | 555           | 1.1         |
| <b>Total other unsuccessful</b>          | <b>2,858</b>  | <b>5.7</b>  |
|  |               |             |
| <b>Total unproductive</b>                | <b>12,990</b> | <b>25.9</b> |
|  |               |             |
| <b>Achieved interviews</b>               | <b>37,213</b> | <b>74.1</b> |

#### 4.7.2 Core response rates by Government Office Region

Response rates differ by area. In particular, response rates across all surveys tend to be lower in inner city areas compared with suburban and rural areas. This is reflected in the fact that the current BCS data includes a weight to correct for differential response rates between those areas defined as inner city and non-inner city (see section 7).

Using this same definition, in the 2003-4 survey the response rate in areas categorised as inner city was 69% compared to 75% in those areas categorised as non-inner city. It is

interesting to note that refusal rates in inner city and non-inner city areas were the same (14% and 15% respectively) and that the differential non response was almost entirely due to the non-contact rate in inner city areas being almost twice that compared with non-inner city areas (9% and 5% respectively).

Response rates also differ by region. Table 4.9 below shows the different response rates achieved by Government Office Region. This shows that across all regions the response to the survey was broadly the same, except for London where it was noticeably lower than average at 61%. As with inner city areas, this was primarily due to a much higher than average non-contact rate (14%). Lower response rates in London are a problem that is common to most major surveys.

**Table 4.9 2003-4 Core sample response rate by Government Office Region**

| <b>GOR</b>             | <b>Response rate (%)</b> |
|------------------------|--------------------------|
| North East             | 79                       |
| North West             | 78                       |
| Yorkshire & Humberside | 74                       |
| East Midlands          | 74                       |
| West Midlands          | 76                       |
| East of England        | 74                       |
| London                 | 61                       |
| South East             | 76                       |
| South West             | 75                       |
| Wales                  | 77                       |

#### 4.7.3 Core response rates by type of property and type of area

As mentioned in section 4.4.2, part of the BCS assignment involved the interviewer collecting some details about the area and about the specific issued address. Since this information was collected for all residential addresses, whether or not an interview was obtained, it is possible to analyse response rates according to this data. Of most interest are how response varies first, by the type of property and second, by the type of area.

Table 4.10 below shows how response rates on the 2003-4 survey varied according to the type of property, ranging from 80% response among addresses identified as detached or semi-detached houses to 65% among addresses identified as flats.

**Table 4.10 2003-4 Core sample response rate by type of property**

| <b>Property</b>                          | <b>Response rate (%)</b> |
|--|--------------------------|
| Detached or semi-detached houses         | 80                       |
| Terraced houses                          | 74                       |
| Maisonettes                              | 70                       |
| Flat with own entrance                   | 72                       |
| Flat with non-lockable communal entrance | 65                       |
| Flat with lockable communal entrance     | 61                       |
| <b>All types of flat</b>                 | <b>65</b>                |

The table also shows the impact of two particular aspects of flats on response, namely whether a property has a communal entrance or not and whether the entrance to the communal entrance is lockable or not (e.g. controlled entry phone system). Flats with communal entrances that had controlled entry systems were the most difficult type of property to gain response. In 2003-4, the response rate at these types of property was 61% compared to 72% for flats with their own (non-communal) entrances. Not surprisingly, this lower response rate was due primarily to the relatively high level of non-contact (15%). This illustrates the difficulty faced by interviewers in trying to gain an interview at an address where they are unable to make direct face-to-face contact with people, often having to communicate via intercom systems.

Taken together these figures go a long way to explaining the lower than average response rate in London. For the country as a whole, flats represented only 13% of the issued sample, while flats with communal entrances that have controlled entry represented 8% of the issued sample. However, in London these types of properties represented 35% and 23% of the issued sample respectively. Therefore, one important reason for the lower response rate in London and inner city areas in general is the composition of the housing stock.

Apart from the actual type of property, interviewers were also asked to make an assessment about the condition of each property that was sampled. Results suggest that the better the condition of the sampled property the more likely an interviewer was to obtain an interview. Thus, among sampled properties assessed as being in 'very good' condition the response rate was 81%, compared to 73% among properties assessed as being in 'neither good or bad' condition and 64% among properties assessed as being in 'very bad' condition. This pattern held true across all types of property (i.e. the response rate among both houses and flats assessed as being in 'very good' condition was higher than among those assessed to be in 'very poor' condition).

Interviewers were also asked to record their general observations about the area immediately surrounding each issued address with respect to a number of characteristics including how common rubbish or litter was, how common vandalism and graffiti was and how common run down houses were. Although these observations were clearly open to a high degree of subjectivity, Table 4.11 below shows that the overall response rates tended to be higher among issued addresses where the interviewer had recorded these types of disorder as being not at all common.

**Table 4.11 2003-4 Core sample response rate by area characteristics**

|   | <b>Very common</b> | <b>Fairly common</b> | <b>Not very common</b> | <b>Not at all common</b> |
|---|--------------------|----------------------|------------------------|--------------------------|
|   | %                  | %                    | %                      | %                        |
| Litter or rubbish lying around            | 70                 | 73                   | 74                     | 79                       |
| Vandalism, graffiti or damage to property | 71                 | 75                   | 73                     | 78                       |
| Homes in poor condition or run down       | 68                 | 71                   | 74                     | 79                       |

#### **4.8 Response rates and reasons for non-response: Non-white boost sample**

Table 4.12 shows the screening outcomes and the response rates for the 2003-4 non-white boost sample. Of the total addresses issued for screening, 6.5% were not screened primarily because the interviewer was unable to identify an adjacent address for screening or because no information could be obtained on a household despite every effort being made. Of the 155,870 addresses screened over the year where a valid outcome was obtained only 4.8% of households had an eligible respondent.

Among households identified as containing an eligible respondent a response rate of 50% was achieved. Although refusal rates were slightly higher compared with the core sample (19% and 15% respectively), the main reason for the lower response rate on the non-white boost sample was a significantly higher non-contact rate (18% compared with 5% on the core sample).

This can be explained partly by the nature of the population. The non-white population tends to be concentrated in inner city areas and especially in London, where non-contact rates tend to be higher. It can also be explained partly by the nature of the methodology since eligible households tend to be identified later in an assignment and so interviewers have, on average, less time to make repeated calls.

**Table 4.12 2003-4 Non-white boost sample screening outcomes, response rate and non-response breakdown**

|   | N              | %           |
|---|----------------|-------------|
| <b>Total addresses issued</b>                   | <b>166,752</b> | <b>100</b>  |
| Information not obtained at address             | 2,452          | 1.5         |
| Unable to identify address for screening        | 8,247          | 4.9         |
| Screened address part of core sample            | 183            | 0.1         |
| <b>Addresses not screened</b>                   | <b>10,882</b>  | <b>6.5</b>  |
| <b>Total addresses screened</b>                 | <b>155,870</b> | <b>100</b>  |
| No non white adults at address                  | 148,415        | 95.2        |
| <b>Total eligible addresses identified</b>      | <b>7,455</b>   | <b>100</b>  |
| Address subsequently identified as non-eligible | 703            | 9.4         |
| Other deadwood                                  | 134            | 1.8         |
| <b>Total ineligible addresses</b>               | <b>837</b>     | <b>11.2</b> |
| <b>Total eligible addresses</b>                 | <b>6,618</b>   | <b>100</b>  |
| No contact made with household                  | 1,061          | 16.0        |
| No contact made with selected respondent        | 137            | 2.1         |
| <b>Total non contact</b>                        | <b>1,091</b>   | <b>18.1</b> |
| Office refusal                                  | 31             | 0.5         |
| Refused all information                         | 480            | 7.3         |
| Personal refusal                                | 589            | 8.9         |
| Proxy refusal                                   | 188            | 2.8         |
| <b>Total refusal</b>                            | <b>1,288</b>   | <b>19.5</b> |
| Broken appointment                              | 210            | 3.2         |
| Temporarily ill/incapacitated                   | 15             | 0.2         |
| Physically or mentally unable                   | 27             | 0.4         |
| Away/in hospital                                | 104            | 1.6         |
| Inadequate English                              | 336            | 5.1         |
| Other unsuccessful                              | 122            | 1.8         |
| <b>Total other unsuccessful</b>                 | <b>814</b>     | <b>12.3</b> |
| <b>Total unproductive</b>                       | <b>3,300</b>   | <b>49.9</b> |
| <b>Achieved interviews</b>                      | <b>3,318</b>   | <b>50.1</b> |

To illustrate the extent to which the non-white population is concentrated in particular areas, Table 4.13 shows the proportion of total eligible address and achieved interviews by Police Force Area. This shows that 79% of the total eligible addresses were identified in just 8 out of the 42 Police Force Areas, with London accounting for half (51%) of all eligible addresses and the West Midlands a further 9%. The pattern for achieved interviews was broadly similar.

**Table 4.13 Proportion of Non-white boost sample eligible addresses and achieved interviews by Police Force Area**

| Police Force Area  | No. of eligible addresses identified (n) | Proportion of total (%) | No. of achieved interviews (n) | Proportion of total (%) |
|--------------------|--|-------------------------|--------------------------------|-------------------------|
| Metropolitan       | 3,363                                    | 50.8                    | 1,569                          | 47.3                    |
| West Midlands      | 600                                      | 9.1                     | 327                            | 9.8                     |
| Leicestershire     | 275                                      | 4.2                     | 165                            | 5.0                     |
| Bedfordshire       | 275                                      | 4.1                     | 144                            | 4.3                     |
| West Yorkshire     | 222                                      | 3.4                     | 118                            | 3.6                     |
| Thames Valley      | 203                                      | 3.1                     | 102                            | 3.1                     |
| Greater Manchester | 139                                      | 2.1                     | 85                             | 2.6                     |
| Lancashire         | 127                                      | 1.9                     | 67                             | 2.0                     |
| All other PFAs     | 1,414                                    | 21.3                    | 741                            | 22.3                    |
| <b>Total</b>       | <b>6,618</b>                             | <b>100</b>              | <b>3,318</b>                   | <b>100</b>              |

Across the whole issued sample, the average number of eligible non-white respondents identified per PSU was 3.7 while the average number of additional non-white boost interviews per PSU was only 1.9. However, in 47% of issued assignments no eligible non-white respondent were identified, while in 56% of all issued assignments no additional non-white boost interviews were achieved. In only 10% of assignments were more than 5 non-white boost interviews achieved.

#### **4.9 Response rates and reasons for non response: Youth boost sample**

Table 4.14 shows the response rates for the youth boost sample. At one in five addresses (20%) no screening was actually carried out because the core address was deadwood, a non-contact or a refusal by the selected respondent on the core sample<sup>19</sup>. At addresses where screening was carried out, an eligible respondent was identified at 8.0% of addresses.

The overall response rate achieved on the youth boost sample was 72%. The level of non-contact and refusal was broadly in line with what was achieved on the core sample.

<sup>19</sup> Interviewers were instructed not to carry out screening at households where the core sample selected respondent refused either in person or by proxy. This was done to maximise the chances of a core sample interview being achieved at a reissue stage.

**Table 4.14 2003-4 Youth boost sample screening outcomes and response rate**

|  | N             | %           |
|--|---------------|-------------|
| <b>Total addresses issued</b>            | <b>54,736</b> | <b>100</b>  |
| No screening attempted                   | 11,064        | 20.2        |
| <b>Total screened addresses</b>          | <b>43,672</b> | <b>100</b>  |
| Core sample respondent aged 16-24        | 3,470         | 7.9         |
| No 16-24 adult at address                | 36,467        | 83.5        |
| All information refused                  | 562           | 1.3         |
| <b>Total ineligible addresses</b>        | <b>40,499</b> | <b>92.7</b> |
| <b>Total eligible addresses</b>          | <b>3,173</b>  | <b>100</b>  |
| No contact made with selected respondent | 157           | 4.9         |
| <b>Total non contact</b>                 | <b>157</b>    | <b>4.9</b>  |
| Office refusal                           | 17            | 0.5         |
| Personal refusal                         | 265           | 8.4         |
| Proxy refusal                            | 172           | 5.4         |
| <b>Total refusal</b>                     | <b>454</b>    | <b>14.3</b> |
| Broken appointment                       | 63            | 2.0         |
| Temporarily ill/incapacitated            | 7             | 0.2         |
| Physically or mentally unable            | 27            | 0.9         |
| Away/in hospital                         | 75            | 2.4         |
| Inadequate English                       | 9             | 0.3         |
| Other unsuccessful                       | 87            | 2.7         |
| <b>Total other unsuccessful</b>          | <b>268</b>    | <b>8.4</b>  |
| <b>Total unproductive</b>                | <b>879</b>    | <b>27.7</b> |
| <b>Achieved interviews</b>               | <b>2,294</b>  | <b>72.3</b> |

#### **4.10 Response to the self-completion questionnaire**

The self-completion questionnaire was presented to respondents at the end of the interview. In 2003-4 the self-completion module consisted only of exposure to illegal drugs and drinking behaviour (16-30 year olds only). As such, the self-completion was considerably shorter than on the previous two surveys.

Table 4.15 shows that on the core sample more than 95% of eligible respondents accepted the self-completion. Respondents in the non-white boost sample were less likely to do the self-completion, with 88% of eligible respondents completing it. The lower response rate among the non-white boost sample was primarily due to language

problems, with only 5% of respondents from the core sample refusing the self-completion due to language problems compared with 27% of respondents from the non-white boost sample.

For respondents who completed the self-completion, about one in eight (11%) of the core sample and about one in five (20%) of the non-white boost sample asked the interviewer to administer it for them.

As might be expected, response to the self-completion among the youth boost sample was high, with 95% of respondents doing the self-completion themselves and a further 4% asking the interviewer to complete it for them.

**Table 4.15 Response to self-completion questionnaire by type of sample**

|   | Core sample   |             | Non-white sample |             | Youth sample |             |
|---|---------------|-------------|------------------|-------------|--------------|-------------|
|   | N             | %           | N                | %           | N            | %           |
| Refused                                 | 1,078         | 4.3         | 354              | 12.0        | 16           | 0.7         |
| Completed by interviewer                | 2,808         | 11.2        | 599              | 20.4        | 92           | 4.0         |
| Completed by respondent                 | 21,123        | 84.5        | 1,989            | 67.6        | 2,186        | 95.3        |
| <b>Overall self-completion response</b> | <b>23,931</b> | <b>95.7</b> | <b>2,588</b>     | <b>88.0</b> | <b>2,278</b> | <b>99.3</b> |
| <i>Base (All 16-59 year olds):</i>      | <i>25,009</i> |             | <i>2,942</i>     |             | <i>2,294</i> |             |

Table 4.16 below shows how response to the self-completion questionnaire varied according to the demographic characteristics of respondents. This shows that there was little difference by gender or by age in terms of the proportion of respondents who completed or refused the self-completion questionnaire. However, older respondents were slightly more likely to ask the interviewer to complete the section for them (15% of 45-59 year olds compared with 8% of 16-29 year olds). The largest differences were among respondents from different ethnic groups, with non-white respondents being more likely than white respondents to refuse to do the self-completion and more likely to ask the interviewer to do it for them. This was especially noticeable among Asian respondents where 12% of respondents refused the self-completion and a further 22% of respondents asked the interviewer to do it for them.

When broken down by these demographic characteristics there were no obvious differences in response between respondents from the core sample and respondents from the non-white boost sample.

**Table 4.16 Response to the self-completion questionnaire by demographic characteristics – combined core and non-white boost sample**

|                    | Completed by respondent | Completed by interviewer | Refused  | Base          |
|--------------------|-------------------------|--------------------------|----------|---------------|
|                    | %                       | %                        | %        | N             |
| <b>Sex</b>         |                         |                          |          |               |
| Male               | 83                      | 12                       | 5        | 12,804        |
| Female             | 82                      | 13                       | 5        | 15,147        |
| <b>Age</b>         |                         |                          |          |               |
| 16-29              | 87                      | 8                        | 5        | 6,352         |
| 30-44              | 83                      | 12                       | 5        | 11,774        |
| 45-59              | 79                      | 15                       | 6        | 9,825         |
| <b>Ethnicity</b>   |                         |                          |          |               |
| White              | 86                      | 10                       | 4        | 23,157        |
| Mixed              | 81                      | 14                       | 5        | 347           |
| Asian              | 66                      | 22                       | 12       | 2,177         |
| Black              | 71                      | 19                       | 10       | 1,553         |
| Other ethnic group | 66                      | 17                       | 17       | 715           |
| <b>Total</b>       | <b>83</b>               | <b>12</b>                | <b>5</b> | <b>27,949</b> |

Table 4.17 shows the reasons given by respondents either for refusing the self-completion questionnaire or for asking the interviewer to do it for them. This shows that a dislike of computers (38%) was the most common reason why respondents asked the interviewer to do it for them, while running out of time (48%) was the most common reason given for refusing to do it. Language problems were a reason given by 17% of respondents who refused the self-completion and 6% of those who asked the interviewer to do it for them. Among Asian respondents 32% of those who either refused or had the interviewer do it for them, did so because of language problems.

**Table 4.17 Reasons for refusing self-completion questionnaire or for completion by interviewer –combined core and non-white boost sample**

|                                      | Refused | Completed by interviewer | Total |
|--------------------------------------|---------|--------------------------|-------|
|                                      | %       | %                        | %     |
| Don't like computers                 | 18      | 38                       | 33    |
| Ran out of time                      | 48      | 28                       | 34    |
| Couldn't be bothered                 | 5       | 11                       | 9     |
| Language problems                    | 17      | 6                        | 9     |
| Children present/caring for children | 6       | 8                        | 7     |
| Could not read/write English         | 3       | 3                        | 3     |
| Eyesight problems                    | 2       | 4                        | 3     |
| Other disability                     | 3       | 3                        | 3     |
| Other people in room                 | 3       | 1                        | 2     |
| Other reasons                        | 8       | 8                        | 8     |
| Base (N) <sup>20</sup>               | 1,267   | 3,407                    | 4,674 |

<sup>20</sup> Percentages add up to more than 100% since more than one answer could be coded at this question

#### **4.11 Full and Partial Interviews**

An interview was only counted as a full interview for the core sample and the non-white boost if the respondent had completed to the end of the demographic section of the questionnaire. Anything less than this was returned as a partial interview. Full and partial interviews were recorded separately in the field figures. In 2003-4, 99.5% of interviews achieved on the core sample were full interviews and only 0.5% (172 interviews) were partial interviews. On the non-white boost sample 98.7% of interviews achieved were full interviews and 1.3% (44 interviews) were partial.

On the youth boost sample the respondent had to complete the survey to the end of the questionnaire (including the self-completion) for it to count as a full interview. Anything less than this was returned as a partial interview. In 2003-4, only 0.7% (16 interviews) of interviews achieved on the youth boost sample were classified as partial.

## 5. Data Processing

### 5.1 Offence coding

The BCS Offence Coding System was developed for the 1982 BCS to match, as closely as possible, the way incidents were classified by the police. Apart from some minor changes, the code frame and the instructions to coders have remained stable since 1982.

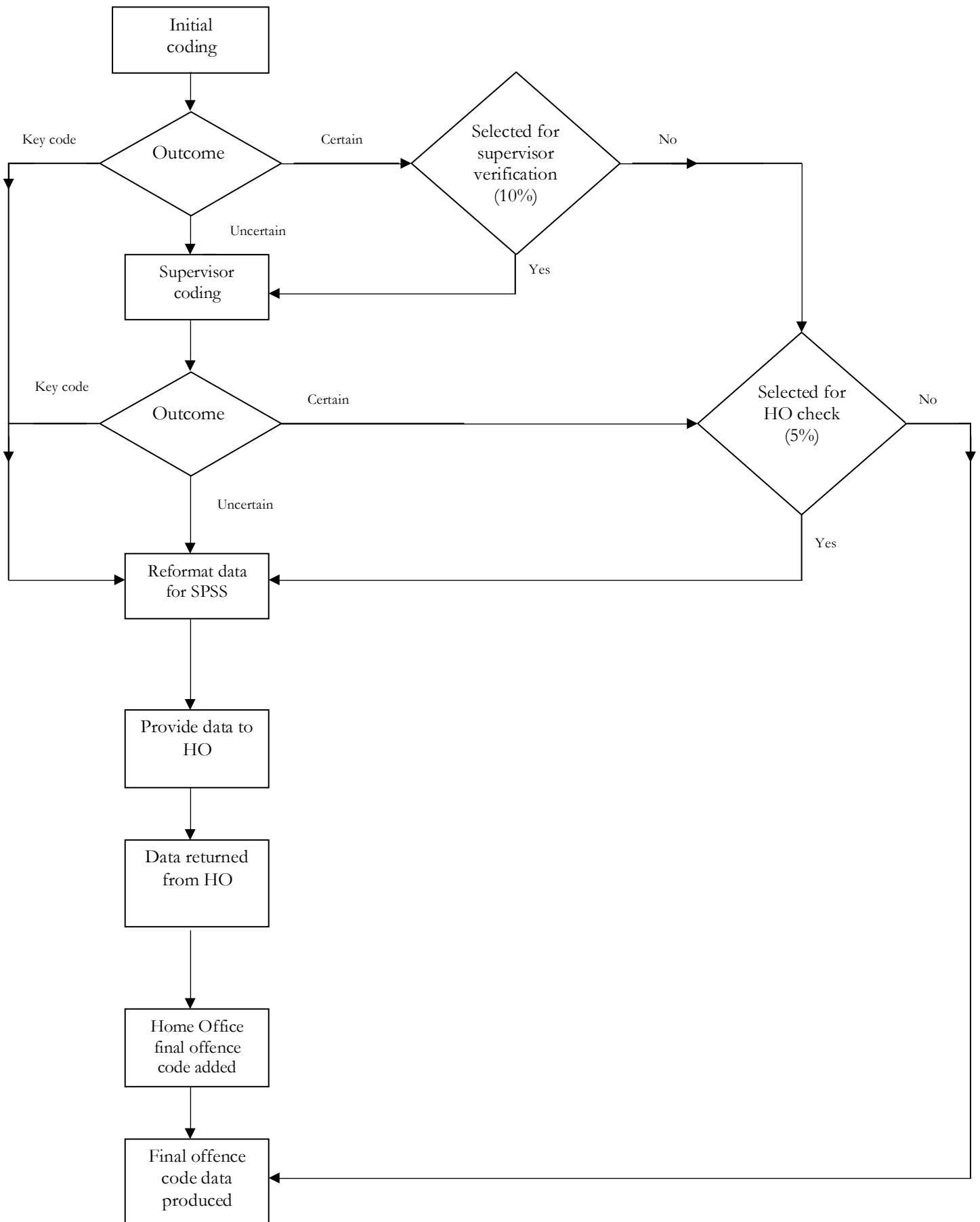
In 2001, new operational procedures were introduced for coders, although the code frame and the way any given code was reached remained consistent with previous waves.

During 2003-4, the Offence Coding System consisted of the following steps:

1. An offence code was generated automatically, based on the answers given to a number of pre-coded variables in the Victim Form.
2. Coders used a specially developed questionnaire to produce an offence code for each Victim Form. The coders also had to record for each case how certain or uncertain they were about the code.
3. A supervisor checked any codes that the original coder was uncertain about, plus at least 5% codes where the coder was certain of the outcome as a quality check.
4. Researchers at the Home Office checked:
  - Any codes that BMRB were uncertain about
  - Certain types of incident that were automatically referred (e.g. arson)
  - A proportion of certain codes as part of a quality check

The result of this process was that every Victim Form had a final outcome code assigned to it. A flow chart of the Offence Coding System is shown in Figure 5.1 and the offence coding system is explained in more detail below.

**Fig. 5.1 British Crime Survey Offence Coding Flowchart**



### 5.1.1 The automatically generated offence code

In 1996 a programme was introduced that automatically generated an offence code based on the answers to a number of pre-coded variables in the Victim Form. The programme that was used for the 2003-4 survey was identical to that used in the 2001 and 2002-3 surveys.

An automatic code cannot be generated in all cases, and in 2003-4 no automatically generated code was produced for about a third (33%) of all Victim Forms (due to missing codes or to some inconsistency between the different variables used). Irrespective of the suggested automatic code, the coder has the responsibility of producing an offence code, and coders are instructed to see the generated code as only a starting point.

On the 2003-4 survey for Victim Forms where a code was automatically generated, it was the same as the final offence code in 77% of cases.

### 5.1.2 The coding task

Coders are provided with a paper-based print out of the key variables from each Victim Form, and this information forms the basis of the coding. This document also provides coders with the offence code that had been generated by the automatic generation programme. An example of this paper form can be found in Appendix G in Volume 2.

Coders used a specially designed computer assisted questionnaire to carry out the coding. The questionnaire asked the coders certain questions about the nature of the offence. The questionnaire takes account of the major rules that apply to offence coding (such as the priority of codes), and by answering the questions on the basis of the information provided in the Victim Form, the coders reach an offence code.

All coders were personally briefed about the offence coding. The coders were also provided with a coding manual. This manual is similar to the one used in previous years of the BCS, and contains all the rules that govern offence coding. The manual also provides flow-charts that show how the coding questionnaire works, so that coders can see how they reached a particular offence code on the basis of the answers that they input. A copy of this manual is provided in Appendix G in Volume 2.

When the coder reaches an offence code, they can say whether they are certain or uncertain that this is the right code. Any Victim Forms which the coder is uncertain about are automatically referred to their supervisor for checking. In addition, the supervisor checks 10% of codes which coders were certain about.

### 5.1.3 Home Office coding

Victim Forms were referred to the Home Office research team if the incident involved:

- deliberate damage by fire
- the police as offenders
- the offender being mentally ill

The Home Office was also sent all offence codes that the supervisors were unsure about.

In 2003-4, 5% of codes which BMRB were certain about were randomly selected to go to the Home Office for quality control checking.

A list of Victim Forms to be checked by researchers at the Home Office was sent every two weeks. This consisted of an Excel spreadsheet that contained the serial number of each Victim Form, the code that the coder (and supervisor if applicable) had given the incident, how certain the coder (and supervisor) was about the incident, and any notes that the coder added about why they were uncertain. An electronic version of the paper-based document providing the key variables from the Victim Form was also provided.

Researchers at the Home Office coded each of the Victim Forms sent to them (using the paper-based document) and returned the spreadsheet with their code added and any comments. These codes were then manually added into the coding file (so that the coders could see the changes that had been made).

Particular attention was paid to cases where the Home Office changed a code that BMRB coders had marked as “certain”. If the BMRB coders disagreed with such a coding decision, this was fed back to both BMRB researchers and Home Office researchers for further consideration and discussion.

#### 5.1.4 Offence coding checks in 2003-4

In total 2,635 cases were sent to the Home Office for checking as part of the 2003-4 survey.

Of the Victim Forms sent to the Home Office:

- 442 were code 01s which were automatically referred to Home Office. This covers cases of arson, cases where the offender was thought to be mentally ill, and cases where the offender was believed to be a police officer.
- 7 were code 02s (suspected duplicate cases, or cases where the BMRB coder was not certain about the code) which were also automatically referred to the Home Office for checking.
- 880 were part of the quality check.
- 1305 were related Victim Forms. To ensure that those checking offence codes had complete information all the Victim Forms belonging to an individual respondent were sent to the Home Office, rather than just the single Victim Form under consideration.

Of the 2,635 Victim Forms sent to the Home Office 443 cases had their code changed by the Home Office, representing 17% of all cases sent. However, this included all of the code 01's which had to be changed to a valid offence code. When these cases are removed, a total of 143 Victim Forms sent to the Home Office had their codes changed, representing only 5% of all cases sent. This level of change was fairly static across the survey year suggesting a degree of stability in the offence coding process.

The breakdown of codes changed by the Home Office according to the categories outlined above were as follows:

- Of 7 cases where BMRB was uncertain or duplicates identified 4 (57%) were changed.
- Of 880 sent for quality control 45 (5%) were changed
- Of 1,305 related cases 94 (7%) were changed

In all cases where the Home Office changed a code that BMRB coders or supervisors had been certain about, this was double checked and verified by BMRB upon return of the coding from the Home Office. Where BMRB coders and researchers did not agree with the Home Office decision these were referred back again for re-checking. Of the 139 cases that BMRB had originally been certain about and the Home Office had changed, 49 were referred back for re-checking. In 36 cases, the original BMRB code was deemed to be correct and was reinstated as the final code.

### 5.1.5 Final Offence Code

The SPSS data set includes all the offence codes that have been given to each Victim Form at every stage of the coding process. This allows a complete history of each case to be maintained at all times. The final offence code is done using a priority ordering system whereby the Home Office code takes priority over the supervisor, which takes priority over the coder. The variables on the data set are:

|          |  |
|----------|--|
| OFFSUG   | Suggested offence code (generated by computer) |
| VOFFENCE | Code assigned by the original coder            |
| SOFFENCE | Code assigned by the supervisor                |
| FINLOFFC | Code assigned by the Home Office research team |
| OFFENCE  | Final offence code                             |

### 5.1.6 Checks on final offence code

During the creation of the SPSS data sets some further consistency checks are run on the final offence codes, checking these against key pre-coded variables in the Victim Form. The purpose of this is to highlight cases where some of the pre-coded data seems potentially anomalous with the final offence code.

The checks carried out are as follows:

- Assaults where no force or violence was used
- Burglary where entry to the property was authorised
- Car thefts where no car was listed as being stolen, or where the police were not informed
- Sexual assaults where there was no sexual element to the assault
- Snatch thefts where the item stolen was not being held or carried
- Other thefts where the item stolen was being held or carried

All cases that fail these checks are examined individually by a researcher and, if necessary, are referred to the Home Office. Experience of running these checks shows that most

flagged cases do have the correct offence codes, but a few may be amended each quarter as a result of this additional check.

## **5.2 Other coding**

In addition to the offence coding, coders also looked at all questions where an “other – specify” had been given as an answer. The aim of this exercise, commonly known as back coding, was to see whether the answer given could actually be coded into one of the original pre-coded response options. Coding was done in Verbastat, part of the quantime suite of programmes.

Coders were provided with the code frames used in the questionnaire as a starting point. Since most of the questions have been used in previous years of the survey, the code frames were already well developed and there was little need to add new codes to the frames. However, if the coding supervisor felt an extra code was needed, this was flagged up to researchers who approved any changes before they were implemented.

Since the BCS tends to have no fully open-ended questions there was no requirement on the 2003-4 survey to develop or code any questions from verbatim answers given by respondents.

## **5.3 Coding of occupation and social class**

Occupation details were collected for all respondents either relating to their current job or to their last job if the respondent was not currently employed but had worked at some time in the past. Occupational details of the Household Reference Person were also collected, if this was not the same person as the respondent.

Occupations were coded using the Standard Occupational Classification 2000 (SOC2000). All occupational coding was done centrally once the data were returned by interviewers.

In addition to coding SOC2000, the old SOC1990 codes were also recorded using Volume 2 of the Standard Occupational Classification 2000. This contains a look-up table that allows SOC2000 codes to be matched to their SOC1990 equivalents.

As well as occupation codes, social class for all respondents and Household Reference Persons was added to the file. The new social class categories –NS-SEC – were derived automatically using an algorithm which was developed from the documentation provided by ONS. Both the NS-SEC operational categories and the NS-SEC analytical categories were derived. Additionally, the old Socio-Economic Group (SEG) was derived using another algorithm based on Volume 3 of the Standard Occupational Classification 1990.

Details of the NS-SEC and SEG categories can be found in Appendix H of Volume 2.

## 6. Data Output

### 6.1 Introduction

The main output associated with the British Crime Survey is two SPSS data files that are delivered to the Home Office on a quarterly basis.

One file, the **Non Victim File (NVF)**, is produced at the level of the individual respondent and contains all questionnaire data, except for that collected in the Victim Forms. Interviews where interviewers had reached the end of the main demographic section were classified as **full interviews**, while anything less than this were classified as **partial interviews** (see section 4.11). Interviews where the interviewer did not reach the end of the screener questions were regarded as unusable interviews and not included on the SPSS data files.

The second file, the **Victim File (VF)**, is produced at the level of the individual incident or crime and contained all the data collected in the Victim Forms. All generated Victim Forms were included on the file, including cases where the Victim Form has either been suspended or the reference period was out of scope. Although such records contain no information and are not used for analysis, it is useful to keep these on the file to monitor the number of Victim Forms that fall into these categories.

During 2003-4 data from the core sample only was analysed on a 12 month rolling basis for each quarter. Data from the non-white boost sample was supplied on a quarterly basis for reference, although this was only weighted and analysed annually. In fact, a separate data file containing **all** non white respondents, whether from the core sample or the non-white boost sample, was supplied to the Home Office separately. Similarly, data for all 16 to 24 year olds, whether from the core sample or the youth boost sample, was supplied as a separate data file on a quarterly basis.

In addition to the achieved sample, a data file of the entire 2003-4 issued sample was also supplied to the Home Office. This contained information on every issued address such as the final outcome, the screening outcomes, the observational data collected by interviewers, sample variables, and geo-demographic variables.

### 6.2 Delivery of data output

Data was delivered to the Home Office approximately six weeks after the end of each quarterly fieldwork period. Each quarterly data delivery included interviews that were **achieved** in a specific time period, rather than those that were **issued** in a specific time period.

Because of the need to deliver data quickly after the end of a quarter this meant that interviews were included on each quarterly file that came from issued sample that had not yet been exhausted. As such with each delivery of data, the most recent quarter of data represented preliminary data in terms of the issued sample and in terms of the weighting.

Each quarter a full 12 months data file was supplied to the Home Office. This was done on a rolling basis meaning that each quarter the oldest quarter of data was removed and a new quarter of data was added.

Due to the continuous nature of the survey and the fact that data files are based on achieved and not issued sample, data from different questionnaires were supplied as part of the same data files. Thus, the first quarter of 2003-4 included data from cases issued as part of the 2002-3 sample as well as cases issued as part of the 2003-4 sample. Similarly, some data from the last quarter of 2003-4 will be included in the first quarter of the 2004-5 dataset.

### 6.3 Content of SPSS data file

The SPSS data files delivered to the Home Office contain various types of variables. The main types of variables contained on the files are:

- **Questionnaire variables** (NVF and VF).
- **Geo-demographic variables** (NVF only). All interviews had a set of pre-specified geo-demographic variables attached to them (see Appendix H in Volume 2 for complete listing).
- **Observational variables** (NVF only). All interviews had the observational data collected by interviewers on the Address Contact Sheets attached to them (see Appendix H in Volume 2 for complete listing). Due to the way in which the Observational data was processed it was difficult to do this on a quarterly basis. Consequently it was agreed that Observational variables only be supplied on an annual basis.
- **Coding variables** (NVF and VF). On the Non Victim File, SOC2000 and SOC1990 codes are included for both the respondent and the Household Reference Person. Additionally, NS-SEC and SEG for both the respondent and the Household Reference Person are included. On the Victim File, a full set of offence codes was attached as outlined in section 5.1.5.
- **Derived variables** (NVF and VF). Many derived variables were also added to the file. These consisted primarily of 2 types:
  - **Flag variables** that identify, for example, the type of sample (Core, Non-white or Youth boost), the follow-up module split and follow-up module sub-split, the date of interview, the month of issue, whether a partial or full interview, whether a victim or non-victim, etc. On the Victim File, flag variables include whether the record was a Long or Short Victim Form, whether it was a Series or a Single incident, and whether it was inside or outside the reference period.
  - **Classificatory variables** derived from the data. These included standard classifications such as ONS harmonised variables, banded age groups, ethnic groups, income groups, etc.
- **Weighting variables** (NVF only).

### 6.4 Conventions used on SPSS Data Files

In creating the 2003-4 data files great attention was paid to ensuring as much consistency as possible was maintained with previous years of the survey.

### 6.4.1 Case identifier

The case identifier was required to be similar to that used on previous years of the survey but also had to be designed to meet the requirements of a continuous survey.

On the Non-Victim File, where each individual case or record represents an individual respondent, the unique case identifier (ROWLABEL) is an 8-digit number constructed as follows:

|                   | <b>Column position</b> | <b>Values</b> |
|-------------------|------------------------|---------------|
| Year of issue     | 1                      | 1-9           |
| Area point number | 2-5                    | 1000-9999     |
| Address number    | 6-7                    | 01-99         |
| Screen number     | 8                      | 0-9           |

On the Victim File, where each individual case or record, represents a Victim Form or incident, the unique case identifier (MATCH) is a 9-digit number, which is identical to ROWLABEL with the addition of the Victim Form number:

|                    | <b>Column position</b> | <b>Values</b> |
|--------------------|------------------------|---------------|
| Year of issue      | 1                      | 1-9           |
| Area point number  | 2-5                    | 1000-9999     |
| Address number     | 6-7                    | 01-99         |
| Screen number      | 8                      | 0-9           |
| Victim Form number | 9                      | 1-6           |

### 6.4.2 Naming conventions

Variable names were kept the same as on the previous surveys wherever possible. For some questions, small changes to the code frame (such as the adding of an extra code) meant that questions which appear similar to previous years are not in actual fact identical. In such situations, the variable name on the 2003-4 data file was changed to reflect this. Examples of variables which were different in 2003-4 compared with 2002-3 included VEHPAR2 (previously VEHPAR), OFFREL2 (previously OFFREL1), RACEOFF2 (previously RACEOFF1) and TYPEVS2 (previously TYPEVS). In all these cases the questions in 2003-4 were the same as those in 2002-3 but the coding frame had changed between surveys.

### 6.4.3 Labelling variables

The changing nature of the 12-month reference period over the course of the year creates a difficulty in labelling certain variables. In the Quancept script, dates were automatically calculated based on the date of interview and appropriate text substitution was used to ensure that the question always referred to the correct period. In the SPSS data files, which contain data from interviews achieved over the whole year, it is difficult to attach meaningful labels to certain variables since the label is different depending upon the month of interview. This issue affects the following variables (all on the Victim File):

- DATESERA-DATESERH
- NQUART1-NQUART5
- QTRRECIN

- QTRINCID

Details of how the code frames for these specific questions relate to the month of interview can be found in Appendix F of Volume 2.

#### 6.4.4 Don't Know and Refused values

The convention for Don't Know and Refusal codes used in the most recent surveys was maintained on the 2003-4 data. This meant that on the SPSS file the code for Don't Know was '9' for code frames up to 7, '99' for code frames up to 97, and so on. The code for Refused was 8, 98, and so on. Since these are standard codes used throughout the SPSS files, Don't Know and Refused codes are not labelled.

#### 6.4.5 Multiple response variables

Prior to the 2001 survey, multiple response variables have been created as a set of variables equal to the maximum number of answers that could be given. The first variable holds the first answer given by the respondent, the second variable holds the second answer given by the respondent, and so on.

After discussions with the Home Office it was agreed from 2001 onwards to present multiple response variables differently from previous years. Instead, multiple response variables were set up as a set of variables equal to the total number of answers possible (including Don't Know and Refused). Each variable was then given a value of '0' or '1' depending on whether the respondent gave that particular answer or not. To denote this change all multiple response variables in 2001 were all named with a letter suffix, rather than the number suffix that was used in previous years of the survey.

An example of a multiple response variable where there are seven possible answer categories, and so seven separate variables, is shown below:

AGEOFFA-  
AGEOFFG

[ASK IF NumOff IN (2..4)]

How old were the people who did it? Would you say they were...READ  
OUT  
CODE ALL THAT APPLY

- |    |                               |           |
|----|-------------------------------|-----------|
| 1. | children under school age     | (AGEOFFA) |
| 2. | children of school age        | (AGEOFFB) |
| 3. | people aged between 16 and 23 | (AGEOFFC) |
| 4. | people aged between 25 and 39 | (AGEOFFD) |
| 5. | or people aged over 40?       | (AGEOFFE) |
|    | Don't Know                    | (AGEOFFF) |
|    | Refused                       | (AGEOFFG) |

This practice of defining multiple response variables was maintained on the 2003-4 data files.

## 7. Weighting the Data

### 7.1 Reasons for weighting

There are three main reasons for calculating weights on the BCS:

- To compensate for unequal selection probabilities. In the BCS, different units of analysis (households, individuals, instances of victimisation) have different probabilities of inclusion in the sample due to factors such as over sampling of small PFA's, the selection of one dwelling unit at multi-household addresses, the selection of one individual within a dwelling, and the inclusion of a single victim form to represent a series of similar incidents.
- To compensate for differential response. Differential response rates can arise both between different geographic units (e.g. differences in response between inner city and non-inner city areas) and between different age and gender sub-groups.
- To ensure that quarters are equally weighted for analyses that combine data from more than one quarter.

A variety of different weights were computed to meet the different analyses requirements. The 2003-4 weighting schedule was broadly similar to the weighting schedule applied on the 2002-3 survey.

All weights include a component to compensate for unequal selection probabilities, while weighting components to compensate for differential response and to equally weight quarters are included in some weights but not in others. Weights were calculated separately for the core sample, the non-white sample and the youth sample.

### 7.2 Component weights

Although a variety of weights were calculated for the 2003-4 BCS, they were all based on a number of key component weights. In constructing all the different weights for the core sample, the non-white sample and the youth sample the following conventions have been used for the component weights:

- $w_1$  : weight to compensate for unequal address selection probabilities;
- $w_2$  : inner city versus non inner-city non-response weight;
- $w_3$  : dwelling unit weight;
- $w_4$  : individual selection weight;
- $numinc$  : series of incidents weight

#### 7.2.1 Police Force Area weight ( $w_1$ )

As already described in Chapter 2, small PFAs were over sampled to ensure a minimum of 600-700 interviews. It was therefore necessary to down-weight these PFAs and up-weight the large ones to correct for the differences in selection probabilities of addresses.

As also outlined in Chapter 2, addresses selected in the 2003-4 sample were selected from two different types of PSU, those that had been rotated forward from 2002-3 (the

‘rotated sample’) and those that were selected for the first time in 2003-4 (the ‘fresh sample’). Since the rotated and fresh samples were drawn independently from each other, any eligible address could have been selected in either sample. Therefore, in computing  $w_1$  for each address it was necessary to take account not only of the probability associated with the method by which an address was actually sampled, but also the probability associated with the method by which it might have been selected but was not. In other words,  $w_1$  was constructed by taking account of both the rotated sample address selection probability and the fresh sample address selection probability.

The PFA weight was computed on a quarterly basis for each Police Force Area. All the ‘small’ PFAs were treated as stratum in their own right, while all the ‘large’ PFAs formed a single stratum.

$W_1$  was computed by the following steps:

1. The rotated sample address selection probability (PADROT) was calculated separately for each quarter in each PFA stratum as follows:

$$p_{1i} = n_{t_i}^g / N_{1i}$$

where  $N_{1i}$  was the total number of delivery points in PFA stratum  $i$  taken from the 2002-3 Postcode Address File (i.e. the sample frame from which the 2002-3 sample was drawn); and  $n_{t_i}^g$  was the **total** number of addresses issued in the 2002-3 quarter  $g$  sample in stratum  $i$ .

2. The fresh sample address selection probability (PADFRESH) was calculated separately for each quarter in each PFA stratum as follows:

$$p_{2i} = n_{f_i}^g / N_{2i}$$

where  $N_{2i}$  was the total number of delivery points in PFA stratum  $i$  taken from the 2003-4 Postcode Address File (i.e. the sample frame from which the 2003-4 sample was drawn); and  $n_{f_i}^g$  was the total number of addresses issued in the 2003-4 quarter  $g$  sample in stratum  $i$  in the **fresh sample only**.

3. Having calculated the two probabilities for each address the value for  $w_1$  was calculated as:

$$w_1 = 1 / (p_{1i} + p_{2i})$$

### 7.2.2 Inner city weight ( $w_2$ )

In some previous years of the BCS, inner city areas were over sampled. However, the 'inner city' weight applied in previous surveys compensated not only for this difference in selection probabilities but also for the differential response rates between inner city and non-inner city areas. Therefore, to be consistent with previous years of the BCS it was necessary to continue applying a weight to correct for differential response rates. In essence, the inner city weight was simply the reciprocal of the achieved response rate in inner city and non-inner city areas (after weighting by  $w_1$ )

The definition of inner city or non-inner city has been kept consistent since it was first used on the BCS and is based on 1981 census data. Details of how the inner city weight is constructed can be found in previous BCS Technical Reports.

### 7.2.3 Dwelling unit weight ( $w_3$ )

At addresses which had more than one dwelling unit, the interviewer made a random selection of one dwelling unit. The dwelling unit weight is therefore simply equal to the number of dwelling units identified at the address. In over 99% of cases, the dwelling unit weight is 1.

### 7.2.4 Individual weight ( $w_4$ )

At dwelling units that had more than one eligible adult, the interviewer made a random selection of one adult. Thus, the probability of any one individual being selected was inversely proportional to the number of adults in the household. The individual weight is therefore simply equal to the number of adults in the household. In the case of the non-white sample,  $w_4$  is equal to the number of eligible adults in the household (i.e. non-white adults only).

### 7.2.5 Series weight (numinc)

This weight is applied when estimating victimisation rates. For single incidents, numinc is always 1. For series incidents, where only details are collected about the most recent incident in the series, the weight equals the number of incidents in the series that fall within the reference period, subject to a maximum limit of 5.

In estimating victimisation rates, the household or individual weights are multiplied by the numinc weight, according to which offence classification code has been assigned to the incident(s).

## 7.3 Core sample weights

The main units of analysis used on the BCS are households, individuals, and incidents of victimisation. Different weights are used depending upon the unit of analysis. In particular, some crimes are considered household crimes (e.g. burglary, vandalism to household property, theft of and from a car) and therefore the main unit of analysis is the household, while others are personal crimes (assault, robbery, sexual offences) and the main unit of analysis is the individual.

For the core sample two weights were constructed to take account of this difference, namely the **core household weight** and the **core individual weight**. These were calculated as follows:

$$\mathbf{wtm2hhu} = w_1 * w_2 * w_3$$

$$\mathbf{wtm2inu} = w_1 * w_2 * w_3 * w_4$$

Once the unscaled weights had been calculated the frequencies were examined and extreme values were capped where necessary. Although capping of extreme weights may introduce a small amount of bias this is more than compensated for by the improvement in precision that results. The capped weights were called **wtm2hhf** and **wtm2inf** respectively.

Finally, the weights were scaled to a notional sample size of 9,250 interviews per quarter. Although an approximately equal number of addresses were issued each quarter during 2003-4, the number of interviews actually achieved per quarter inevitably varied to some extent. Consequently, although the survey is ultimately designed to achieve a notional number of core interviews each quarter (i.e. 9,250), there will always be a degree of variation around this. Thus, for analyses based upon a 12 month period, the weights were constructed to adjust for quarterly differences in sample size by equalising the quarterly achieved sample sizes. The final scaled weights were called **wtm2hhs** and **wtm2ins** respectively.

#### 7.4 Non-white sample weights

The 2003-4 non-white sample comprised all non-white respondents who were interviewed between the start of April 2003 and the end of March 2004, irrespective of whether the respondents were interviewed as part of the core sample or as part of the Non-white Boost sample. Both household and individual weights were computed for the non-white sample. In addition, to the design weights a non-response multiplier was added to the weights. This was designed to adjust the ethnic group, age and sex distribution of the achieved sample using Labour Force Survey data.

The basis of the non-white weighting was the component weights calculated for the core sample. Over the last three years of the survey, the proportion of addresses at which focused enumeration was carried out has varied across the survey year (see 2002-3 Technical Report for further details). However, in 2003-4 the proportion of addresses at which focused enumeration was carried out was constant throughout the survey year as follows:

- 75% of core sample addresses were screened in assignments where the proportion of non-white households was less than 20% (stratum y);
- 100% of core sample addresses were screened in assignments where the proportion of non-white households was 20% or more (stratum z).

In calculating  $w1_{em}$  the first step was to start with the  $w1$  weight derived for the core sample address. This unscaled weight was then scaled as follows:

$$w1_s = w1_u / \text{mean value across the whole sample of } w1_u$$

Once the scaled core address weight was calculated the difference in the volume of screening in the two different strata described above had to be taken into account. This was done by weighting the 2003-4 issued core sample by  $w1_s$  and calculating the total number of addresses issued in the different strata as shown in the table below.

**Table 7.1 Total number of addresses in different strata over 2003-4**

|               | Stratum y                            |                           | Stratum z                            |                           |
|---------------|--------------------------------------|---------------------------|--------------------------------------|---------------------------|
|               | PSUs with non-white population < 20% |                           | PSUs with non-white population >=20% |                           |
|               | No. of core addresses                | No. of screened addresses | No. of core addresses                | No. of screened addresses |
| Apr-June 2003 | <i>i</i>                             | <i>i</i> *4               | <i>j</i>                             | <i>j</i> *5               |
| <b>Total</b>  | <b>A</b>                             | <b>B</b>                  | <b>C</b>                             | <b>D</b>                  |

The above table takes account of the fact that the weights are applied to all non-white respondents identified from both the core sample and the non-white boost sample.

Thus, in stratum y, where the proportion of screening was 4 adjacent addresses at 75% of issued core addresses, the number of screened addresses was equal to:

$$75\% \times (4 \times \text{no. of core addresses}) + \text{core addresses} = 4 \times \text{no. of core addresses}$$

Similarly, in stratum z, where the proportion of screening was 4 adjacent addresses at 100% of issued core addresses, the total number of screened addresses was equal to:

$$100\% \times (4 \times \text{no. of core addresses}) + \text{core addresses} = 5 \times \text{no. of core addresses}$$

The non-white address weight was then calculated as follows for addresses in each of the two different strata:

$$w1_{em} = (A*w1_s)/B \text{ for PSUs in stratum y}$$

$$w1_{em} = (C*w1_s)/D \text{ for PSUs in stratum z}$$

No inner city weight was applied to the non-white weighting.

The dwelling unit weight ( $w3_{em}$ ) was computed in the same way as the core sample. For core sample addresses this was defined as the number of dwelling units at the address and for non-white boost addresses this was defined as the number of dwelling units at the address containing one or more eligible non-white individuals.

The individual selection weight ( $w4_{em}$ ) was also calculated in the same way as the core sample. For core sample addresses this was defined as the number of adults in the household, while for non-white boost addresses this was defined as the number of non-white adults in the household.

Once all these components had been calculated a household weight and individual weight for the non-white sample were then computed in exactly the same way as the core sample, as follows:

$$emwthhu = w1_{em} * w3_{em}$$

$$emwtinu = w1_{em} * w3_{em} * w4_{em}$$

Before the design weights were capped and scaled an attempt was made to compensate for differential non-response among different age and sex sub-groups within each ethnic group. This was done by matching the profile of the achieved sample to the profile of the Labour Force Survey. Once this had been done the weights were then capped and scaled so that the weighted and unweighted sample sizes across the whole year were equalised.

## 7.5 Youth weights

The 2003-4 young people sample comprised all respondents aged 16 to 24 who were interviewed between the start of April 2003 and the end of March 2004, irrespective of whether the respondents were interviewed as part of the core sample or as part of the Youth Boost sample. For the youth data only an individual weight was computed. As with the non-white sample, the basis of the youth weighting was the core sample weights.

In terms of computing a youth weight all respondents came from households that were selected as part of the core sample. This meant that the main household weight (unscaled and uncapped) calculated for the core sample was applied to the youth sample as the starting point for the youth weights.

The youth boost weight was calculated as follows:

$$ypwtu = wtm2hhf * r$$

where,  $r$  was the number of individuals aged between 16 and 24 in the household.

Once the unscaled weight was created the distribution of weights were examined and extreme values capped where necessary ( $ypwtf$ ). Finally, the weights were scaled to ensure the weighted and unweighted sample sizes were the same ( $ypwts$ ).

## 7.6 Calibration Weights

From 2001 onward the Home Office have calculated and applied additional calibration weights to counter the effect of differential response rates between age, gender and regional sub-groups. Results for BCS surveys from 1996 onwards have all been re-weighted using this technique<sup>21</sup>.

Calibration weighting is designed to make adjustments for known differentials in response rates between different age by gender subgroups and households with different age and gender composition. For example, a 24 year old male living alone may be less likely to respond to the survey than one living with a partner and a child. The procedure therefore gives different weights to different household types based on their age and sex

---

<sup>21</sup>Calibration weights are applied to the data by the Home Office after the application of the design weights.

composition in such a way that the weighted distribution of individuals in the responding households matches the known distribution in the population as a whole.

The effects of applying these weights are generally low for household crime, but are more important for estimates of personal crime, where young respondents generally have much higher crime victimisation rates than average but also lower response rates to the survey. However, crime trends since the 1996 survey have not been altered to any great extent by this new system of weighting.

## 8. Comparing key survey variables with the population

The achieved sample was weighted in order to be representative of the population in private households in England and Wales. A series of comparisons is presented in the following tables, showing to what extent the achieved 2003-4 BCS sample reflected the population as a whole, both before and after applying the appropriate weights.

The regional distribution of the adult population aged 16 or over by Government Office Region is shown in Table 8.1. This shows that the main discrepancy in the achieved sample was the under-representation of London, reflecting the lower response rates achieved here. The proportion of the achieved sample achieved in London was still about 2% less than might be expected. Consequently many of the other regions were slightly over represented compared with their population distribution.

**Table 8.1 Distribution of the 2003-4 BCS achieved sample by Government Office Region compared with the population**

| <b>Government Office Region</b>                                      | <b>Weighted Core Sample<sup>22</sup></b> | <b>Mid-Year 2002 population estimates (16+)</b> | <b>Difference (Weighted core % less population %)</b> |
|--|--|---|---|
|  | <b>%</b>                                 | <b>%</b>  | <b>%</b>  |
| North East   | 5.3                                      | 4.8   | +0.5  |
| North West   | 14.1                                     | 12.8  | +1.3  |
| Yorkshire & Humberside   | 9.4                                      | 9.5   | -0.1  |
| East Midlands  | 8.1                                      | 8.1   | 0   |
| West Midlands  | 10.3                                     | 10.0  | +0.3  |
| East of England  | 10.4                                     | 10.3  | 0.1   |
| London   | 12.2                                     | 14.0  | -1.8  |
| South East   | 14.9                                     | 15.3  | -0.4  |
| South West   | 9.5                                      | 9.6   | -0.1  |
| Wales  | 5.8                                      | 5.6   | +0.2  |
| <i>Base(All)</i>   | <i>37,213</i>                            | <i>42,056,500</i>                               |   |
| Source: Mid-2003 Population Estimates, Office of National Statistics |  |   |   |

<sup>22</sup> Prior to the calibration weighting applied by the Home Office

Table 8.2 shows similar comparisons between the achieved sample and the population in relation to age and sex distribution. The key feature here was a slight under representation of people aged under 35. Men were also under represented in the achieved sample compared with the population. All of these patterns are fairly common in large scale surveys and reflect the slightly lower response rates achieved among these particular groups.

Although not reported here the age and sex distribution of the achieved sample are further corrected by the Home Office at the analysis stage through the application of calibration weights (see section 7).

**Table 8.2 Age and sex distribution of 2003-4 BCS achieved sample compared with the population**

|  | <b>Weighted<br/>Core Sample</b> | <b>Mid-Year 2000<br/>population<br/>estimates<br/>(16+)</b> | <b>Difference<br/>(Weighted core %<br/>less population %)</b> |
|--|---------------------------------|---|---|
|  | <b>%</b>                        | <b>%</b>  | <b>%</b>  |
| <b>Age Group</b>   |                                 |   |   |
| 16-19  | 5.4                             | 6.4   | -1.0  |
| 20-24  | 6.4                             | 7.7   | -1.3  |
| 25-34  | 15.0                            | 17.0  | -2.0  |
| 35-44  | 18.8                            | 19.0  | -0.2  |
| 45-54  | 17.6                            | 15.9  | +1.7  |
| 55-64  | 16.5                            | 14.1  | +2.4  |
| 65-74  | 11.6                            | 10.4  | +1.2  |
| 75-84  | 7.2                             | 7.2   | 0   |
| 85 and over  | 1.5                             | 2.3   | -0.8  |
|  |                                 |   |   |
| <b>Sex</b>   |                                 |   |   |
| Male   | 47.1                            | 48.4  | -1.3  |
| Female   | 52.9                            | 51.6  | +1.3  |
| <i>Base::</i>  | <i>37,213</i>                   | <i>42,412,900</i>   |   |
| Source: Mid-2003 Population Estimates, Office of National Statistics |                                 |   |   |

Other comparisons between the achieved sample and the population are summarised in Table 8.3. This shows that private rented households and households with no cars were slightly underrepresented in the achieved sample, as were non-white households. Similarly, single person households were slightly underrepresented. As with the age and sex distribution these small differences probably reflect differential response rates on the survey.

**Table 8.3 Other characteristics of 2003-4 BCS achieved sample compared with the population**

|  | <b>Weighted<br/>Core<br/>Sample</b> | <b>2001<br/>Census<br/>estimates</b> | <b>Difference (Core<br/>sample % less<br/>population %)</b> |
|--|-------------------------------------|--------------------------------------|---|
|  | %                                   | %                                    | %   |
| <b>Tenure</b>                                      |                                     |                                      |   |
| Owners   | 70.6                                | 69.0                                 | +1.6  |
| Social rented sector                               | 19.5                                | 19.0                                 | +0.5  |
| Private rented sector                              | 9.9                                 | 12.0                                 | -2.1  |
|  |                                     |                                      |   |
| <b>Car availability</b>                            |                                     |                                      |   |
| None   | 24.4                                | 26.8                                 | -2.4  |
| One  | 42.7                                | 43.8                                 | -1.1  |
| Two or more  | 32.9                                | 29.4                                 | +3.5  |
|  |                                     |                                      |   |
| <b>Household size</b>                              |                                     |                                      |   |
| One person household                               | 27                                  | 29                                   | -2  |
| Two person household                               | 37                                  | 35                                   | +2  |
| Three person household                             | 16                                  | 15                                   | +1  |
| Four + person household                            | 20                                  | 21                                   | -1  |
|  |                                     |                                      |   |
| <b>Ethnic Group<br/>(Grouped)</b>                  |                                     |                                      |   |
| White  | 92.4                                | 92.1                                 | +0.3  |
| Mixed  | 0.6                                 | 1.2                                  | -0.6  |
| Asian  | 3.7                                 | 4.0                                  | -0.3  |
| Black  | 2.1                                 | 2.0                                  | +0.1  |
| Other  | 1.3                                 | 0.8                                  | +0.5  |
| <i>Base: All</i>                                   | <i>37,213</i>                       |                                      |   |
| Source: 2001 Census, Office of National Statistics |                                     |                                      |   |