British Crime Survey: options for extending the coverage to children and people living in communal establishments

Kevin Pickering, Patten Smith, Caroline Bryson and Christine Farmer
National Centre for Social Research and Ipsos MORI

This report presents findings from methodological research carried out to examine the feasibility of covering children (under 16s) and people living in communal establishments as part of the British Crime Survey (BCS) or as separate surveys. The main aim of the research was to outline options for obtaining nationally representative estimates of crimes against these groups.

Children

- The research has concluded that including a sample of children in the BCS would be feasible and that this should be done by including children aged 10-15 identified in the households selected for the core BCS. Interviews should only be attempted in households where an interview with the main adult was achieved.
- There are two approaches that could be used to obtain information from these children: a mix of CAPI, CASI and audio-CASI or a paper questionnaire. The decision depends on the balance between costs and quality/flexibility. If the higher quality computer interviewing is used, then the number of children selected in each household should be limited to one. If a paper questionnaire is used, then more than one child in a household could be included.
- Information from the children should not be combined with the core BCS data – the two should be analysed separately. Therefore, questions for the child sample could be specifically written for that age range.

Communal establishments

- The research also concluded that it would not be feasible to cover communal establishment residents in the BCS in a statistically reliable manner without incurring very substantial additional costs. If a decision were made to cover this population, a properly resourced feasibility study would be essential.
- Because lists of communal establishment residents do not exist, it would be necessary to draw a sample of institutions first. Communal establishments might be sampled either from the Postcode Address File (PAF) or directly from lists of types of communal establishments e.g. lists of educational institutions or care homes.
- The PAF does not identify communal establishments and identifying a useable sample would require considerable address screening, which would be time-consuming and costly.
- Available lists of communal establishments are insufficiently comprehensive to form the basis of a full sample frame, although it might prove feasible to construct a frame covering some populations, in particular institutions for students, nurses and older people.

Contents

Keywords

British Crime Survey (BCS)

Children

Under 16s

Communal establishments

Institutions

Group residences

Smith review

Sampling

Methodology

Home Office

British Crime Survey: options for extending the coverage to children and people living in communal establishments

Kevin Pickering, Patten Smith, Caroline Bryson and Christine Farmer
National Centre for Social Research and Ipsos MORI

This report presents findings from methodological research carried out to examine the feasibility of covering children (under 16s) and people living in communal establishments as part of the BCS or as separate surveys.

This research was commissioned in response to recent reviews of crime statistics carried out by the Statistics Commission and an independent cross-party group led by Professor Adrian Smith. The reviews highlighted criticism about the coverage of the BCS, which is currently restricted to measuring crimes experienced by adults resident in private households and excludes crimes committed against: under 16s; those living in institutions, communal establishments or on the streets; and businesses.

The main aims of the methodological work were:

- to examine the feasibility of covering children and those living in communal establishments as part of the main BCS or as separate surveys; and
- to outline different options for obtaining nationally representative estimates of crimes against these groups, as well as estimates of crime-related perceptions.

Children

One of the aims of the research was to assess the approaches that could be taken to obtain a sample of children and to investigate features of the design, in particular: how to obtain the sample; the appropriate age range; the survey mode; ethical considerations and sample size.

Children aged younger than 10 years should not be included in the BCS child sample, as they are less likely to have the appropriate cognitive skills and the non-coverage of this group is likely to have only a marginal impact on victimisation estimates for the under 16s. Restricting the survey to those aged 10-15 means that a single questionnaire could be produced to cover the whole age range, thus reducing complexity and costs. This would not be the case if children younger than 10 were included. The information collected from the child sample would need to be analysed separately from the main BCS data, so the questions

could be written to ensure that they could be completed by children aged 10-15.

The best approach to obtain a sample of under 16s would be to select them from the households that were selected for the main BCS survey rather than from alternative sources such as child benefit (CB) records or via schools. Children in the eligible age range (10-15) would be identified in 15 per cent of households in the main BCS sample. By selecting children in these households, adequate sample sizes to obtain nationally representative estimates by age group and sex would be achieved.

The children should either be interviewed using computer-assisted interviewing (CAI) a mix of computer-assisted personal interviewing (CAPI), computer-assisted self interviewing (CASI) and audio-CASI or the children should complete a paper questionnaire. The former would produce higher quality data with great flexibility for the interview, but at a higher cost. The authors recommend that only one child should be interviewed in the household if CAPI, CASI or audio-CASI is used, so as to minimise the burden on households. If a paper questionnaire is used, then more than one child in a household could be included. This choice does, therefore, also impact on sample sizes. With a paper questionnaire it would be possible to achieve a sample size of over 6,000 children. With the computer interviewing, the sample size would be about 5,500.

Communal establishments

To date, BCS has not covered the population of communal establishment residents[1] because it comprises a small proportion of the adult population and would be costly to cover. This research examined whether this conclusion remains justified for the BCS.

The communal establishment resident population made up about 2.1 per cent of the adult population at the time of the 2001 Census; the BCS already covers nearly 98 per cent of the population aged 16 and over. Extending the BCS to cover communal establishments would, therefore, have little impact on the overall estimates of the prevalence of victimisation produced from the BCS.

---

1   We use the terms communal establishment and institution interchangeably.

Home Office

No lists of people living in communal establishments currently exist and any sampling method would therefore require identification of residents at a previously drawn sample of institutions. The authors identify two feasible approaches to sampling communal establishments:

- The first is to screen a Postcode Address File (PAF) sample for the presence of institutions. Small-user PAF addresses would be screened during main BCS fieldwork. A larger-user PAF sample would be drawn and screened using the Inter-Departmental Business Register (IDBR). This approach proved to be workable in a pilot study by the Office for National Statistics (ONS). However, because institutions vary widely in size, this approach will only give statistically efficient estimates if large institutions are sampled at higher rates than small institutions. It may be possible to achieve this to a limited extent by adjusting the relative small-and large-user PAF address selection probabilities, although recent changes to the structure of PAF may render the method ineffective. Efficient samples of residents could be drawn if size information were available for a large initial sample of institutions,

as this would enable a sample of institutions that was disproportionately stratified by size to be drawn. The initial institution sample would need to be very large, and because each member of the initial sample would need to be contacted for size information, the cost of this approach would be considerable.

- The second approach would be to compile a frame of communal establishments from available listings. This would require considerable work, and would probably not result in a fully comprehensive frame. It may prove feasible to construct comprehensive frames of residential and care homes, educational establishments and nurses' homes for BCS use, although this would also require considerable time and effort. These would cover about two-thirds of the institutional population.

In conclusion, it would not be feasible to cover communal establishment residents in BCS in a statistically reliable manner without incurring very substantial additional costs. If a decision were made to cover this population, a properly resourced feasibility study would be essential.

British Crime Survey: options for extending the coverage to children and people living in communal establishments

Kevin Pickering, Patten Smith, Caroline Bryson and Christine Farmer
National Centre for Social Research and Ipsos MORI

1. Introduction

The Research, Development and Statistics Directorate at the Home Office commissioned the National Centre for Social Research and IPSOS-MORI to carry out methodological research to advise on possible strategies for extending the British Crime Survey (BCS)[2] to cover some populations currently not included in the BCS: children (under 16s) and those living in communal establishments.

This research was commissioned in response to recent reviews of crime statistics carried out by the Statistics Commission and an independent cross-party group led by Professor Adrian Smith. The reviews highlighted criticism about the coverage of the BCS, which is currently restricted to measuring crimes experienced by adults resident in private households and excludes crimes committed against: under 16s; those living in institutions, communal establishments[3] or on the streets; and businesses.

The main aims of the methodological work were:

- to examine the feasibility of covering under 16s and those living in communal establishments, either as part of the main BCS or as separate surveys; and
- to outline different options for obtaining nationally representative estimates of crimes against these groups.

More specifically, the research examined possible sample designs, data collection methods and other related issues (e.g. any ethical issues).

One of the key requirements was that the proposed options had to take into account the context of the main BCS, and to consider possible effects the proposed options might have on the main survey if implemented. The research was also required to take into account how feasible possible strategies would be in practice, and value for money considerations.

This report presents findings of the methodological work. Section 2 examines the feasibility and options for covering children, and Section 3 communal establishments.

2. Children

Background

This section investigates the approaches that could be taken to obtain a sample of children[4] for the BCS. Key components of the design that are covered include:

- an appropriate sampling frame to obtain a sample of children;
- the appropriate age range and its implications for question design;
- the relative advantages and disadvantages of different survey modes;
- practical and ethical issues to consider when implementing the survey among children;
- the required sample size.

In considering the optimal approach for the survey of children in BCS, a number of factors and constraints had to be considered.

- There should be minimal impact on the design of, and estimates from, the core adult sample to ensure consistency of trend data;
- A boost sample of young people (aged 16-24) is also required for the BCS; currently, this sample is selected in a subsample of households participating in the BCS.

---

2   For more information about the BCS see
    http://www.homeoffice.gov.uk/rds/bcs1.html
3   The Smith review used the term 'group residences' but here the term 'communal establishments' is used as it is the more usual term used in the context of social surveys.

4   Note that in this report the word 'children' is used to refer to people younger than 16. When the authors talk about 'young people' in this context, they are referring to the additional sample of 16-24-year-olds that is currently collected for the BCS.

Home Office

- The sample size for the survey of children will be either 2,000, 4,000 or 6,000. [5]
- There was no expectation that the estimates from the child and core surveys would be combined (i.e. the two samples would be analysed separately).

Surveys of children

The British Crime Survey has included a sample of children[6] only once (Hales, 1993). Although the survey was carried out 15 years ago, aspects of its design and performance will still be relevant for a survey carried out now and hence are reviewed in this section. In addition to the 1992 BCS, the authors have reviewed the design for a number of other national surveys that have included children. These are summarised in Appendix A.

For the 1992 BCS survey, all children between the ages of 12 and 15 who were identified in households where an adult interview took place were eligible for the child survey. The administration of the child questionnaires was carried out after the victim forms (if any) had been completed in the adult interview. The children were given a paper questionnaire to fill in while the remaining part (about 20 minutes) of the adult interview was completed. The completed questionnaires were sealed in envelopes by the children and taken away by the interviewers. If any child was not present in a household at the time of the initial interview with the main adult respondent, then an appointment was made for a further visit to administer the questionnaire, rather than leave a questionnaire to be completed. This was so that the questionnaires were always completed in the presence of the interviewer and taken away straight afterwards.

The first reason given in the technical report (Hales, 1993) for using a self-completion questionnaire was "the lack of privacy in many households, and the need to ensure that parents did not overhear the interview". It should be noted that the 1992 BCS was carried out before computer-assisted interviewing (CAI) was used on the BCS (it was introduced for BCS 1994), and, therefore, there was no scope for using computer-assisted self-interviewing (CASI), which would give children as much privacy as a paper questionnaire (see Section on survey mode). The second reason given for using a self-competition questionnaire was that there "was also the likelihood that more than one young person [child] would be eligible for the survey in some households, and this would involve interviewers spending an unacceptably long time in the household." This would still be the case for a survey carried out in the future (see Section on sampling design).

In the 10,059 participating households, 1,733 eligible children were identified for the survey and 1,350 (77.9%) questionnaires were completed, with non-response resulting from: parental refusal[7] (10.5%); child refusal (3.2%); and other unproductive outcomes, such as non-contact (8.3%).

The questionnaire included questions on victimisation outside the home, contacts with the police and self-reported offending and, where possible, the questions replicated those in the adults' version. The child questionnaire was 28 pages long and took between 20 to 40 minutes to complete, depending on cognitive ability and the number of relevant questions. The authors would be concerned about issuing a questionnaire with that many pages for a current survey. It is likely that it would either put off potential respondents or mean that the questionnaire was not completed – this might explain the relatively high proportion of other unproductive interviews (8.3%) in 1992.

Sampling frames for a sample of children

This section describes three potential sampling frames for samples of children and considers whether they would be appropriate for the BCS. A summary of the advantages and disadvantages of each is included in Appendix B.

Postcode address file

The issued sample of addresses for the BCS is selected from the Postcode Address File, a sampling frame of (nearly) all addresses in the United Kingdom, which is maintained by the Royal Mail. The core sample for the BCS is designed to achieve interviews with one randomly selected adult in approximately 46,000 households. For the 2005/06 BCS, core adult interviews were achieved in 47,479 households. In these participating households the total number of children of any age younger than 16 identified was 23,214, a sufficient number to obtain adequate samples of children for the BCS without any need for additional boost samples of children (see Section on sampling design).

There are a number of advantages of selecting the sample of children from the participating households in the BCS. The main advantage is the efficiency for fieldwork – children in participating households are identified as part of the BCS anyway, so no additional work is required to identify the sample of children. This approach also allows more complex analyses of the data to be carried out, as selecting the sample of children within the same households as the core adult sample would allow analyses of the associations between the children's and parents' (or other adults') findings.

The main disadvantage of sampling additional children in the core BCS households is the additional burden to the households and the interviewers. At about 15 per cent[8] of households, additional interviews would be required compared with the current design of the BCS. This additional burden would have an impact on the (composite) response rate for child interviews compared with an approach that used households not selected for the core BCS. The impact of this burden does depend on other aspects of the design – for example, a paper questionnaire for children completed at the same time as the core interview, as used for the 1992 BCS survey, would be less burdensome than carrying out CAPI/CASI interviews with all selected household

---

5   These sample sizes were selected to obtain robust national estimates based on cost constraints.
6   The sample was referred to as a sample of young people in the BCS 1992.
7   Note that parental refusal will include parents refusing on behalf of the child.

8   Assuming children aged 10-15 are to be included in child sample (see Section on the age range for the children's survey and its implications for question design).

members (see Section on survey mode). Note that, if the child interviews were introduced during the core BCS interview, as was the procedure for the 1992 BCS (see Section on surveys of children), then it is unlikely that there would be an impact on the response rates for the core sample. It is, therefore, recommended that this is the approach used.

It would be possible to obtain additional samples of children by carrying out screening and/or focused enumeration at additional addresses selected from the PAF. However, at present there is no requirement for samples of children for the BCS additional to those identified from the core household sample (as these would be sufficient for providing nationally representative estimates for the key groups), so these methods are not developed further in this report.

Child benefit records

An alternative sampling frame for children are child benefit (CB) records. CB records are maintained by the Department for Work and Pensions (DWP) on behalf of HM Revenue & Customs (HMRC) and include the names and addresses of parents, usually mothers, registered to receive child benefit. Because the CB records include the date of birth for all children, in theory, a sample of addresses that have a child in the required age range listed in the CB records could be obtained.

One advantage of sampling from CB records is that the sample would consist of children whose household had not already been selected for the core BCS sample. This would reduce the burden within households selected for the BCS. However, to make contact with these additional households and to persuade them to take part would mean the fieldwork burden and costs incurred would be higher.

Before a sample can be obtained from the CB records, HMRC require that an opt-out procedure[9] is carried out – this usually removes about five to 10 per cent of the sample. In addition, for most parents, child benefit is paid straight into their bank accounts; therefore, there is little impetus for them to update their addresses on the CB records. Because of this, a relatively large proportion of addresses (about 10%[10]) are found to be inaccurate in the field. These two factors tend to have a negative impact on the composite response rate for samples from CB records.

Because of the additional fieldwork burden and the likely higher composite non-response rates, the authors do not recommend sampling from CB records in preference to selecting the children in the core BCS households (see Section on Postcode Address File). One additional issue to note with using the CB records is that permission would be required from HMRC to use them for sampling.

Sampling via schools

Another approach to obtain a sample of children would be to sample via schools. This is the approach used to obtain a sample of 11-15-year-olds for the Information Centre for Health and Social Care's *Smoking, Drinking and Drug Use among Young People in England* (SDD) survey (Fuller, 2006). A random sample of schools is selected and then within the selected schools a fixed number of pupils sampled at random. The paper questionnaire is then completed by the selected pupils in a classroom under the supervision of an interviewer (see Appendix A for more detail).

The authors do not recommend that this sampling method is used to obtain a separate sample of children for the BCS. A lot of work is required to get the schools to participate in the survey and to obtain a register of pupils required for the second stage of sampling – even with this effort, about 30 per cent of selected schools do not participate in the SDD survey. There are some additional issues with this approach. First, the survey would have to be carried out during either the winter or autumn school terms, so the interviewing period would not match that of the adult survey. Second, it would not be feasible to use computer-assisted interviewing – the survey would have to be administered as a paper questionnaire (see Section on survey mode).

Summary

Obtaining the child sample from households already participating in the main BCS sample is the best, and therefore the recommended approach from those considered. Even if the additional burden on the households did impact on participation rates for children, the composite response rate is still likely to be highest for this sample – this would reduce the likelihood of bias compared to using the other sampling methods. There is a low risk that the additional interviewing with children would impact on the response rate of the core sample at the 15 per cent of households that would be eligible for a child interview.

Survey mode

In theory there are a range of possible survey modes but the conclusion from the review of sampling frames in the Section on sampling frames for a sample of children, i.e. to select the sample of children from the core sample of households, suggests two feasible options: either a paper self-completion questionnaire, or a mixture of computer-assisted personal interviewing and computer-assisted self-interviewing.

Computer assisted interviewing

In terms of the quality of the data collected, computer-assisted interviewing would be the preferred option. Using CAI reduces the amount of missing data, and checks can be included in the programme for the feasibility, eligibility and consistency of the responses entered, thus reducing response error compared with a paper questionnaire. Also, CAI can automatically route the respondent to the correct place in the questionnaire based on their previous responses, thus preventing respondents from skipping questions by mistake and allowing for more complex routing/questionnaires. In addition to improved data

---

9   An opt-out procedure is usually carried out by post and is an opportunity for parents to remove themselves from the sample – effectively, it is in lieu of seeking permission from parents for their details to be passed onto the survey organisation. Once parents have opted not to take part in the survey, they cannot be contacted again.

10 This estimate (10.2%) is based on a sample of parents of children aged 14 and under for the *Childcare and Early Years Provision 2007* survey. Note that this figure had increased from 7.5 per cent for the 2004 survey (Bryson *et al.*, 2006).

quality, CAI also allows automatic adjustment of question text based on the answers to previous questions or dates, and gives greater flexibility for having different versions of the questionnaire if required.

The negative consequences of using CAI for the child survey are the increase in interviewer time, and hence costs, of the additional CAI interviewing[11] and also the added burden on the household. Because of the added burden on the households from carrying out additional CAI interviews, a limit is often set on the number of children (and adults) that can be selected within a household. For the *Health Survey for England*, this limit is set at two children per household. For the *Offending, Crime and Justice Survey* (OCJS), the limit was one respondent aged 10-25. Depending on the length of the interview, the authors recommend that a limit of either one child or, if necessary to achieve the required sample sizes, two children per household is set for the BCS.

In the past few years, increasing use has been made of CASI whereby the computer is handed to the respondent with the questions displayed on the screen and the respondent enters their answers directly into the computer. Self-completion methods such as CASI are recommended for asking sensitive questions which respondents might not want to disclose to an interviewer in a standard face-to-face interview. For example, CASI is used in the BCS adult interviews for sensitive questions related to drug use and drinking, stolen goods and interpersonal violence.

Most survey organisations' good practice guidelines would tell interviewers to ensure that a responsible adult (usually a parent) was at home during the interview. This is, amongst other reasons, to ensure the safety of the child and to reassure the parent. In addition, from work carried out recently with children about their views in taking part in survey research (Reeves *et al.*, 2007), it is known that younger children, in particular, choose to have their parents near at hand for security and/or support. So, although parents would not necessarily be in the room during the interview, they are likely to be within earshot. While this is, of course, a crucial part in making families comfortable in taking part in the study, it is necessary to ensure that children can give uninhibited replies. *Therefore, the authors recommend that any modules which contained questions which a child would not be comfortable answering if their answers were to be seen or overheard by an interviewer or their parent would need to be administered using CASI.*

Audio-CASI
There has been an increasing body of evidence indicating the potential benefits of using audio-CASI, as opposed to conventional CASI, to help respondents who may otherwise find self-keying problematic, for example because of reading difficulties (Gatward, 2002; Tourangeau and Smith, 1996; O'Reilly et al, 1994). Audio-CASI involves respondents hearing questions and the possible set of answers through headphones, so they do not need to read the questions and answers on screen (although they do need to recognise numbers on the key pad to enter their response).

A large component of the *Offending, Crime and Justice Survey* was carried out with audio-CASI, although respondents still had the option of turning off the audio-CASI and simply reading the questions on the computer screen. Based on data from the 2005 OCJS, the younger respondents were more likely to find audio-CASI useful and to listen to the questions: 43 per cent of 10-11-year-olds found audio-CASI very useful, compared with 22 per cent of 14-15-year-olds; and 64 per cent of 10-11-year-olds listened to all the questions, compared with 35 per cent of 14-15-year-olds. Of the 14-15-year-olds, 26 per cent listened to none of the questions compared with just 10 per cent of 10-11-year-olds.

Given it is likely that the lower age limit for the BCS would be 10 (see Section on the age range of children), the authors recommend that audio-CASI be considered. However, there should be an option for older children to have the audio component turned off. One of the key drawbacks of audio-CASI is that it takes longer for the respondents to complete, as the rate at which they can listen to questions being read out is slower than the rate at which they can read questions on screen for themselves without audio assistance. Therefore, there is no gain in using audio-CASI if respondents are able to read and answer CASI questions themselves.

Paper self-completion questionnaire
The main advantage of using a paper questionnaire over CAI for the BCS children survey is that the paper questionnaire(s) can be completed by the child(ren) while the selected adult is being interviewed. This means that data can be collected from children in the household, but for only a minimum increase in interviewer time compared with carrying out an adult interview only (in most cases). As interviewer time is a major component of survey costs, issuing the child survey as a paper questionnaire would be significantly cheaper than using CAI. However, it should be recognised that the interviewing task may become more difficult, for example handling questions from a child respondent whilst continuing with the adult interview.

Another advantage of using paper questionnaires would be to reduce the burden on the selected households compared with CAI interviews for children. This would mean that there is less impetus to set a limit on the number of eligible children selected in a household. Indeed, for some surveys where paper questionnaires were used for the child sample, all children are included, for instance the *Young People's Social Attitudes Survey 2003* and all waves of the *Welsh Health Survey* (see Appendix A).

It should be noted, however, that for most surveys the questionnaires can be left for the children to complete in their own time and then picked up by the interviewer at a later date. Such a design would be problematic for the BCS. Given the sensitive nature of the questions, the questionnaire would have to be completed by the child with the interviewer present and taken away immediately afterwards, as was the design for the 1992 BCS child sample (see Section on Surveys of children). If children were not present at the time of the interview with the adult respondent, appointments would need to be made so that questionnaires could be completed at a later date. Therefore,

---

11 Each interviewer carries only one laptop so any CAI would have to be done sequentially.

one might want to impose a limit on the number of children sampled for the BCS child survey just to reduce the number of repeated visits to the same household.

Summary

The two most effective approaches to collecting the survey data are: either a paper self-completion questionnaire, or a mix of CAPI, CASI and audio-CASI. The main competing factors for deciding between these two survey modes are cost and the quality of data. CAI would collect higher quality data and allow for more flexibility in the information collected. However, the amount of time the interviewer would spend at a household with eligible children would mean that the cost would be increased compared with using paper questionnaires. It is, therefore, a decision that needs to be reached based on whether obtaining improved quality of data represents value for money.

The age range for the children's survey and its implications for question design

A decision made about the lower age range of the children who could or should be included in the BCS needs to be based on:

● the age at which children have the cognitive ability to answer the questions;
● the age at which the questions and topic areas become relevant to children;
● the age at which parents would be willing to allow their children to take part.

It is increasingly accepted that, from a cognitive perspective, children can participate in survey research from around the age of eight. However, that age is only a rough proxy for verbal, cognitive, intellectual and emotional capacity. This variation needs to be accounted for in the design and implementation of the study.

Of course, the questions would need careful design and cognitive testing[12] to ensure that they are asked and presented in an appropriate fashion for the age of the children. In addition, one must also ensure that the topic areas are ones which are relevant and appropriate to ask of children. The authors question the rationale for including children as young as eight and nine for a number of reasons. First they were concerned, ethically, about the potential to worry younger children by asking about crimes that could possibly be committed towards them. (Of course, the same issues about the potential to worry apply to older children and this needs to be taken into account during the design.) Second, the age of 10 would be a reasonable lower age limit, given that one expects to see children having a greater amount of independence from this age. Third, including children younger than 10 would require either that questions in the survey were asked in such a way that they were relevant and could be understood by these young children or for a different version of the questionnaire to be produced.

---

12 Cognitive testing was used in the development of questions for the *Offending, Crime and Justice Survey* (Phelps *et al.*, 2006; Home Office, 2005). For a review of the issues related to developing questions and carrying out cognitive interviewing with children, refer to de Leeuw *et al.* (2002).

The authors therefore recommend including children from the age of 10 in the BCS children's survey.

The authors were not asked to make recommendations about the exact content of the children's questionnaire. However, they did see it as part of their remit to comment on:

● the extent to which a single version of the questionnaire is appropriate for children aged between 10 and 15;
● and (partially linked to the first point) the optimal length of the questionnaire.

The cognitive range of children aged from 10-15 is wide, particularly taking into account the fact that some 10-year-olds will have a lower than average cognitive ability and some 15-year-olds higher. The authors have, therefore, considered the appropriateness of using one instrument for the whole age range. The arguments for having two different versions of the questionnaire, say one for 10-11/12-year-olds and another for 12/13-15-year-olds, would be:

● being able to word questions (or, if on paper, use more complex routing) that are appropriate for children of different ages;
● being able to ask older children about issues which may not be relevant to the younger children;
● being able to conduct a longer interview with older children, rather than being bound by an interview length that is suitable for the youngest respondents;
● (for self-completion questionnaires) being able to design questionnaires which are attractive and appropriate for children of different ages.

But the disadvantages are:

● the difficulties in combining data from the two age groups if the question or answer categories are worded differently for children of different ages;
● the administrative complexities of different versions, particularly if on paper.

Taking these pros and cons into account, the authors recommend having a 'core' interview where the same questions are asked of all children. Although this means 'pitching' them at the lowest age range, the authors think this is doable for those aged 10-15 and preferable compared to the disadvantages regarding combining data. (Certainly, if the authors had recommended an age limit lower than 10 [e.g. eight], they would have had to advocate more than one version of the questionnaire, which was one of the arguments for not using that lower age limit.) If taking this approach means that there are key areas omitted that would be appropriate with the older children (aged 12 or 13 upwards), the authors suggest the inclusion of additional 'modules' asked only of these older children.

In the Section on survey mode, the authors discussed the survey mode and the relative advantages of CAI and paper questionnaires. Administratively, it would be easier to implement this 'modular' approach using CAI. If the survey was on paper,

two versions of the questionnaire would have to be produced. However, even if one did have one standard set of questions asked of all children, one might still want to consider having two versions of a paper questionnaire – exactly the same questions but different designs. Beyond the wording and appropriateness of the actual questions, a further difficulty of having one instrument for a wide age range is making it attractive and interesting to the children. What attracts younger children may seem patronising or 'babyish' to older children. It would be a relatively inexpensive exercise to produce two designs for the different age groups.

Decisions around the number of instruments is linked to the recommendations about interview length. As mentioned above, the older children would be able to cope with a longer interview than the younger children. *An interview length of around 20 minutes for the younger children is recommended by the authors*. CAI would facilitate a somewhat longer interview than paper (but still no more than 30 minutes) because of the ability to mix interviewer-administered and self-completion sections. *If it was decided to give older children additional modules, then around 30 minutes, but no longer, would be recommended.*

Fieldwork sampling procedure

Currently, for the BCS, the screening for the additional young person interview is carried out by listing the eligible young people in the sample contact sheet and then making a selection using a Kish grid. *It would seem sensible to screen for children in the same way as with the main survey and, assuming that the sampling is fairly straightforward, the authors recommend that this be done*. Screening using the sample contact sheet does restrict the complexity of any disproportionate sampling that can be carried out, so might not be feasible if the sampling design was more complex than could be simply carried out by an interviewer.

Any complex sampling procedure would probably require the relevant information to be entered into a computer first. The procedure for this would need to be considered so that it could be incorporated into the design of the main BCS survey. One approach that is used for the *Health Survey for England* (see Appendix A) is for the CAPI programme to select the children to be interviewed once the household grid has been entered in the main household interview. *Therefore, if complex sampling procedure is adopted, the authors recommend that it is administered within the CAPI programme.*

Sampling design

Above the authors have concluded that the most cost-effective approach is to sample children from within households responding to the main BCS. Without more expensive boost samples via methods such as focused enumeration, the size of a child sample is naturally constrained by the number of eligible children resident in households responding to the main BCS and cost. The Home Office has indicated that it would consider three options for the achieved sample sizes for the children survey, which would be sufficient for providing nationally representative

estimates for key groups. When developing the sampling design, the following target samples have been considered:

- 2,000 in total – 1,000 girls and 1,000 boys;
- 4,000 in total – 1,000 in four subgroups defined by gender and age group (10-12 and 13-15);
- 6,000 in total – 1,000 in six subgroups defined by gender and age group (10-11, 12-13 and 14-15).

Obtain a child sample that is independent of the young person sample

The authors have considered four sampling designs to obtain the child samples for the BCS. For some, the sampling procedure varies depending on the number of eligible children (i.e. 10-15-year-olds) identified.

- Design A: for households with two or more children,[13] sample one child per household. For households with one child, randomly screen out half of the households.
- Design B: for households with two or more children, sample one child per household. For households with one child, screen in all households and include the child.
- Design C: for a household with three or more children, randomly sample two children. For households with one or two children, include all children in the household.
- Design D: for all households, select all children.

The first two Designs (A and B) would be particularly appropriate if the children were to be interviewed using CAI. This is because with CAI it would be sensible to limit the number of child interviews in the households to one in order to minimise the time the interviewer spends in any one household (see Section on CAI). The last two Designs (C and D) would be more suitable if the survey was to be administered using a paper questionnaire. In which case, the limit of the number of children interviewed per household has less of an impact on the time the interviewer spends in the household (see Section on paper self-completion questionnaires).

The estimated sample sizes[14] for the four competing designs from issuing all identified eligible children are shown in Table 2.1. In addition to the actual sample sizes, the effective sample sizes (NEFFs)[15] have also been shown. It is important to consider the effective sample size when designing the sampling. First, it gives an indication of the true precision of estimates that one does not obtain from the sample size alone. Second, it can also help to identify if a design is suboptimal. (Sampling efficiency is described in more detail in Appendix C).

---

13 When describing the sampling designs, the authors are actually referring to eligible children (i.e. aged 10-15).

14 The estimates of the number of eligible children are taken directly from the households from BCS 2005/6 at which a core BCS interview was completed, with an assumed response rate for children of 80 per cent.

15 The child weights were generated from the household weights in the BCS 2005/06, with an additional component for the selection of children in the household. They, therefore, allow for all selection (design) weights and include a component for household non-response. To estimate the clustering effect within PSUs, the authors have assumed the ROH for PSUs to be 0.01 and the number of PSUs to be 2,000. To estimate the clustering effect within households, they have assumed the ROH for households to be 0.05.

Table 2.1: Estimated sample sizes based on current main BCS sample

	Issued	Sample Size Achieved	NEFF	Design Effect
(A) Sample 1 child per household, subsample households with 1 child	4,695	3,755	2,945	1.28
(B) Sample 1 child per household	6,985	5,590	3,775	1.48
(C) Sample 2 children per household	9,180	7,340	5,610	1.31
(D) Select all children in households	9,475	7,580	5,880	1.29

Table 2.1 shows that the estimated number of interviews that would be achieved by limiting the number of children interviewed in each household to one (Design B) is 5,590. One negative consequence of restricting the number of interviews in a household is that a design effect is incurred, resulting from the selection within the household and the weights required to correct for it. As most households contain either one or two children, an efficient sampling approach is one that gives an (approximately) equal weight to households that contain one or two eligible children. Limiting the number of children in a household to one requires a selection weight equal to the number of eligible children in the household (i.e. the selection weight is equal to one for a household with one eligible child and two with two eligible children).

To make the sample more efficient, one could add an additional component of selection, whereby half of households with one eligible child are included and the other half screened out. This would imply that the selection weight for a household with one eligible child was now two, equal to the weight for households with two eligible children. This is the rationale for considering Design A – it generates a sample that is more statistically efficient[16] compared with a design that limits the number of children to one but includes all households (Design B). Of course, with Design A some potential sample is being discarded, so it would not generate a larger effective sample size. It should only be considered, therefore, if the achieved sample size from Design A was adequate for the analyses required.

With the assumptions made, it is clear that it would be possible to obtain a sample of 2,000 children with any of the four designs. In order to achieve a sample of 2,000, however, some

households would have to be screened out. There are several approaches for doing this, but it is proposed that it should be done while considering the impact on statistical efficiency (see below). It would not be possible to use Design A to achieve a sample size of 4,000. If the required sample size was 6,000, then it would be necessary to sample more than one child in households with two or more eligible children (i.e. Designs A and B would not be feasible).

Other sampling options
As described above, it is recommended that the sampling design considers statistical efficiency and aims to minimise design effects as much as possible, given other constraints. There are additional sampling designs that could be considered to improve the statistical efficiency further.

A large component of the inefficiency of the sample is a result of the disproportionate sampling of the PAF addresses by Police Force Area (PFA) for the core BCS sample. The impact of this on the child sample could be reduced by subsampling the households with eligible children with probability as close to the inverse of the original address selection probability as possible. For example, if an address in PFA 1 had twice the chance of being selected as an address in PFA 2, then if one included all eligible households in PFA 1, but sampled 50 per cent of eligible households in PFA 2, the composite selection weights for PFA 1 and PFA 2 would then be equal, and hence the sample more efficient.

A second sampling approach which would reduce the impact of clustering would be to ensure that the subsampling of households was done within the primary sampling units (PSUs) as much as possible. This would reduce the average cluster size and hence reduce the design effect due to clustering. The allocation of addresses to the child sample could be done at random when the BCS sample is selected, i.e. a proportion of addresses in each PSU would be eligible for the child sample. However this approach would add to the complexity of the fieldwork – interviewers would have a mix of core only, and core and child sample addresses in their assignments.

Sample sizes for subgroups
Table 2.2 shows the numbers that would be achieved in each age/sex group from the four designs. This table shows that the sample would be distributed fairly evenly across the age/sex groups, although it does suggest that the sample would be larger in the older age group. *Even so, the authors do not recommend disproportionately sampling children based on their age and/or sex.*

---

16 The component of the design effect for the selection of children within households already selected for the BCS for Design A is DEFF = 1.03, compared with DEFF = 1.18 for Design B.

Table 2.2: Achieved sample sizes by age group and sex				
	Design A	Design B	Design C	Design D
Boys:				
10-11	470	895	1,155	1,190
12-13	495	905	1,215	1,250
14-15	560	1,040	1,370	1,415
All	1,530	2,840	3,735	3,855
Girls:				
10-11	470	875	1,145	1,180
12-13	485	885	1,180	1,220
14-15	525	990	1,280	1,325
All	1,475	2,750	3,605	3,725
Total	3,755	5,590	7,340	7,580

Note that with the sample sizes proposed, it would not be possible to obtain estimates by some subgroups such as ethnic group and region.

Sampling the children and young people dependently

The BCS currently selects an additional sample of young people aged 16-24. In households containing young people aged 16-24, if the selected respondent is aged 25 or over, one of the young people is selected at random for the young person sample. The aim for the 2005/06 BCS was to obtain interviews with 2,000 young people. In fact, the design had to be changed for the final six months as the numbers achieved were greater than selected. Of the 3,211 young people issued for the survey, 2,301 (72%) responded.

If the child sample was to be interviewed by CAI, it would make sense to combine the sampling for the child sample and young people sample to reduce the burden on participating households that contained both children and young people. The authors have considered a number of designs for combining the sampling and have compared these against a design that keeps the sampling of young people and children independent (Design 1).

- Design 1: sample young people and children separately. Sample one child from all identified. Sample one young person from all identified, unless the core respondent is a young person.
- Design 2: randomly select one child in a household if the core respondent is a young person (aged 16-24), or randomly select one child or one young person if the core respondent is not a young person.
- Design 3: as for Design 2, but select an additional child in households from any not previously selected.

Table 2.3: Sample sizes for options for young person and child sampling			
	Design 1 (independent)	Design 2 (combined)	Design 3 (combined)
Children	5,590	5,085	6,840
Young people	3,025	2,555	2,555

The first design (Design 1) separately selects one child and one young person within participating households. This would generate a sample of 5,590 children and 3,025 additional young people (Table 2.3). If a sample of 2,000 young people were required (as for the 2005/06 BCS), then the additional young person could be screened out for one household in three. This design does imply that there is a chance of carrying out three interviews in a household – one with an adult (aged over 25), one with a young person and one with a child. However, this would actually be fairly uncommon. Of the 47,796 households in the 2005/06 BCS, the authors estimate that three interviews would have been attempted at about 1,300 households (3%) under this design. If one-third of the young person sample was screened out, this would have been about 850 households (2%).

The authors do not think that Design 1 should be dismissed out of hand, just because some households would be selected for three interviews. There are several advantages to Design 1:

- it is easy to implement the sampling, so it could be done by the interviewer using a Kish grid on the sample contact sheet;
- the weighting would be less complicated than using a combined sampling strategy;
- it would be easier to change the sampling design, for example, to sample additional children at addresses.

The simplest approach to combining the sampling, to ensure that a maximum of two interviews were attempted in any household would be to randomly select one child in a household if the core respondent is a young person (aged 16-24), or randomly select one child or one young person if the core respondent is not a young person (Design 2). This would give an achieved sample of 5,085 children and 2,555 young people (Table 2.3).

If a sample of 6,000 children – one of the target sample sizes – was required, it would be necessary to select a second child for interview in some households. One method for doing this that ensures that there is a maximum of three interviews in any household (with a maximum of two interviews with adults) would be to carry out the sampling as for Design 2 and then randomly sample an additional child from all those left unselected (Design 3). This would achieve a sample of 6,840 children and 2,555 young people. With this design, three interviews would be attempted at 2,200 households.

One of the disadvantages of sampling children and young people dependently is that the sampling probabilities (and hence selection weights) become quite complex. This complexity limits the amount of adjustment one can do with the sampling in order to make it more efficient. For example, it would be difficult to adjust Design 2 to subsample households to make the sample of children more efficient (as was done for Design A in the Section on obtaining a child sample that is independent of the young person sample), when the sampling is dependent on both the number of young people and the number of adults in the household. However, it would still be possible to disproportionately sample based on PFA to reduce the variance

of the selection weights (see Section on obtaining a child sample that is independent of the young person sample).

Summary

If the child sample was to be selected independently of the young person sample, then the choice of design would depend on the sample size required. An achieved sample size of 6,000 or more could only be achieved by issuing more than one child interview per household (see Table 2.1). Smaller sample sizes could be achieved by issuing either one per household or even subsampling households to achieve a more efficient design.

To restrict the maximum number of additional interviews per household to two would require the sampling for young people and children to be carried out dependently (see Table 3). This would give achieved sample sizes of 5,085 for children and 2,555 for young people (Design 2). To obtain 6,000 interviews would require additional child interviews to be carried out (Design 3). This would mean that some households would have three interviews carried out, although there would be a maximum of two interviews with adult household members.

Practical and ethical issues to consider when implementing the survey among children

Whichever survey organisation carries out the BCS fieldwork, it will have its own procedures for conducting research with children. The authors have, therefore, not felt it appropriate to provide firm recommendations within this report about the practical and ethical procedures that should be put in place. However, they have highlighted what they see to be some key issues. Some of these are generic to many studies among children; others are of particular relevance given (a) the approach to sampling and (b) the content of the interview.

- **Parents' role in the consent process.** Of course, any research with children needs to be done with the full knowledge and consent of their parents. Conversely, the final choice about whether or not to take part in the survey lies with the children, as does their right to have their answers kept confidential, including from their parents. It is therefore important to consider how best to approach parents – to ensure that they are happy for their child to take part and for their 'buy in' to their child's right to privacy and confidentiality. The survey organisation needs to ensure that parents are giving informed consent. This means that – when the survey is introduced and explained to parents either by letter or verbally – it explains key information about the topics covered in the interview, the interview length, together with the child's right to give his or her own consent, to privacy and to confidentiality (all expanded in the bullets below). Due consideration needs to be given to the appropriate level of detail that parents should be given about the exact content of the interview (from giving them the full (blank) questionnaire to providing only broad topic areas).

- **Ensuring that children are active in the consent process.** In thinking about parental consent, it is important to recognise that permission is being sought to approach a child, rather than asking for consent for a child to participate. Efforts should be made to minimise parental refusal without a child's involvement in the decision process. The 1992 BCS had a reasonably high level of parental refusal, and attempts should be made to reduce this as far as possible. Conversely, procedures should be put in place which ensures that children give informed consent rather than take part by default given the involvement of their parent in the main survey. Careful consideration should be given to providing child-specific information to the children about the survey and to briefing the interviewers about approaching the children. From the authors' research (Reeves *et al.*, 2007), children of all ages thought that parents had a role to play in the consent process. However, there were strong views about the individual choice that children should be able to make around whether or not to participate in a survey.

- **Presence of the parent during the interview.** This was raised above (see Section on CAI) in relation to its effect on the recommendations around survey mode. The potentially sensitive nature of some of the questions points both to the importance of the presence of parents and to the need to maintain confidentiality for the child. Thought needs to be given to discussing with the parent and the child the appropriateness (or otherwise) of the parent being in the same room (as opposed to in the house but not in the same room).

- **The interviewer: respondent dynamics.** As with all interviews with children, it will be important to establish the rules of engagement between the interviewer and child. A key issue which emerged from the Reeves *et al.* (2007) study was the importance to the children of their 'relationship' with the interviewer (leading to feelings of obligation, of comfort/discomfort, etc.). Careful briefing of interviewers is required to ensure that children feel comfortable communicating with the interviewer if, for example, they want to end the interview/questionnaire early or want to refuse a particular question or set of questions. This is of particular importance given the potentially sensitive nature of some of the questions.

- **Confidentiality and disclosure of harm:** All survey organisations will have their own procedures in place around confidentiality and disclosure of harm (i.e. if a respondent reveals that they are, or are in danger of, being at serious risk of harm). This will be an important element of the procedures for the main element of the BCS, given the topic area, as well as for the children's survey. The survey organisation will have to think through carefully how it briefs children about what will be kept confidential, and its policy for the extreme circumstances when they might need to implement their policy on disclosure of harm.
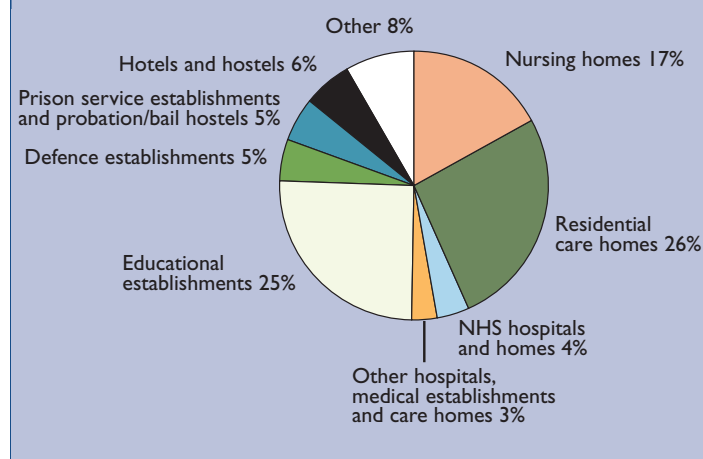
Conclusions

Including a sample of children in the BCS would be feasible and the authors believe that the best approach to doing this would be to include children identified in the same households that participate in the BCS for the core sample. Given their cognitive abilities, it is recommended that children younger than 10 should not be interviewed for the BCS. Therefore, the sample would be of children aged 10-15 – it is estimated that adequate sample sizes of children in that age range would be obtained from the households selected for the core BCS.

The children should either be interviewed using a mix of CASI and audio-CASI or should complete a paper questionnaire. The former would produce higher quality data with great flexibility for the interview, but at a higher cost. It is recommended that only one child should be interviewed in the household if CASI or audio-CASI are used. This is to minimise the burden on households. If a paper questionnaire is used, then more than one child in a household could be included.

The information collected from the child sample would need to be analysed separately from the main BCS data. Therefore the child interview would not be limited by the questions asked in the main BCS interview.

3. Communal establishments

In common with all the other major population surveys in Britain, the BCS does not cover the population of communal establishment residents. [17] To date, it has been judged that the benefits of covering this relatively very small population do not justify the additional effort and costs of so doing. This section investigates whether this conclusion remains justified for the BCS.

Size of communal establishment population

The 2001 Census showed 934,263 people, including staff and their families, as living in communal establishments. It is estimated that about 870,799 of these were aged 16 or over. This is 2.1 per cent of the population of England and Wales aged 16 or over.

The largest number of residents aged over 16 is to be found in nursing homes (143,887), residential care homes (227,657) and educational establishments (220,854). Together these categories comprise 68 per cent of the institutional population aged over 16. The nursing homes and care homes house predominantly elderly residents (79% and 74% respectively of their resident populations comprise non-staff residents aged 65 and over), whereas, educational establishments primarily house younger people (95% are under 35).

This indicates that if one was able to cover fully just nursing homes, residential care homes and educational establishments, two-thirds of the communal establishment population aged 16 or over would be covered. This would improve coverage of the population aged over 16 from 97.9 per cent to 99.3 per cent.

Figure 3.1: Communal establishment population aged 16 and over

Other 8%
Hotels and hostels 6%
Prison service establishments and probation/bail hostels 5%
Defence establishments 5%
Educational establishments 25%
Nursing homes 17%
Residential care homes 26%
NHS hospitals and homes 4%
Other hospitals, medical establishments and care homes 3%

Source: Census 2001

Drawing a sample of communal establishment residents

In the following the authors first discuss how the communal establishment population might be defined, and then discuss two general approaches that could be taken to drawing a sample of communal establishment residents. They then briefly discuss methods for sampling individuals in sampled establishments.

Defining the population
The definition of the survey population of residents of communal establishments should be complementary to the definition of residents of private households. If this is not done, some classes of individuals may either be included in both private residential and communal establishment samples or be omitted from the survey altogether; in either case some bias would be likely to ensue.

In general terms communal establishments are defined as establishments which provide professionally managed residential accommodation (Gatward *et al.*, 2002). However, the distinction between communal establishments and private households is not always clear cut, and it is therefore necessary to provide specific instructions to deal with borderline cases. For example NatCen brief their interviewers to treat addresses in which four or more unrelated or partly unrelated people are catered for communally as communal establishments, but to treat similar addresses in which three or fewer people are catered for communally as private households.

The BCS documentation provides a rough definition of institutions/communal establishments by means of examples (see Appendix D for extract from interviewer instructions). This contrasts with most large government-funded surveys which

17 Note that in the discussion the authors use the terms *communal establishment* and *institution* interchangeably.

offer more precise definitions. In practice, the exact definition of institutions used by interviewers will have no discernible impact on the main BCS results because they are encountered so rarely. *However, the authors recommend that as a matter of good practice a clearer definition of communal establishment should be used in future rounds of the BCS whether or not the sample is extended to cover communal establishment residents.*

Outline of possible sampling approaches

The Office for National Statistics (ONS) has previously investigated possible methods for including communal establishments in population surveys. The results of these investigations are well documented (Bruce, 1993; SSD, 2002), and ONS also provided the authors with additional material.

It is clear from the ONS work that no lists of communal establishment residents exist and that, as a result, the communal establishment population cannot be sampled directly. This points to a multi stage sample design whereby residents are selected only at previously sampled communal establishments (in other words where communal establishments are used as cluster units). Key to the success of this approach, of course, is the availability of good coverage sampling frames of communal establishments.

On the basis of the ONS work there are two general types of approach that might be taken on sampling communal establishments. The first is to draw a large sample from one or more general establishment/address sample frames which cover, but do not separately identify, communal establishments, and to screen this sample for communal establishments. The second is to collate one or more sampling frames on which communal establishments are *already identified as such*, and sample directly from this list. The two approaches are discussed in the next two Sections.

Screening a general sample frame

This approach involves drawing a sample from a general sample frame on which institutions are *not* separately identified and which is known or assumed to have good coverage of institutions. Sampled addresses are screened to establish whether or not they are communal establishments. Such screening might be done in the office or in the field by interviewers.

From the previous ONS work it appears that, in practice, PAF is the only candidate for a general frame. The ONS have successfully piloted an approach to sampling communal establishment residents based upon PAF.

In the ONS pilot, a sample of communal establishments was selected from PAF, and residents were sampled within them. Communal establishments were sampled from both small-user and large-user PAFs.[18] A different procedure was used to identify communal establishments in each of the two PAF component

files. The small-user file component of the sample comprised addresses which had initially been selected for the Labour Force Survey (LFS) but had then been coded by interviewers as out-of-scope because they were institutions. The large-user file component was sampled by: selecting a large sample of addresses from a large-user PAF (on the assumption that only 2.5 per cent of these would be eligible); looking up these addresses on the Inter-Departmental Business Register (IDBR);[19] and screening out ineligible addresses on the basis of IDBR Standard Industrial Classification (SIC92) codes. It is understood that there may also have been an element of manual checking of the organisation, department or address fields in the PAF file.

Interviewers sampled residents in each co-operating and eligible establishment. This involved:

1. visiting the establishment and identifying all residents aged 16 and over and who satisfied standard address residency requirements, namely that they lived or intended to live at the institution for six months or more, or that they had no other usual address or that the communal establishment administrator considered the institution to be their main address;

2. sampling up to 20 residents, where the exact number depended upon the number of residents living in the establishment.

Sixty-six per cent of the issued communal establishment sample was found to be eligible (small-user addresses: 66%; large-user: 71%). Although this is slightly unclear in the technical report, the establishment response rate appears to be 57 per cent (small-user addresses: 56%; large-user ones: 59%) and the response rate amongst selected individuals was 81 per cent. The final unweighted net response rate can therefore be estimated as 57 per cent x 81 per cent = 46 per cent. Although this is not high,[20] the authors feel that it is high enough to warrant exploring the methodology further for the BCS.

If the ONS pilot approach were adopted for the BCS without major change it would involve a methodology along the following lines:

- identifying all the addresses that were included in the main BCS sample (drawn from the small-user Postcode Address File) that were subsequently coded by interviewers as being ineligible because they were institutions;
- drawing a sample from the large-user Postcode Address File using a selection probability that was similar to or larger than that used for the main BCS sample addresses;
- screening the large-user file sample against IDBR to remove all cases that can be clearly recognised as not being institutions;

---

18 Postcodes on PAF are categorised into large and small. Small-user postcodes identify groups of addresses (on average around 15), whereas larger-user postcodes identify single addresses receiving large quantities of mail. To qualify for inclusion in the large-user file an address must receive a minimum number of mail items per day. This minimum has increased over the years and is currently set at 500.

19 The Inter-Departmental Business Register is a list of UK businesses maintained by the Office for National Statistics (ONS) which combines the former Central Statistical Office (CSO) VAT-based business register and the former Employment Department (ED) employment statistics system. It is used for selecting samples for surveys of businesses and to produce analyses of business activity.

20 It is worth noting part of the small-user file sample was issued late in fieldwork and this reduced the institution response rate somewhat.

- manually inspecting the organisation/department/address fields of remaining large-user PAF cases, and removing further addresses that can be clearly identified as ineligible;
- having interviewers visit sampled establishments to confirm eligibility and select a sample of residents.

According to the ONS study, institutions vary widely in size (see Table 3.1). This means that if the ONS pilot methodology were to be adopted, without allowing institution selection probabilities to vary substantially, the sample of communal establishment residents would necessarily be statistically inefficient.[21] In theory, one would have to choose between a sample that was inefficient because it contained a number of very large clusters and one that was inefficient because it necessitated using a wide range of weights. In practice, one would not wish to interview a large number of individuals at any one institution because of the burden this would place on it, and one would, therefore, have to use a design requiring use of a wide range of weights. A design of this sort might cap the number of individuals selected at any one institution at, say, 20. With moderate cluster homogeneity (roh = 0.04), one would anticipate that such a design would deliver a design effect of about 10.5. This would mean that the effective sample size would be only 9.5 per cent of the actual sample size.

Table 3.1: Size distribution of establishments included in ONS pilot study by PAF component from which sampled

Establishment size (total persons)	Small-user file		Large-user file	
	n	%	n	%
0-10	54	45	2	20
11-20	24	20	0	0
21-30	10	8	1	10
31-50	16	13	1	10
51-100	8	7	0	0
101-200	4	3	1	10
201-500	4	3	2	20
501-1000	0	0	1	10
1001-2000	1	1	1	10
2001+	0	0	1	10
All	121	100	10	100

The statistical efficiency of this design would increase if the sample of larger institutions could be sampled at higher rates than smaller ones. If one then sampled *residents* at lower rates in the larger sampled institutions than in the smaller ones, the design would place some control on variation in both cluster sizes and weights, and this would reduce design effects relative to a sample in which all institutions were selected with equal probability. The difficulty with implementing such a design, however, is that it requires one to have a reasonable idea of institution size before the sample is selected.

The simplest, but crudest, method for sampling larger institutions at higher rates than smaller ones would be to apply a larger sampling fraction to addresses on the large-user PAF than to ones on the small-user PAF. Efficiency gains would arise to the extent that, on average, institutions present on the large-user PAF were larger than those present on the small-user PAF. Table 3.1 demonstrates that this was indeed the case at the time of the ONS study. The authors estimate that if the number of establishments sampled from large-user PAF was increased fourfold and the number of residents selected in institutions containing 50 or fewer residents were reduced to one in 10 (i.e. between 1 and 5) residents, the effective sample size would be about a third of the actual sample size (again assuming a roh value of 0.04).

Little improvement beyond this could be made by adjusting relative small-user and large-user frame sampling fractions because the association between an establishment's size and the frame it appears on is relatively poor (see Table 3.1) and selection probabilities will not vary by size within each frame.

A major limitation of this approach is that the maximum attainable effective sample size would be limited by the fact that small-user PAF institutions would be identified as part of main BCS fieldwork, that it to say the number of available small-user PAF institutions identified would depend on the main BCS sample address selection probabilities. The authors estimate that it would not be possible (under current assumptions) to increase the effective size of the communal establishment resident sample to more than about 240 using this design.

The authors are also concerned that changes to the make-up of the PAF files in recent years[22] may mean that many larger institutions may have migrated from the large-user PAF to the small-user one. If this is the case, the efficiency gains made by boosting numbers selected from the large-user PAF would be diminished.

If one wishes to achieve significantly greater increases in efficiency and/or effective sample size, they need good information on the sizes of the institutions appearing on the large-and small-user PAFs, as this will enable them to draw a sample of institutions that is disproportionately stratified by size. In Table 3.2, for example, a disproportionately stratified sample is presented in which: the sampling fraction used to select the largest institutions is 20 times the size of the one used to select the smallest institutions; no more than 20 residents are selected per institution; and the authors set out to achieve around 500 interviews after taking account of an institution response rate of 70 per cent and a within-institution resident response rate of 80 per cent. The authors estimate that the design effect for this sample would, assuming moderate cluster homogeneity (roh =0.04), be around 1.41, meaning that the effective sample size (around 350) was 71 per cent of the size of the actual sample size (494).

---

21 More efficient samples of a given size produce smaller standard errors than less efficient ones. Formally, the efficiency of sample A of size n relative to that of sample B of size n is equal to the ratio: sampling variance for sample B/sampling variance for sample A.

22 Since the ONS study was done the criterion for inclusion in the large-user file has increased from receiving 50 items of mail a day to receiving 500 items per day. The authors have not been able to obtain any definitive information from the Post Office as to what this means in practice, but it may well mean that the small-user PAF now includes a good many more large institutions than it did at the time of the ONS study.

Institution size stratum	No. institutions from which sample selected	Institution relative sampling fraction	Total no. institutions selected	Total no. institutions participating (70%)	Within-institution resident sampling fraction	Mean no. residents selected in each institution	Mean resident sample fraction	Total no. residents selected in stratum	Total no. residents participating (80%)
0-10	241	0.05	12	8	0.30	2	0.30	13	10
11-20	100	0.05	5	3	0.30	5	0.30	16	13
21-30	48	0.06	3	2	0.24	6	0.30	13	10
31-50	73	0.10	7	5	0.15	6	0.30	31	25
51-100	35	0.19	7	5	0.08	6	0.30	28	22
101-200	23	0.38	9	6	0.04	6	0.30	37	29
201-500	31	0.88	28	19	0.02	6	0.30	116	93
501-1,000	8	1.00	8	6	0.02	11	0.30	65	52
1,001-2,000	13	1.00	13	9	0.01	20	0.27	185	148
2,001+	8	1.00	8	6	0.01	20	0.13	115	92
All	**580**		**100**	**70**				**618**	**494**

Table 3.2: *Example of sample disproportionately stratified by size*

It should be noted that with this design, size information is required for considerably more institutions than will appear in the sample. This means that, compared with the design where each institution has the same selection probability, considerably more PAF addresses have to be screened for the presence of an institution. For the large-user PAF, more addresses would have to undergo IDBR and manual checks. For the small-user PAF more addresses would have to be checked during main BCS fieldwork. The latter would have major resource implications.

The authors estimate that to implement the design presented in Table 3.2 they would need to have successfully collected size information for some 580 institutions, 498 of which would have come from the small-user PAF. On the basis of figures in the 2005/06 BCS technical report (Grant *et al.*, 2006) and the ONS pilot report it is estimated that the 2005/06 issued BCS sample of 68,848 addresses would generate about 89 eligible institutions. Therefore, if one were to implement this disproportionately stratified design one would need to have identified (and obtained size information for) 498 instead of the 89 or so institutions identified during main BCS fieldwork. This would mean that for every issued main sample BCS address an additional 4.6 other addresses (over 300,000 additional addresses in all) would need to be screened in the field for the presence of an institution.

This additional address screening could take place during main BCS fieldwork by issuing interviewers with additional PAF addresses to screen. Another approach might be to seek details of institutions identified in other large random probability small-user PAF-based surveys.

The above sample design would deliver an effective sample size of around 350. Increasing the number of residents selected from institutions would have very little impact on this figure: although the sample size would increase, this would be counterbalanced by a reduction in efficiency. Real increases in effective sample size would require a larger sample of institutions, which, in turn, would require more PAF addresses to be screened.

A statistically efficient sample based on the ONS pilot design would, then, require considerable in-field screening of small-user PAF addresses. This level of screening could be considerably reduced if one could identify lists of institutions which provided good coverage of medium and large institutions (say those with 50 or more residents) and use these to construct a sample frame. If such lists could be identified, it would be possible to design an efficient sample in which little, or no, additional screening of PAF addresses was required. The idea would be:

1. to draw a sample from this institution frame using a relatively large sampling fraction (say 20-50 times larger than that used in main BCS);

2. contact these (probably by telephone and post) to establish the number of residents at each;

3. identify all institutions at main (small-user PAF) BCS issued sample addresses;

4. draw an initial large-user PAF sample using similar selection probabilities to those used for addresses in the main BCS sample;

5. identify which members of the large-user PAF initial sample were institutions using IDBR;

6. check the list of small-and large-user institutions identified from PAF in the preceding steps against the institution frame and remove all found to be on it;

7. contact the institutions remaining on the PAF after step 6 and establish the number of residents;

8. combine the institution frame addresses with the residual PAF-identified institutions, and draw a sample disproportionately stratified by size from this combined list.

Whether or not this would be possible would depend entirely on whether a suitable institution sample frame with good coverage of medium and large institutions can be identified.

In conclusion, we note that a method has been successfully tested by ONS in which a PAF address sample is screened for communal establishments, and in which those identified are approached for a sample of residents. However, this method cannot be used to generate large, statistically efficient samples of communal establishment residents without modification. Any such modification would need to be preceded by further feasibility work, and would almost certainly prove costly to implement.

Sampling from lists of communal establishments

This method involves the compilation of one or more sample frames of institutions from available lists of communal establishments (e.g. residential care homes). A sample of communal establishments would be drawn from this frame, following which a sample of residents living within the institutions would be drawn.

In the following discussion, how existing lists of communal establishments might be used to create a general sample frame for *all* types of institutions is examined. Much of this is guided by Bruce's (1992) review of sample frames conducted in the early 1990's. Following this the authors focus on three institution types which, between them, cover a large proportion of the communal establishment population and discuss procedures for compiling sample frames for these.

There are three lists identifying a range of types of communal establishments which might be regarded as candidate sample frames for institutions in general: ONS List C, Census Communal Establishment file and the Yellow Pages. Unfortunately none of these provides adequate coverage of the total communal establishment population (see box below).

Existing sample frames

ONS List C
List C is a list of communal establishments in England and Wales. It is used for coding births and deaths that occur outside a private household and is updated on a regular basis. Whilst the list provides good coverage for hospitals, nursing homes and homes for the elderly, its coverage for other types of communal establishments (e.g. hotels, military establishments, student halls) is poor since there are fewer births or deaths in these.

Census Communal Establishments files
These files are produced to help enumerators in the field, and although they provide relatively good coverage at the time of the Census, they quickly become out of date.

Yellow Pages
The Yellow Pages provide contact details for communal establishments but the coverage is likely to be poor because it is limited to establishments listed in the Yellow Pages, some listings will be imprecisely coded, and some establishments may also be listed under more than one category.

In the section on screening a general sample frame it was noted that IDBR has been successfully used to screen large-user PAF addresses for institutions. Although the authors have not explored the possibility in this work, they feel it may be worth investigating whether IDBR might be used *directly* as another general list of communal establishments.

There are also several lists of specific institution types. The availability and quality of these varies by type of institution as can be seen in the box below.

Hotels, guest houses and (tourist) bed and breakfasts
The main sources are tourist board lists of tourist accommodation (VisitBritain, VisitWales and Enjoy England). We are unsure of how good the coverage of these lists is and would recommend that, at a minimum, they be cross-checked against lists of tourist accommodation and local authority lists in the Yellow Pages.

Student halls of residence/flats/boarding schools
There are no comprehensive lists of student accommodation. However, the Department for Children, Schools and Families hold lists of all schools and further and higher educational establishments, from which accommodation lists might be obtained (see further discussion below).

Hospitals, nursing/dual registered homes and residential homes
ONS List C would be the most suitable for this group since it provides good coverage of these types of institutions (as opposed to institutions in general). Two other options are (i) Department of Health (DH) and Welsh Office (WO) lists which provide information on the number of beds within establishments and (ii) the CareSearch database which holds a range of other information on care homes. These lists are discussed further below.

DWP bed and breakfast accommodation, miscellaneous hotels, lodging houses and accommodation for homeless people
There are no comprehensive lists available for these establishments. One possibility may be to construct partial lists from hostel providers such as the YMCA. Any lists would need cross-checking against other lists (e.g. tourist accommodation) and the small-user PAF.

Children's homes
There are no comprehensive lists available for these establishments.

Accommodation provided by employers
There is no readily available list apart from one for nurses' accommodation, held by the LFS (discussed below).

A considerable amount of work would be required to construct good quality sample frames for a number of these establishment types. It is also likely that this would prove impossible for some establishment types (e.g. employers' accommodation and hostels).

It may, however, be feasible to use available lists to cover some of the larger institutions, in particular nurses' accommodation, educational establishments, and nursing and residential homes. Each of these is discussed further below.

Nurses' accommodation
The LFS has a sample frame of nurses' accommodation[23] which they have used since 1992. The sample frame was constructed by contacting all district health authorities and NHS trusts and asking them to supply a complete list of their accommodation (Collins, 1992). Sample frame coverage was good at the time of compilation – of the 455 authorities and trusts contacted, information was received from 417 providing 92 per cent coverage of nurses' accommodation. The sample frame lists all rooms in nurses' accommodation ordered by postcode.

The LFS still use the original sample frame which is some 15 years old. In 2005, they started work to update the list which was originally expected to last six months but has since been extended. It is expected to be complete for the 2008 survey.

If the Home Office wished to include individuals in nurses' accommodation the authors recommend that ONS should be contacted with a view to being given access to the updated LFS sample frame when it becomes available.

Educational establishments
The Department for Children, Schools and Families (DCSF) holds a register, 'EduBase'[24], of all educational establishments in England and Wales. The information on the register is provided by local authorities, educational establishments and specialist agencies. The database is available to buy and can be filtered to include particular establishment types (e.g. universities, higher education colleges and boarding schools). The database covers all educational establishments in England and Wales (approximately 10,000), although establishments can choose to opt out of the list made available to the public. Of the 10,000 establishments listed, about 700 have opted out, giving 93 per cent coverage. It may be possible for the Home Office to negotiate with DCSF to obtain access to the complete list including establishments that have opted out.

To compile a sample frame of educational communal establishments, it would be necessary to contact all educational establishments (including schools) providing accommodation for students aged 16 and over. The office or individual responsible for student accommodation within each establishment would need to be contacted and asked to provide full details of student accommodation. This would probably be a fairly labour-intensive exercise.

Nursing, residential and dual-registered homes
The authors are aware of three available sample frames to cover this population: ONS List C, the Department of Health and corresponding Welsh Assembly lists, and the CareSearch database. ONS List C includes information on the name and address of each communal establishment, along with institution type codes and is known to have good coverage of this population. The DH and the Welsh Assembly hold lists of nursing and residential homes, including information on the number of beds, based on returns from registered homes. The CareSearch database, produced by Laing & Buission[25] contains the names and addresses of registered care homes, as well as information on the number of beds and detailed resident profile data (e.g. aged over 65, whether residents have learning disabilities, whether residents are terminally ill). The database providers were not able to give any information on the database coverage.

The authors suggest exploring the use of ONS List C in conjunction with the DH and Welsh Assembly lists, and possibly also in conjunction with the CareSearch database. Whilst List C provides good coverage of nursing and residential homes, the other lists provide additional information (notably, number of beds) which would be useful for sampling.

Selecting individuals

The foregoing discussion has focused upon selecting samples of institutions from which residents may then be selected. One would expect that, in most cases, residents would be selected by interviewers in the field. Probably the only exception to this would arise if the ONS sample frame of nurses' homes were used, because this lists rooms rather than institutions.

Interviewers would select individuals by identifying and listing, in a systematic order, all *eligible* residents in the institution (or by identifying all eligible residents on pre-existing lists of residents), and then using a look-up table to select a systematic sample from across the list sample. In reality the look-up table might take the form of a CAPI program; this would have the advantage over paper-based methods that it would make it less administratively onerous to randomise systematic sample start numbers across the sample of institutions.

Conclusions

It is estimated that 2.1 per cent of the population of England and Wales aged 16 and over live in communal establishments. This means that, although BCS does not cover the communal establishment population, it still covers 97.9 per cent of the population as a whole, and that it is most unlikely that this non-coverage bias will be significant relative to other potential sources of bias such as measurement error and non-response.

---

23 In fact this NHS accommodation also houses some non-nurse residents such as doctors.
24 See http://www.edubase.gov.uk

25 See http://www.laingbuisson.co.uk

This also indicates that extending the BCS to cover communal establishments would have little impact on the overall estimates of victimisation prevalence rates produced from the BCS.[26]

This argues against expending major resources to cover the communal establishment population, unless there is a requirement to collect separate estimates for it.

Because no list of communal establishment residents exists, any sample design will require a minimum of two sample stages such that: an initial sample of communal establishments is drawn; and a residents' sample is drawn from these.

A method has been successfully tested by ONS in which a PAF address sample is screened for communal establishments, and in which those identified are approached for a sample of residents. Although the method has been demonstrated to work procedurally, it cannot be used to generate statistically efficient samples of communal establishment residents without modification. It may prove possible to select a small, relatively inefficient but useable sample of communal establishment residents without large additional resource input by adjusting the ONS pilot design, but this is by no means guaranteed given recent changes to the structure of the PAF files.

More efficient samples based upon the PAF approach will require very large numbers of PAF addresses to be screened, something which would be time consuming and costly. It would, however, prove possible to reduce substantially the level of PAF screening required if a sample frame with good coverage of medium and large institutions could be found or constructed.

In theory, it would be possible to compile a sampling frame from available administrative lists of communal establishments. However, the authors have concluded that any attempt to construct a *comprehensive* frame of establishments in this way would require very substantial resource investment and would almost certainly not meet with total success. It may, however, prove feasible to construct sample frames for educational establishments, nursing/residential homes and nurses' homes, which between them cover over two-thirds of the communal establishment population. This would still, however, be a labour-intensive and costly exercise.

It is likely that it would not be feasible to cover communal establishment residents in BCS in a statistically reliable manner without incurring very substantial additional costs. *Furthermore, even if a decision were made, in principle, to cover this population the authors recommend that a properly resourced feasibility study should be conducted.*

---

26 For example, if one were to assume that those in communal establishments had around twice the level of victimisation as the population resident in private households, because they form a relatively small subgroup of the overall population, the inclusion of those in communal establishments would increase prevalence of victimisation in England and Wales by only around half a percentage point (BCS 2006/07, Nicholas *et al.*,2007).

References

Bryson, C., Kazimirski, A. and Southwood, H. (2006) *Childcare and Early Years Provision: a Study of Parents' Use, Views and Experience*. Nottingham: Department for Education and Skills.

Bruce, S. (1993) *An investigation of sampling frames of communal establishments for the Family Expenditure Survey*. London: OPCS.

Collins, D. (1992) *Building a sampling frame of institutions – nurses' home*. Survey Methodology Bulletin. London: Social Survey Division, OPCS.

de Leeuw, E., Borgers, N. and StrijBos-Smits, A. (2002) *Children as Respondents: Developing, Evaluating, and Testing Questionnaires for Children*. International Conference on Questionnaire Development and Testing Methods, Charleston, South Carolina.

Fuller, E. (ed) (2006) *Smoking, drinking and drug use among young people in England in 2005*. Leeds: Health and Social Care Information Centre.

Gatward, R. (2002) *Interviewing children using audio-CASI*. Social Survey Methodology Bulletin, No 50, pp 16-26.

Gatward, R., Lound, C. and Bowman, J. (2002) *A pilot survey of people living in communal establishments*. Labour Market Trends, March 2002, pp 141-149.

Grant, C., Bolling, K. and Sexton, M. (2006) *2005/6 British Crime Survey (England and Wales) Technical Report Volume 1*. BMRB Social Research.

Hales, J. (1993) *British Crime Survey 1992 (England and Wales)*. London: Social and Community Planning Research.

Home Office (2004/5) *Crime and Justice Survey: General Population Feasibility Study*. London: Home Office Online Report.

Nicholas, S., Kershaw, C. and Walker, A. (2007). *Crime in England and Wales 2006/07*. Home Office Statistical Bulletin 11/07. London: Home Office.

O'Reilly, J., Hubbard, M., Lessler, J., Biemer, P. and Turner, C. (1994) *Audio and video computer assisted self interviewing*. Journal of Official Statistics, 10, 2, pp 198-214.

Phelps, A., Maxwell, C., Anderson, T., Pickering, K. and Tait, C. (2006) *2004 Crime and Justice Survey (England and Wales) Technical Report*. London: National Centre for Social Research/BMRB Social Research.

Reeves, A., Bryson, C., Ormston, R. and White, C. (2007) *Children's perspectives on taking part in survey research*. London: National Centre for Social Research.

Smith, A. (2006) *Crime statistics: An independent review*. Independent Report. London: Home Office.

Social Survey Division, Office for National Statistics (2002) *Survey of people living in communal establishments (pilot): technical report*. London: ONS.

Statistics Commission (2006) *Crime statistics: User perspectives. Statistics Commission Report No. 30*. London: Statistics Commission.

Tourangeau, R. and Smith, T. (1996) Asking sensitive questions: the impact of data collection mode, question format and question context. *Public Opinion Quarterly*, 60, pp275-304.

Appendix A: Children samples for other surveys

1. Health Survey for England (core sample)

- Two children (aged 0-15) selected in participating households. (All adults are included)
- CAPI interview carried out for children aged 0-15, completed by the parent if child aged 12 or younger and child if aged 13 or older.
- A self-completion paper interview is also completed by children aged 8-15. There are two versions: one for children aged 8-12, the other for children aged 13-15.
- If child is not present in the household for visit (to complete one of the adult interviews), then information is not collected.
- Response rate for children's interviews is 93 per cent for all ages and 87 per cent for the age range 13-15 for which the children complete the interview.
- CAPI interview length: 20-40 minutes.
- Self-completion: 10-15 minutes.

2. Welsh Health Survey

- All children (aged 0-15) in participating households included.
- Paper self-completion questionnaire, filled in by all children aged 13-15. For children aged 12 or younger, questionnaire completed by either the child or the adult.
- There are three different versions of the paper questionnaire for the age groups 0-3, 4-12 and 13-15.
- Interviewers introduce the questionnaires to the appropriate person and collect them around three to four days later.
- Of the 4,829 eligible children, there were 4,114 (85%) completed questionnaires. Refusal rate was two per cent and 11 per cent were not returned.
- Questionnaire length: 10-12 pages. (Adult version: 21 pages)

3. National Travel Survey

- All children (aged 0-15) in participating households included.
- CAPI interview for children, completed by child if aged 11-15 and by a parent if 10 or younger (but preferably with the child present if aged 5-10).
- A travel diary is also left for children, completed by child if

aged 11-15 and by a parent if 10 or younger. This is picked up by the interviewer at the end of the 'travel week'.
- Interview length: 50 minutes for whole household.

4. Offending, Crime and Justice Survey

- Young person sample defined as people aged 10-25.
- For wave 1:
    core sample of one adult per household – include those aged 10-25;
    boost sample of 10 to 25-year-olds identified by focused enumeration.
- For wave 2:
    one young person (aged 10-25) selected in each household;
    households including a young person identified by screening/focused enumeration.
- CAPI/CASI interview. CASI used for sensitive questions.
- Children (aged 10-15) complete all questions, except the household grid. Note that questions asked vary by age group.
- Response rate for eligible households estimated to be about 75 per cent.
- CAPI/CASI interview length: 60 minutes.

5. Young People's Social Attitudes Survey

- All children/young people aged 12-19 in the household of a British Social Attitudes Survey (BSA) respondent. Note if the BSA respondent was aged 16-19, they were excluded from the young person's sample.
- Only young people in households for which a core BSA interview had been obtained were included.
- BSA respondent was not aware of the young person interview until after completing the interview.
- Interviewer had to make contact with every eligible young person in the household. Therefore, if a young person was not present, interviewers had to make additional visit(s).
- Young people eligible for interview = 997. Interview obtained = 993 (66.5%). (Refused = 24.1%; non-contact = 3.9%; other = 5.5%.)
- CAPI.
- Length: 30 minutes.

6. Smoking, drinking and drug use among young people in England

- Random sample of schools selected. All types of school are eligible (comprehensive, secondary modern, grammar and private schools), except special schools.
- Approximately 35 pupils selected from years 7 to 11, so selected pupils are mostly aged 11-15.
- Paper questionnaire.
- A lesson is set aside for the selected pupils to complete the questionnaire in examination conditions, under the supervision of an interviewer.
- The completed questionnaires are taken away by the interviewers.
- Participation rate for schools = 70 per cent. Response rates for selected children in participating schools = 89 per cent.
- Length: approximately 30 minutes.

Appendix B: Advantages and disadvantages of different sampling frames

	PAF	CB Records	Via Schools
Obtaining the sampling frame	Households already identified for core sample.	Would need to obtain permission to use from HMRC.	Would need to obtain permission from DCSF.
Selecting the sample	No additional work – already identified as part of the core households.	Quite complex sampling based on anonymised records. Extra field effort to contact selected households.	A lot of extra work is required to get selected schools to participate and to supply lists of students.
Impact on response rates	Households already participating – additional non-response for children likely to be low.	Additional opt-out stage (about 5-10% opt out). Problems with inaccurate addresses add to non-contact.	School participation rate <70 per cent for SDD 2005. Would need permission from parents.
Interview location	Home	Home	Schools – would be problematic to set up.
Feasible mode	CAI/Paper.	CAI/Paper.	Paper.
Analysis	Would be able to look at relationships between adults and children in the same household.		
Representativeness	Nationally and seasonally representative.	Nationally and seasonally representative.	Excludes certain schools (e.g. special needs) and children that are excluded or truants. Would not be seasonally representative.
Burden (interviewers)	Additional interviews within 15 per cent (if aged 10-15) of BCS households.	Additional burden of making contact and gaining response from fresh sample.	Would require extra interviewers to supervise at schools.
Burden (BCS households)	Additional burden for BCS households.	No additional burden for BCS households.	No additional burden for BCS households

Appendix C: Design effects & sample efficiency

There are three main factors that impact on the efficiency, and hence the effective sample size, of a sample: weighting, clustering and stratification. The impact of these factors is measured using design effects (DEFFs) or design factors (DEFTs). A related term is the effective sample size (NEFF).

The design factor is defined as the relative loss or gain in precision from having a complex survey design (CS) compared to a simple random sample (SRS) (i.e. one that has no clustering or stratification and is sampled with equal probability) with the same number of cases:

$$DEFT = S.E.(CS) / S.E.(SRS)$$

The effective sample size is interpreted as the sample size that would have been required for a simple random sample to give the same precision as the complex sample. The design effect is equal to the square of the DEFT and can be considered to be the relative loss or gain in the effective sample size from employing a complex survey design compared with a simple random sample with the same number of cases.

These three terms can therefore be derived from each other:

$$DEFF = DEFT^2$$

$$NEFF = N / DEFF = N / DEFT^2$$

where N is the size of the sample.

Three factors impact on the effective sample size..

1. In general, the more variable the weights, the higher the design effect and hence the smaller the effective sample size. Therefore, when designing the sampling, one aims to reduce the variability of the weights.

2. Clustering also impacts negatively on the efficiency – in general, the more clustered the sample, the smaller the effective sample size. The design effect for clustering is calculated from two components: the rate of homogeneity (ROH) within the cluster (effectively how similar the measures are within the cluster) and the average size of cluster (M):

    $$DEFF \text{ (clustering)} = 1 + (M - 1) \times ROH$$

3. Stratification, on the other hand, usually has a positive impact on the efficiency. If the sample is stratified by an external variable that is correlated with the survey measures of interest, then the sampling variance is reduced and hence the sample is more efficient.

Appendix D: Excerpt from BSC interviewer instructions on identifying institutions or communal establishments

Examples include nursing or residential care homes, hotels, hostels, NHS nursing accommodation and college halls of residence. Although these types of addresses might be thought of as residential, we are only interested in private residential addresses.

One must be aware of how to distinguish a communal establishment from a private residential establishment. In some cases the distinction between the two can be subtle. Three examples illustrate the potential difficulties.

- While residential care homes for the elderly are usually classed as communal establishments, sheltered accommodation for the elderly would normally be considered private residential addresses (even where there is a warden).
- While most hostel-type establishments are usually classed as communal establishments, bedsits would normally be considered private residential addresses.
- While army barracks are usually classified as communal establishments, private residences located on an army base would normally be considered private residential addresses.

In making these distinctions one should always try to think in terms of how people actually live at an address and the extent to which people live independently. Communal living is generally taken to be situations where people share meals together and also share communal living space. Where there is a degree of independent living with people generally cooking for themselves and having their own living space, this is generally regarded as private residential.