
Evaluation of survey data quality using matched Census-survey records

Amanda White, Stephanie Freeth and Jean Martin

1. Introduction

In England and Wales the Office for National Statistics (ONS) is responsible for conducting the decennial Census of Population. Following the censuses in 1971, 1981 and 1991 investigations into survey nonresponse on major government household surveys were carried out by the Social Survey Division (SSD) of ONS which carries out many of the major government household surveys. Addresses sampled for surveys taking place around the time of the censuses were linked with individual census records for the same addresses. Since all the Census variables are available for both responding and nonresponding addresses, this provides a very powerful means of investigating the characteristics of nonrespondents, measuring nonresponse bias and evaluating methods of adjusting for the bias.

For the studies carried out following the 1971 and 1981 censuses, the actual matched datasets were not available to the methodologists investigating nonresponse, only specified aggregate tables, which limited the analysis to descriptive comparisons of the characteristics of respondents and nonrespondents, and measurement of bias in terms of census characteristics. However, in 1991 matched micro records were made available under strict confidentiality arrangements which allowed much more detailed statistical modelling to be undertaken to investigate the interrelationship between different variables which relate to nonresponse bias and to assess different methods of re-weighting data to compensate for the bias. Five surveys were included in the study in 1991, allowing comparison of results for surveys with very different designs. Key results are presented in Foster (1998).

A more ambitious study is planned in connection with the 2001 Census. Some 8-10 surveys are likely to be included, with very varying survey designs and response rates. The response rates are expected to range from just under 60% to over 80%, based on current levels. More importantly, our knowledge of factors contributing to survey nonresponse has advanced considerably in the past ten years suggesting enhancements to the basic study design, while more powerful and sophisticated statistical modelling techniques are now available for the analysis.

This paper outlines our plans for the study following the 2001 Census and presents the results of the pilot for some aspects of the study.

2. Design of the 2001 Census-linked Study of Survey Nonresponse

The 2001 study builds on the previous census-linked studies of survey nonresponse carried out by ONS and by Groves and Couper (1998). The key feature of the 2001 study is the collection and use of a significant amount of auxiliary data to supplement the information available from the census. Advances in research on non-response have led to a better understanding of the probable causal mechanisms that determine likelihood of response. Groves and Couper (1998) have put forward conceptual models of the factors determining likelihood of contact and likelihood of co-operation given contact. They list four broad categories of influence:

- area characteristics
- household characteristics

- survey design features
- interviewer characteristics

Each of these combines with the others to affect both likelihood of contact, the interaction between the household and the interviewer and hence the likelihood of co-operation given contact. Their work emphasises that it is not demographic characteristics per se which determine non-response; rather that people with certain characteristics are likely to lead lifestyles or hold attitudes which determine how easy they are to find at home or persuade to take part in a survey. Interviewers have a huge influence on response outcomes. The attitudes and strategies they bring to their work and their detailed behaviour at the household have been shown to be major determinants of response outcome. The 2001 census-linked study of survey non-response plans to measure as many factors as possible which are likely to influence response outcomes in order to explore how area, household, interviewer and survey design characteristics interact to impact on non-response.

3. The data to be collected

The data to be used in the study will be drawn from census records and data of survey interviews carried out around the time of the 2001 Census (which took place on 29 April). The census data for households sampled for the participating surveys will be extracted and matched with the corresponding survey data and other information relevant to the analysis. The study will collect the following information:

- Information about the areas sampled for the surveys
- Information about households
- Survey design features
- Information about the interviewers
- Information about interviewer behaviour and outcome of visits to each sampled address.

Figure 1 Design of matched dataset for 1991 study

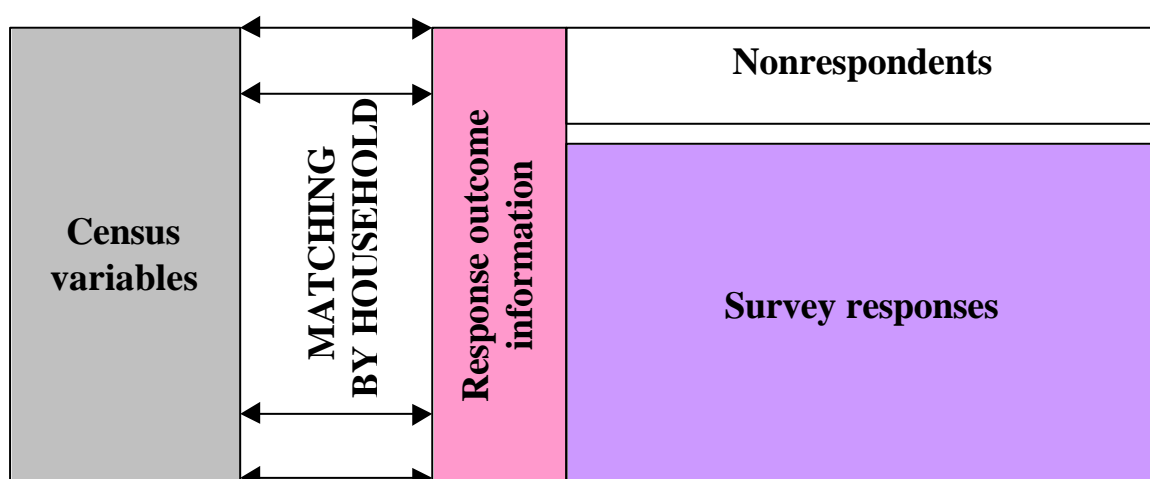
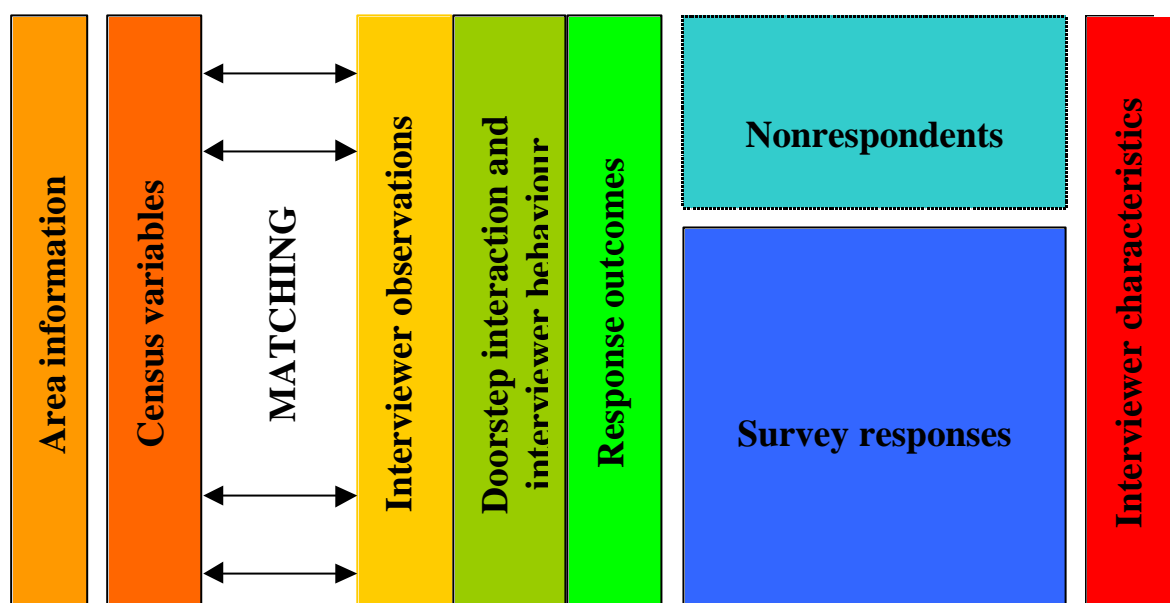


Figure 2 The design for the 2001 study



3.1. Information about the area

Much of the information available about the areas sampled for the surveys in the study will come from the census itself but some other information may also be available. We plan to use the following area level information:

- population density
- whether urban or rural
- crime rate
- unemployment rate
- proportion of owner occupiers
- proportion of multi-occupied units

We will also ask interviewers to record information about the area that might be predictive of ease of contact or gaining co-operation. In addition, we will investigate various area classifications that are currently available. Area level information can be linked to samples for any general population survey we carry out and is therefore available for routine non-response adjustment (and for sample stratification and analysis). This study will allow us to assess how well it compensates for non-response bias and how it may best be used.

3.2. Information about households

The study's database will have all the variables included in the census. We also aim to ask interviewers to record their observations about characteristics of the selected addresses/households which might be predictive of ease of contact (eg presence of entry phone or other barriers to access) or gaining co-operation.

3.3. Survey design features

The surveys that are likely to be included in the study are all carried out by face-to-face interview but vary with respect to the following design features:

- whether information is required about all the individuals in the household or a selected adult
- whether all adults are to be interviewed in person or whether proxy information is allowed
- the response rules which determine the response rate
- the average length of interview
- the length of the field period
- the survey topics covered
- whether non interview data collected is included (e.g. diaries)

We will record survey design information on the database for use in the analysis.

3.4. Interviewers characteristics

We plan to carry out a survey of interviewers, repeating an earlier survey of 1998 (Martin and Beerten, 1999) which ONS carried out as part of an international project (Hox, 1999). Based on the earlier results we expect to collect the following data:

- socio-demographic characteristics
- length and nature of survey experience
- performance grade (based on response and other factors over a year)
- confidence in ability to gain response
- knowledge of techniques which encourage response
- reports of calling strategies
- reports of strategies used to persuade people to respond

3.5. Interviewer behaviour and outcome of calls

As with the earlier studies, we will record the information about the visits interviewers make to each sampled address and the outcome of each call. We will also ask interviewers to record other potentially useful information about their own behaviour and the interaction with respondents. We expect to have the following information:

- time of day and day of week of each call
- outcome of each call
- number of calls to make initial contact
- number of calls to complete interview after making contact

- reasons given for not granting an interview
- questions and comments of respondent

4. Sample size

The sample size for each participating survey will be similar to the 1991 study - around 5,000 addresses. We plan to carry out a lot of analyses separately for each survey and a sample of 5,000 addresses will yield a large enough number of non-responding households for survey level analyses. As in previous years, responding households will not be sub-sampled because this will have adverse effects on the statistical models to be developed. Addresses in Scotland and Wales will be included for surveys that cover these countries.

5. Timing of the study

Ideally we want to match addresses which will be contacted as near to the census night (April 29 2001) as possible. This was carried out in previous years by including addresses selected for survey interviews in the months on either side of census night. In 2001, we propose to match census data with the data of addresses selected for interview from April and to start collecting the data outlined in section 3 in May. Matching addresses selected for interview before April is not feasible due to data compatibility problems linked to the adoption in April of the National Statistics Socio-economic Classification (NS-SEC) and the modification of a number of classificatory questions used on government surveys in Britain.

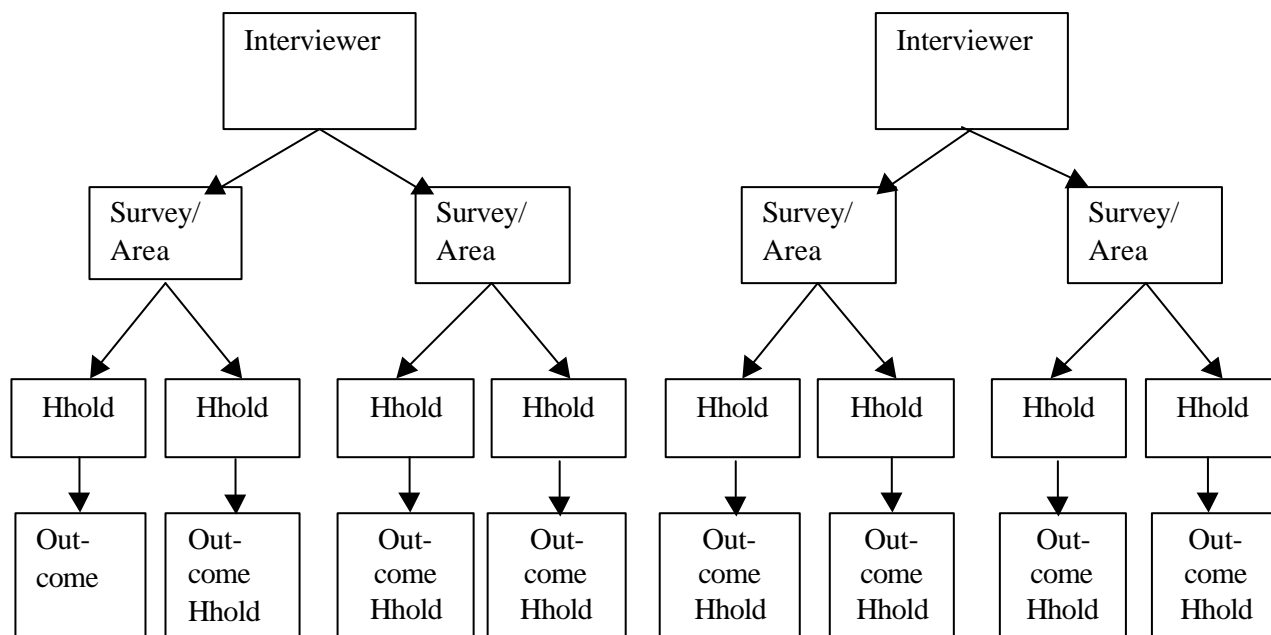
6. Analysis plan

Organisations sponsoring the surveys included in the study will be paying for their participation and will commission outputs for their particular survey. As with the 1991 study they will be offered a range of analyses from which to select. On the 1991 study all survey sponsors wanted basic descriptive analysis of the characteristics of nonrespondents and extent of nonresponse bias but only one commissioned detailed investigation of alternative methods of nonresponse adjustment.

Apart from survey specific analyses we plan to use this rich dataset to examine influences on nonresponse across the surveys. This follows the approach we used for the analysis of the effect of interviewer characteristics on nonresponse carried out recently (Martin and Beerten, 1999). For that study we did not have any information at household level whereas for this we will have a great deal. In planning these analyses we will take account of the hierarchical structure of the dataset, using multilevel modelling to distinguish effects at the different levels. Essentially we have three levels of information (Figure 3):

- a) interviewer – characteristics and attitudes
- b) assignment – survey design features, area characteristics
- c) household – interviewer observations, interviewer behaviour, household characteristics

Figure 3 Data structure



Our plan is to explore two aspects of non-response separately: whether the interviewer made contact or not; and for each household where contact was made, whether the household co-operated or not. Multi-level modelling is most suitable for investigating the determinants of response. With so many variables potentially available we need to have clear hypotheses about the likely relationships between them in order to decide in which order to enter variables and to interpret the results.

7. Output options

Reports giving the results relating to a particular survey will be prepared for the organisations sponsoring each participating survey. Other methodological papers and reports comparing the results across surveys and drawing more general conclusions about the nature of non-response will also be produced.

The analysis options we will offer include:

a) Descriptive comparisons of characteristics of non-respondents:

comparisons of characteristics in terms of census variables of responding, non-contacted and refusing households.

b) Measurement of non-response bias in terms of census characteristics:

summaries of over or under-representation of different subgroups in terms of census characteristics.

c) Comparison of the above with results from previous years (if available) to identify changes over time.

d) Logistic regression modelling to determine relative influence of different census variables on response outcomes.

e) Multilevel modelling to determine the relative effects of area, household and, if available or appropriate, survey design and interviewer level variables on response outcomes.

f) Development of weighting schemes based on the above analyses. Ideally these will incorporate information that could be collected routinely by interviewers as well as the information used in current non-response adjustment.

8. Development and implementation of the study

Development work for the study started last year. We have had discussions with colleagues in ONS Census Division about the arrangements required to match records. We have also carried out a lot of work to design the procedures for collecting the auxiliary data required for analysis. These were piloted in January and February 2001. The main purpose of the pilot was to:

- check that the procedures operate satisfactorily when used simultaneously on a number of large continuous surveys.
- assess the time interviewers need to collect the information to ensure that the module can be fitted into interviewers' assignments.
- obtain suggestions for improving the design to minimise burden on interviewers and to help develop training material for the main stage.

Nine interviewers and one interviewer supervisor working on a number of government household surveys carried out by ONS were involved in the pilot. The pilot interviewers provided feedback and advice on question wording and layout, and practical matters such as how to handle the material in the field. The key findings of the pilot were:

- interviewers can obtain the information required from observation or from their introductory conversation with the informant.
- the information required must be recorded as soon as practicable because the details can easily be forgotten. Interviewers working on the pilot recommended recording the information immediately after the contact with the household or on the same day if immediate recording is not possible.
- the procedure for recording the information has to be designed to encourage interviewers to record the information immediately after the contact. This means that a compact and user-friendly paper interviewer observation form has to be developed because it is not always safe or convenient for interviewers to use their laptop computer in their car.

We incorporated the findings of the pilot in re-designing the procedures and circulated the revised interviewer observation form to three of the original pilot interviewers, eleven other interviewers and an interviewer supervisor. This second round of consultation produced some suggested simplification to the lay-out of the form which were incorporated into the final design. (Appendix A).

The collection of auxiliary information has begun and most of the surveys will have completed the data collection by the end of September 2001. The census data will be extracted for matching in 2002 after Census Division has completed processing the census data. We expect to begin analysis in 2003 and to produce the first survey specific outputs later that year.

References

- Barton, J. (1999) Effective calling strategies for interviewers on household surveys. *Survey Methodology Bulletin*, 44, London: Office for National Statistics
- Campanelli, P., Sturgis, P. and Purdon, S. (1997) Can you hear me knocking? An investigation into the impact of interviewers on survey response rates. London: Social and Community Planning Research.
- Foster, K. (1998) Evaluating nonresponse on household surveys. *GSS Methodology Series No. 8*. London: Government Statistical Service.
- Groves, R.M. and Couper, M. P. (1998) *Nonresponse in household interview surveys*. New York: Wiley.
- Hox, J. J. (1999) The Influence of Interviewer Attitudes and Behaviour on Household Survey Nonresponse: an International Comparison. Invited paper presented at the International Conference on Survey Nonresponse, October 28-31 1999, Portland, Oregon.
- Lynn, P., Clarke, P., Martin, J. and Sturgis, P. (1999) The Effects of Extended Interviewer Efforts on Nonresponse Bias. Invited paper presented at the International Conference on Survey Nonresponse, October 28-31 1999, Portland, Oregon.
- Martin, J. and Beerten, R. (1999) The Effect of Interviewer Characteristics on Survey Response Rates. Paper presented at the International Conference on Survey Nonresponse, October 28-31 1999, Portland, Oregon.
- Morton-Williams, J. (1993) *Interviewer Approaches*. Cambridge: Dartmouth.

Appendix A

The Interviewer Observation Form

The Interviewer Observation Form is a double-sided A5 booklet. We have only reproduced the key pages here; the pages for recording calls 02 to 10 are broadly the same as the ones for recording call 01.

Interviewer Observation Form

Survey

Area

<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
----------------------	----------------------	----------------------	----------------------	----------------------

Address

<input type="text"/>	<input type="text"/>
----------------------	----------------------

Household

<input type="text"/>

Month

Is this a Reissue?

Yes 1

No 2

Interviewer Name

Interviewer Number

<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
----------------------	----------------------	----------------------	----------------------	----------------------

Please complete this form for **every** telephone or visiting call made at the selected address.

This includes the contact when an interview is obtained.

The questions refer to the **introductory conversation** with the member of the household. This is the conversation that took place between the time you introduced yourself and the time you either started the interview, or ended the contact.

Call 01

Q1. Date of call
(ddmmyyyy)

Q2. Time of call
(24h: hh.mm)

Q3. Outcome of this call
Code ALL that apply

HQ refusal	1
Ineligible: not built /demolished/derelict	2
Ineligible: vacant/empty	3
Ineligible: non-residential	4
Ineligible: residential but no resident hhd	5
Ineligible: communal estab / institution	6
Unable to establish eligibility	7
Address temporarily inaccessible due to weather or other causes	8
No answer on the phone or got answerphone	9
No face-to face contact with anyone at address	10
No contact with anyone at the address but spoke to a neighbour	11
Contact made with sampled household but not with a responsible resident	12
Contact made with responsible resident but not with selected respondent	13
Refusal at intro/bef int - by hhd member	14
Refusal at intro/bef int - by proxy	15
Refusal at intro/bef int – DK hhd or proxy	16
Refusal at intro/bef int - by selected resp	17
Refusal during interview	18
No interview due to language, age, infirmity, disability etc.	19
Appointment made	20
Interviewer withdrew to try again later	21
Appointment broken	22
Placement interview / checking /reminder call completed (diary surveys only)	23
Partial household interview completed	24
Hhd interview completed but non-contact with one or more elements	25
Hhd interview completed but refusal or incomplete interview by one or more elements	26
Full co-operation / interview	27
Other outcomes	28

*If Q3 = 1 to 8, ® Q11.
If non-contact (Q3 = 9 to 13), ® Q4.
Otherwise, ® Q5.*

Q4. Did you leave a card or message at the address / on the phone / on the answerphone?

- Card/message left at address 1
- Message left on the phone 2
- Message left on the answerphone 3
- No card/message left 4

*If non-contact with anyone at address (Q3 = 9 to 11), ® Q10.
Otherwise, ® Q5.*

Q5. How did you first contact the household at this call?

Code ALL that apply

- Spoke through entryphone/ intercom 1
- Spoke through closed door/ window letter box/ door on a chain 2
- Spoke through open door/ window 3
- Spoke to informant who was outside the household unit 4
- Went /invited inside the household unit and spoke to informant 5
- Spoke over the telephone 6

Q6 to Q8 refer to the MAIN PERSON you made contact with.

Q6. Was the main person you talked to:

- a man/boy? 1
- a woman/girl? 2
- Don't know, not sure 3

Q7. What is the approximate age of the main person?

- Less than 16 1
- 16-34 2
- 35-59 3
- 60 and over 4
- Don't know 5

Q8. Did the main person make any of the following comments during your introductory conversation?

Code ALL that apply

- Main person did not comment** 99
- Positive/neutral comments:**
- Received / remembered advance letter 1
- Expecting someone to call 2
- Make an appointment and come back 3
- I'll think about it 4
- Survey topic is important / or other positive comments about the survey topic 5
- Enjoy doing surveys 6
- Other positive /neutral comments 7

Negative comments:

- Not interested / can't be bothered 8
- I'm too busy / Bad time / Just going out / About to go away 9
- I don't know anything 10
- We are not typical 11
- Not capable / too sick/ old/ infirm 12
- Waste of time 13
- Waste of money 14
- Government knows everything 15
- Don't trust study is confidential 16
- Invasion of privacy / too many personal questions 17
- Don't trust surveys 18
- Never do surveys / I hate forms 19
- Already participated in surveys 20
- Negative comments about the survey topic 21
- Other negative comments 22

Q9. Did the main person ask any of the following questions during your introductory conversation?

Code ALL that apply

- Main person did not ask questions 99
- What is the purpose of the survey / What's it all about? 1
- Who is paying for this/who is the sponsor? 2
- What will happen to the information How will the results be used? 3
- Why/how was I chosen? 4
- How long will the interview take? 5
- Who's going to see my answers? 6
- Can I be identified? 7
- Is it confidential? 8
- Is this compulsory? 9
- Can I get more information? 10
- Can I get a copy of the results? 11
- What's in it for me? 12
- Do I get an incentive? How much is the incentive? When / how will I get paid? 13
- Other questions 14

Q10. Diary surveys only. Non-diary surveys ® Q11.

Was this call a:

- call to make an appointment? 1
- placement call? 2
- checking call? 3
- reminder call? 4
- collection call? 5

Q11. When did you fill in the details about this call?

- Immediately after the call 1
- On the same day 2
- On a different day from the call 3

Go to Accommodation Section (p.22)

Accommodation

Please complete for all addresses including ineligible.

A1. What type of accommodation is it?

House or bungalow

- Detached 1
Semi-detached 2
Terrace/end of terrace 3

Flat or maisonette

- In a purpose built block 4
Part of a converted house / some other kind of building 5

- Room or rooms 6

- Caravan, mobile home or houseboat 7

- Some other kind of accommodation 8

- Don't know / not applicable / unable to code 9

If flat, maisonette or rooms (A1 = 4 to 6), ® A2. Otherwise, ® A4.

A2. Is the household's accommodation self-contained?

- Yes, all rooms are behind a door that only this household can use 1
No 2
Don't know 3

A3. What is the floor level of this household's accommodation?

- Basement or semi-basement 1
Ground floor (street level) 2
1st floor (floor above street level) 3
Second floor 4
Third floor 5
Fourth floor 6
Fifth to ninth floor 7
Tenth floor or higher 8
Don't know 9

A4. Are there any physical barriers to entry to the house/flat/building?

Code ALL that apply

- Locked common entrance 1
Locked gates 2
Security staff or other gatekeeper 3
Entry phone access 4
None 5
Don't know 6

A5. Which of the following are visible at the sampled address?

Code ALL that apply

- Burglar alarm 1
Security gate over front door 2
Bars/grills on any windows 3
Other security device(s) e.g. CCTV 4
Security staff or security lodge on estate or block 5
None of these 6
Don't know 7

Neighbourhood

The term "area" in the following questions refers to the area that you can see from the address.

N1.* Is the sampled house/flat/building in a better or worse condition than the others in the area?

- Better 1
Worse 2
About the same 3
Unable to code 4

N2.* Are the houses/blocks in this area in a good or bad state of repair?

- Mainly very good 1
Mainly good 2
Mainly fair 3
Mainly bad 4
Mainly very bad 5
Unable to code 6

N3.* How many boarded-up or uninhabitable buildings are there in this area?

- None 1
One or two 2
A few 3
Several or many 4
Unable to code 5

N4.* Are most of the buildings in the area residential or commercial / non-residential?

- All residential 1
Mainly residential with some commercial or non-residential 2
Mainly commercial or other non-residential 3
Unable to code 4

N5.* Is the house/flat part of a council or Housing Association housing estate?

- Yes, part of a large council estate 1
Yes, part of a council block 2
No 3
Unable to code / not applicable 4

N6.* How safe would you feel walking alone in this area after dark?

- Very safe 1
Fairly safe 2
A bit unsafe 3
Very unsafe 4

N7. Is this your final call to this household?

Yes 1 → **Go to Household Section (p.24)**

No 2 → **You are now finished until the next call**

Household information

H1. Interviewer code final outcome:

Full or partial interview / co-operation	1	} ® H2
Refusal	2	
No interview due to language, age, infirmity etc	3	
Non-contact	4	} ® H8
Ineligible	5	
HQ refusal	6	

H2. Please enter household details. Use "DK" for don't know and "Ref" for refused.

No. of adults (aged 16 or over) No. of children (less than 16)

	Sex	Age band
	M - Male	1 16 - 34
	F - Female	2 35 - 59
		3 60+
	DK - Don't know	DK - Don't know
	Ref - Refused	Ref - Refused
Adult 1		
Adult 2		
Adult 3		
Adult 4		
Adult 5		
Adult 6		
Adult 7		

H3. Are there any children aged 5 or under?

- Yes, definitely 1
- Possibly 2
- No 3
- Don't know 4
- Refused 5

H4. Is any adult in paid work?

- Yes 1
- No 2
- Don't know /Refused 3

H5. How long has the household lived at this address?

Months ®

OR
Years ®

Don't know/ refused 99

H6. Do you know or think the occupants are:

Code from observation Code ALL that apply

- White 1
- Mixed 2
- Asian (Indian, Pakistani, Bangladeshi, other) 3
- Black (Caribbean, African, other) 4
- Chinese and other ethnic group 5
- Don't know 8

National Travel Survey ® H7.
Other surveys ® H8.

H7. Does the household at present own or have available for use any cars or vans?

Include any company cars or vans if available for private use.

- Yes 1
- No 2
- Don't know /refused 3

H8. Please check that you have completed the Accommodation & Neighbourhood Section (p.22-23) and enter the number of calls made to this address →

You have finished at this address.
Thank you for completing the questionnaire.
Please key the information into your laptop at home and return the questionnaires to the office when you have completed this quota.