



The Magenta Book

Guidance Notes for Policy Evaluation and Analysis

Chapter 6: How are the data collected?

Published: March 2004

Government Chief Social Researcher's Office
Prime Minister's Strategy Unit
Cabinet Office
Admiralty Arch
The Mall
London SW1A 2WH

Contact: phil.davies@cabinet-office.x.gsi.gov.uk
Telephone: 020 7272 1862



Methods of data collection

The selection of a particular method of data collection will be influenced by the nature of the research question(s). Specifically the research question or questions will determine the most appropriate methodology to be employed.

When deciding on how the data should be collected (the methodology) it is useful to consider the types of questions the research is attempting to address. The following illustrates the range of questions one might wish to ask about a particular issue, such as teenage pregnancy or homelessness.

- 1) How many people are in this situation, are affected by this problem, have been helped by this initiative?
- 2) How prevalent is this problem?
- 3) Which groups are most affected by these issues/ are most at risk?
- 4) How much of a difference does the initiative/programme make to the prevalence of these problems?
- 5) Why are people affected by this problem?
- 6) How do people end up in this situation?
- 7) How does the initiative / programme work?
- 8) Why does it work, not work?
- 9) What do people think about the intervention?/How could it be improved?

If we were principally concerned with knowing the answers to questions 1 to 4 then a quantitative methodology would be more appropriate. If, however, we are principally concerned with knowing the answers to questions 5 to 9 then a qualitative methodology may be more suitable.

Strengths and limitations of qualitative and quantitative methods

The strengths and limitations of qualitative and quantitative are outlined in the following table:

	Qualitative	Quantitative
<i>Strengths</i>	<ul style="list-style-type: none">• Flexible• Enables exploration of the meaning of concepts, events• Produces valid data as issues explored in sufficient depth to provide clear understanding• Enables study of motivations	<ul style="list-style-type: none">• Produces statistical data• Where random probability samples are used, survey estimates can be defined within specified bounds of precision• Can measure the extent, prevalence, size and strength of observed characteristics, differences, relationships and associations• Can determine the importance of

	and patterns of association between factors	factors in influencing outcomes
<i>Limitations</i>	<ul style="list-style-type: none"> • Provides a detailed understanding of how individuals interact with their environment, cope with change etc. 	<ul style="list-style-type: none"> • Uses standardised procedures and questioning, enabling reproducibility of results
	<ul style="list-style-type: none"> • Sample sizes are often small 	<ul style="list-style-type: none"> • Can be costly, particularly if population rare or 'hard to reach'
	<ul style="list-style-type: none"> • Need to be able to anticipate factors associated with issues to be studied, to design 'good' sampling strategy 	<ul style="list-style-type: none"> • Sampling frame may not be available
	<ul style="list-style-type: none"> • Interviewing methods rely on respondents being reasonably articulate 	<ul style="list-style-type: none"> • Structured interview hinders detailed exploration of reasons underpinning decisions or views
	<ul style="list-style-type: none"> • Analysis of data to generate findings is not always transparent or replicable 	<ul style="list-style-type: none"> • Standardised questionnaire design and administration means there is little flexibility to be able to deal with respondents' misunderstanding the question (or its intention), leading to problems of validity
<ul style="list-style-type: none"> • Generalisability of findings can be an issue 	<ul style="list-style-type: none"> • Requires key concepts to be clearly defined and translated into meaningful survey questions. 'Fuzzy' concepts are difficult to measure 	

Combining qualitative and quantitative methods

Qualitative and quantitative methods can be combined and this is a useful strategy for both measuring the topic of interest and providing a detailed understanding of its nature or origins.

The following table illustrates how the different types of evidence obtained from quantitative and qualitative methods would contribute to a study about bullying among school children and the effectiveness of an intervention.

Contribution of qualitative and quantitative evidence to answering research questions: bullying among school children[#]

Qualitative methods investigate/ understand	Quantitative methods measure
<ul style="list-style-type: none"> The nature of different forms of bullying The experience of being bullied and being a bully The events leading to bullying/ the circumstances in which it occurs Why bullying continues Appraisal of any interventions experienced Influential factors in bringing periods of being bullied/ being a bully to an end Suggestions / strategies for supporting those bullied/ bullies 	<ul style="list-style-type: none"> The extent to which different forms of bullying exist The characteristics of those bullied and of bullies Factors associated, statistically, with being bullied/ being a bully Characteristics/ circumstances that correlate with length of time being bullied/ bullying Extent to which schools have anti-bullying policies Extent to which policies have an impact on levels of bullying in school Prediction of future levels of bullying Prediction of resources required to deal with bullying effectively

[#] Based on Ritchie, 2003.

When deciding whether to combine qualitative and quantitative (survey) methods of data collection, it is important to consider what types of evidence or information are required and at what stage in the research process this evidence or information will be needed.

To get the most out of combining qualitative and quantitative methods requires:

- a clear set of research questions;
- a reasonable timeframe for the research;
- close working relationships between the qualitative and survey researchers (if they are different people); and
- Sufficient funding to allow this close working.

Quantitative (survey) methods of data collection

There is a range of different types of data collection methods that can be employed in collecting quantitative survey data. These are outlined in the following table.

Types of data collection methods

Interviewer-administered methods

- Face-to-face
- Telephone

Self-completion methods

- Postal
- Web/email

In broad terms there are three sets of factors that will influence the decision over which data collection method to employ:

- survey administration and resource issues;
- questionnaire issues; and
- data quality issues

The difference between face-to-face, telephone and postal surveys, in terms of these factors, are outlined in the following table:

Summary of the strengths and weaknesses of different modes of data collection[#]

	Face-to-face	Telephone	Postal
Design parameter			
<i>Cost of data collection</i>	Usually most expensive method	Usually around 50-70% of face-to-face cost for same interview	Relatively cheap (but q'naires need to be kept short and simple)
<i>Amount and type of resources required</i>	Specialised fieldworker skills and field-force management resources needed	Specialised interviewer skills and management resources needed	For samples < 1,000 normal office resources suffice
<i>Timetable considerations</i>	May require several months unless respondents are easily accessible or 'captive'.	Usually the fastest mode of data collection, but depends on respondent availability	With response reminders, may require several months
<i>Operational control</i>	Best for control of field sampling and data collection	Good for interviewer supervision, but respondent tolerance may be limited	Few means of controlling how q'naires are completed

<i>Amount/complexity of data to be collected</i>	Best/mandatory for long and complicated questionnaires	Limitations on length and data collection complexity compared with face-to-face	Weaker for groups with poor literacy or motivation, but can be good for experts
<i>Likely quality of the data</i>	Best for complex topics and issues. Computer assistance improves quality. May incur interviewer effects	Good for simple factual and attitudinal questions. Computer assistance improves quality. Interviewer effects less likely	Worst for missing data, routing errors, misunderstandings
<i>Statistical efficiency</i>	To reduce fieldwork costs less efficient clustered samples needed for national surveys	Does not require clustered samples, but may have sampling problems	Does not require clustered samples
<i>Expected response rate</i>	Usually gets highest rate	Likely to be 10-40% lower than face to face	Usually lowest rate. Can be well below 50% for less literate/motivated

Based on Lynn & Thomas, 2003.

Combining methods

There are times when it may be appropriate to combine different methods of data collection, for example to:

- save money;
- improve geographical coverage;
- overcome sample frame bias;
- improve data quality, such as response rates or item non-response;
- speed up data collection; or
- Overcome resource problems, such as a lack of face-to-face interviewers.

Despite these advantages there are potential pitfalls to combining data collection methods.

- Development time may need to be extended, as two or more data collection instruments will need to be designed.
- Survey management costs will be increased, as different groups of people will receive different treatments (although these may be offset against savings in the overall cost of the survey).
- Keeping track of the outcome (interview, refusal, non-contact etc) for each case will be required at each stage of the data collection process, particularly if the design involves a follow up of non-responders using an alternative data collection method.

- Particular care will be required to avoid data being lost or duplicated, as a result of having to stitch together data collected from different sources.
- Mode effects can impact on the reliability and validity of the data collected. These effects result from differences in the way in which questions are presented to respondents, for example whether the answer options are presented on a card or are read aloud, and are a particular problem for attitudinal questions. Mode effects can also be observed when asking about sensitive behaviour, when social desirability factors can influence how respondents decide to answer the question.

Computer Assisted Survey Information Collection (CASIC)

It is now commonplace for large-scale face-to-face interviews to be conducted using Computer Assisted Personal Interviewing (CAPI), whereby the questionnaire is a computer program loaded on to a lap-top computer that an interviewer takes out into the field. Respondents' answers are entered into the laptop and interviews transmitted back to the office via modem.

Paper and Pencil Interviewing (PAPI) methods, in contrast, require questionnaires to be posted back to the office, where the information has to be converted into an electronic format, either by being keyed or scanned, which takes longer.

Computer Assisted Telephone Interviewing (CATI) is widely used. Here the questionnaire is accessed via a computer terminal located in a centralised telephone unit.

Computer Assisted Self-Interviewing (CASI) enables respondents to complete a self-completion questionnaire using a laptop computer. Audio-CASI enables respondents to hear the questions rather than relying on them being read.

CASIC methods offer many advantages for surveys.

They:

- Automatically direct the interviewer or respondent to the appropriate questions based on answers given earlier in the questionnaire. It therefore prevents interviewers (or respondents, if CASI) from making routing errors.
- Allow complex routing that would be impractical on paper questionnaires and potentially error-prone (e.g. missing data, answers to inapplicable questions).
- Interviewer can concentrate on the actual questions and respondents' answers, if CAPI or CATI, or if CASI the respondent can concentrate on answering the questions.
- In CAPI and CATI warnings can be triggered if improbably extreme values are entered, or if there is an inconsistency between answers at different questions.
- Substantial timesavings after the completion of the fieldwork (although more time is needed in the beginning to set up CASIC questionnaires compared to paper ones).

- Data are entered directly into a computer during the interview, so data entry as a separate task is eliminated,

Web and email-based data collection

Web and email-based methods of data collection are not widely used in social research at present. This is mainly because of difficulties over sampling; specifically about being able to select random probability samples for general population surveys.

The advantages of CASIC methods can be realised, such as a reduction in routing errors and speed of data transfer. However, to achieve this a significant programming effort is required, using a Web-based language such as Java, which in turn is costly and time-consuming.

Survey instruments

There are a number of different types of survey instrument that can be used to collect the information required to answer the research questions. These include:

- structured questionnaires;
- diaries;
- measurements (e.g. height, weight);
- tests (e.g. reading, memory); and
- Observations (e.g. quality of house conditions).

Questionnaires collect information by means of pre-scripted questions. The questionnaire can be either administered by an interviewer or completed by the respondent. In the case of the former, the question order is predetermined. Questionnaires can collect factual, behavioural and attitudinal information as well as measuring respondents' knowledge, although the latter can only be reliably collected if an interviewer administers the questionnaire or the respondent completes it in a controlled environment. The mode of data collection can influence the reliability and accuracy of the information obtained. For example, the accuracy of information on 'sensitive' behaviours, such as drug taking, may be influenced by whether the data are collected by an interviewer or using a self-completion method.

Diaries allow information to be collected prospectively, that is at the time of the event. They are a form of self-completion questionnaire, with respondents being asked to record details of the behaviour of interest over a specified time period. In this way it is hoped that details of respondents' usual behaviour are captured. Diaries can capture much more detailed information about behaviour than is often possible in a questionnaire, and can be used alongside structured questionnaires.

Measurements can be taken to collect factual information such as respondents' height, weight, blood pressure, blood iron levels and so on. As with diaries, these measurements can be collected in conjunction with information obtained from a questionnaire (and diary). Protocols need to be developed to ensure these are taken in a standardised way. Ethical approval may be required.

Tests can be administered, as part of the survey interview process, to measure respondents' ability to perform certain tasks, such as reading or walking. Such tests are often standard assessment tools that have been developed for a particular setting, such as a clinical or educational assessment in a hospital or school. As with the collection of measurements, protocols will need to be developed that ensure the tests are administered in a consistent way and that they can be administered (reliably) in a survey interview situation.

Observations can be made of factual information, such as the condition of the respondent's accommodation. Observers need to be carefully trained to record information in a consistent way. Observational data can be collected alongside other types of information to provide a more detailed picture of respondents' circumstances.

The choice of data collection instrument will be influenced by the nature of the research questions, the type of information required, the level of detail required, the level of accuracy of data required, the characteristics of the target population, time and money.

Survey questions

Survey questions can be asked that seek different types of information:

- factual;
- behavioural;
- attitudinal; and
- knowledge

Factual questions

Whilst some factual information can be obtained through observation or by reference to official records, they are often collected by asking people questions to obtain it. This is because surveys often offer the only practical and affordable way of collecting such information, and in some cases there is no other source or other way of measuring the attribute of interest.

Behavioural questions, as the name implies, are concerned with measuring respondents behaviour, and can be seen as being a particular type of factual question. We often want to know information about behaviour because we want to understand what people do and or what impact government policy has on them. The following are typical behaviour-type questions:

- What do people do?

- How much do they do it?
- How often do they do it?
- When do they do it?
- Where do they do it?
- Who do they do it with?
- Why do they do it?

Attitudinal questions seek to measure respondents' opinions, beliefs, values and feelings. These are subjective attributes, which cannot be verified by reference to observation or external data sources, and as such they can be difficult to measure and validate.

The following indicates the stages involved in developing an attitude scale.

- 1) Determine clearly what it is you want to measure
- 2) Generate an item pool
- 3) Determine the format of measurement
- 4) Construct the scale
- 5) Evaluate the scale

Knowledge questions are used to assess what respondents know about a particular topic. An example of their use is in welfare benefits research, where they can be used to assess people's awareness of particular benefits and tax credits and the qualifying rules for them. Answers to knowledge questions can be affected by the wording of other survey questions. For example, if we want to measure what respondents' know about the qualifying rules for Working Families Tax Credit, we should ask these questions before questions that give an indication of what the rules are.

Open and closed questions

Closed questions constrain answers to a set of pre-scripted answer alternatives. *Open* questions have no such restrictions.

The pros and cons of open and closed questions are detailed in the following table:

Pros and cons of open and closed questions

Question type	Pros	Cons
<i>Open</i>	<ul style="list-style-type: none"> • Respondents can answer in their own words • Are not leading so can, 	<ul style="list-style-type: none"> • Responses have to be coded to allow statistical analysis, which is costly, time-consuming and subject to error • If an interviewer-administered

potentially allow measurement of salience – how important issue is to respondent; indicate respondent’s level of knowledge; indicate strength of feeling

survey, can be difficult to get interviewers to probe consistently

- Can avoid format effects, such as primacy (the tendency to endorse the first option seen) or recency (the tendency to endorse the last option heard)
- Are required for the development of response options for closed questions
- Respondents may not provide sufficient detail when answering to capture key differences
- Can generate irrelevant responses

Closed

- Can help with respondent-recall
- Level of detail, areas of interest can be conveyed to respondent
- Little or no coding of answers required, thus quicker and cheaper
- Less likelihood of coder bias or inconsistent probing
- If not a complete list of answer-options, can introduce bias
- Cannot capture detailed information, for example occupation details
- Categories may not be recognisable to respondents
- Answer options can influence way in which respondents’ interpret question

Designing survey instruments

There are three golden rules that are useful to consider when writing survey questions.

- Can the respondent understand the question?
- Is the respondent able to answer the question?
- Is the respondent willing to answer the question?

The key rule to remember in designing questions that can be understood is to keep them simple. That way the complexities and ambiguities described above can be avoided.

Respondents’ ability to answer the question

It should not be assumed *a priori* that respondents will have the information necessary to be able to answer the survey questions being posed. Rather we need to consider whether respondents have been exposed to the event(s) or experiences we are asking them about, and if they have, whether they will be able to remember them or not.

Even if the respondent experienced the event of interest, she or he may not be able to answer the question because:

- the information required never got stored in her or his (long-term) memory (Willis et al, 1999) ;
- the retrieval (survey) context is different to the original encoding context, so the respondent may not recognise that the event took place or be able to recall the event correctly (Tulving and Thompson, 1973);
- the item may be difficult to distinguish from other, similar, events (Anderson, 1983); or
- The memory and or the cues associated it with it, have faded over time (Sudman and Bradburn, 1974).

The following table summarises the key factors affecting recall:

Figure 6.6 Summary of factors affecting recall

Variable	Finding	Implication for survey design
<u>Characteristic of event</u>		
Time of event	Events that happened long ago harder to recall	Shorten reference period
Proximity to temporal boundaries	Events near significant temporal boundaries easier to recall	Use personal landmarks, life events calendars to promote recall
Distinctiveness	Distinctive events easier to recall	Tailor length of the reference period to properties of target event; use multiple cues to single out individual events
Importance, emotional impact	Important, emotionally involving events easier to recall	Tailor length of the reference period to properties of target event
<u>Question characteristics</u>		
Recall order	Backwards search may promote fuller recall	Not clear whether backward recall better in surveys
Number and type of cues	Multiple cues typically better than single cues; cues about the type of event (what) better than cues about participants or location (who or where), which are better than cues about time (when)	Provide multiple cues; use decomposition
Time on task	Taking more time improves recall	Use longer introductions to questions; slow pace of the interview

(Tourangeau et al, 2000: 98)

Respondents’ willingness to answer the question

Even if respondents understand the question and are in possession of the necessary information to be able to answer it they still might not provide an answer because they are unwilling to do so.

Respondents need to be motivated to engage in the necessary cognitive (thinking) effort to answer the question.

Studies have identified a number of factors that can impact on respondents’ willingness to answer survey questions. These include whether they perceive the request for information to be:

- legitimate;

- reasonable;
- beneficial; and
- Not to have any negative consequences (Kahn and Cannell, 1957; Cannell et al, 1979).

Salience and length

Research has shown that the more salient respondents find the questions the more likely they are to answer them.

An excessively lengthy questionnaire can impact on survey data quality in a number of ways, including:

- lowering survey response rates;
- increasing item non-response for questions later in the questionnaire;
- Increasing respondent errors.

Question wording and order effects

In designing questionnaires particular attention is required to the way in which questions are worded and to the order in which they are asked, as these things can have an impact on respondents' answers.

Question wording effects

Changes in question wording, even what appear to be small ones, can have an impact on the way in which respondents answer them.

It is often difficult to predict whether changes in question wording have any effect on response. They can be detected by conducting a split-panel experiment.

In a split-panel experiment one version of the question is sent to half the respondents (group a) in the sample and the other version to the other half (group b). The allocation of respondents to group a) or b) would be random. Apart from the wording of the question all other aspects of administration of the survey question would be the same. This approach allows us to compare the results from the two question variants, and if we observe a difference in the distribution of answers between them, be certain that this difference is the result of the change in question wording alone.

Question order

Question order can affect the way in which survey respondents interpret survey questions and thus answer them. This is because the wording of preceding questions can help to shape the context in which respondents interpret the current question. There are two main types of context effects:

- assimilation; and
- Contrast.

Assimilation effects (sometimes known as consistency effects) occur when respondents infer that the current (target) question is directly related to the preceding (context) questions. For example, in a study conducted by Schuman and Presser (1981) the order of the following two questions was varied.

- a) Do you think the United States should let Communist reporters from other countries come in here and send back to their papers the news as they see it?
- b) Do you think a Communist country like Russia should let American newspaper reporters come in and send back to their papers the news as they see it?

Support for statement a) varied by 20 per cent according to whether it was asked before or after question b). The explanation for this variation, put forward Schuman and his colleagues is that when question a) is asked first, many answers reflect attitudes towards communism or Russia. When a) is asked after b), answers are based on the notion of even-handedness: American reporters should be allowed in to Russia and be able to report on the news there as they see and thus the same principle should be applied to Russian reports.

Contrast effects occur when respondents infer that the target question should be compared with the context question(s). For example, answers to general questions can be influenced by whether they are asked before or after questions about specifics.

Things to avoid

The following provides a summary on the “do’s and don’ts” of questionnaire design:

1. Do not ask people for information they do not have
2. Do not ask more than one question at a time: avoid double-barrels (i.e. including two different concepts in one question)
3. Avoid double-or implicit negatives

This is particularly an issue for attitude statements involving the use of agree/ disagree response options. Consider the following example:

I am much less confident now than I used to be?

If the respondent feels more confident now than she used to then she has to disagree with the statement, which can cause problems because she has to engage in a double-negative, that is she does not agree that she feels less confident now.

4. Long lists of response choices

These can suffer from primacy or recency effects. If long lists are presented to respondents in a visual format there is a tendency for people to only look at the first few items in the list rather than read all items. If the list is read out, then respondents are more likely to only remember the last few items.

5. Beware of questions that include hidden contingencies, as these may only be relevant to a subsection of the population. For example the question *“How often do you drive your car to work?”* clearly only applies to people who a) can drive, b) have access to a car that they can drive, c) who work and d) who do not work at home. Thus the question should not be asked of all respondents, but only of those to whom it is relevant. Asking questions that are not relevant to respondents is likely to annoy them, making them less likely to want to continue to answer questions and makes the answers to such questions difficult to interpret.

6. Questions that start with response choices such as *“Would you say that you often, sometimes, rarely or never [buy a newspaper]?”*. Response options should always come after the question.

7. Vaguely worded questions as these encourage or permit vaguely worded answers. For example, the question *“How happy are you with the way things are at the moment?”* is vague. What does “the way things are” mean? Respondents could interpret this phrase in any number of ways and give answers such as “OK” or “They could be better”, which are equally vague and uninformative.

8. Ambiguous words and questions. For example, the question *“Do you have a car?”* is ambiguous. Is it asking about whether I own a car or have access to one? The question asks about ‘you’ but who is ‘you’, the respondent, the household, the family or the company?

8. Using jargon, technical terms, acronyms or abbreviations as these may not be familiar to respondents or may be interpreted in different ways by different people. For example, terms such as “social exclusion” or “hypertension” may be commonly used by researchers or health practitioners but may be meaningless or poorly understood by members of the public.

9. Using colloquialisms or words with alternative usage, as different respondents may interpret these in different ways. For example, the term ‘dinner’ has different meanings: to some it denotes a cooked meal, to others an evening meal. Vegetables can be known by different names in different parts of the country, for example a turnip in parts of Scotland and the north of England is a large root vegetable with orange flesh, whereas in other parts of the country it is a small root vegetable with white flesh.

10. Beware of leading questions, for example *“When did you last telephone your mother?”* Asking this question without any prior knowledge about the respondent’s mother assumes that a) the respondent has a mother who is alive, b) who has a telephone, and c) that the respondent is in contact with her. Similarly the following question asked of those who indicated that they had once been told they had

cancer ‘Which type of cancer do you have?’ – assumes that the respondent currently has it, which might not be the case.

11. Not including don’t know and not applicable codes. It is often useful to know that someone does not know the answer to a question, as this can be an important piece of information. It is also important to differentiate between reasons for no answer to a question, such as between those who don’t know, refuse to answer or cannot answer because the question is not applicable to their circumstances.

12. Avoid proverbs or using stereotypes, especially when measuring attitudes, as such sayings can lead to unconscious agreement. It is better to get the respondent to think afresh about the issue.

13. Using loaded terms, such as free, democratic, natural, modern and so on.

Sources of measurement error in surveys

In a questionnaire a question may not measure the factor it was designed to detect. This is referred to as *measurement error*.

Traditionally measurement errors have been classified into two broad categories: those connected with survey questions and those connected with survey interviewers.

More recently there has been a shift in emphasis, from viewing errors as being the product of either the questionnaire or the interviewer, to being related to the nature of the tasks the actors in a survey interview have to perform (Oksenberg et al, 1991). This task-focused classification is useful in helping us to understand the potential sources of measurement error as it focuses on the specific components of the question-and-answer process. The task-focused model on the other hand, would help to identify the cause of the problem by enabling the researcher to identify whether the problem is one of comprehension, processing or communication. These problems are summarised in the following table:

Components of measurement error

Traditional Model	Task-focused model
<ul style="list-style-type: none">• Problems with survey questions that<ul style="list-style-type: none">- are misunderstood- cannot be answered, either at all or accurately- respondents will not answer	<ul style="list-style-type: none">• Comprehension problems resulting from:<ul style="list-style-type: none">- use of vocabulary- complex sentence structure- not understanding the nature of the task and the rules about how to respond

- Problems with survey interviewers
 - do not read the questions as worded
 - probe directly
 - bias answers as a result of the way interviewers relate to respondents (for example, differences in ethnicity, age, social class, gender)
 - record answers inaccurately
 - Validity problems resulting from:
 - respondents interpreting the same question in different ways, or
 - in the same way but not in the way the researcher intended
 - Processing difficulties
 - respondents may be unwilling or unable to retrieve the information necessary to answer the question
 - Pronunciation or communication difficulties
 - these may affect both interviewers and respondents
-

Interviewer error

Interviewers can be a source of error in surveys. There are different components of interviewer error:

- interviewer characteristics, such as gender, age or ethnicity (e.g. lower rates of anti-Semitism were reported by respondents who had been interviewed by someone who appeared to be Jewish than those who did not appear Jewish (Robinson and Rhode, 1946 cited in Fowler and Mangione, 1990);
- interviewer expectations and attitudes; and
- interviewer behaviour such as
 - not reading the question as worded
 - directive probing
 - relating to the respondent in a way that affects his /her behaviour
 - Inaccurate recording of answers.

The following actions, put forward by Fowler and Mangione, 1990, can be taken by the researcher to mitigate against interviewer-error.

- a) Questions must be carefully designed so that:
 - as written they fully prepare respondents to provide answers;
 - they mean the same thing to every respondent; and
 - The kinds of answers that constitute an appropriate response to the question are communicated to all respondents.

- b) Interviewers are given specific instructions or guidelines about how to conduct the interview. As a minimum, these guidelines should include:
- reading the question exactly as worded;
 - probing inadequate answers non-directively;
 - recording answers without interviewer discretion;
 - Maintaining neutral, non-judgemental relations with the respondent.
- c) Interviewers need to be trained and this training needs to be on going.
- d) Interviewers need to be supervised to ensure that they follow the guidance in b) above.

Evaluating survey questions and instruments

Due to the various sources of error that can occur when developing questionnaires it is important to evaluate the survey questions and instruments. The following table summarises the different methods available for pre-testing questionnaires, including cognitive interviewing methods.

Methods of reviewing and testing questionnaires¹

Method	Description
<i>Focus groups / depth interviews / other flexible qualitative methods</i>	Qualitative techniques widely used to explore the concepts, viewpoints and vocabulary used by the population which is to be sampled for a quantitative survey on a given topic. Flexible, exploratory approach not bound by a fixed questionnaire that incorporates preconceptions that may be false. Not fully replicable. Depends on judgement and intuition of qualitative researchers. Does not directly lead to questions suitable for a quantitative survey instrument.
<i>Field rehearsal piloting</i>	Rehearsal of the field data collection process as a whole. Researchers often mainly preoccupied with response, length of questionnaire and other operational issues not directly bearing on questions as measures. Coverage of question performance often sketchy. Personal or telephone debriefing of interviewers is sometimes conducted by the research team after rehearsal pilot. Interviewers may sometimes be allowed to try limited rewordings of questions that appear not to work well and to report back on the effects of rewording. Debriefing may be backed by interviewers' notes and/or tape recordings of some interviews. Can capture the observations and intuitions of experienced interviewers; but typically each interviewer sees few cases. Debriefing may be dominated by reports on a few examples, which prove to be atypical.

¹ For a more detailed review of techniques, see Esposito and Rothgeb (1997).

Method	Description
<i>Pilot respondent debriefing</i>	<p>Personal debriefing of respondents after they have responded to a trial version of a questionnaire, in the field or as a 'hall test'. May be conducted by interviewers or researchers.</p> <p>Applied shortly after the event can capture some of respondent's impressions and thought processes in responding to questions/ completing a questionnaire. Relies on respondent to identify 'problems'.</p> <p>Tends to be at level of whole questionnaire because of lack of time to cover all questions in detail. May be difficult for respondent to understand purpose of re-interrogation.</p>
<i>Dynamic piloting</i>	<p>An intensive, informal, small scale, iterative process. Initial version of a questionnaire is tried out by a small interactive research team on a small sample of respondents from the population. Wording or other problems are identified, changes are rapidly made and a revised version is again tested.</p> <p>Very time-effective. Question designers interact directly with respondents. Leads directly to a final questionnaire. Results and decisions, other than the final questionnaire, often not recorded. Based on small convenience samples. Method not fully replicable.</p>
<i>Split panel comparisons</i>	<p>Experiment comparing 2 or more set versions of quantitative question(s) or questionnaire. Allows statistical comparison of results. Experiment needs to be on a large scale for results to be useful. Sponsors may reject because only a fraction of the sample will be asked the 'best' question. Method may reveal differences in response distributions, but not reasons or which question is 'better'.</p>
<i>Interview re-interview</i>	<p>Some time after initial interview using test questionnaire respondents are re-interviewed. In the 'with feedback' version the re-interviewer has access to original response and can probe discrepancies. Good in principle for assessing reliability of response but tends to be contaminated by recall of first interview. Respondents may misunderstand and think they have given 'wrong' answers at first interview.</p>
<i>Cognitive interviewing</i>	<p>Term includes a number of different techniques, including 'think aloud', card sorting etc. Based on theory of the question comprehension and answering processes. Can be used to explore and delineate respondent's conceptual space and/or to study understanding of the question, recall processes, formulation of response, internal review and articulation of response, etc.</p>

Method	Description
<i>Behaviour coding</i>	Requires sound and preferably also video recording of test interviews. Utterances and exchanges between interviewer and respondent are coded according to a strict protocol through observation of the interview and/or review of the recordings. Questions can be scored according to the number of ‘symptoms of difficulty’ (e.g. request for repeat or clarification of question, pauses, expressions of uncertainty etc). Very time-consuming and laborious to carry out.
<i>Expert review</i>	Requires a small panel of ‘experts’ (usually researchers with appropriate experience) to critically review draft questionnaires and make comments and suggestions.

[Extract from: Thomas, R., Collins D., Rowlands, O. (2000) *Question Testing and Survey Quality*. WAPOR Seminar Proceedings on Quality Criteria in Survey Research III.]

Types of survey

There are two main types of survey, those that are concerned with providing information about a *cross-section* of the population of interest at a particular point in time and those that are concerned with providing *longitudinal* information about individual sample members over time.

There are different types of cross-sectional surveys and these are described below.

Continuous surveys take place ‘continuously’. Fieldwork takes place in each month of the year, with the sample in any one month being broadly representative of the target population. Such surveys are designed to measure net annual change at the aggregate rather than individual level. Thus there is nothing in the design of such surveys that requires an overlap in sample units at different points in time. Examples of continuous surveys are the National Travel Survey, Family Resources Survey and Health Survey for England.

Repeat surveys take place at scheduled regular points in time, such as every year or every two years. Fieldwork is concentrated into a few months. Such surveys enable net change at the aggregate level to be measured, as estimates from one survey can be compared with another in the series. However, unlike continuous surveys, for repeat surveys is not possible to determine whether the observed change took place gradually or not. As with continuous surveys, there is nothing in the design of repeat surveys that requires an overlap of sample units at different points in time. Examples of repeat surveys are the

Repeat Study of Parents' Demand for Childcare, the National Diet and Nutrition surveys, the National Adult Learning Surveys and the British Social Attitudes surveys.

Ad hoc surveys are one-off studies: there is no plan to repeat them at regular intervals. Although they may use questions used on other surveys it is important to bear in mind that if findings from an ad hoc survey are compared with another survey, differences observed may be the result of differences in the methodologies used by the two surveys, rather than indicating real change. Ad hoc surveys include Attitudes Towards and Experiences of Disability in Britain Survey, the National Study of the Lottery and other Gambling Activities and the Temporary Employment Survey.

Rotating panel surveys are scheduled to take place at regular intervals, or continuously, and include rotating panels; that is, people are introduced into the survey, surveyed a number of times, and then rotated out of the survey. There is no attempt to follow respondents or sample units that move or link records for individuals or sample units over time to make longitudinal estimates. Rotating panel survey designs are used where estimates of change are required to be accurate for small time periods, such as three-month periods as panel designs reduce the variance of estimates of change compared with a system of independent samples over a given time period. Furthermore, such designs enable the identification of gross change between groups or states, which may be masked by aggregate net change data because all changes at the micro level cancel each other out. An example of a rotating panel study is the Labour Force Survey.

Longitudinal studies are concerned with measuring change at the individual level.

In **longitudinal studies without rotation**, individuals or sample units are followed over time, to create a longitudinal record. Analysis is at the individual level. Such data, over time, are not suitable for generalising to the wider population. Examples of such longitudinal studies are the birth cohorts, such as the National Child Development Survey, and the English Longitudinal Study of Ageing.

Longitudinal studies can be designed to include **rotation**, which means that they follow a particular group for a specified period and introduce new sample units at specified periods, to create a longitudinal record. Data can be analysed longitudinally but also each data collection period, including new sample units, can be analysed cross-sectionally, as the study sample remains representative of the survey target population. An example of a longitudinal study with this design is the Families and Children Survey (FACS).

Further sources of information

A great deal of information about specific surveys, such as the General Household Survey or the British Household Panel Survey is available over the Internet.

What follows is a summary of some of the key Internet sites that contain useful information on government surveys and other important studies.

GHS reports can be accessed via

<http://www.statistics.gov.uk/statbase/Product.asp?vlnk=5756>

Other National Statistics reports relating to social and welfare topics, such as the FRS, and EFS can be accessed via

<http://www.statistics.gov.uk/onlineproducts/default.asp#social>

By scrolling up and down this page you can gain access to a range of other reports on different topics such as transport, employment and health.

For information on the British Crime Survey visit:

<http://www.homeoffice.gov.uk/rds/bcs1.html>

For information on the Health Survey for England visit:

<http://www.doh.gov.uk/public/summary.htm>

For information on the British Household Panel Survey visit:

<http://www.iser.essex.ac.uk/bhps/index.php>

For information about Birth Cohort Studies visit:

<http://www.cls.ioe.ac.uk/Cohort/mainncds.htm>

For more information about the Families and Children Survey go to:

<http://www.dwp.gov.uk/asd/asd5/facs/>

To find out more about the ESRC data archive go to:

<http://www.data-archive.ac.uk/>

To find out more about the Question Bank go to:

<http://qb.soc.surrey.ac.uk/docs/home.htm>

To find out more about the Scottish Household Survey go to:

<http://www.scotland.gov.uk/about/SR/CRU-SocInc/00016002/SHShome.aspx>

Further reading

Anderson J. (1983) *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.

Bailar B.B. (1989) "Information Needs, Surveys and Measurement Errors" in Kasprzyk D., Duncan G., Kalton G., Singh M.P. [eds] *Panel Surveys*. New York: John Wiley & Sons; 1-24.

Bechhofer F., Paterson L. (2000) *Principles of Research Design in the Social Sciences*. London: Routledge.

Brannen J. [Ed] (1992) *Mixing Methods: Qualitative and Quantitative Research*. Aldershot: Gower.

Cantril, H., Fried, E. (1944) "The meaning of questions". In Cantril H., *Gauging Public Opinion*. Princeton, NJ: Princeton University Press.

Centre for Longitudinal Studies Research Review 1999.

Clark H.H., Schober M.F. (1992) "Asking questions and influencing answers." in Tanur JM (ed.) *Questions About Questions: Inquiries into the cognitive bases of surveys*. New York: Russell Sage Foundation; 15-48.

Collins D., White A. (1995) "Making the next Census form more respondent-friendly." *Survey Methodology Bulletin*: 37; 8-14.

Converse J., Presser S. (1986) *Survey Questions: Handicrafting the Standardized Questionnaire*. Newbury Park: Sage:63.

Czaja R. (1998) "Questionnaire testing comes of age." *Marketing Bulletin*; 9: 52-66.

Czaja R., Blair J. (1996) *Designing Surveys: a guide to decisions and procedures*. Thousand Oaks, CA: Pine Forge Press.

DeVellis, R.F. (1991) *Scale development: theory and applications*. Thousand Oaks, CA: Sage.

DeLamater J. (1982) "Response-effects of question content" in Dijkstra W., van der Zouwen J. (eds) *Response Behaviour in the Survey-Interview*. London: Academic Press.

de Leeuw E.D., van der Zouwen J. (1988) "Data quality in telephone and face-to-face surveys: a comparative meta-analysis." in Groves R.M., Biemer P.N., Lyberg L.E., Massey J.T., Nichols II W.L., Waksberg J. [eds] *Telephone survey methodology*. New York: John Wiley & Sons; 267-286.

Dillman D (1999) *Mail and Internet Surveys: the tailored design method*. New York: Wiley & Sons Inc.

Dillman D (1978) *Mail and Telephone Surveys: the total design method*. New York: John Wiley & Sons.

Edwards P, Roberts I, Clarke M, DiGiuseppi C, Pratap, S Wentz R, and Kwan I. Increasing response rates to postal questionnaires: systematic review. *British Medical Journal*, May 2002; 324: 1183.

Flately F. (2001) "The Internet as a mode of data collection in government social surveys: issues and investigations." *Survey Methodology Bulletin*. 49: 1-10.

Foddy W. (1993) *Constructing Questions for Interviews and Questionnaires: Theory and practice in social research*. Cambridge: Cambridge University Press.

Forsyth B.H., Lessler J.T. (1991) "Cognitive laboratory methods: A taxonomy." in Biemer P.P., Groves R.M., Lyberg L.E., Mathiowetz N.A., Sudman S (eds.) *Measurement Errors in Surveys*. New York: Wiley: 393-418.

Fowler Jr F.J. (1995) *Improving Survey Questions*. Thousand Oaks, CA: Sage.

Fowler Jr F.J. (2002) *Survey Research Methods: third edition*. Applied Social Research Methods Series, 1. Thousand Oaks, CA: Sage.

Fowler Jr. F.J., Mangione, T.W. (1990) *Standardized Survey Interviewing: minimising interviewer-related error*. Applied Social Research Methods Series: 18. Newbury Park, CA: Sage.

Gribble, J.N., Miller H.G., Rogers S.M., Turner C. (1999) "Interview Mode and Measurement of Sexual Behaviours: Methodological Issues." *The Journal of Sex Research*, 36 (1):16-24.

Groves R.M., Biemer P.N., Lyberg L.E., Massey J.T., Nichols II W.L., Waksberg J. [eds]. *Telephone survey methodology*. New York: John Wiley & Sons.

Groves R.M., Kahn R.L. (1979) *Surveys by Telephone: a national comparison with personal interviews*. Orlando, FL: Academic Press.

Hakim C. (2000) *Research Design: successful designs for social and economic research. Second Edition*. London: Routledge.

Jenkins C., Dillman D. (1997) "Towards a theory of self-administered questionnaire design". in Lyberg L., Biemer P., Collins M., de Leeuw E., Dippo C., Schwarz N., Trewin D (eds) *Survey Measurement and Process Quality*. New York: John Wiley & Sons Inc.: 165-196.

- Knight I (1994) "Changes in the sample design for the Labour Force Survey." *Survey Methodology Bulletin*. 34: 25-27.
- Krosnick J.A. (1991) "Response strategies for coping with the cognitive demands of attitude measures in surveys." *Applied Cognitive Psychology*; 5: 213-236.
- Krosnick J.A., Alwin D.F. (1987) "An evaluation of a cognitive theory of response-order effects in survey measurement." *Public Opinion Quarterly*; 51: 201-219.
- Martin J., Manners T. (1995) Computer assisted personal interviewing in survey research. In: Lee R.M. (ed) *Information Technology for the Social Scientist*. London: UCL Press: 52-71.
- Mathiowetz, N.A., McGonagle K.A. (2000) "An assessment of the current state of dependent interviewing in household surveys". *Journal of Official Statistics*; 16 (4): 401-418.
- Nicholls II, W.L., Baker R.P., Martin, J. (1997) "The effect of new data collection technologies on survey data quality." in Lyberg, L., Biemer P., Collins M., de Leeuw E., Dippo C., Schwarz N., Trewin, T. [Eds] *Survey Measurement and Process Quality*. New York: John Wiley & Sons Inc.: 221-248.
- Oksenberg L., Cannell C, Kalton G. (1991) "New strategies for pretesting survey questions." *Journal of Official Statistics*; 7 (3): 349-365.
- Oppenheim, A.N. (1996) *Questionnaire Design, Interviewing and Attitude Measurement. New Edition*. London: Printer Publishers Ltd.
- Patton, M. Q. (2002). *Qualitative Research and Evaluation Methods. Third Edition*. Thousand Oaks, CA: Sage Publications.
- Payne S.L. (1951) *The Art of Asking Questions*. Princeton: Princeton University Press.
- Ritchie J. (2003) "The application of qualitative research methods to social research" in Ritchie J., Lewis J. [eds] *Qualitative Research Practice: a guide for social science students and researchers*. London: Sage Publications: 24-46.
- Schuman H., Presser S. (1981) *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording and Context*. New York: Academic Press.

Schwarz, N., Hippler, H. (1991) "Response alternatives: the impact of their choice and presentation order" in Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A., Sudman, S. (eds) *Measurement Errors in Surveys*. New York: John Wiley and Sons, Inc. 41-56.

Sudman S., Bradburn N.M, (1982) *Asking Questions: A practical guide to questionnaire design*. San Francisco: Jossey-Bass.

Sudman S., Bradburn N.M., Schwarz N. (1996) *Thinking About Answers: the application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass.

Thomas R., Lynn P (eds) [forthcoming] *Survey Methods and Practice*. 2nd Edition. London: Sage.

Tourangeau R. (1984) "Cognitive sciences and survey methods." in Jabine T, Staf M, Tanur J, Tourangeau R (eds). *Cognitive Aspects of Survey Methodology: Building a bridge between the disciplines*. Washington DC: National Academy Press: 73-100.

Tourangeau R., Rips L., Rasinski K. (2000) *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

Tourangeau R., Smith T.W. (1996). "Asking sensitive questions." *Public Opinion Quarterly*. 60: 275-304.

Tulving E., Thompson D.M. (1973) "Encoding specificity and retrieval processes in episodic memory". *Psychological Review*. 80: 352-373.

Turner, C.F. (1984) "Why Do Surveys Disagree? Some Preliminary Hypotheses and Some Disagreeable Examples", in Turner, C.F. and Martin, E. (eds) *Surveying Subjective Phenomena*, 2. New York: Russell Sage.

Willis G.B. (1994) *Cognitive Interviewing and Questionnaire Development: A training manual*. Hyattsville, MD: National Centre for Health Statistics.

Willis G.B., Brittingham A., Lee L., Tourangeau R., Ching P. (1999) *Response errors in surveys of children's immunizations*. Vital and Health Statistics; 6 (8). Hyattsville, M.D: National Centre for Health Statistics.

Willis GB, deMaio T, Harris-Kojetin B. (1999) "Is the bandwagon headed to the methodological promised land? Evaluating the validity of cognitive interviewing techniques." in Sirken MG, Herrmann DJ, Schechter S, Schwarz N, Tanur JM, Tourangeau R (eds) *Cognition and Survey Research*. New York: Wiley: 133-153.

