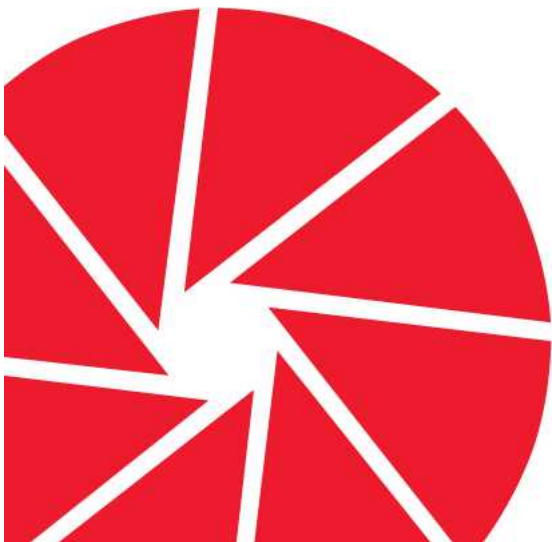


The British Crime Survey: A Review of Methodology

Peter Lynn (National Centre for Social Research) and
Dave Elliot (Office for National Statistics)



The British Crime Survey: A Review of Methodology

Peter Lynn and Dave Elliot

Prepared for The Home Office

P.1974

March 2000

Contents

List of tables.....	2
List of figures.....	2
Acknowledgements.....	2
1 SUMMARY.....	3
2 PRINCIPLES.....	5
2.1 Components of Accuracy.....	5
2.1.1 Variance.....	6
<i>Population variability</i>	6
<i>Sample size</i>	6
<i>Design effects due to clustering</i>	7
<i>Design effects due to stratification</i>	7
<i>Design effects due to variable sampling fractions</i>	7
<i>Interviewer variability</i>	7
<i>Coder variability</i>	8
2.1.2 Bias.....	8
<i>Sampling bias</i>	8
<i>Non-response bias</i>	8
<i>Other sources of bias</i>	9
2.2 Key Estimates.....	9
2.3 Targeting of Resources.....	10
2.4 Current BCS Design.....	11
3 SAMPLING OVER TIME.....	13
3.1 Sampling over a Year.....	13
3.1.1 Continuous field work.....	13
3.1.2 A spliced design.....	15
3.1.3 Implications of continuous field work for reporting.....	16
3.2 Sampling between Years.....	17
3.2.1 Panel design to increase precision.....	17
<i>Choice of panel units</i>	18
<i>What proportion of units to retain?</i>	19
<i>The retention pattern</i>	20
3.2.2 Panel design for longitudinal analysis.....	21
4 SAMPLE SELECTION.....	23
4.1 Sampling Frame.....	23
4.2 Sampling Stages.....	23
4.2.1 Clustering households within PSUs.....	23
4.2.2 Sampling individuals within households.....	27
4.3 Sample Stratification.....	28
4.3.1 Proportionate Stratification.....	28
4.3.2 Disproportionate Stratification.....	29
<i>Police Force Areas</i>	29
<i>Ethnic Groups</i>	30
5 DATA COLLECTION.....	33
5.1 Recall Errors.....	33
5.2 Seasonal Effects.....	35
5.3 Coding.....	35

6	NON-RESPONSE WEIGHTING	37
6.1	The Nature of Non-Response on the BCS	37
6.2	Non-Response Weighting Methods	38
6.3	Construction of Initial Weights.....	39
6.4	Calibration Options	40
6.5	Recommendations	41
7	GEOGRAPHICAL ANALYSIS.....	43
7.1	Defining Areas.....	43
7.2	Building Models.....	45
7.3	Using the Models	46
8	RECOMMENDATIONS.....	47
8.1	Further Methodological Work	47
8.2	Immediate Recommendations	48
8.3	Radical Recommendations	48
	REFERENCES	51

List of tables

Table 1:	Distribution of interviewing over months of the year (BCS 1996 and 1998).....	13
Table 2:	Example publication schedule for a continuous survey.....	17
Table 3:	Estimated reduction in standard errors from 50% rotation of PSUs	20
Table 4:	Estimated design effects associated with different degrees of sample clustering	25
Table 5:	Summary of impact of three sample designs on standard errors	26
Table 6:	Estimated design effects for alternative sample distributions over PFAs.....	30
Table 7:	Impact on precision of including ethnic boost sample for total population estimates (BCS 1996).....	32

List of figures

Figure 1:	A 2-period 50% sample rotation pattern.....	20
-----------	---	----

Acknowledgements

The authors are grateful to the Research Development and Statistics Directorate of the Home Office for funding this review, and particularly to Chris Kershaw and Pat Mayhew for their encouragement and support. We are also grateful for the advice and help of colleagues, especially Jon Hales and Paul Clarke.

1 SUMMARY

This report reviews some aspects of the methodology of the British Crime Survey (BCS). The review was commissioned in the context of plans to move the survey to an annual cycle and to increase substantially the sample size for the survey. The move to an annual cycle was originally proposed as a means of increasing the frequency with which survey-based estimates could be published, while the increase in sample size was motivated primarily by a need to increase the precision of certain estimates, notably those of rates of violent crimes. Subsequent to those proposed changes to the BCS design, the DETR published (in December 1999) the performance indicators to be used under the Best Value initiative (DETR, 1999). For Police Forces, these include measures of level of crime, fear of crime, feelings of public safety and confidence in the criminal justice system and its component parts. The DETR document specifies that these measures are to be provided by the BCS.

This review was therefore to consider the implications of all of these demands upon the BCS (more frequent estimates, more precise estimates of levels of violent crime, estimates of sufficient precision to provide performance indicators at the Police Force level). To structure the review, the Home Office identified a set of issues upon which they particularly wanted advice. Broadly, these were:

- Sample clustering
- Geographic analysis
- Sample stratification
- Adjustment for non-response
- Sampling of ethnic minorities

The review was carried out within a limited timetable and budget and is therefore not comprehensive. Rather, the intention was to focus upon aspects of the survey design that appeared to require urgent attention or to offer the possibility of significant cost-efficiency improvements. The discussion is limited mainly to sample design and estimation. Only a few aspects of data collection and data capture are mentioned. Questionnaire content and design, for example, was deemed to be outside the remit of the review.

However, the review has attempted to place the issues discussed within a unifying framework. That framework is described in section 2 of this report. In summary, the approach adopted is to consider the effects of design alternatives upon the accuracy of estimation and upon costs, seeking to identify design features which maximise the accuracy relative to the cost.

The specific design features considered are discussed in the subsequent five sections of the report. Section 3, *Sampling Over Time*, covers sampling within and between years. This raises the possibility of continuous fieldwork and discusses options for a

panel element. Section 4 addresses other aspects of sample design, including the stratification of primary sampling units (PSUs), the clustering of sample households within PSUs and the selection of individuals within households. Section 5, Data Collection, discusses the impact of the survey design on response errors – particularly recall errors and seasonal effects. Coding errors are also discussed. Section 6 states the case for the introduction of non-response weighting on the BCS and describes how that might be done, while section 7 considers the possibilities for geographical analysis. Then, recommendations arising from the review are summarised in section 8, where they are divided into three groups: recommendations for further methodological work, for immediate design changes, and for design changes that might require more discussion and development work.

2 PRINCIPLES

This section describes the principles underlying this review. We believe that the Home Office should be seeking to identify ways of optimising the trade-off between the cost of the BCS and its value. This can be thought of as maximising the value for a given cost, or minimising the cost for a given value.

One of the main components of value of the survey is the accuracy of the survey estimates. There are other components such as relevance, validity and timeliness, but these are mainly influenced by factors outside of the remit of this review. Thus, these components of value will not be discussed in detail, but comment will be made whenever issues discussed in this review impinge upon them. The main focus of the review will be upon the accuracy of estimates.

Within the constraints of achieving the survey's aims, the review is therefore seeking to identify changes to the design and implementation of the survey which should either:

- Increase the accuracy of estimates to an extent which outweighs any increase in cost; or
- Reduce costs to an extent which outweighs any loss in accuracy.

It has also been assumed that the budgetary constraint is equivalent to the ability to carry out approximately 40,000 interviews per annum.

With this in mind, the review has examined a number of aspects of the survey design and implementation. Each is discussed in turn, culminating in a set of recommendations (section 8). First, we describe what is meant by the accuracy of estimates.

2.1 Components of Accuracy

The accuracy of a survey estimate can be thought of in simple terms as the expected magnitude of the difference between the estimate and the true population value. The larger the difference, the less accurate an estimate. For a fuller description and discussion see, for example, Groves (1989). In the survey methodological literature, this quantity - the magnitude of the difference between the estimate and the true population value - is usually referred to as the total survey error.

A large number of factors contribute to the total survey error. Some of these factors can be thought of as sources of bias. A bias is an effect which, if the survey design was implemented repeatedly many times, would cause the estimates to be systematically different (i.e. in a particular direction) from the true value. Other

factors are a source of random variance. If a bias-free survey design was implemented many times, the estimates would not differ systematically from the true value, but they would vary from one another. In other words, the average of all the possible estimates would equal the true value, though any one estimate (from one implementation of the survey) would not necessarily do so. With a biased survey design, even the average of all the possible estimates would not equal the true value. Total survey error (i.e. the accuracy of a survey estimator) can be measured by a statistical quantity known as the mean square error, which is equal to the sum of the variance and the squared bias.

2.1.1 Variance

The main sources of variance in survey estimates are the following:

- Population variability
- Sample size
- Sample design effects due to:
 - clustering
 - stratification
 - variable sampling fractions
- Interviewer variability
- Coder variability

Population variability

For a given survey design, the greater the variability in the population of the concept being measured, the greater will be the variance of survey estimates based upon those measures. For example, suppose that a survey includes the following two measures: A. household income from all sources, B. household income from state benefits. If in the population there is greater variability in A than in B, then estimates of the mean of A will be subject to greater variance (larger standard errors) than estimates of the mean of B. It is (partly) for this reason that different estimates from the same survey, based on the same (sub-)sample, have different standard errors. In designing a survey, it is therefore important to know something about the nature of the estimates to be made (see section 2.2).

Sample size

Other things being equal, larger samples result in survey estimates having smaller variance (smaller standard errors). Variance is inversely proportional to sample size, and hence standard errors and confidence intervals are inversely proportional to the square root of sample size. For example, doubling a sample size will tend to reduce standard errors by around 29% (the multiplying factor being $1/\sqrt{2}$).

Design effects due to clustering

The BCS sample, like those for most large-scale surveys, is clustered within a sample of postcode sectors. This clustering tends to result in larger variance for survey estimates than would be the case with a simple random sample of the same size. This is because people living within a postcode sector tend to be more similar to one another (in terms of survey variables) than people within the population as a whole. In other words, an extra interview within the same sector as other interviews provides less new information than would be provided by an extra interview from a random selection within the whole population.

Design effects due to stratification

The BCS sample design involves stratification of the sampling frame at two levels – postcode sectors and addresses within sectors. This ensures that the sample correctly reflects the population distribution in terms of the variables used to define the stratification. Stratification has a beneficial effect of reducing the variance of estimates, but the magnitude of the effect depends on the choice of strata. Careful construction of strata is therefore important.

Design effects due to variable sampling fractions

If some population elements have selection probabilities which differ from those of others, this will tend to have an effect on the variance of estimates. In some circumstances, the probabilities can be controlled in a way which reduces standard errors, but usually variation in selection probabilities has the effect of increasing standard errors.

This only affects estimates that involve combining cases that had different selection probabilities. On the BCS, between 1992 and 1998 selection probabilities for households have only differed between the “inner-city” and “other” strata. For individuals, probabilities have additionally differed between households containing different numbers of individuals. Thus, for example, an estimate based on all households in inner cities would not have attracted any design effect due to variable sampling fractions, whereas an estimate based on all individuals in inner cities would have.

Interviewer variability

Interviewers can be a source of variance due to more or less subtle differences between them in the ways they read questions, the tone of their voice, and many other factors. The interviewers who work on a particular round of BCS could be thought of as a sample from some conceptual population of all interviewers. A different sample of interviewers, interviewing the same sample of respondents in identical circumstances, might have obtained slightly different data. Special studies have been set up to measure interviewer variability and have sometimes found it to be non-negligible (e.g. Bailey et al, 1978; Hartley and Rao, 1978; Kish, 1962).

Coder variability

Coders can also introduce variance due to differences in the way they perform their task (Biemer and Stokes, 1991; Kalton and Stowell, 1979). This is potentially important for the BCS, given its reliance on office coding of offences, which is central to the survey's key estimates.

2.1.2 Bias

The main sources of bias in survey estimates are the following:

- Sampling bias
- Non-response bias
- Respondent bias
- Interviewer bias
- Coder bias

Sampling bias

This can result from some elements or subgroups in the population being given a zero probability of selection (either deliberately or by accident) or from the use of differential selection probabilities which go uncorrected by weighting (for example, when the frame contains duplicate entries, but these cannot be identified). An example of giving some cases a zero probability of selection is the practice of excluding from a sampling frame addresses in very remote locations, where data collection costs might be very high.

Non-response bias

Survey non-response is important because it may introduce bias into survey estimates. In other words, all the desirable qualities of a carefully designed and selected sample can be compromised by the fact that data cannot be obtained for every unit in the selected sample. Survey non-response can be thought of as an extension of sampling: sample design, sample selection and survey response together form a process which determines the constitution of the final sample of cases for which data are observed and upon which survey estimates are based. Specifically, a sample may be designed so as to provide precise, unbiased estimates, but non-response will then introduce unwanted bias unless the non-respondents happen to be a completely random subset of the selected sample. The developing theories of household survey non-response (Groves, Cialdini and Couper, 1992; Groves and Couper, 1998) emphasise that survey participation is not at all random. Indeed, most empirical studies of non-response have found evidence of bias. However, the exact nature and magnitude of the bias is likely to vary between surveys and to depend on a multitude of factors.

Other sources of bias

Having already argued that the behaviour and characteristics of both interviewers and coders can affect the data collected, it should be obvious that these effects need not be random. They could be systematic. For example, if interviewers tend to probe insufficiently at an open-ended question which asks a respondent to list all occurrences of X in the past year, they will tend to get under-reports. In consequence, survey estimates of the mean number of occurrences of X will tend to be under-estimates, i.e. to be downwardly biased.

In fact, this is an example of how interviewer behaviour (probing) can affect respondent behaviour (reporting). There are also other factors which may affect respondent behaviour in a systematic way, thus causing respondent bias (also sometimes referred to as response bias). Social desirability effects are one example. Respondents may feel embarrassed or uncomfortable about reporting certain behaviour or experiences that may be perceived as sensitive or unacceptable. Obvious examples include illegal drug use, excessive alcohol consumption, taking part in criminal activities, and certain sexual activities. The effect will tend to be an under-reporting, again leading to downwards bias in estimates of the prevalence of such behaviour or experiences (Tourangeau and Smith, 1996).

2.2 Key Estimates

The concept of total survey error, and all of its components, applies to each and every estimate based upon a survey data set. Even though the sample design, implementation and procedures are the same for all sample cases, the way in which the various error sources affect estimates will differ across the estimates. Some of the concepts measured by the survey will cluster within postal sectors to a greater extent than others, and will thus attract larger design effects due to clustering, even though the nature of the clustering is the same. Some of the concepts will be more susceptible than others to interviewer errors, and so on.

Thus, estimation of survey errors must be relative to particular survey estimates. One cannot talk about errors in the abstract unless one is prepared to make some gross simplifications and assumptions. It is therefore important to consider the nature of the key survey estimates, so that errors can be considered in the context of those estimates.

We believe the key BCS estimates to fall into the following groups:

- Victimisation rates for main offence groups (nationally and for PFAs)
- Victimisation rates for detailed categories of offences (nationally)
- Changes in victimisation rates (year-on-year, for both the above)
- Levels of fear of crime¹ (nationally and for PFAs)
- Differences in levels of fear of crime (between PFAs)
- Changes in levels of fear of crime (year-on-year)

¹ "Fear of crime" is used here as a shorthand for fear of crime, feelings of public safety and confidence in the police.

These different groups each require rather different considerations to be taken into account in terms of the relationship between survey design and accuracy of estimation. Some of these differences are relevant to the discussion at various points in this review. Within each of the groups, a number of different specific estimates are made. Differences between these estimates are also important.

2.3 Targeting of Resources

As stated earlier, we believe that ideally the BCS – like any survey – should seek to optimise the trade-off between survey cost and survey accuracy. Essentially this means targeting the available survey resources at those aspects of the design and implementation that are most likely to produce worthwhile improvements in accuracy. Based on our knowledge of the BCS and of survey errors more generally, we have initially focussed attention on the following areas:

- ◆ For national estimates, the bias component of error could be more important.

The BCS sample size is large and the design reasonably efficient, so the random variance component of total survey error for national estimates is relatively small.

- ◆ For PFA estimates, variance is likely to be more important.

This is simply because the sample sizes are much smaller for PFA estimates than for national estimates.

- ◆ The main bias components are likely to be non-response bias and respondent bias

Earlier studies have demonstrated the presence of non-response bias, at least in certain respects. The BCS has never implemented any correction strategy for non-response bias. Recall errors in particular are likely to be a source of respondent bias. This has been demonstrated on other surveys, including the US National Crime Victimization Survey (MacKenzie et al, 1989). Sampling bias should be trivial given what we know about the sampling frame and its relationship with the defined target population for the survey. It is also possible that coder bias could be significant.

- ◆ The main variance components are likely to be sample size, clustering, the effect of variable sampling fractions, and coder variance.

(The effect of variable sampling fractions between households only applies to national estimates, whereas the effect of variable sampling fractions between individuals in households of different sizes affects both national and PFA estimates.)

- ◆ For estimates of change, the independence of the samples being compared is also important.

Whenever estimates from two samples are compared, there is potential to reduce standard errors if the samples can be made non-independent. This is one of the motivations behind the use of panel survey designs.

2.4 Current BCS Design

This review has taken the design of the BCS in 2000 as the benchmark against which to compare possible design alterations. The BCS 2000 design is broadly similar to the design of the previous surveys in the series (Hales and Stratford, 1997; Hales and Stratford, 1999). The one major change since the 1998 survey is that the over-sampling of “inner city” areas has been discontinued. Instead, primary stratification for the 2000 survey was by PFA, with disproportionate over-sampling of the smallest PFAs, in order to ensure a minimum number of achieved interviews per PFA. The sample design therefore involved a core sample of 28,992 addresses in 906 postal sectors (i.e. 32 in each). In addition, an ethnic minority boost sample was selected, with a design similar to that used on the 1996 survey (there was no ethnic minority boost in 1998).

3 SAMPLING OVER TIME

Survey samples are selected in time as well as space, but the choice of time point(s) is often arbitrary at best. In this section, we review the representation of the population over time by BCS.

3.1 Sampling over a Year

The current design involves selecting the sample and collecting the data within a relatively short period of two or three months early in the calendar year (Table 1). The main reason for this is to be consistent with earlier rounds of the survey. It is likely that the original reasons for this design were connected with the (perceived) advantages of offering respondents a calendar year as a reference period, which is easiest if the interview is taking place early in the following year. However, the sample size has grown considerably during the life of the BCS. It is becoming increasingly difficult to constrain the field work to the same short period of time and will become significantly more difficult when the sample size is substantially increased further in 2001. We are seriously concerned about the feasibility of tackling a sample of well over 50,000 addresses during the traditional BCS period.

Table 1: Distribution of interviewing over months of the year (BCS 1996 and 1998)

	Percentage of core sample interviews			
	January	February	March	April or later
BCS 1992	25	53	15	7
BCS 1994	6	37	37	21
BCS 1996	21	46	26	7
BCS 1998	12	44	26	17

Spreading the sample over more of the year would reduce the problem. This could bring an associated cost saving, as fewer interviewers would be needed (as many could work on a number of sample points at different times in the year). For example, it might be beneficial to split the field work into two tranches, one to be undertaken between January and May (as in recent rounds) and the other between July and November.

3.1.1 Continuous field work

A more radical suggestion would be not merely to spread the field work out a bit more, but to move to completely continuous field work. Distributing the sample evenly over the whole year in this way would have a number of advantages. It would provide a guarantee against possible seasonal effects in the survey estimates.

It would also provide a mechanism for the production of rolling estimates relating to any desired time period. This would improve the timeliness and frequency with which survey results could be produced and the flexibility of the survey for providing estimates. For example, the survey could continue to produce estimates each autumn, based on the most recent twelve months of field work, but could also publish estimates based upon financial years (to be consistent with the published reported crime rates) and/or intermediate updates, say each spring. We therefore think that moving the survey to a continuous fieldwork basis would have considerable advantages.

However, we also recognise that in the short-term such a move would cause concerns about the possible impact that the change in design might have upon the survey estimates themselves and, in particular, upon comparisons of years before and after the change. To address these concerns, we recommend that in the first year of continuous fieldwork a “spliced design” should be used, which would enable direct estimation of the impact of the change upon estimates. An adjustment factor could then be incorporated into any time series analysis if that proved necessary.

Any impact on survey estimates of victimisation rates is likely to be caused by differences in the way the old and new designs affect respondent recall. These differences could have two components: differences in elapsed time between the end of the reference period and the interview date and differences in reference period itself. (It should be noted that the first of these two effects could already be present in the BCS data, as the mean elapsed time since the end of the reference period has varied between rounds of the survey - Table 1.)

The impact of either of these two sources of measurement effects could also be affected by any seasonal trends in victimisation. (The issue of recall periods is discussed further in section 5.1.)

In addition to victimisation rate estimates, some of the fear of crime and other measures could also be subject to seasonal effects. For example, it is possible that questions such as, “How safe do you feel walking alone in this area after dark?” and “how often do you usually go out after dark?” could be interpreted rather differently – and therefore produce different responses – depending on whether they are asked in January, when it might get dark around 4pm, or in June, when it might not get dark until 10pm.

We therefore propose that the survey should move to a fully continuous operation, involving a monthly cycle of field work, the sample each month being a random subset of the annual sample. Victimisation questions should then refer to the twelve calendar months ending with the month before the interview² (e.g. an interview on 11th June 2002 would ask about the period June 2001 – May 2002)

² In fact, it could also be worthwhile reviewing whether a 12-month recall period is the most appropriate for all the victimisation questions, given that some crimes are likely to be far more salient and memorable than others.

3.1.2 A spliced design

As suggested above, we propose that an experimental (“spliced”) design should be built in to the first year of continuous operation, to allow assessment of any effects on the survey estimates. This design would thus provide a “bridge” between the old and new designs, allowing compatible comparisons to be made in either direction. Here we describe how that design would be implemented and what analyses it would enable.

The basic design would be as follows. (There is also an optional variant that we describe subsequently.)

- The annual sample of PSUs would be randomly (systematically) allocated to each of the twelve months of the year;
- For each of the six months January to June, sample addresses should be randomly (systematically) allocated to two equal-sized subgroups – group A and group B. This allocation should be done *within* PSUs (in order to control for area and interviewer effects);
- At group A addresses, the victimisation questions should continue to refer to the previous calendar year, as currently;
- At group B addresses, the questions should refer to the twelve months up to and including the last completed month.

Group A addresses could then be used to make estimates comparable to previous years (weighting to the previous distributions of interviews over months if necessary). Group B addresses, plus all interviews from July to December, could be used to make estimates comparable to future years. Comparison of these two sets of estimates would provide an assessment of the impact of the change in methodology. If any significant differences are found, adjustment factors could be calculated. If these prove necessary, we would recommend that they take the form of an adjustment to the estimates published in previous years to make them comparable with current and future ones (rather than *vice versa*), as this reflects the likelihood that the new design provides less biased estimates than the old one.

In addition, this design would allow any overall differences to be partitioned into the difference due to the change in timing of the interviews and the difference due to the change in the reference period. Comparisons of estimates from A and B addresses (weighted to the previous distribution over months if necessary) would provide an estimate of the latter, while comparison of B interviews with interviews from July to December could provide the former.

A refinement to this design could be envisaged, as follows:

- Sample addresses could also be randomly allocated to two subgroups for each of the six months July to December – group C and group D. (Again, this allocation should be done *within* PSUs);

- At group C addresses, the victimisation questions should refer to the twelve month period from July of the previous year until June of the current year;
- The treatment of group D addresses is identical to that of group B.

This variant could be desirable if there is any uncertainty about asking about a reference period of the twelve months ending with the last completed month – in other words, if it seemed possible that the group A/C treatment could prove to be a more reliable data collection method than the group B/D approach.

This extra component of the randomised design would allow an overall comparison to be made between:

- Asking about the twelve months ending with the last completed month (groups B and D); versus
- Asking about the twelve months ending last December/June (whichever more recent) (groups A and C).

This spliced design would lend itself to this type of estimation of the effects on any survey estimates – not only levels of crime but also fear of crime measures, feelings of public safety, and so on.

3.1.3 Implications of continuous field work for reporting

The continuous field work design would have a number of positive implications for reporting of the survey data.

First, it would in principle be possible to produce estimates at whatever frequency was required. Second, these estimates could be produced as rolling averages of any number of months of survey data – i.e. quarterly estimates, six-monthly estimates, annual estimates, etc. In practice, we would suggest that frequency of publication should probably be limited to no more frequent than quarterly and that the number of different time periods for which estimates are produced should probably be limited to two (quarterly and annual). One example of a possible publication schedule is shown in Table 2, though we would not recommend that the BCS should move immediately to such an ambitious publication schedule. The first year of continuous data collection might be treated as a “dry-run” for such publication, with results initially released only internally.

It should be noted that the publication scheduled for September 2002 in the example would be equivalent to the current autumn publication of annual estimates relating to the previous calendar year, though not identical in construction. These estimates would be based upon data collected between July 2001 and June 2002. But each of these twelve months of interviews will relate to a different reference period. The reference periods will range from July 2000 – June 2001 to June 2001 – May 2002. Thus, the reference periods will be centred around June 2001 and the estimates might reasonably be thought of as estimates for the calendar year 2001. But contributing to the estimates will be twelve monthly samples reporting on June 2001, eleven that

include May 2001 and eleven that include July 2001, and so on, down to just one monthly sample that includes July 2000 and one that includes May 2002.

Table 2: Example publication schedule for a continuous survey

Approximate date of publication	Quarterly estimates – data collection period	Annual estimates – data collection period
June 2002	Jan-Mar 2002	Apr 2001 – Mar 2002
September 2002	Apr-Jun 2002	Jul 2001 – Jun 2002
December 2002	Jul-Sep 2002	Oct 2001 – Sep 2002
March 2003	Oct-Dec 2002	Jan 2002 – Dec 2002

There are, of course, a number of alternative ways to construct estimates based upon moving averages and, though perhaps the simplest, the one described here may be considered undesirable. Estimates based on twelve monthly samples each reporting on each of the twelve months of the calendar year 2001 could also be made, of course, but not until approximately February 2003. Alternatively, an estimate based solely on crimes reported as happening in 2001 could be made in autumn 2002, but would have to involve an up-weighting of crimes reported from August onwards, which would be based upon a smaller reporting sample size than the earlier months.

This greater frequency of reporting that would be made possible by a continuous BCS could be of particular value now that the recorded crime statistics are to be published twice a year (relating to April-March and October-September periods, neither of which coincide with the current BCS reference period).

3.2 Sampling between Years

Given that a main interest of the BCS is in monitoring changes over time, we could usefully consider whether a panel design might offer some potential improvement in the precision of estimates of change. There are two fundamental reasons for considering panel designs. One is to enable genuine longitudinal analysis. The other is to increase the precision of estimates of change between one survey reference period and the next. Both of these considerations are relevant to the BCS, though it is perhaps the latter that is particularly important. However, a panel design can also have other advantages and disadvantages which must also be taken into account.

3.2.1 Panel design to increase precision

To increase the precision of estimates of change between one survey reference period and the next it is necessary to retain a proportion of sample units from one period to the next, so that the two samples are no longer independent. If there is a correlation over time within sample units in the measures being compared, then this dependence of the samples will reduce the standard errors of measures of change. Thus, the higher the correlation, the more likely it is that a panel design would be worthwhile.

There are three key dimensions which define a panel design: the units to retain, the proportion to retain, the retention pattern over time. Thus, there are three corresponding design questions to be addressed:

- ◆ Which units should form the panel elements?
- ◆ What proportion of these units should be retained from one survey period to another?
- ◆ For how many survey periods should these units be retained?

We now address each of these questions in turn, with reference to the aim of increasing the precision of estimates of change over time.

Choice of panel units

For the BCS, there are three options worth considering. One is to re-interview the same people who were interviewed in an earlier survey period (an earlier round of the survey). Another is to retain the same addresses, but make a fresh random selection from amongst the residents at the address, regardless of whether these are the same people who were present at the time of the earlier survey. The third option is to retain the primary sampling units (postcode sectors) but make a fresh selection of addresses within each sector (which may or may not be controlled so as to avoid selecting any of the same addresses that were selected on the earlier occasion).

Retention of the sample of selected respondents would maximise the precision gains, as correlation over time is almost certain to be higher within persons than within addresses or postcode sectors. On the other hand, this option has some disadvantages. It is likely to be the most expensive option, owing to the need to trace movers and follow them up to new addresses (so the sample will become less geographically clustered than it was originally). It is also possible that non-response at the second time period, on top of the initial non-response, would cause the non-response bias to be considerable under this option. Furthermore, there are reasons to suppose that this second stage of non-response could be correlated with the key survey variables. Propensity to refuse at the second stage is likely to be related to the experience of the earlier interview, and this in turn is closely related to the respondent's victimisation experience(s). Interviews last longer with victims of crime than with non-victims, and are considerably longer still for multiple victims (Hales and Stratford, 1999). However, we do not consider this to be the most important consideration as a wealth of relevant information (the questionnaire data from the first survey period) is available to inform non-response weighting in this situation.

Another possible negative aspect of this design is that it introduces the possibility of conditioning effects. In other words, the experience of being interviewed previously on the BCS could affect a sample member's responses on a subsequent occasion. This would have to be carefully evaluated. In addition, there is a risk of the sample becoming "out-of-date." In other words, at the second survey period, the sample will be biased against recent immigrants and 16-year-olds.

In addition to producing precision gains for estimates of change, retention of a panel of individuals would also have other benefits. These are discussed below in section 3.2.2.

Retention of addresses rather than people will avoid the extra cost of tracing and interviewing movers, but it is likely also to produce a smaller precision gain. The precision gain could still be substantial, however, as in many cases it will be the same household, and even the same person, who is interviewed. For precisely this reason, however, the effect of the extra burden of a second interview on non-response could be almost as great as under the first option. This option also has a risk of the sample becoming biased over time, though the nature of the possible bias is rather different. In this case, the bias would be against people living in newly-built or newly-converted properties.

Retaining areas but selecting fresh addresses avoids all of the above disadvantages. It is also, however, the option likely to produce the smallest precision gains. The magnitude of the likely precision gains could be estimated using the data from the 1996 and 1998 surveys. A proportion of the PSUs in 1998 consisted of postal sectors first selected for the 1996 survey. However, the calculation of the impact of this design on the standard errors of measures of change is not straightforward and we have not been able to perform these calculations within the constraints of this review. We would recommend that the Home Office considers commissioning an assessment of the impact of this design on standard errors.

We believe that this third option, of retaining areas but selecting fresh addresses, may be the most promising one for the BCS. Information from BCS on repeat victimisation suggests that there will indeed be within-person correlation between years, but that this may not be high enough to provide a precision gain from retaining persons or addresses that would outweigh the disadvantages discussed above.

What proportion of units to retain?

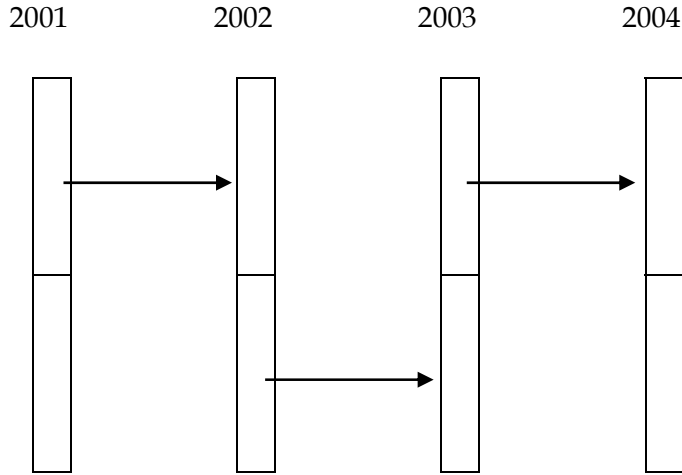
The higher the proportion retained, the greater the potential precision gain. On the other hand, the higher the proportion retained, the greater the risk of bias in the estimates of levels due to the sample becoming "out-of-date." In the BCS context, this bias issue is unlikely to be important, however. If the retention is only at the level of postcode sectors, the bias would only be against new addresses in newly-created sectors (new addresses in other sectors would still have a chance of inclusion as the addresses would only be selected at the time of the second survey period). Such addresses form a trivial proportion of the total.

We would therefore propose retention of the maximum proportion that is sustainable over time. If PSUs are to be retained for only one survey period, the maximum proportion would be 50%; if two periods, 33%, and so on (assuming a constant total sample size).

The retention pattern

It would be unduly complex to retain PSUs for more than one survey period. We would therefore recommend a 2-period 50% rotation design, the principle of which is illustrated in Figure 1.

Figure 1: A 2-period 50% sample rotation pattern



A 2-period 50% rotation design maximises the precision gains for estimates of changes between consecutive survey periods, but will not bring gains for longer-term comparisons.

We have estimated the likely precision gains to result from such a design (Table 3). It can be seen that the gains might be quite worthwhile, particularly for estimates with a ρ -value greater than 0.03. For example, assuming a similar average cluster size to the BCS 2000, standard errors might be reduced by around 10% for an estimate with $\rho = 0.03$ and 13% if $\rho = 0.05$.

Table 3: Estimated reduction in standard errors from 50% rotation of PSUs

	PSUs (mean interviews per PSU)		
	2000 (20.0)	1800 (22.2)	1500 (26.7)
0.01	3.9	4.2	4.9
0.015	5.4	5.9	6.7
0.02	6.7	7.3	8.3
0.03	9.0	9.6	10.7
0.05	12.3	13.0	14.1
0.10	16.9	17.6	18.5

Note: The figures in the body of the table are estimates of $(se_N - se_R) / (se_N)$, where se_N is the standard error of a difference between two years assuming no sample rotation and se_R is the standard error with 50% rotation of PSUs. These estimates are based upon a standard variance model for 2-stage repeated sample designs (Holt and Farver, 1992) and assume area-level year-on-year correlation of around 90%.

We propose that “survey period” should be interpreted to mean a year, as the most important BCS estimates are probably annual estimates. However, it would be

possible to simultaneously improve the precision of estimates of change between equivalent quarters, if the PSUs are re-used twelve months after their initial use. For example, half of the PSUs sampled and issued in Jan-Mar 2001 should be re-used in Jan-Mar 2002, half of those first used in Apr-Jun 2001 should be re-used in Apr-Jun 2002, and so on. Thus, the precision gains estimated in Table 3 above would apply to estimates of change between consecutive annual estimates (e.g. from Apr 2001-Mar 2002 to Apr 2002-Mar 2003) and between quarterly estimates one year apart (e.g. Apr-Jun 2001 to Apr-Jun 2002) but not between other combinations of time periods.

3.2.2 Panel design for longitudinal analysis

A panel of individual respondents would permit longitudinal analyses that are not currently possible. One obvious example would be the ability to examine the extent and nature of repeat victimisation over a period of two years. Another would be the ability to separate out the components of change in measures of fear of crime or feelings of safety (i.e. to estimate gross flows as well as net flows). Such extensions potentially represent major advances in our knowledge of the nature of crime and perceptions of crime.

An individual panel would also allow some evaluation of recall effects. For example, suppose a set of respondents were asked in January 2001 about victimisation during 2000 and again in January 2002 about victimisation during 2001. Then, "telescoping" might result in some incidents from November and December 2000 being reported (incorrectly) on the second occasion. Careful cross-checking with the incidents reported on the first occasion should enable estimation of the extent of the phenomenon. Alternatively, over-lapping recall periods could be used. These have the advantage of providing a direct measure of the extent of net relative over- or under-reporting, though less data overall. Such evaluation of recall effects could aid interpretation and understanding of the main BCS findings.

We suggest that there might be advantages in mounting a panel sample of individuals. Initially, this could consist of a limited sub-sample of BCS respondents. Once the feasibility of the methodology and the utility of the results had been established, then the panel might be incorporated as a regular feature of the BCS.

Our recommendation is that initially a random subset of respondents to BCS 2001 should be re-interviewed twelve months later. This subset need not necessarily be more than perhaps 10% - 20% of all respondents. (This sample size would not support any analysis at the police force area level, but should be adequate for most total sample analysis.) The follow-up interview could be carried out by telephone where possible, in order to minimise the data collection costs (though that would introduce the possibility of mode effects). The success and utility of this exercise should then be reviewed before deciding whether or not to adopt a panel of respondents as a regular element of the BCS.

4 SAMPLE SELECTION

4.1 Sampling Frame

The Postcode Address File (PAF) has been used as the sampling frame for the BCS since 1992 (Lynn, 1997a). PAF continues to be the most suitable sampling frame for general population surveys in the UK, so its use is not discussed in this report. It is assumed that the BCS will continue to use PAF as the sampling frame.

4.2 Sampling Stages

Since 1994, the BCS sample has been selected in three stages: quarter postal sectors, addresses and individuals. From 1994 to 1998, selections were made separately within each of two strata – “inner cities” and other areas. From 2000, selections were made separately within each of a number of Police Force Areas. (See section 4.3.2 below for a discussion of these explicit strata.)

In this section, we review the number and nature of these stages of clustering.

4.2.1 *Clustering households within PSUs*

The reason for using a multi-stage sample design is to reduce the unit cost of data collection by “clustering” the sample within a number of small areas. The BCS is designed so that the sample addresses in one (quarter) postal sector form an efficient workload for one interviewer. For any given budget, the size of sample that can be afforded is much larger with a clustered design than with an unclustered one. This increase in sample size tends to increase the precision of survey estimates. However, clustering also results in a “design factor” that tends to reduce the precision of estimates for any given sample size. A clustered design is the most cost-efficient one if and only if the increase in precision due to the increase in sample size outweighs the loss in precision due to the design effect. In practice, this is usually the case for national surveys.

However, samples can be clustered to lesser or greater extents. There therefore remains a choice of the appropriate geographical units to be used to form the clusters, and the size of sample to select within each. Ideally, the choice should be made based upon good estimates of the relative precision per unit cost of each alternative design. This requires estimation of both the design factor and the unit cost of data collection associated with each design. Within the limited scope of this review, we have chosen to examine four alternative degrees of clustering. These were chosen on purely intuitive grounds, as being likely to be close to optimum.

The designs examined are:

- 1) 32 selected addresses per quarter sector, as in the 2000 BCS
- 2) 24 selected addresses per quarter sector
- 3) 32 selected addresses per full sector
- 4) 24 selected addresses per full sector

We have assumed an annual selected sample size of around 58,000 addresses³, though the comparison is in fact relatively insensitive to the sample size assumption. For each of designs 2, 3 and 4, we have attempted to estimate both the unit cost of data collection and the design factor relative to design 1.

We estimate that the increase in data collection costs relative to design 1 should be relatively small for all three of the alternative designs, but particularly small for design 3. We think that it could be misleading for us to make our own cost estimates, as these are likely to be organisation-specific.

The design factor, DEFT, depends on two quantities, the intra-cluster correlation (ρ) and the average number of interviews per cluster (b), thus:

$$DEFT = \sqrt{1 + (b - 1)\rho}$$

DEFT is the ratio by which the standard errors of an estimate produced from a survey using the clustered design will exceed those of a simple random sample of the same size. (The square of the design factor is referred to as the design effect, DEFF, which is the ratio of the sampling variances.)

We have estimated design factors in three ways. First, for design 2 we have estimated design factors directly for key BCS estimates, based upon those published in the BCS technical reports. We have estimated ρ for all of the victimisation rates for which design factor estimates are published and thus the design factors expected under designs 1 and 2. We estimate that the design factors under these two designs would average 1.19 and 1.15 respectively, a ratio of 0.96. In other words, standard errors might be reduced by 4%. This would be a very small precision gain and probably not worth the extra data collection cost. In other words, we would judge that design 2 is likely to be inferior to design 1 in terms of precision per unit cost.

For designs 3 and 4, we cannot directly estimate design factors for BCS key variables as there is no BCS data available for whole postal sectors. Instead, we have to use other data sources.

The first other source of data that we have used is a set of estimates of P for both sectors and quarter sectors, produced by Social Survey Division of ONS, based upon

³ Allowing for around 12% of PAF addresses being ineligible and a response rate of around 80%, this sample size should provide 40,000 completed interviews.

Table 4: Estimated design effects associated with different degrees of sample clustering

Variable (Survey estimate)	¹ / ₄ -sectors	sectors	DEFF		Design 4	DEFF-ratio		DEFT-ratio	
			Design 1	Design 3		Design 1 vs 3	Design 1 vs 4	Design 1 vs 3	Design 1 vs 4
% owner-occupiers (households)	0.190	0.125	5.09	3.69	2.99	1.38	1.70	1.17	1.30
% sharing accommodation (households)	0.064	0.037	2.38	1.80	1.59	1.32	1.50	1.15	1.22
% with > 1.5 persons per room (households)	0.017	0.010	1.37	1.21	1.16	1.12	1.18	1.06	1.09
% HoH in SEG 1-5 or 13 (households)	0.094	0.060	3.02	2.29	1.95	1.32	1.55	1.15	1.24
% Non-white head (households)	0.176	0.151	4.78	4.25	3.40	1.13	1.41	1.06	1.19
% no car (households)	0.130	0.100	3.80	3.15	2.59	1.20	1.47	1.10	1.21
% single parent households	0.024	0.015	1.52	1.32	1.24	1.15	1.22	1.07	1.11
% with limiting long-standing illness (persons)	0.016	0.011	1.34	1.24	1.17	1.09	1.14	1.04	1.07
% above pensionable age (persons)	0.030	0.018	1.65	1.39	1.29	1.19	1.28	1.09	1.13
% economically active (persons)	0.019	0.012	1.41	1.26	1.19	1.12	1.18	1.06	1.09
% unemployed (economically active males)	0.014	0.010	1.30	1.22	1.16	1.07	1.12	1.03	1.06
% with degrees (persons)	0.057	0.038	2.23	1.82	1.60	1.22	1.39	1.11	1.18
Means:						1.19	1.34	1.09	1.16

1991 Census of Population data⁴. These therefore refer only to variables available from the Census and not directly to key BCS estimates. The estimated values of ρ appear in Table 4, along with the implied estimates of DEFF under the three designs considered here. It can be seen that design factors, across the estimates studied, are on average 9% greater under design 1 than they would be under design 3 and 16% greater under design 1 than they would be under design 4. In other words, designs 3 and 4 would produce reductions in standard errors of around 8% and 14% respectively.

The third method that we have used to estimate design factor ratios utilises estimates of ρ from an experimental study carried out at the *National Centre* in 1990. For this study, a random half of the sample points for a national (England-only) survey were postal sectors while the other half were quarter sectors. However, the survey concerned attitudes and knowledge regarding local authority finance and services, so again no crime variables are available. The estimates of ρ obtained for a range of attitude and knowledge variables imply that the design factor ratios would average around 1.05 and 1.12 respectively for designs 3 and 4 (relative to design 1). Equivalent estimates from a small number of demographic variables suggest ratios of 1.06 and 1.13 respectively. These estimates are therefore broadly in line with those from the first method of estimation⁵. The average reductions in standard errors that would be implied by these design factors are summarised in Table 5.

Table 5: Summary of impact of three sample designs on standard errors

		Victimisation rates	Demographics		Attitude and knowledge measures
		Method 1: BCS-based Estimates	Method 2: Census-based Estimates	Method 3: Experimental study	Method 3: Experimental study
Reduction in standard errors relative to design 1	Design 2:	4%			
	Design 3:		8%	6%	5%
	Design 4:		14%	12%	11%

Note: The reduction in standard errors is defined as $(SE_n - SE_1)/SE_1$, where SE_1 is the average standard error under design 1 and SE_n is the average standard error under design n. The reductions are estimated for method 2 from the DEFTs in Table 4, and for methods 1 and 3 from the figures presented above in the text.

Based upon these estimates, it would seem that a move to design 3 is likely to represent a cost-efficient improvement. The standard error reductions should be worthwhile, while the cost implication is likely to be very small. Design 4 represents

⁴ The estimates are based upon modelling data published for enumeration districts, wards and local authority districts to provide estimates for sectors and quarter sectors. The models appear fairly robust as the estimates are similar whether modelled upwards from smaller areas or downwards from larger areas.

⁵ We have not presented the estimates for individual variables here, as the estimates of ρ are themselves subject to large standard errors, due to the small sample sizes per sector. The estimates of averages across a number of variables should, however, be more stable.

an even greater reduction in standard errors, but the cost implication is likely to be greater than for design 3. The trade-off between cost and precision could be organisation-specific, depending upon the way that field work is organised and managed. Our conclusion, then, is that a move to whole postal sectors as PSUs would almost certainly be cost efficient, while the question of how many addresses to select per sector (and therefore how many sectors to select) is perhaps one that should be left open for future tenderers to comment upon.

Although we cannot be sure that the precision gains indicated here will necessarily apply to key BCS estimates, we can be reasonably confident that the gains should be similar (as the range of DEFTs observed for BCS estimates of victimisation rates is similar to the range observed for variables examined here). If a design based upon whole sectors as PSUs were adopted in the future, it would be possible crudely to assess the impact on precision post-hoc, by comparing estimated design effects with those for the equivalent variables in earlier years.

While a move from quarter sectors to whole sectors will clearly improve the precision of virtually all of the range of BCS estimates published by the Home Office, there is one form of analysis on which it could potentially have an adverse effect - the analysis of the geographical clustering of crime victimisation. However, we feel that there are far more important factors to take into account when considering small area analysis and we think that there are better ways of improving the nature of such analysis based upon BCS data. We explore these issues further in section 7.

4.2.2 Sampling individuals within households

There is likely to be a loss of precision for most survey estimates due to the use of variable selection probabilities of individuals within households. Obviously, this would affect only individual-level estimates and not household-level estimates. However, there are ways of reducing the range of selection probabilities and hence improving precision (Butcher, 1988). A simple one is to interview only a random half of people identified in 1-adult households. This produces equal selection probabilities for all adults in 1 or 2-adult households (who account for a large proportion of all adults) and is thus likely to reduce the average design effect due to variable selection probabilities of individuals from around 1.27 to 1.15.

However, it is not obvious that this design would necessarily be beneficial in the case of the BCS. There are three reasons for this:

1. The BCS is simultaneously a sample of both households and individuals. Reducing the range of selection probabilities for individuals would imply increasing the range for households. Thus, any increase in precision for individual-level estimates might be accompanied by a loss in precision for household-level estimates.
2. For some crimes, persons in single adult households are more likely than others to be victims. Estimation of the level of such crimes (and of its characteristics) will be more precise if single person households are over-sampled relative to others (which is what happens under the current design). It would be possible to assess the extent to which it is efficient to over-

sample single-person households by analysis of existing BCS data. This may be worthwhile.

3. The design of interviewing only a random half of people identified in 1-adult households brings with it an extra field cost (for a given number of completed interviews) as contact needs to be made at extra (ineligible) addresses. Approximately 30% of households contain just one adult, so 15% of all households would now be ineligible for the survey. Precision gains thus need to be large enough to outweigh this extra screening cost in order for the design to be cost-effective.

We do not recommend changing the design in this way at present, though investigation of the likely effects on precision may be warranted.

4.3 Sample Stratification

4.3.1 Proportionate Stratification

The BCS sample design currently involves proportionate stratification at two stages. Postal sectors are implicitly stratified prior to systematic selection within major strata. This stratification involves the use of census data. Up until 1998 the factors used were standard region, population density and proportion non-manual household heads within the inner city and non-inner city major strata. In 2000, population density and proportion non-manual were used within PFA. Then, once sectors have been selected, addresses are ordered in alpha-numeric order of postcode within sector prior to systematic selection of addresses.

Both of these stages of stratification are for the purpose of increasing the precision of estimates. Implicit or proportionate stratification will tend to have this effect, to the extent that the stratification factors are correlated with the survey variables.

It is possible that the stratification factors used currently can be improved upon. In other words, there may exist other available factors that correlate more highly with key BCS measures and would therefore produce greater precision gains. It is possible to estimate the precision gains likely to be produced by any possible stratification scheme, by aggregating BCS data to sector level and then linking to the potential stratification factors. Analysis can then be performed to estimate the partitioning of variance within and between strata, for any definition of strata. Analysis of this sort has been performed on a number of surveys in recent years (e.g. Barton, 1996; Bruce, 1993; Clemens and Lynn, 1995) and in each case has identified a substantial improvement in the stratification. We have in the past proposed, while tendering for the BCS, analysis of this kind as an optional add-on to the main survey, but the Home Office have not taken up this suggestion. We believe that such analysis is relatively low cost compared with the potential gains in precision that could be made and recommend that consideration is given to the commissioning of such analysis.

4.3.2 Disproportionate Stratification

In all rounds of the BCS up to and including 1998, inner cities have been over-sampled. This practice was discontinued in 2000 and we understand that it is unlikely to be reinstated in the future. We do not, therefore, consider the over-sampling of inner cities in this review. In the 2000 BCS there are two elements of disproportionate stratification. First, some police force areas are over-sampled relative to others in order to provide adequate sample sizes for separate estimation. Second, ethnic minorities are over-sampled via “boost” samples, for the same reason.

Police Force Areas

Relative to the BCS 2000 design, it would be possible to further equalise the sample sizes across PFAs. This would necessitate over-sampling the smallest PFAs to an even greater extent and thus increasing the range of sampling fractions in the sample. One impact of this is likely to be a reduction in precision for national estimates. The extent of this effect can be estimated. The benefit, on the other hand, would be an increase in precision for the estimates for the smallest PFAs. These are precisely the estimates that are likely to have the lowest precision under the current design.

A decision to change the design in this way can therefore be seen as a trade off between precision for national estimates and precision for PFA estimates. The relative importance and value of these two types of estimates must thus be taken into consideration. At present, the design is very much skewed towards maximising the possible precision (within a sample size constraint) for national estimates. Given the shift of policy interests towards PFA-level issues and the imminent increase in annual sample size, it would seem desirable to consider a further over-sampling of small PFAs. By doing this, it should be possible to increase considerably the precision of estimates for PFAs and of estimates of differences between PFAs, while also increasing precision for national estimates (because of the larger overall sample size).

We have estimated the likely design effects and effective sample sizes for a range of sample design options. For a sample size of 40,000 achieved interviews, even the extreme option of having an identical sample size in each PFA (i.e. 952 interviews in each PFA) would still result in a larger effective sample size (and therefore smaller standard errors) for national estimates than under the BCS 2000 design with 20,000 interviews. If we compare this with the opposite extreme – proportionate sampling – standard errors would only be increased by 38% for national estimates, but would be reduced by 38% for the smallest PFAs.

We recognise that this extreme design of equal sample sizes per PFA may be undesirable for other reasons (for example, the metropolitan police area may be important enough to warrant rather more precision than other PFAs). Nevertheless, it serves to illustrate the scope for increasing the sample sizes in small PFAs without undue detriment to the precision of national estimates. In Table 6 we present the estimated effective national sample size under a number of different sample distributions, alongside that for the BCS 2000 design, for comparison. It can be seen that increasing the minimum number of interviews per PFA as high as 800 has only a

Table 6: Estimated design effects for alternative sample distributions over PFAs

n (interviews)	Minimum per PFA	Effective sample size for national estimates	Design effect for national estimates
20,000	300	19,640	1.02
40,000	None ⁶	40,000	1.00
40,000	500	39,840	1.00
40,000	600	39,230	1.02
40,000	700	38,025	1.05
40,000	800	36,070	1.11
40,000	952	20,920	1.91

Notes: The figures in this table are based upon a simplifying assumption that population variances are approximately equal across PFAs. The extent to which this is true might vary across survey estimates. This could be assessed, and the design effect estimates thus refined, via analysis of BCS 1998 data. Also, it should be noted that these design effects are only the component due to variable sampling fractions. Overall survey design effects will be the product of this and other components, such as those due to clustering and stratification.

small impact on the precision of national estimates. Beyond this number, the reduction in precision soon becomes dramatic. We would recommend consideration of a minimum PFA sample size of around 600 or 700 (with the remaining interviews distributed proportionately across the larger PFAs, as was done in 2000).

Ethnic Groups

The design employed in 1996 and 2000 to boost the numbers of interviews with ethnic minority respondents consisted of two elements. The first was a variant of focussed enumeration (Brown and Ritchie, 1984), whereby two addresses on either side of each core sample address were screening to identify any resident ethnic minority persons. The second element was a separate sample of addresses in a sample of sectors where the overall proportion of ethnic minority persons was expected to be relatively high (based upon 1991 Census figures).

The focussed enumeration (FE) component has the advantage of being relatively low-cost (because interviewers are calling at the core sample addresses in any case), but it produces a highly clustered sample of ethnic minority persons. With 32 core addresses per PSU, the FE adds a further 128 addresses per point. In a sector where 25%, say, of addresses contain a person of ethnic minority, this could lead to as many as 40 ethnic minority interviews in the sector. This method also appears to produce some under-coverage, as the number of ethnic minority persons identified at FE addresses is rather less than four times the number identified at core addresses. This may partly be due to situations where there is no adjacent address to screen and partly due to other deficiencies of the method.

The separate sample in dense ethnic sectors is certainly subject to under-coverage as ethnic minority people living in areas of low ethnic density are excluded. The extent

⁶ Imposing no minimum is likely to lead in practice to a minimum of around 370 interviews.

to which this leads to bias could be assessed using existing BCS data (for example, from the 1996 survey), by comparing key survey estimates for ethnic minorities between those living in sectors included in the boost sample and those living in sectors with lower ethnic densities (according to Census figures). If this bias appeared substantial, it would be possible to reduce it in future by covering a broader range of sectors (i.e. having a lower cut-off density). This would be costly, but this might be offset to some extent by creating a number of strata and using sampling fractions which declined with declining ethnic density.

If the BCS continues with both components of the ethnic minority sample, there may be scope for improving the efficiency of the design of the boost sample of sectors. Estimation of likely coverage bias should be undertaken in order to inform the design. Similarly, it might be possible to improve the cost efficiency of the FE sample by further limiting the proportion of sectors in which it is carried out. This suggestion too should be informed by estimation of any possible coverage bias that might result.

However, there may also be other options. First, it is worth considering carefully the sample size requirement for ethnic minority interviews in future years. The 1996 and 2000 rounds of the survey have both aimed to achieve around 4,000 interviews with ethnic minorities. With an increase in the core sample to around 40,000 interviews envisaged for 2001, the use of focussed enumeration alone is likely to result in a total (core plus focussed enumeration) of over 6,000 ethnic minority interviews. This is likely to be adequate, given that separate analysis by Police Force Area will still not be possible for the ethnic minority sample (except perhaps for two or three of the PFAs with the largest ethnic minority populations). The precision of estimates for ethnic minorities as a whole will in any case be greater than in the past. Thus, there may no longer be any need for a method of further boosting the numbers. The dense sector boost sample could therefore be discontinued. Alternatively, the FE sample could be limited to one address either side of each core sample address instead of two.

Also, methods other than the two approaches used on recent rounds of the BCS may have advantages. Name-match sampling has obvious limitations, but has been used with some success on other surveys (e.g. Smith, 1991; Sproston et al, 1999 – see also Harding et al, 1999) and could form part of a dual-frame approach. It cannot be used for black caribbeans, but may be useful for southern asians, for example. Also, for some survey organisations it may be possible to follow-up samples of ethnic minorities previously identified on other surveys. For this reason, we suggest that the Home Office should not prescribe too tightly the method to be used for sampling ethnic minorities, but rather should perhaps specify a minimum effective sample size in future invitations to tender.

It may also be unnecessary to boost the number of interviews with ethnic minorities every year. For example, the boost could take place once every two years. This would of course represent a substantial cost saving.

We would also like to take the opportunity to comment on the use of the ethnic minority samples in analysis. On past BCS rounds, a weight has been calculated and included in the data file, to permit all ethnic minority interviews to be included in

total sample analysis. However, it has been recommended that data users should not use this approach. Rather, it is suggested that total population estimates be based upon the core sample only. We feel that it may be more efficient to include all the ethnic interviews. The precision of total population estimates should be improved by this approach. We have estimated the precision gain in Table 7 below. For the 1996 design, adding in the extra 2,598 ethnic boost interviews would increase the effective sample size by 866. It can be seen that this provides a small reduction in standard errors.

Table 7: Impact on precision of including ethnic boost sample for total population estimates (BCS 1996)

Sample	Sample size	Effective sample size	Predicted design effect	Standard error (10%)	Standard error (50%)
Core only	16,335	12,731	1.28	0.27%	0.44%
Core + FE + boost	18,933	13,597	1.40	0.26%	0.43%

Note: Effective sample sizes are estimated under the assumption of equal population variance with weighting classes.

On the other hand, some bias would be introduced due to the limited coverage of the ethnic boost samples, described above. We suggest that a decision as to whether the ethnic boost interviews should be included in total sample analysis should be made in the light of the bias assessment suggested above. All interviews should be included only if the variance reduction appears to outweigh any bias introduced (see section 2.1). As the variance reduction appears modest indeed, it would appear sensible to include the ethnic boost interviews in total sample analysis only if there is no evidence at all of any bias in the boost samples.

There is also a third possibility, which would be to include the focussed enumeration sample but not the area boost sample. The focussed enumeration sample is likely to suffer less bias (if any) than the area boost sample, as the FE sample covers all sectors with 0.3% or more non-white, whereas the area boost sample covers only sectors with 19% or more non-white. However, we have not been able to formally assess the precision gain within this review, as the BCS data sets do not contain a weight developed for this purpose.

5 DATA COLLECTION

5.1 Recall Errors

The issue of the timing of interviews and the recall period for questions about crime victimisation has already been raised in section 3.1 above in the context of discussion of the distribution of the sample over time. However, the impact of respondent recall on survey estimates is a topic deserving of attention in its own right.

A considerable amount of methodological research has examined the role of recall in contributing to survey errors and has examined ways in which survey design can be used to reduce the impact of inaccurate recall on survey errors (Bradburn and Sudman, 1979; Converse and Presser, 1986; Eisenhower et al, 1991; Fowler, 1993; Moss and Goldstein, 1979; Schuman and Presser, 1981; Sudman and Bradburn, 1973). It is quite clear that this is not a trivial issue and that the effects of recall errors can be considerable.

On the BCS, it is clearly the questions on crime victimisation for which this issue is most pertinent, though it also affects some other retrospective questions, such as those on fires in the home.

Two main types of recall errors are likely to occur. One is "telescoping." This is a tendency on the part of respondents to report events which in fact took place outside the reference period. Research has shown that this tendency is greater the nearer the event was in time to the reference period and that the tendency is greater the more salient or memorable the event. The other main type of recall error is omission, i.e. failure to report an event which happened within the reference period and therefore should have been reported. This can happen due to simple forgetfulness, in which case it is more likely to happen the less salient and memorable the event. It can also happen because of respondent sensitivity regarding the question or the answer, for example because of social desirability concerns. Overall, then, recall errors tend to lead to over-estimation of the frequency of memorable, non-sensitive events, and under-estimation of the frequency of events which are either sensitive or not memorable. The magnitude of these effects will of course depend upon the recall task, the reference period, the survey population and other factors.

It is likely that events of each of these types are represented amongst the BCS measures. For some people, reporting of some types of crime victimisation is likely to be a sensitive issue. This may include, for example, sexual crimes and violent crimes committed by friends, relations or household members. Other events asked about in the BCS interview may be not particularly memorable for some people, such as vandalism of property exteriors or threats of violence. There may, therefore, be a risk that these crimes are under-reported. On the other hand, some events – such as burglary, perhaps – are likely to be highly memorable and salient. It is easy to imagine a respondent wanting to tell to an interviewer about a recent burglary even if it happened just outside the reference period. It is therefore quite likely that

response errors cause bias in the BCS estimates of some or all crime victimisation rates.

However, if the nature of these biases is constant over time, then there may not be any bias in estimates of change over time. This hope is perhaps the main reason that the reference period and field work timing has been held fairly constant over rounds of the BCS. There has, however, been no assessment of whether or not the biases are constant, nor indeed of any aspects of the nature and extent of possible bias.

We believe that it is highly desirable to understand more about the nature of recall bias on the BCS. As already argued in section 3.1, this would be particularly important in the context of a move towards continuous field work. But even in the absence of such a move, we cannot be at all sure that any bias is constant over time unless we attempt to estimate the bias. And it would be very useful to know a little about the nature of bias in the cross-sectional estimates, particularly those of victimisation rates.

We would recommend that some experimental studies are built in to a forthcoming round of the BCS in order to assess the nature and magnitude of recall bias. We would also suggest that these studies should be designed with a view to identifying alternative survey methods that could reduce the bias. In particular, the optimum recall period is likely to be different for different types of events. While it could be counter-productive for the BCS to use a large number of different recall periods for different questions, there could be benefit in considering more variation than at present. Furthermore, the specific implementation of a given recall period could have an effect. For example, a 12-month recall period could be implemented as any of the following:

- ◆ The last complete calendar year
- ◆ The twelve months up to and including the last completed month
- ◆ The twelve months up to and including yesterday

In addition, the effects of both the recall period and the implementation of the period could be subject to seasonal effects.

Specifically, we propose the following:

1. Analysis of existing BCS data by date of interview in order to detect any effects of the elapsed time since the reference period. As this analysis would stem from a non-experimental design, the results can only be indicative. There are two possible confounding factors. One is that persons interviewed late in the survey period may have different characteristics to those interviewed earlier. This can be controlled to some extent in the analysis. The other is that the elapsed time is confounded with the length of the reference period, as respondents are asked about crimes since the 1st of January the previous year.
2. Consideration of a factorial experiment, involving the random allocation of a sample to different reference periods, different implementations of the

reference period and/or different months of the year for interview. The number of alternative treatments must be constrained, however, in order for the sample sizes to be large enough that effects are likely to be detected. The choice of treatments should depend upon the likely future format and timing of the BCS.

5.2 Seasonal Effects

The experimental design proposed above incorporates a seasonal control as we believe that recall effects could interact with seasonal effects. But aside from their possible impact on recall effects, seasonal effects should be of concern in their own right. As with recall effects, understanding of seasonal effects would be particularly important if the BCS were to move to a continuous design, but is still of some importance even in the absence of such a change. Obvious examples of BCS questions that could be subject to seasonal effects are the following series of questions: feeling safe, worries, problems in the area, going out.

Continuous field work would in itself provide the data necessary to assess seasonal effects (assuming that the sample were randomly distributed across seasons in a balanced way). Analysis based upon a year's worth of data from a continuous BCS would provide the information necessary to produce "correction factors" to be used in any analysis comparing with earlier rounds of BCS – if indeed such correction factors turned out to be needed.

In the absence of continuous fieldwork, a modest-scale experiment would be needed in order to assess seasonal effects. This would involve the random allocation of a sample to different seasons and could be run as an adjunct to a future round of the BCS. Again, if significant effects were found, adjustment factors could be estimated in order to provide estimates of survey measures which related to the average over a year rather than the particular period in which the field work was carried out. However, if the general design of the BCS were to remain unchanged this sort of analysis will probably not be of high priority.

5.3 Coding

For the reasons described in section 2.1.1, the variance potentially introduced by coders should be a concern to the BCS team. It may be worthwhile to set up a controlled coder variance experiment to ascertain the extent of the problem with respect to victimisation codes. This would not affect field work, but would require that a sample of interviews is randomly assigned to coders, controlling for interviewer and area, and that those interviews are then also subsequently coded by the other coders in the experiment (four or five coders would be needed in the experiment).

In the meanwhile, we would recommend that a maximum limit should be imposed upon the volume of coding carried out by any one coder. The main coding exercise on BCS relates to victimisation coding, so we suggest that the maximum work load should perhaps be set in terms of a number of victim forms rather than a number of interviews. We would suggest something like 600. On BCS 98, the mean number of

victim forms per interview was 0.69, so 600 victim forms would imply 870 interviews. With a sample size of 40,000 interviews, even this modest limit (the effect of coder variance could still be very large with this limit) would necessitate a minimum of 46 coders working on the survey. The Home Office should thus be aware that there may be a cost implication. This suggested maximum figure is of course arbitrary and it would be possible to optimise this if information about coder variance were available from the study proposed above.

6 NON-RESPONSE WEIGHTING

6.1 The Nature of Non-Response on the BCS

The 1996 and 1998 BCS Technical Reports present *prima facie* evidence of lower response by adults aged 16-34 and (in 1998) by males. Similarly the 1994 Technical Report showed a shortfall in the achieved sample by males aged 20-29. There was some shortfall in the achieved sample in some regions in the different rounds of the survey with London, perhaps, being more consistently under-represented. There is also some evidence that all three surveys under-represented renters, compared to owner-occupiers. Also in the latest two surveys (it was not examined in the earlier one), households with no access to a car were under-represented.

The proportion of non-response which takes place at the household level – i.e. before the respondent has been identified – has been declining rapidly over recent rounds of the survey. Amongst the core sample, this proportion was 36% in 1992, 21% in 1996 and 13% in 1998. This change raises concerns that the nature of non-response, and possibly, therefore, of any non-response bias, could be changing over time.

The analysis of non-response on the 1996 survey (Lynn, 1997b) found several factors that were important in predicting the final response status:

- condition of the property
- dwelling type
- region
- ACORN group
- population density
- the general condition of properties in the area.

The first 2 of these are address-level variables from interviewer observation and the last 4 are area-level variables. However, the data that was known or could be readily collected about the full sample of both respondents and non-respondents did not include any individual or household-level demographic variables.

A follow-up of a sample of the non-respondents, which achieved a 25% response rate found no large differences between survey respondents and non-response follow-up respondents on a number of key survey measures and Lynn (1997b) concluded that “... the preliminary findings of this study reject the hypothesis that non-respondents experience higher victimisation rates than respondents. If anything they report less victimisation ...”. However this analysis and its conclusion were limited by the fact that the non-respondents follow-up only achieved a 25% response rate. Nothing is known about the other 75% of BCS non-respondents.

An earlier study of BCS non-response had drawn a different conclusion, based on data available for the whole sample of respondents and non-respondents. In an

analysis of data from the 1992 survey (Lynn, 1996), the selected sample was matched to the Electoral Register and the number of electors, a proxy for number of adults in the household, was thus available and was shown to be a useful predictor of both refusal and non-contact. That paper concluded that "It is very likely that the achieved sample ... under-represents people living alone, those living in purpose-built flats, particularly flats above the first floor, and those in blocks with a common entrance that cannot be locked. These household size and type measures are likely to be correlated with some of the key measures ... such as experience of crime and attitudes towards crime".

A more recent analysis of factors affecting non-response to the 1998 survey (Laiho & Lynn, 1999) was able to examine the factors that affect the different stages of survey non-response. This showed that besides factors under the direct control of the survey organisation, the type of dwelling and several area-level variables were useful predictors of response at the household level. However, for individual response, the area characteristics were not important. Two different features of the dwelling, the presence of a security device and the number of floors in the building were important but the most important factor by far was the number of adults in the household. Again, no person-level factors were available in this analysis.

Each of these analyses permits the construction of weights to adjust for the explained variation in response rates by means of a response propensity model of greater or lesser complexity. However the effectiveness of this procedure in reducing non-response bias will depend both on how much of the variation in response rates is explained by the available factors and on how far the response rates correlate with the key survey measures.

The Home Office Bulletin on the 1998 survey (Mirrlees-Black, 1998) identifies risk factors for 3 types of crime – burglary, car crime and violence against the person. This shows that the age group 16-24, both at the individual and head-of-household (HOH) level, single parents, private renters, the unemployed (personal and HOH) and people living in inner city areas are at substantially higher than average risk. At the same time people (and HOHs) aged 65 and over, owner-occupiers, people living in detached houses and people living in rural areas have a substantially lower than average risk. These then seem important factors to consider in any weighting strategy for the BCS.

6.2 Non-Response Weighting Methods

There are three main approaches to the construction of non-response weights on surveys: sample-based, population-based and model-based methods. Sample-based methods compare the responding sample with the selected sample on an appropriate (and available) set of factors and then align the two by weighting – this is the approach taken in all three of the studies referred to above, although the weights are not derived explicitly. With population based methods, the responding sample is compared with the population for which inferences are to be made, which sometimes but not always coincides with the population from which the sample was selected. Finally, a third approach is to posit some explicit model of the non-response process

and to derive response probabilities and weighting factors from the model without recourse to any other source of data. This third approach is risky since in cases where the model is inappropriate (and its appropriateness is usually not testable) the results of the model-based approach will be invalid.

Elliot (1991) reviewed the various options and concluded that, other things being equal, the population-based weighting methods were preferable since they are able to provide partial correction for non-coverage bias (where the sampling frame omits some members of the population of interest) and also produce more precise survey estimates than the sample-based methods. However, they suffer from the drawback that there is often very little data available at the population level and the available data is often collected by administrative or other means and so is unlikely to be comparable with data collected from survey respondents by personal interview. Thus, in practice it is rarely the case that the same variables are available at the sample and population level so the choice between approaches is likely to be driven by the available data.

In the case of the BCS, although the analysis of the sample-based measures has identified several relevant factors, the paucity of the demographic indicators is a limitation. With a population-based approach, age and sex at an individual level and household type defined in terms of the number of people (including children) in different age by sex groups can largely be corrected⁷ for by means of a calibration algorithm. Age, sex and household composition have already been shown to be closely related to victimisation risk and seem likely also to be related to the survey's attitude measures.

However, it is important to realise that in searching for an effective non-response weighting strategy, it is not necessary to choose between a population-based and a sample-based approach – the best strategy is likely to involve both elements. The calibration methods noted in the last paragraph can accommodate initial weights derived from other sources. Thus both the weights necessary to correct for the different selection probabilities that are used on the survey (to select one person per household and to over-sample some Police Authority Areas) and the weights derived from the most recent response propensity research can be used. Unlike the sample-based methods, the calibration method does not guarantee that the weighted sample distributions will match those of the selected sample, but aims to minimise the “distance” between the initial weights and the final weights while guaranteeing that the weighted sample distributions match the known population distributions exactly.

6.3 Construction of Initial Weights

The latest research into BCS response propensities (Laiho and Lynn, 1999) represents an important conceptual advance over the previous work but it is not yet clear if this will provide a better model for response rates or not. One drawback of using complex models of response propensity is that more parameters need to be fitted and

⁷ Evidence from the Family Expenditure Survey and the General Household Survey, where this method has been used, supports this conclusion.

so the risk of over-fitting the models cannot be ignored. One way around this is to use the models to indicate which factors are important in predicting response rates and to inform the construction of weighting classes but then to use real rather than modelled response rates to construct the weights.

Two different methods of doing this that have been suggested in the literature. The first is to use a hierarchical splitting algorithm, such as *SPSS/CHAID (AnswerTree)*, to partition the sample into a number of groups on the basis of combinations of the relevant factors. The groups are designed to maximise the variation in response rates and one can specify a minimum group size to improve the stability of the weights. A second method is to rank the sample on the modelled response propensities and then partition the sample into a small number of quantile groups. Within the groups, actual response rates can be calculated and used to construct initial weights.

6.4 Calibration Options

Data from the survey is analysed at both a household and a person level. When creating person-level estimates, an extra weighting factor is needed to correct for the different selection probabilities associated with selecting one adult per household. The sample based methods suggested by the previous research use data only available at the household or area levels and so provide a nonresponse adjustment method for both household and person-level estimates.

For the household-level estimates the calibration method proposed by Lemaitre and Dufour (1987) seems the most appropriate. This is the method currently used on the Family Expenditure and Family Resources surveys as well as the household file from the Labour Force survey. The aim of the method is to match the weighted distribution of age by sex of all responding household members (adults and children) with that in the population by means of a single household weight. As originally formulated, the method uses a type of Generalised Regression (GREG) Estimation but variants of the method have now been developed to limit the range of weights and these are usually preferred.

The method has performed well in tests and is remarkably successful in recovering the correct (but usually unknown) distributions of types of household/family in the population while at the same time making worthwhile reductions in sampling error, compared with unweighted estimates (Elliot, 1997, 1999). It will, incidentally, provide weights to gross the sample to the population level so that no separate grossing step will be required.

This calibration method using only age, sex and region should therefore reduce the risk of non-response bias for young and for elderly adults and also to some extent for single parents. Similarly the response propensity corrections should account for differences at an area level (inner city, urban, rural) and in accommodation type. However, for several of the factors known to be associated with higher or lower than average crime risk, it is not known whether response propensities vary, so a cautious approach would be to control these factors to population levels. Unfortunately up-to-date population data are not available for employment status and tenure so

survey or other estimates would need to be substituted, with attendant risks of introducing other biases. The Labour Force Survey is possibly the best source for both these distributions because of its very large, unclustered sample in England and Wales and its relatively high response rate. One way to proceed would be to derive weights omitting these factors and then compare the weighted and LFS distributions before deciding whether or not to incorporate these as population controls. If this approach were to be considered, it would be important for the BCS to amend the questions that it asks about economic status and housing tenure to the harmonised versions that are asked on the LFS (GSS, 1998).

For the person-level estimates, one could either stick to the household-level non-response weights (with the additional design weight for selecting one person in the household) or introduce person-level calibration weights for age, sex and region. This choice depends on which factors are most important in determining response by the selected individual. The work done so far on response propensity gives no guidance on this matter but it seems more natural to use individual-level factors.

6.5 Recommendations

For household-level analysis, we recommend a 3 step approach:

- i. Design weights to account for different sampling rates in different areas.
- ii. Weights derived from the response propensity models, either directly or by grouping cases with similar response propensities. Design weights should be used⁸ when calculating the group response rates.
- iii. Household-level calibration weights based on age group, sex and region of all household members. Consider adding survey-based distributions of tenure and economic status of the head of household (based, e.g. on LFS) to the set of control totals. Combined weights from steps i. and ii. should be used as initial weights in the calibration.

For person-level analysis, a 3 step approach is again recommended:

- i. Design weights to account for different sampling rates in different areas and different selection probabilities for adults in different sized households.
- ii. Weights derived from the response propensity models, either directly or by grouping cases with similar response propensities. Once again, weighted response rates should be used.
- iii. Person-level calibration weights based on age group, sex and region of the responding adult. Consider adding a survey-based distribution of economic status of adults (based on the LFS) as a control variable. Combined weights from steps iv and v should again be used as initial weights in the calibration.

Once the non-response weights have been developed, the impact of the weighting on survey estimates, and hence upon the time series, should be assessed and reported.

⁸ The response rate is thus the weighted number of respondents divided by the weighted sample size.

7 GEOGRAPHICAL ANALYSIS

The Home Office is interested in identifying areas of high crime risk. Specifically, there is interest in estimating the level of crime that might be expected in an area, given its socio-demographic characteristics. These issues are important for determining how to target crime reduction initiatives and for interpreting observed levels of crimes for areas.

To achieve this, one could consider three conceptual stages. The first would be to establish the appropriate definition and level of areas for the analysis. The second would then be to build a model for the prediction of crime levels for those areas. A third would be to interpret and use the model appropriately. A number of issues are raised at each stage.

7.1 Defining Areas

From a sociological and criminological perspective, it would be revealing to be able to identify the geographical level which best discriminates in terms of crime rates. This would help to explain the factors involved in determining crime levels and the ways in which crime reduction initiatives might be most likely to be productive. In theory, such information should also reveal the areas at which crime reduction initiatives could most effectively be targeted. However, there may also be relevant practical limitations. For example, there may be pre-existing initiatives which are only appropriate to apply to areas of a certain approximate size. If this were the case, it would seem sensible to constrain the search for the best-discriminating areas to definitions which met that size criterion. Additionally, the size and complexity of the analysis task to identify the appropriate definition of areas is greater the fewer constraints that are imposed.

A practical approach might be to consider a couple of alternative definitions of areas and to perform analysis of BCS data in order to identify how best to predict crime levels for each of those definitions of area and for which of the definitions the explanatory power appears greatest. The choice of area definitions should be influenced partly by knowledge of the practical implementation of crime reduction initiatives (which we do not possess) and partly by the nature of the available data.

Clearly, the definitions must be ones which can be applied to the BCS. In other words, it is necessary to be able to classify every BCS sample address to an area under each definition. But furthermore, it is inherently more likely that it will be possible to obtain good predictions if the definitions are ones which match closely the areas for which potential explanatory variables are available. These variables include (but are not limited to) the following:

- ◆ Census Small Area Statistics
- ◆ Geo-demographic classification systems, e.g.
 - ACORN*,
 - Mosaic*;
- ◆ Deprivation and poverty measures, e.g.
 - the Index of Local Deprivation (DETR, 1999),
 - the Jarman Index,
 - the Townsend Index,
 - the Carstairs Index;
- ◆ Administrative data, e.g.
 - DSS income support and benefit statistics,
 - Insurance premium data,
 - Etc.

Some of these sources of data are available for very small area units such as Census enumeration districts or even unit postcodes. But most are only available for larger areas such as postcode sectors or electoral wards. It is possible in principle to assign data defined at a postcode sector level, say, to unit postcodes, by attaching to each postcode the data for the sector to which the postcode belongs. And similarly data for unit postcodes could be aggregated to provide estimates for sectors. But analysis is most likely to find strong relationships if the data is used at the level at which it was defined. Aggregation to higher levels can be successful if appropriate measures of size are available to act as weights in the aggregation procedure.

The most promising approach might be to use geographical areas such as postcode sectors or electoral wards as the units of analysis. Almost all of the variables mentioned above are available at those levels. However, because the BCS sample is based upon sectors it is more likely that relationships will be observed for sector-level data.

It could also be worthwhile exploring lower geographical levels such as enumeration districts, as their much smaller size offers the potential for greater discrimination. Although a smaller range of variables are available at this level, they include some key ones such as Census Small Area Statistics and *ACORN*. Above the level of postcode sector or ward it is not until one reaches considerably larger areas such as Local Authority Districts (LADs) that extra useful information becomes available. LADs are probably too large to form useful analysis units.

An effective starting point might therefore be to attempt analysis for a couple of pre-existing definitions of geographical areas, such as enumeration districts and postcode sectors.

7.2 Building Models

Appropriate modelling techniques should be used to develop models to predict crime levels. It may be sensible to develop models separately for each of a small number of different categories of crime (e.g. personal crime, violent crime, property crime). Initial exploratory analysis should be able to determine whether the area-level correlates of each of these categories are broadly similar.

Careful attention should be paid to the construction of the crime measures which will form the dependent variables in the models. One possibility would be to use recorded crime statistics for this purpose. Unlike BCS estimates, recorded crime levels would not suffer from sampling error. The sampling error in BCS estimates would be large even for postal sectors, and considerably larger still for enumeration districts. There may however be practical difficulties in making estimates for small areas defined by home address of victim(s) from the recorded crime data. We are not sufficiently familiar with the nature of crime records to be able to comment further. BCS estimates, however, have the obvious advantage that they are defined in a consistent way over time (which is, after all, a main rationale for carrying out the survey). A careful evaluation of the errors inherent in each data source might usefully inform the choice of dependent variable.

If BCS estimates are to be used, there may be advantage in combining data from more than one year in order to reduce the sampling error of estimates. For example, if the rotating panel design proposed in section 3.2 were adopted, then the samples from within the rotated sectors in consecutive years could be combined, thus doubling the postcode sector sample size. However, even then the sample sizes per sector would be small (perhaps around 45 interviews). Previous attempts to use survey data to estimate small area effects have found that the inherent variance in PSU-level estimates makes it very difficult to identify well-fitting models (e.g. Korovessis and Purdon, 1999). More sophisticated small area estimation techniques may be needed (Rao, 1999).

Having derived the dependent variables (crime rates) for each postcode sector available for analysis (e.g. all sectors included in the BCS sample over two years), the full range of available data can be used to form the basis of useful predictor variables. One could either take a theoretical approach – i.e. testing primarily variables that should be expected to correlate with crime rates at an area level on theoretical grounds – or a purely empirical approach where all available variables are treated equally. If the main aim of the analysis is simply to develop a tool that can be used to predict crime levels (rather than to explain them) then the latter approach could be the more appropriate.

We would recommend that a broad range of Census indicators be included in the analysis, along with at least one of the geo-demographic classifications. ACORN is already known to correlate with crime measures, and is easily available, and would therefore seem the obvious candidate. In addition, measures of poverty or deprivation should be worth including. The Jarman and Townsend indexes are based entirely upon Census data, so it may be possible to replicate whatever explanatory power they might have simply by incorporating the appropriate Census variables in the analysis. The DETR Index of Local Deprivation, on the other hand,

draws upon data from a wide range of sources, and there is evidence that it discriminates well in terms of some area characteristics. It would seem to have considerable potential as a predictor of crime levels.

As the purpose of the model is predictive, the predictive power of the fitted model is important. Only if the overall goodness-of-fit of the model is high will the model be likely to provide meaningful estimates for sectors for which no survey data is available. In other words, for the model to be useful in terms of its intended purpose, it is necessary that it should explain a large proportion of the between-area variation in crime rates for the sample of areas in the data set. As stated above, we are doubtful that it will be possible to achieve this.

Multi-level modelling may well have much to offer in this situation. It will not necessarily provide a better-fitting model, but by taking into account the hierarchical nature of the data it should provide useful information about the variance structure and patterns in the data. This will help to describe the relationship of crime risk with area factors, but it will not help to identify which areas specifically are at high risk.

Modelling of this sort requires considerable expertise and should be carried out by experts. Once a model has been developed, its properties and predictive behaviour should be carefully explored and understood before it is used "for real."

7.3 Using the Models

A model developed as outlined above – if sufficiently well-fitting – could be used to derive estimates of predicted crime rate for all areas. For the areas included in the data set upon which the model was based, the modelled estimates will not be equal to the direct estimates. This simply reflects unexplained variation. If the model is appropriate and well-fitting the differences should not be either large or systematic, however. For precisely this reason, the predictions for areas not included in the data set should be treated as general indications rather than precise estimates. A model may be able to sort areas reasonably well into a small number of broad bands ("high", "medium" and "low" crime risk) but is rather less likely to be able to sort areas into rank order with any precision. For this reason, it may be desirable to publish model-based predictions (even within the Home Office) only as broad bands or levels, rather than estimates of rates per thousand residents, say.

8 RECOMMENDATIONS

A number of suggestions for changes to the BCS methodology and/or areas worthy of further exploration have been identified in the course of this review. Some of these suggestions have relatively minor implications. Others, we recognise, have significant implications and clearly require careful consideration and evaluation. In this section, we have attempted to bring together and summarise our recommendations. It should be noted that where we think the current (BCS 2000) methodology does not require alteration, we have not explicitly recommended maintaining the current design. That should be assumed. Our recommendations relate to changes to the BCS 2000 design. We have crudely sorted them into three categories.

The first category consists of issues where further methodological exploration is necessary before a firm recommendation for redesign can be made. For each issue we have also indicated approximately how many researcher-days we feel would be needed to carry out the required work, in order to provide an indication of the relative size of the tasks.

The second category consists of recommendations which we feel have few if any negative implications and should therefore be implemented as soon as possible. Rather more radical recommendations are reserved for the third category. We feel there is a very strong case for the changes suggested here but there may be a need for further discussion and consideration before these changes are adopted, not least because considerable planning may be necessary prior to implementation in some cases.

8.1 Further Methodological Work

Direct estimates should be made of the precision gain in key measures of change from a rotating PSU design. This can be based upon data from the 1996 and 1998 BCS rounds. (6 researcher days)

Optimal PSU stratification factors should be identified via a postcode sector level analysis of existing data. (6 researcher days)

Estimates should be made of the impact on precision of variable selection probabilities of persons within households and of alternative designs. (5 researcher days)

Estimates should be made of the coverage bias in the ethnic boost sample (dense sectors) and that which would arise under alternative designs. (3 researcher days)

Estimates should be made of the coverage bias in the focussed enumeration sample and that which would arise under alternative designs involving an increase in the proportion of PSUs excluded from this part of the sample. (2 researcher days)

Estimates should be made of the effect of date of interview on recall data, using existing data. (5 researcher days)

A coder variance experiment should be built in to a forthcoming round of BCS.

A study of alternative reference periods and implementations of reference periods could be built in to a forthcoming round of BCS.

8.2 Immediate Recommendations

The primary sampling units (PSUs) should be changed from quarter postcode sectors to whole sectors.

In each survey year, 50% of PSUs should be rotated forward from the previous year, but fresh addresses selected. Thus, each selected PSU will be retained in the sample for two survey years.

It should be ensured that no one coder codes more than around 900 interviews in any BCS year.

The sample should be distributed over Police Force Areas so as to provide a minimum of 600 or 700 interviews per PFAs.

Non-response weighting should be introduced to the BCS. This should incorporate both sample weighting and calibration to population totals.

To aid area analysis, the DETR Index of Local Deprivation should be added to the BCS data, along with a range of Census Small Area Statistics

8.3 Radical Recommendations

The survey should move to a continuous field work basis.

With continuous field work, the reference period for the crime victimisation questions should be the twelve months ending with the most recent completed calendar month.

In the first year of continuous field work, half the sample in each month should be asked about the "new" reference period, while the other half are asked about the twelve months ending in December/June (the "spliced design" described in section 3.1.2).

Analysis of the above experimental design should be performed to determine if the design change has had any significant impact on key survey estimates and, if so, to provide adjustment factors for trend analysis.

REFERENCES

- Bailey, L., Moore, T.F. and Bailar, B.A. (1978) An interviewer variance study for the eight impact cities of the National Crime Survey cities sample. *Journal of the American Statistical Association*, Vol.73, 16-23.
- Barton, J. (1996) Selecting stratifiers for the Family Expenditure Survey (FES), in *Survey Methodology Bulletin* 39, 21-26.
- Biemer, P. and Stokes, S.L. (1991) Approaches to the modeling of measurement error, in Biemer et al (eds), *Measurement Errors in Surveys*. New York: Wiley
- Bradburn, N. and Sudman, S. (1979) *Improving Interview Methods and Questionnaire Design*. San Francisco: Jossey-Bass.
- Brown, C. and Ritchie, J. (1984) *Focussed Enumeration: The Development of a Method for Sampling Ethnic Minority Groups*. London: SCPR (National Centre for Social Research).
- Bruce, S. (1993) Selecting stratifiers for the Family Resources Survey (FES), in *Survey Methodology Bulletin* 32, 20-25
- Butcher, B. (1988) Individual sampling using the postcode address file, in *Survey Methodology Bulletin* 22, 1-6
- Clemens, S. and Lynn, P. (1995) A Review of Sample Stratification for the Health Survey for England. Paper submitted to the Department of Health.
- Converse, J., and Presser, S. (1986) *Survey Questions: Handcrafting the Standardized Questionnaire*, Sage Series No 63. Thousand Oaks, CA: Sage Publications, Inc.
- DETR (1999) *Best Value and Audit Commission Performance Indicators for 2000/2001. Volume 1: The Performance Indicators including The Publication of Information Direction 1999 (England)*. London: DETR. Also published on the web at <<http://www.local-regions.detr.gov.uk/bestvalue/indicators/bvaudit/index.htm>>
- Eisenhower, D., Mathiowetz, N.A., and Morgenstein, D. (1991) *Recall Error: Sources and Bias Reduction Techniques*, in Biemer et al (eds), *Measurement Errors in Surveys*. New York: Wiley
- Elliot, D. (1991) *Weighting for Nonresponse - a survey researcher's guide*. London: OPCS (ONS).
- Elliot, D. (1997) *Software to weight and gross survey data*. GSS Methodology Series No. 1. London: ONS.
- Elliot, D. (1999) *Report of the Task Force on Weighting and Estimation*. GSS Methodology Series No. 16. London: ONS.
- Fowler, F.J. Jr., (1993) *Survey Research Methods, 2nd Edition, Applied Social Research Methods Series Volume 1*. Thousand Oaks, CA: SAGE Publications
- Government Statistical Service (1998), *Harmonised Concepts and Questions for Government Social Surveys*. Available at <<http://www.statistics.gov.uk/harmony/harmonfp.asp>>
- Groves, R.M. (1989) *Survey Errors and Survey Costs*. Chichester: John Wiley.

- Groves, R.M., Cialdini, R.B. and Couper, M.P. (1992): Understanding the decision to participate in a survey, *Public Opinion Quarterly*, 56, pp. 475-495.
- Groves, R.M. and Couper, M.P. (1998): Non-Response on Household Interview Surveys. Chichester: John Wiley & Sons.
- Hales, J. and Stratford, N. (1997) 1996 British Crime Survey (England and Wales): Technical Report. London: SCPR (National Centre for Social Research).
- Hales, J. and Stratford, N. (1999) 1998 British Crime Survey (England and Wales): Technical Report. London: SCPR (National Centre for Social Research).
- Harding, S., Dews, H. and Simpson, S.L. (1999) The potential to identify south asians using a computerised algorithm to classify names. *Population Trends*, no.97, 46-49.
- Hartley, H.O. and Rao, J.N.K. (1978) Estimation of nonsampling variance components in sample surveys. In *Survey Sampling and Measurement* (ed. N.K. Namboodiri), 35-43. New York: Academic Press.
- Holt & Farver (1992) The use of composite estimators with 2-stage repeated sample designs. *Journal of Official Statistics*, Vol.8, 405-416.
- Kalton, G. and Stowell, R. (1979) A study of coder variability. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, Vol.28, 276-289.
- Kish, L. (1962) Studies of interviewer variance for attitudinal variables. *Journal of the American Statistical Association*, Vol.57, 92-115.
- Lemaitre G & Dufour J (1987) An integrated method for weighting persons and families. *Survey Methodology* Vol.13, 199-207.
- Lynn, P. (1996) Who responds to the British Crime Survey? Paper presented to the Fourth International ISA Conference on Social Science Methodology, Colchester.
- Lynn, P. (1997a) Sampling frame effects on the British Crime Survey. *Journal of the Royal Statistical Society Series A (Statistics in Society)*, Vol.160, 253-269.
- Lynn, P. (1997b) Collecting data about non-respondents to the British Crime Survey. Report to Home Office.
- Laiho, J. & Lynn, P. (1999) Separating out the stages of the survey participation process. Paper presented to the International Conference on Survey Nonresponse, Portland, Oregon.
- MacKenzie, D.L., Baunach, P.J. and Roberg, R.R., ed.s (1989) *Measuring Crime: Large-Scale, Long-Range Efforts*. Albany, NY: State University of New York Press.
- Mirrlees-Black, C. *et al* (1998) The 1998 British Crime Survey. Statistical Bulletin, Issue 21/98. London: Home Office.
- Moss, L. and Goldstein, H. (1979) *The Recall Method in Social Surveys*. Slough: NFER Publishing Ltd.
- Rao, J.N.K. (1999) Some recent advances in model-based small area estimation. *Survey Methodology*, Vol. 25, 175-186.
- Schuman, H. and Presser, S. (1981) *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. New York: Academic Press.
- Smith, P. (1991), *Ethnic Minorities in Scotland*. Edinburgh: Scottish Office Central Research Unit.

Sproston, K., Pitson, L., Whitfield, G. and Walker, E. (1999), *Health and Lifestyles of the Chinese Population in England*. London: Health Education Authority.

Sudman, S. and Bradburn, N.M. (1973) Effects of Time and Memory Factors on Response in Surveys, *Journal of the American Statistical Association*, 68

Tourangeau, R. and Smith, T.W. (1996). Asking sensitive questions: the impact of data collection mode, question format, and question content. *Public Opinion Quarterly*, 60, 275-304.