

Computer Assisted Survey Information
Collection: Australian Health Surveys:
Question and module development
principles and practice

Fearnley Szuster

Occasional Paper Series No. 5

DRAFT

Public Health Information Development Unit

Copyright

© Commonwealth of Australia 2003

This work may be reproduced and used subject to acknowledgement of the source of any material so reproduced.

National Library of Australia Cataloguing in Publication entry

Szuster, Fearnley Stefan Pym, 1951- .
Computer assisted survey information collection :
Australian health surveys : question and module development
principles and practice.

Bibliography.
ISBN 0 7308 9236 0.

1. Health surveys - Australia - Methodology. I. Public
Health Information Development Unit (Australia).
II. Australia. Dept. of Health and Ageing. III. Title.
(Series : Occasional paper series (Public Health
Information Development Unit (Australia)) ; no. 5).

614.42940285

Public Health Information Development Unit, The University of Adelaide

This research was produced by the Public Health Information Development Unit (PHIDU), The University of Adelaide, South Australia in November 2001. The research was funded under a grant from the Australian Government Department of Health and Ageing. The views expressed in this paper are solely those of the authors and should not be attributed to the Department of Health and Ageing or the Minister for Health and Ageing.

Suggested citation:

Szuster F. (2003) *Computer Assisted Survey Information Collection: Australian Health Surveys: Question and module development principles and practice*. Occasional Paper Series No. 5. Public Health Information Development Unit, Adelaide.

Enquiries about or comments on this publication should be addressed to:

PHIDU, The University of Adelaide, South Australia 5005
Phone: 08-8303 6239 or e-mail: PHIDU@publichealth.gov.au

This and other publications are available from the PHIDU website (www.publichealth.gov.au).

ISSN 1447-8803 Occasional Paper Series

Published online by Public Health Information Development Unit, The University of Adelaide

Contents	Page
Executive Summary	v
Introduction and Overview	1
Data Requirements and Question Selection	9
- Policy Questions	
- Literature Review	
- Key Informants	
- Question Appraisal	
- Data Requirement Consultation	
Pre-Testing	17
- Focus Groups	
- Cognitive Testing	
Question Design	23
Validity and Reliability	27
Field Testing	33
Usability Testing	37
Reporting	39
Interviewers and Supervisors	41
Additional Issues	45
- Ethical and Related Issues	
- Core Questions	
- Survey Coverage	
- Survey Introduction	
- Order Effect	
- Re-interviewing and Monitoring	
References	49
Abbreviations	57
Glossary	59

	Page
Attachment A: Stage One Template Example for the Collation of CASIC Questions	63
Attachment B: Validity and Reliability	65
Figure 1: Stages in question development	3
Figure 2: National Public Health Partnership performance indicator framework	10
Figure 3: Schematic diagram of the CATI TRG question appraisal process	13
Figure 4: Proposed CATI module standard consultation process	14
Figure 5: Cognitive model of the survey response process	17
Table 1: Selected reliability estimates on demographic, health risk factors and co-morbidity questions	30

Executive Summary

Introduction and Overview

The work of the National Computer Assisted Telephone Interview Technical Reference Group (CATI TRG) under the National Public Health Partnership has emphasized the need to establish question and module development standards in the Computer Assisted Survey Information Collection (CASIC) environment that can be utilised in CATI health surveys across Australia. This has been seen as important so that measurement error in the CASIC environment can be reduced as well as providing greater harmonisation of the results of different CASIC data collection activities [18]. This overview of question/module development principles and practices provides a guide to the various stages that are required to ensure that 'best practice' question development is followed. [page 1]

This document outlines principles and practices for the development of consistent and uniform questions (or sets of questions) in CASIC. It covers the identification and assessment of questions, key processes in the development of questions and the processes to test validity and reliability. In addition, subsequent fieldwork issues, interviewer feedback and reporting on questions are outlined. [page 2]

Traditionally the major contributors to the development of survey questions have been questionnaire designers working closely with subject matter experts, with input from experienced interviewers. In recent years there has been a shift to explicitly define at the beginning of the process the policy and data requirements for health monitoring and surveillance, and later in the process to include an understanding of questions from the perspective of the respondent, with the operational aspects of questionnaire development and testing incorporated into survey planning [29]. Accordingly, this document covers the early stages of question development that include data requirements, question selection, pre-testing and question design, through to quantitative evaluation, field testing and interviewer support. [page 2]

In Australia, the proportion of households with fixed (or cordless) telephones has in recent years risen to levels sufficient for the conduct of CATI population surveys. Currently 97.8% of households have fixed telephones [8]. [page 5]

The recent dramatic changes in the telecommunication industry with the high take-up of mobile phones in Australia could be expected to have a major impact on CATI. In 2000, 58.5% of households had at least one mobile phone [8]. Accordingly, in the future there could be greater reliance on Computer Assisted Personal Interview (CAPI) procedures as well as surveys conducted via the world wide web for particular population groups. [page 6]

It is expected that with the development of questions for CASIC (particularly CATI), the preparation and dissemination of reports that document the development process will be publicly available. The acceptance of the questions into the National Health Data Dictionary would also ensure the broad utilisation of standard (comparable) questions across data collections [9]. [page 8]

Data Requirement and Question Selection

A structured approach to the selection of variables for health survey data collection and analysis is required. Particularly, documentation (in tabular form) of the policy questions that are relevant to the topic and a statement of data requirements for health surveillance is recommended, followed by an assessment of the relevant variables or indicators for each policy question and operationalisation of the data requirements [77]. [page 9]

There are a number of resources that can be utilised in the gathering together of existing questions. In particular, there have been a number of State/Territory CATI Health Surveys that have covered a wide range of health topics. The questions used in the 1995 NHS are published in the ABS reference package [6]. [page 11]

There are also four particular international resources that provide extensive detail for health questions in the US, Canada and the UK, namely:

- BRFSS: the Behavioral Risk Factor Surveillance Survey conducted in each State of the US by the Centers for Disease Control and Prevention
- CCHS: the Canadian Community Health Survey conducted as part of the Canadian Roadmap Initiative
- CASS: the Centre for Applied Social Survey in the UK
- Health Poll Search: Roper Center at the University of Connecticut and the Kaiser Family Foundation, this feature at the Kaiser website is a searchable archive of public opinion questions on health issues. [page 11]

One important consideration when identifying existing questions from other sources is that the same question in a different collection mode is a different question [32]. Accordingly, having gained questions, it is important that the question is tested and evaluated in the environment of the CASIC (particularly CATI) health survey and also in the Australian setting. [pages 11 and 12]

Key informants are usually those individuals who have demonstrated extensive knowledge in the topic, such as researchers, academics, policy makers and representatives of special interest groups. Key informants should also incorporate established expert groups and other national work that is underway such as the national health priority areas. [page 12]

Recently the CATI TRG, a subcommittee of the National Public Health Information Working Group (NPHIWG), developed a detailed appraisal process to evaluate questions that are already in use in State/Territory CATI health surveys. [page 12]

In addition to this systematic appraisal of existing questions, there is also a need to clarify and state the data requirements for health surveillance of the area and consult with national expert and strategy groups. In particular, it will be vital for CATI questions to meet the requirements of the health agencies represented on the CATI TRG and NPHIWG. [page 14]

Pre-Testing

Focus groups can be used in question development and consist of approximately a dozen people who engage in an intensive discussion focused on the question topic. The discussion is lead by a moderator and respondents are given considerable liberty in expressing their opinions on the topic at hand [102]. [page 18]

In addition to using focus groups in the development of survey questions, Statistics Canada have conducted focus groups after the survey has taken place to evaluate the questionnaire and interviewing procedures. [page 19]

Survey design in developed countries has incorporated the use of cognitive testing since the 1980s. With cognitive testing, survey researchers attempt to understand how respondents have interpreted the questions in relation to the identification of potential problems with perception, sensitivity, comprehension, memory, context, format, vocabulary, reference periods, judgment and response categories [51, 107]. Essentially cognitive testing is conducted in a laboratory where volunteers are interviewed by trained cognitive interviewers. The respondents are asked to give details of their understanding of the questions (or sets of questions) along with the processes they used to derive their answers. [page 20]

Question Design

Fowler (1998) presents five basic characteristics of questions and answers that are fundamental to a good self report measurement process, namely:

1. questions need to be consistently understood;
2. questions need to be consistently administered or communicated to respondents;
3. what constitutes an adequate answer should be consistently communicated;
4. unless measuring knowledge is the goal of the question, all respondents should have access to the information needed to answer the question accurately; and
5. respondents must be willing to provide the answers called for in the question [36]. [page 23]

Another key advantage of computer assisted interviewing is the ability to provide additional information for the interviewer to answer additional anticipated queries with the preparation of set introduction/selection procedures and the provision of fallback statements for use by the interviewers [47, 60]. [page 24]

Validity and Reliability

As an integral part of the testing and design stages, there needs to be specific attention to the validity and reliability of the questions in the CASIC environment [2]. Simplistically, validity is the ability to actually measure what it is intended to measure while reliability is the consistency of the measurement. [page 27]

Content Validity - are all relevant concepts represented in the questions?

- Criterion Validity** - how do the questions compare to a “true” value or “gold standard”?
- Construct Validity** - how do the questions match to other related measures?
- Reliability** - are responses to the questions consistent and stable from one survey to another? [page 31]

Field Testing

Testing the questions in the field is a vital stage in the development and assessment of survey questions, along with the development and testing of associated procedures and instructions [75]. There are three main types of testing - skirmishing, pilot testing and dress rehearsals. [page 33]

- **Skirmishing** is the process of informally testing the questions and can be a part of the discussions with key informants and/or a reference group. [page 33]
- **Pilot testing** involves formally testing the questions with a small sample of respondents in the same way that the final survey will be conducted. As part of the pilot testing, debriefing of the interviewers is required to identify possible problems and anomalies with the questions [27, 34, 101]. [page 34]
- A **dress rehearsal** is a trial run of a survey where the chosen sampling methodology is used to select a small sample from the target population. Dress rehearsals are used to conduct a final test of the questions before moving to the final survey. [page 34]

Usability Testing

There is a need to evaluate the ‘usability’ of the CASIC instrument to ascertain how easy or difficult it is for both interviewers and respondents to interact and to determine the accuracy of responses. Usability testing focuses on the interviewer’s interaction with the CASIC system and survey instrument, shifting the focus from system feasibility and functionality to design of instruments from the user’s perspective, and increases the importance of usability testing [45]. [page 37]

Reporting

The wide dissemination of the details of the question development process is required to ensure that extensive value can be gained from the resources used and advances that have been achieved. Not only does this support the input from the key informants but it facilitates the wide use of comparable questions across a number of surveys. [page 39]

Interviewers and Supervisors

One of the strengths of CATI is the centralised supervision and support for interviewers [16]. With regard to the development of questions there is also a

need to provide interviewers and supervisors with additional material related to the questions [109]. [page 41]

Additional Issues

One issue associated with the development of sets of questions for a topic is the determination of which questions should be considered as core questions and which questions can be considered as expanded questions. To a major extent the determination of which question is considered to be part of a core component of an ongoing series of surveys depends on the priority given to the topic and the requirement to monitor this topic. This determination would be made by the body responsible for the survey following consultation with appropriate reference groups. [page 46]

Coverage of CASIC surveys is a major concern, particularly when the characteristics of those not covered by the population being studied can be expected to be substantially different from those covered by the survey methodology. [page 46]

As part of an ongoing assessment of the questions, there should be a component of re-interviews as part of the final survey. The BRFSS incorporates a 5% re-interview of the completed interviews (see <http://www.cdc.gov/nccdphp/brfss>). [page 48]

Introduction and Overview

Computer Assisted Survey Information Collection (CASIC)

This document presents the principles and practices for the various stages that should be addressed in the development of questions/modules for computer assisted survey information collection (CASIC) in Australian health surveys. CASIC is a broad approach that denotes the use of computers for survey data collection, data capture, data preparation, and associated tasks [24]. This approach encompasses computer assisted telephone interviewing (CATI), computer assisted personal interviewing (CAPI), computer assisted self-interviewing (CASI) and surveys conducted via the world wide web [26]. CATI and CAPI are interviewer-based survey methods and this document is oriented predominantly towards these methods, although the principles and practices covered apply equally to all CASIC methods.

In developed countries CASIC health surveys are used for monitoring, surveillance, and research to provide information to support policy. Since the mid-1980s the Behavioral Risk Factor Surveillance Survey has been collecting State CATI health survey information in the United States (US) on an ongoing basis. In 1998, the State and Local Area Integrated Telephone Survey has been introduced by the US Centers for Disease Control and Prevention, with an emphasis on children's health survey data. Recently the Canadian Community Health Survey has been introduced using CAPI procedures and collects ongoing health survey data at regional and provincial level. In Europe a number of countries have introduced CATI health surveys and some harmonisation of surveys across countries is underway.

A number of Australian jurisdictions introduced CATI health surveys in the 1990s and the 2004/5 National Health Survey (NHS) is planned to be the first CAPI health survey to be run by the Australian Bureau of Statistics (ABS). In 1998, a working group of the National Public Health Partnership (NPHP) was established with an aim to standardise the collection of CATI health survey data. The material presented in this document has been drawn from a number of sources, including the work of the CATI Technical Reference Group (CATI TRG), the working group of the NPHP.

The work of the National CATI TRG under the NPHP has emphasized the need to establish question and module development standards in the CASIC environment that can be utilised in CATI health surveys across Australia. This has been seen as important so that measurement error in the CASIC environment can be reduced as well as providing greater harmonisation of the results of different CASIC data collection activities [18]. This overview of question/module development principles and practices provides a guide to the various stages that are required to ensure that 'best practice' question development is followed.

Over the last twenty-five years internationally there have been major developments in the methods and techniques of data collection for sample

surveys and CASIC has predominantly become the standard survey technique [39]. In some government agencies and survey organisations the use of CATI has become the dominant mode of data collection, particularly where there is extensive telephone coverage. The lower cost of data collection, improved sampling methods through random digit dialing and improved technology have been put forward as key reasons for the increased interest in telephone surveys [56]. During the same period there has been recognition that question meaning, wording and ordering have a major impact on responses in survey interviews. In addition, there has been increasing research into the issues of recall biases related to the changing data collection technology [97]. However, in Australia these developments have only recently been introduced.

Purpose

This document outlines principles and practices for the development of consistent and uniform questions (or sets of questions) in CASIC. It covers the identification and assessment of questions, key processes in the development of questions and the processes to test validity and reliability. In addition, subsequent fieldwork issues, interviewer feedback and reporting on questions are outlined.

In practice the development of clusters of questions on related issues could be expected, rather than each question being considered in isolation. The procedures outlined in this document are thought to be essentially the same whether one question or a group of questions is being developed.

Traditionally the major contributors to the development of survey questions have been questionnaire designers working closely with subject matter experts, with input from experienced interviewers. In recent years there has been a shift to explicitly define at the beginning of the process the policy and data requirements for health monitoring and surveillance, and later in the process to include an understanding of questions from the perspective of the respondent, with the operational aspects of questionnaire development and testing incorporated into survey planning [29]. Accordingly, this document covers the early stages of question development that include data requirements, question selection, pre-testing and question design, through to quantitative evaluation, field testing and interviewer support.

A summary of the question development stages is presented in Figure 1. These stages are similar to the processes outlined by Dillman (2000) where he recommends the review of questions by knowledgeable colleagues, qualitative evaluation, small pilot study and a final check [32]. The stages are also similar to those employed by Statistics Netherlands where the following development steps are used [2]:

- project preparation and risk analysis
- qualitative study using data collection tools in a questionnaire laboratory
- qualitative field study
- quantitative pilot in the field (with respondent's feedback)
- incorporation of changes and corrections
- grand rehearsal then implementation

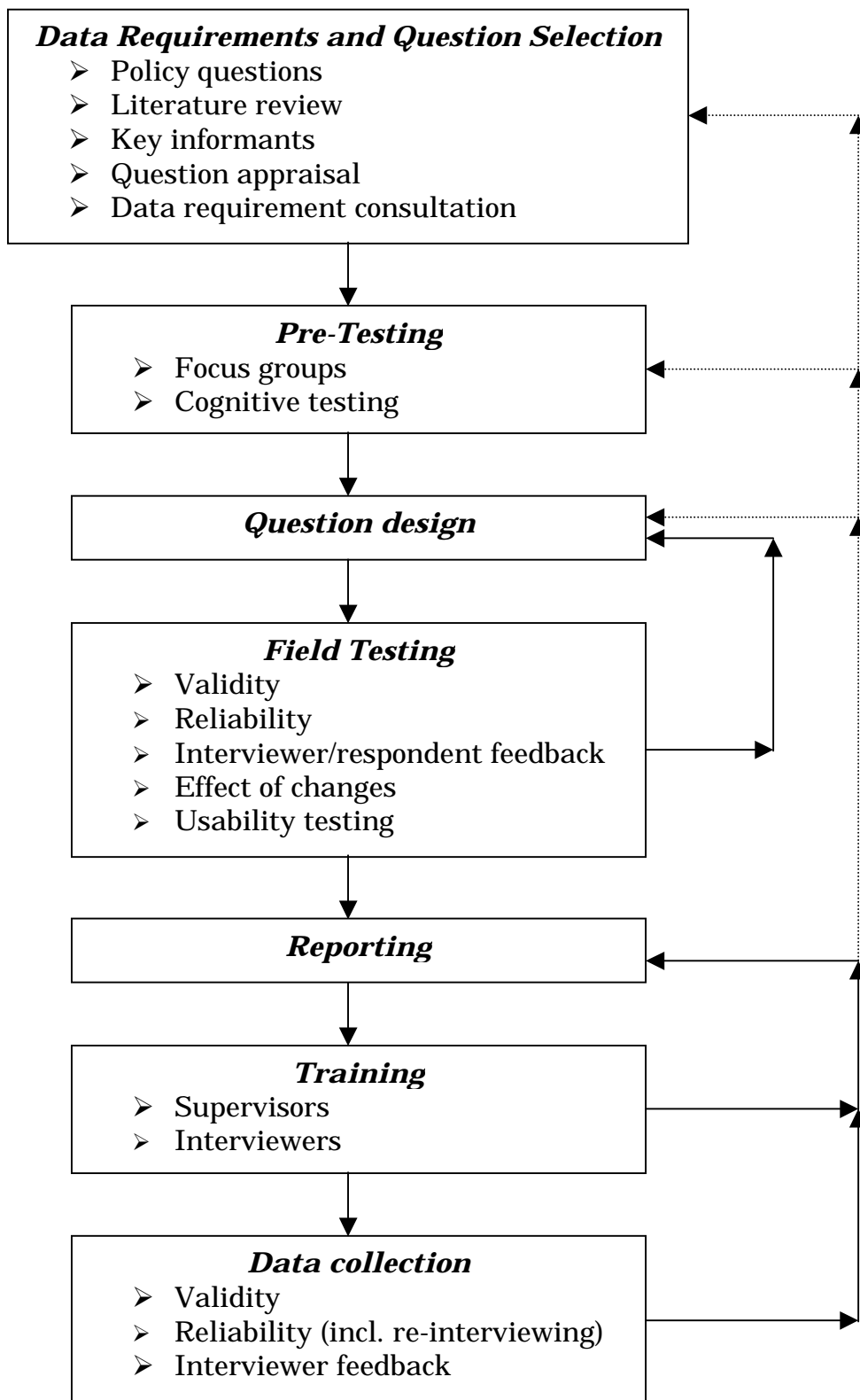


Figure 1: Stages in question development

This document follows the stages presented in Figure 1 and emphasizes the need for a systematic and rigorous approach to question development. By following these stages it can be expected that the resulting questions will be able to be utilised across many CASIC data collections with a high degree of confidence. It should be noted that the later stages provide valuable feedback to the earlier development stages thus emphasizing that the monitoring and assessment of questions is a continual process with a need to monitor the ongoing appropriateness of survey questions. Although the stages presented in Figure 1 are outlined separately in this overview of question development principles and practices, there are obviously linkages across the various stages. In particular, the key aim of pre-testing is to produce questions that are as good as possible (given the context of the survey development) to go to field testing.

Between stages there is a need to consider whether the stage has been successfully completed and whether development can proceed to the next stage. This can be formally achieved by referring to the body responsible for the survey question development or a reference group to assess the outcomes of each stage before proceeding to the next stage.

To some extent this document presents a wide range of options and techniques that can be used in the development of questions (or sets of questions). It does not specify one set of procedures for all questions and topics, although a systematic and rigorous approach is required for the development of CASIC health survey questions. The requirements for different topics will vary depending on a number of factors, particularly the availability of existing questions. Other factors include the time available for testing, the level of funds available, human resources, and the aims and objectives of the researchers (such as consistency over time versus maximizing item quality). For example, in the topic of smoking there are a wide range of questions that have been developed in the Australian setting and it could be expected that most of the stages outlined in this document will have already been conducted. On the other hand questions related to community capacity and gambling have not been widely developed and would require extensive qualitative and quantitative testing.

Unless there are important reasons to the contrary, it is suggested that the development stages outlined in Figure 1 be followed in the order presented. However, if substantial evidence is available that the procedures and practices have already been followed then one or more of the stages might either be omitted or covered in a reduced manner.

CASIC Advantages and Disadvantages

CASIC methods can reduce costs, improve timeliness and improve the quality of collected data compared to traditional paper-and-pen based data collections. CASIC methods have also shown improvements in questionnaire administration, as well as interviewer and respondent acceptance. With regard to questionnaire development and administration effects the shift to CASIC procedures can have a dramatic effect with:

1. more complicated routing patterns of respondents through the questionnaire;
2. on-line range, consistency, and data completeness checks;
3. automated prompting and standardised prompts for error resolution;
4. on-line 'help' utility;
5. customised wording of complex questions based on preloaded data or answers to previous questions;
6. ability to access data from external sources and use it in the interview; and
7. randomisation of question order and response options [55, 103].

Lavrakas (1987) identifies that the major disadvantages are both the limitations in the forms of questions that can be asked and the length of questioning, particularly in CATI interviews. In addition, he believes that “the quality of data that (are) gathered (via CASIC) is directly related to the quality of interviewing that is performed. This in turn is a function of the skill of individual interviewers and the rigor of (the) systematic routine that interviewers are expected to follow”. Accordingly, it is imperative to develop questions and associated interviewer instructions that have been thoroughly tested and evaluated in the CASIC environment [59].

Mode of Collection

The mode of collection is not dealt with in this document. However, there is a need to consider the different collection modes that might be preferable for the questions under development.

The requirement to address issues of errors in surveys has been acknowledged long before national health surveys were conducted. In 1944, Edwards Deming published his landmark paper outlining a classification of the factors affecting the usefulness of a survey. This included the differences related to the mode of collection. He called for an understanding of “the differences in results obtained from mail, telephone, telegraph, and interview canvasses, or the results obtained from different plans of questionnaires” [30].

CATI health surveys conducted by properly structured interviews have an advantage over other in-person and mail surveys with the data collection closely monitored and supervised. However, there is a need for questions (and sets of questions) to be developed for the CATI mode of collection as the quality of CATI surveys is affected by several factors, including the quality of responses obtained from respondents [93].

In Australia, the proportion of households with fixed (or cordless) telephones has in recent years risen to levels sufficient for the conduct of CATI population surveys. Currently 97.8% of households have fixed telephones [8]. Previously Trewin and Lee (1988) reported that Australia was at the lower end of telephone coverage amongst developed countries with 91.3% of households (in 1986) having a fixed telephone [100]. At the same time it was concluded by Steel and Boal (1988) that CATI surveys should not be used where the subject (such as health) was related to areas of low coverage including low income, young people or people

in rented accommodation [92]. Currently there are important groups such as the remote indigenous population, the homeless and those in institutions who would not be covered by CATI health surveys despite this high percentage of Australian households with fixed telephones (see survey coverage, page 44).

The recent dramatic changes in the telecommunication industry with the high take-up of mobile phones in Australia could be expected to have a major impact on CATI. In 2000, 58.5% of households had at least one mobile phone [8]. Accordingly, in the future there could be greater reliance on CAPI procedures as well as surveys conducted via the world wide web for particular population groups.

CASIC Questionnaire Development Issues

This document follows a series of broad headings in order to address the key issues related to the reduction of measurement error in survey data collection. DeLeeuw (2000) explains that the main advantages of computer assisted interviewing are improved data quality, efficiency and lower costs. Particularly with the automatic routing of questions and the use of range and consistency checks, data quality has improved [28]. However, there has been a re-examination of other non-sampling errors, including errors in survey measurement. She observes that

“computer assisted data collection is no panacea for good data quality. It requires one to do almost everything that is needed with a good paper-and-pen interview or questionnaire, and to add extra effort in computer implementation, in testing the questionnaire, ... in extra interviewer training, and in designing a respondent friendly and trustworthy questionnaire.”

Her views are not new but they highlight a number of key areas that require specific attention with the development of questions for CASIC. With regard to the magnitude of bias in responses, questionnaire construction and questionnaire administration are two of the major sources of bias in data collection [38]. These areas are supported by the earlier observations by Platek (1985) who wrote that with “errors resulting from unrealistic demands on the respondent’s knowledge or memory, the use of overly difficult and technical language, or the excessive demands on respondent’s patience are all sources of non-response which have their roots in the questionnaire” [79].

Of relevance to the subject of this document, Deming (1944, page 364) highlighted the imperfections in the design of questions and questionnaires [32].

“Faulty design ... can be the cause of considerable bias. Faulty design often arises from lack of knowledge of the subject matter. It is not sufficient merely to limit answers to questions somehow or other related to the subject. The questions must attack the root of the problem by discovering what are the significant underlying causes. ... An understanding of the subject is accordingly demanded as one of the qualifications for planning a survey.”

‘Best Practice’ has been defined as a “holistic strategy”, a jigsaw with all elements interlocking and connected [10]. In this situation, question

development can be seen as the integration of the selection, evaluation and testing of questions within the CASIC environment, combined with interviewer procedures and supervision. Willis et al. (1999) identified that the use of different techniques in question development (specifically the use of cognitive testing, field testing and the use of 'experts') were complementary rather than conflicting with one another. They found no reason to choose one technique over the others and recommended that the different approaches be used at particular points in the survey development process [105].

The pretesting policy of the US Census Bureau presents mandatory standards for question and questionnaire development. This policy emphasizes that a survey question and questionnaire must 'work' and if there is insufficient evidence about how well a question performs then the question must be subjected to some form of questionnaire pretest along with field interviews [101]. These pretests include:

- Respondent focus groups
- Cognitive laboratory interviews
- Behaviour coding of respondent/interview interactions (usability)
- Respondent debriefing
- Interviewer debriefing
- Split panel tests (where alternative questions are compared)
- Analysis of item non-response rates and response distributions.

This pretest policy also requires quantitative measures of validity and reliability to ensure that question and questionnaire results can be generalised to the population of interest. The policy also encourages the evaluation of data quality of existing questions through re-interviews, as questions become outdated and should be periodically evaluated.

Similarly, Statistics Netherlands has an extensive program of questionnaire development. The Questionnaire Design Resource Center provides design, development and testing that incorporate ordinary interviews, in-depth interviews, cognitive interviews and focus groups [3].

This overview of question development principles and practices does not provide the specific prescription on issues such as question wording and format. It concentrates on the development of uniform and consistent questions, although Platek (1985) observes that a questionnaire should not be too rigid and must be flexible to adapt to respondents of different age/sex groups, language and social backgrounds. Different words or examples may be needed in order to convey the desired meaning to all respondents, and the questionnaire must be able to anticipate all possible answers. It is therefore necessary to develop clear guidelines for interviewers and supervisors, providing the range and extent of prompts and expansion of questions that will assist respondents to provide the required information without leading or introducing bias into the responses [79].

In Australia, Trewin (1987) reported that the ABS was investigating the use of CATI and CAPI systems, particularly as "there is less leeway for interviewers to make errors" [99]. With respect to CATI he noted that there are important quality control processes that can be incorporated into CATI procedures,

including the monitoring of interviewer performance, their interaction with respondents, as well as examining the performance of the questionnaire.

The ABS has been engaged in a comprehensive range of testing strategies for their paper-and-pen interview survey program. The stages outlined in this document are essentially the same as those followed by the ABS, but are extended to highlight the CASIC requirements.

Conclusion

To conclude, the document covers issues related to the wide reporting of the questions (and sets of questions) that have been developed along with recommendations related to issues for interviewers and supervisors and a selection of additional issues that sometimes need to be considered. As the interviewer's primary task is to interview, they must be able to answer any queries from respondents to clarify ambiguous concepts, recognise and correct response errors and 'convert' refusals. In the CASIC environment interviewers learn to rely on the system to provide responses to most situations. Accordingly, an integral addition to the development of questions (and sets of questions) is the development of interviewer and supervisor support, such as set answers to anticipated difficulties some respondents might have [48].

It is expected that with the development of questions for CASIC (particularly CATI), the preparation and dissemination of reports that document the development process will be publicly available. The acceptance of the questions into the National Health Data Dictionary would also ensure the broad utilisation of standard (comparable) questions across data collections [9].

For the broader context of CASIC question development within the conduct and quality of conducting health surveys, the following texts will provide further information:

- Aday LA. Designing and Conducting Health Surveys. [1]
- ABS. An Introduction to Sample Surveys, Cat. No. 1299.0. [7]
- Bickman L & Rog DJ. Handbook of Applied Social Research Methods. [17]
- Couper MP et al. (eds.) Computer Assisted Survey Information Collection. [24]
- Dillman DA. Mail and Internet Surveys. The Tailored Design Method. [32]
- Lessler JT & Kalsbeek WD. Nonsampling Error in Surveys. [62]
- McDowell I & Newell C. Measuring Health. A Guide to Rating Scales and Questionnaires. [66]

Data Requirement and Question Selection

There are a number of steps that can be followed in identifying the data requirements for health monitoring and surveillance, operationalisation of those data requirements, and selection of possible questions. They include the identification of relevant policy questions, a review of the literature, the input from subject matter experts, other stakeholders and key informants, and a systematic appraisal of existing questions.

Policy Questions

The specification of data requirements for health surveillance and selection of subject topics within a health survey can be a less than trivial task. Certainly, the topic (eg smoking) may well have been determined at an early stage, but the more precise data requirements/variables/questions both within the topic and associated with the topic require close attention. Accordingly, there is a need to examine (or re-examine) the concepts to be measured and the subsequent explicit specification of data requirements. From this it could be expected that there may be a need to develop new questions or modify existing questions.

A structured approach to the selection of variables for health survey data collection and analysis is required. Particularly, documentation (in tabular form) of the policy questions that are relevant to the topic and a statement of data requirements for health surveillance is recommended, followed by an assessment of the relevant variables or indicators for each policy question and operationalisation of the data requirements [77].

To assist in this approach, the NPHP has been developing a performance indicator framework for population health that has been based on recent Canadian and US initiatives, see Figure 2 below [72]. Utilising this framework provides a mechanism to identify variables/questions associated with the topic and reinforces the associations between the topic with the other health indicators and outcome measures.

Literature Review

A thorough review of the literature to identify existing data collections and the method of collection used previously, is a key step in the identification of possible questions and variables related to the particular topic and associated variables/questions. It is desirable to identify as broad a range of existing questions as possible, although with time and resource constraints it may be practical to examine the following databases and the international websites presented below. Databases to be searched should include:

- | | |
|--------------------|--------------|
| ➤ AustHealth | ➤ HealthSTAR |
| ➤ AustRom | ➤ Medline |
| ➤ CINAHL | ➤ PsycInfo |
| ➤ Current Contents | ➤ Sociofile |

Health Status and Outcomes				
How healthy are Australians? Is it the same for everyone? Where is the most opportunity for improvement?				
Health Conditions	Human Function	Life Expectancy and Well-Being	Deaths	
Prevalence of disease, disorder, injury or trauma or other health-related states.	Alterations to body, structure or function (impairment), activities (activity limitation) and participation (restrictions in participation)	Broad measures of physical, mental, and social well-being of individuals and other derived indicators such as Disability Adjusted Life Expectancy (DALE).	Age or condition specific mortality rates.	
Determinants of Health				
Are the factors determining health changing for the better? Is it the same for everyone? Where and for whom are they changing for the worse?				
Environmental Factors	Socio-economic Factors	Community Capacity	Health Behaviours	Person-related Factors
Physical, chemical and biological factors such as air, water, food and soil quality resulting from chemical pollution and waste disposal.	Socio-economic factors such as education, employment per capita expenditure on health, and average weekly earnings.	Characteristics of the community and family such as population density, age distribution, health literacy, housing, community support services and transport.	Attitudes, beliefs knowledge and behaviours eg patterns of eating, physical activity, excess alcohol consumption and smoking.	Genetic related susceptibility to disease and other factors such as blood pressure, cholesterol levels and body weight.
Health System Performance				
How well is the health system performing in delivering quality health actions to improve the health of all Australians? Is it the same for everyone?				
Effective	Appropriate		Efficient	
Care, intervention or action achieves desired outcome.	Care/intervention/action provided is relevant to the client's needs and based on established standards.		Achieving desired results with most cost effective use of resources.	
Responsive	Accessible		Safe	
Service provides respect for persons and is client orientated: - respect for dignity, confidential, participate in choices, prompt, quality of amenities, access to social support networks, and choice of provider.	Ability of people to obtain health care at the right place and right time irrespective of income, geography and cultural background.		Potential risks of an intervention or the environment are identified and avoided or minimised.	
Continuous	Capable		Sustainable	
Ability to provide uninterrupted, coordinated care or service across programs, practitioners, organisations and levels over time.	An individual or service's capacity to provide a health service based on skills and knowledge.		System or organisation's capacity to provide infrastructure such as workforce, facilities and equipment, and be innovative and respond to emerging needs (research, monitoring).	

Figure 2: National Public Health Partnership performance indicator framework
Source – NPHP, 2000

The preparation of the General Nordic Questionnaire provides an example of the use of an extensive review of the literature on sets of questions to derive a comprehensive survey program [63]. The review of the literature for each set of questions was lead by a researcher familiar with the subject. A summary of the review, along with a draft questionnaire, was published by the Nordic Council of Ministers for wide consultation prior to testing.

There are a number of resources that can be utilised in the gathering together of existing questions. In particular, there have been a number of State/Territory CATI Health Surveys that have covered a wide range of health topics. NSW Health and the SA Department of Human Services have conducted an extensive number of surveys, see

- <http://www.health.nsw.gov.au/public-health>
<http://www.health.sa.gov.au/pehs/CPSE.html>

The NHS questions can also be a valuable resource in the identification of existing questions. The questions used in the 1995 NHS are published in the ABS reference package [6].

There are also four particular international resources that provide extensive detail for health questions in the US, Canada and the UK, namely:

- BRFSS: the Behavioral Risk Factor Surveillance Survey conducted in each State of the US by the Centers for Disease Control and Prevention, see <http://www.cdc.gov/nccdphp/brfss>
- CCHS: the Canadian Community Health Survey conducted as part of the Canadian Roadmap Initiative, see <http://www.cihi.ca/Roadmap/rdindex.shtml>
[Roadmap Initiative]
<http://www.statcan.ca/english/concepts/quest.htm>
[Questionnaires and data dictionaries]
<http://www.statcan.ca/english/concepts/health/content.htm>
[CCHS questionnaire]
- CASS: the Centre for Applied Social Survey in the UK, see <http://www.scpr.ac.uk/cass/docs/casshome.htm> [CASS]
<http://qb.soc.surrey.ac.uk> [the Question Bank]
- Health Poll Search: Presented by the Roper Center at the University of Connecticut and the Kaiser Family Foundation, this feature at the Kaiser website is a searchable archive of public opinion questions on health issues. The database contains about 40,000 selected questions dating back to 1935, which are drawn from the Roper Center's comprehensive database of more than 350,000 questions.
http://www.kaisernetwork.org/health_poll

One important consideration when identifying existing questions from other sources is that the same question in a different collection mode is a different question [32]. Accordingly, having gained questions, it is important that the

question is tested and evaluated in the environment of the CASIC (particularly CATI) health survey and also in the Australian setting.

From a review of the literature there can be an identification of key questions/variables associated with the topic being investigated. The literature review may result in the elimination of unnecessary questions and the incorporation of new questions. For example, if there has been demonstrated to be no difference in health status between those who are self-employed and those who are salaried workers, then this question might be excluded. However, if new evidence is being reported on the relationship between employment security, control at work and health status then questions on work organisation and health should be investigated and developed.

Key Informants

It is fundamental to include subject matter experts, policy and program developers, other stakeholders and key informants in the process of definition of data requirements for health surveillance and subsequent question selection. Key informants are usually those individuals who have demonstrated extensive knowledge in the topic, such as researchers, academics, policy makers and representatives of special interest groups. They provide an important link to the existing literature and can provide expansion of possible collections that have not been reported in the (predominantly) academic literature. They can also be incorporated in reference groups to assist in the appropriate links to individuals or organisations with special interests in the topics under consideration. In addition, there are advantages if the involvement of key informants can be maintained for the interpretation and presentation of survey results.

Key informants should also incorporate established expert groups and other national work that is underway such as the national health priority areas. In particular, these groups provide an identification of the data requirements, questions and variables that are relevant to policy in the topic areas. They can also provide direction in the establishment of key results and performance indicators relevant to the question (or set of questions) under development.

Question Appraisal

Recently the CATI TRG, a subcommittee of the National Public Health Information Working Group (NPHIWG), developed a detailed appraisal process to evaluate questions that are already in use in State/Territory CATI health surveys. A schematic diagram of the module process is presented below as Figure 3 and shows that this appraisal process provides a systematic mechanism to gather, appraise and cull existing questions used in State/Territory CATI health surveys.

Stage 1 of this appraisal process involves the collation of all available questions: an example of the template developed by the CATI Technical Reference Group is presented in Attachment A. In particular, it can be seen that this collation process incorporates a structured approach that documents specific policy questions relevant to the question topic [21].

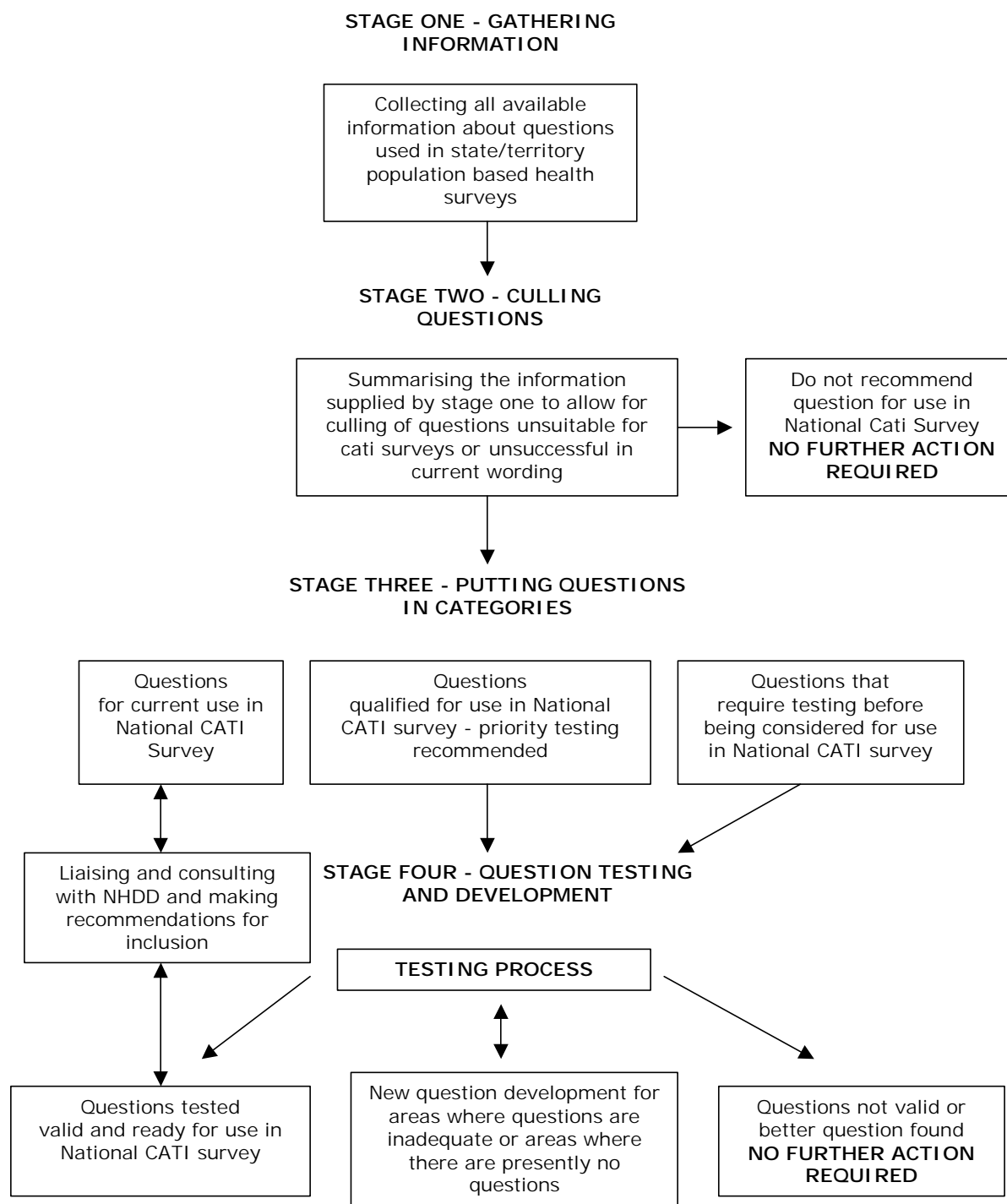


Figure 3: Schematic diagram of the CATI TRG question appraisal process
Source – CATI Technical Reference Group, 2000

Stage 2 of the appraisal process requires the culling of questions that the reference group does not recommend for use in CATI health surveys. This stage utilises the information provided in the collation process template.

Stage 3 requires the categorisation of questions into three groups. Those that are considered to be recommended for current use and have met all standards of validity and reliability testing (and do not require further testing); those that need testing to confirm inclusion in CATI health surveys; and finally those that need more extensive testing before being considered.

Stage 4 of the appraisal process is the testing and development of questions. Essentially, this stage is related to the subject of the remainder of this document.

In addition to this appraisal process, experts in survey design can contribute to the assessment of proposed questions through their analysis and review of questions.

Data Requirement Consultation

In addition to this systematic appraisal of existing questions, there is also a need to clarify and state the data requirements for health surveillance of the area and consult with national expert and strategy groups. When consensus has been reached this statement of data requirements then provides a valuable input into the subsequent qualitative and quantitative testing, as well as being incorporated in the reporting of the question development. In particular, it will be vital for CATI questions to meet the requirements of the health agencies represented on the CATI TRG and NPHIWG. Figure 4 shows a schematic diagram of the CATI module standard consultation process for chronic disease and risk factor module development with a focus on health surveillance.

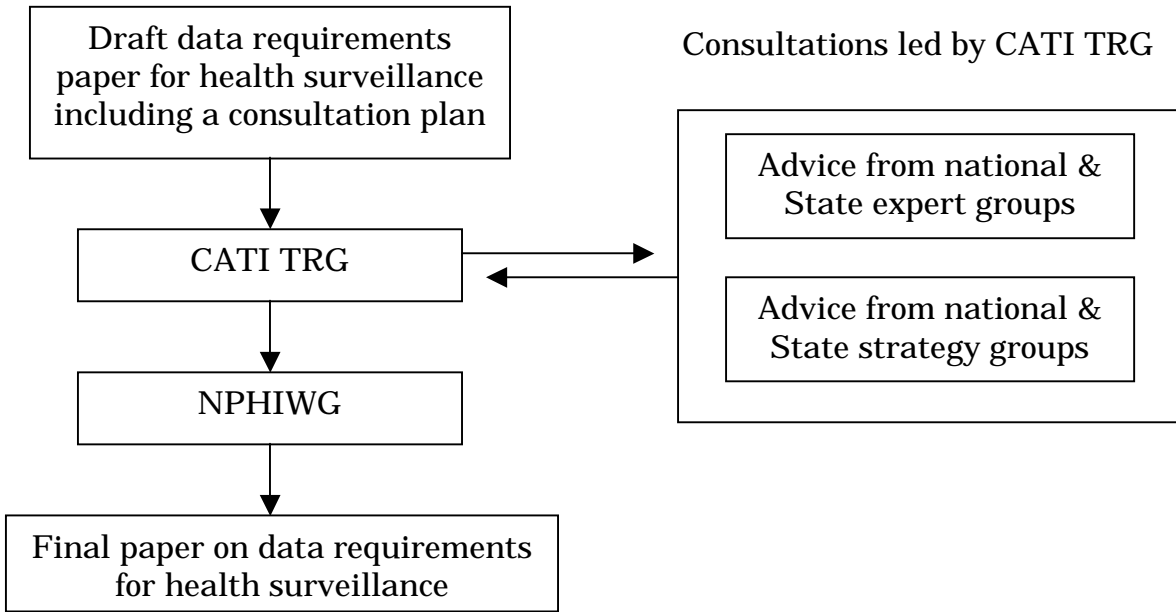


Figure 4: Proposed CATI module standard consultation process

Policy Questions

- what are the policy questions relevant to the topic?
- how does the topic fit into the performance indicator framework?

Literature Review

- what questions have already been used for the topic?
- do questions exist in major collections such as the BRFSS or CCHS?

Key Informants

- are there researchers, academics, policy makers or representatives of special interest groups who can advise on the topic?

Question Appraisal

- can existing questions be used?
- do CATI questions meet the data requirements of the CATI TRG and NPHIWG?

Data Requirement Consultation

- what are the data requirements for health surveillance?
- do they include advice from national expert and national strategy groups?
- have data requirements achieved endorsement of the CATI TRG and NPHIWG?

Pre-Testing

In the conduct of surveys there are four stages of information processing for the respondent:

- the comprehension of the question;
- the retrieval or construction of an answer;
- an evaluation of or judgement about the answer; and
- the reporting of the answer.

There needs to be an understanding of all of these stages in the development of questions, particularly as problems have been found with questions in relation to the comprehension, recall and decision processes respondents use to answer questions [82].

The simple model presented in Figure 5 reinforces this view that the retrieval of information in answering questions is not necessarily automatic and that the development of questions needs to be mindful of the processes involved [104].

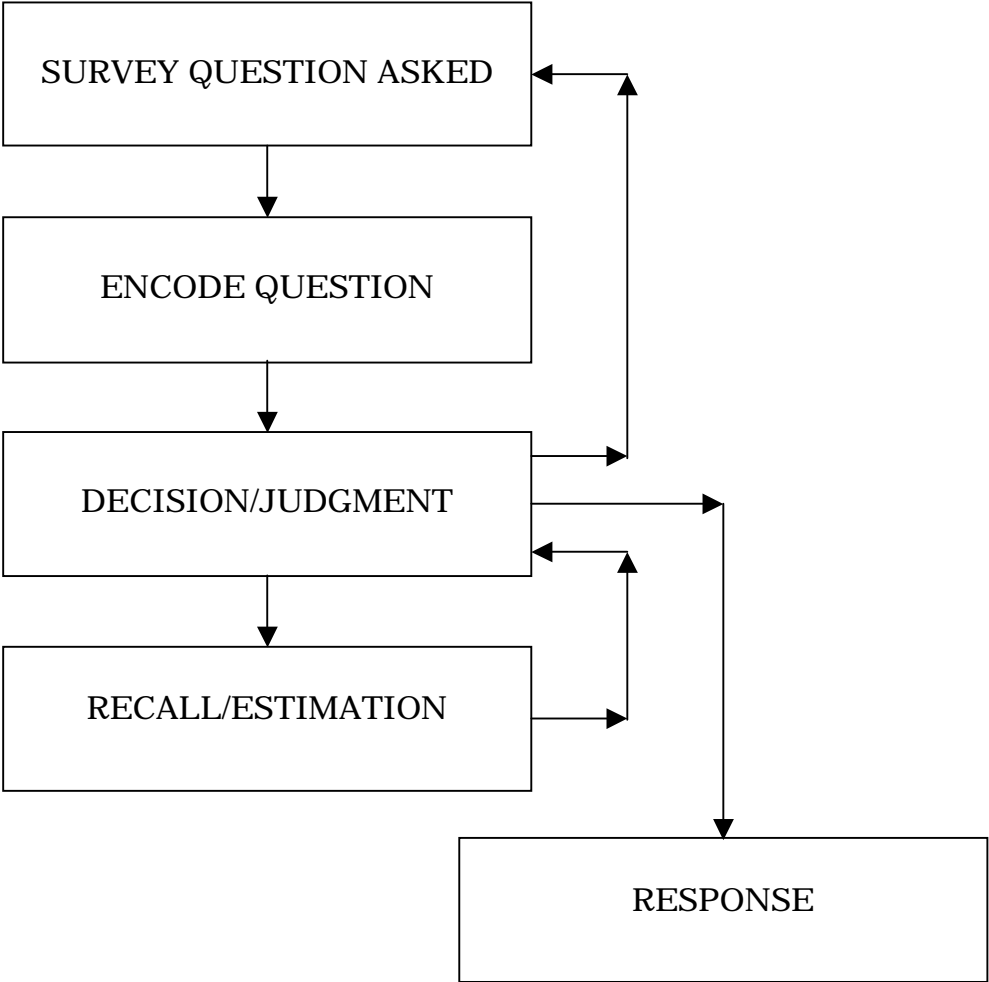


Figure 5: Cognitive model of the survey response process
Source – Willis et al., 1991, page 253

In recent years there have been significant improvements in the art and science of survey question development, with the application of the principles of cognitive psychology, the value of qualitative and ethnographic research methods, and understanding of language intended for respondents. Focus groups and ethnographic interviewing have been used to solicit opinions on research topics, with the particular procedures chosen dependent on the nature of the topic. For example, in-depth interviews may be preferred when the topic relates to more personal issues. The use of design oriented focus groups and cognitive interviewing in a laboratory setting are now seen as complementary processes that can evaluate many aspects of question design [1, 33, 89].

Typically, for pre-testing respondents are recruited through a call for volunteers such as advertisements in the press. Most organisations also provide some form of remuneration for the time they are required in the laboratory to assist in the pre-testing.

Focus Groups

Statistics Canada use focus groups at various phases during the survey development process [41]. They have used them to:

- help define and clarify research objectives and data requirements;
- gain an understanding of concepts and issues from the perspective of respondents prior to developing the questionnaire;
- test questionnaires (including questions and response categories) and data collection procedures;
- evaluate alternative versions of questions and response categories;
- obtain respondents' reactions to questionnaires;
- investigate the wording and vocabulary that respondents use;
- evaluate the respondent-friendliness of questionnaires; and
- discuss respondent relations issues.

With the use of standardised and sometimes technical language in survey questions, it has been recognised that many respondents may have difficulty understanding the meaning and intent of survey questions [1]. Valuable insight into the issues involved in a question topic can be gained through the use of group interviews that focus discussion on the particular topic that the survey questions are endeavouring to measure [54]. Focus groups can be used in question development and consist of approximately a dozen people who engage in an intensive discussion focused on the question topic. The discussion is led by a moderator and respondents are given considerable liberty in expressing their opinions on the topic at hand [102].

The advantages of conducting focus groups include the relative ease in quickly gaining information related to the understanding and background to specific topics from respondents. The moderator of the focus group can quickly explore issues that arise in the focus group and can ascertain misconceptions that respondents might have. In addition, comments from different members of the group can produce insights and stimulate broader group discussion that would not have been gained from individual interviews [54].

The moderator of the focus group needs to be mindful that they can lead the discussion towards their own predispositions and prejudices. Also, dominant individuals can monopolise the focus group discussion at the expense of the input from less dominant members [54].

There are a number of texts that provide greater detail on focus groups. Kerr et al. (1998, pp 527-532) is a readily available text that can provide a succinct insight into the procedures required to conduct focus groups as well as providing reference to further reading [54].

In their pretesting policy, the US Census Bureau has identified that it is critical to identify problems for both respondents and interviewers [101]. They state that:

“Respondent focus groups ... are used early in the questionnaire development cycle ... to assess the question-answering process. ... Focus groups can be used to quickly identify variations in language or terminology, or interpretation of questions and response options.”

Focus group discussions contribute to question design by helping to examine the underlying assumptions about the reality about which people will be asked [36]. They are also valuable in assisting in the evaluation of vocabulary assumptions; specifically the way people understand the particular terms or concepts related to the questions (or sets of questions). There are three basic focus group conversation topics:

1. do the questions appropriately cover what respondents are supposed to describe?
2. are the response tasks that the questions will pose, tasks that respondents are able and willing to perform?
3. do the words or descriptions proposed in the questions convey consistent meaning, so that respondents have a common understanding of what question they are to cover?

In a study of environmental risk factors, Desvousges & Frey (1989) reported that “focus groups (were) effective in developing visual aids, evaluating experimental design alternatives, assessing the order of questionnaire topics, constructing scales and other measures, identifying levels of knowledge among a population, and overcoming problems with troublesome language, words and phrasing” [31]. Although there has been a wide acceptance of the use of focus groups in developing questions and determining public opinions on topics, there has been a paucity of research into the quality of the procedure over and above other question and questionnaire testing. In particular, more research into the training of the moderator is required [19]. However, where a number of issues and concepts related to the questions require clarification, the use of focus groups assists in providing input prior to cognitive testing.

In addition to using focus groups in the development of survey questions, Statistics Canada have conducted focus groups after the survey has taken place to evaluate the questionnaire and interviewing procedures. They report that

these focus groups also provided an opportunity for respondents to comment about the interviewers, response burden and the accuracy of their response, particularly in relation to confusion or misunderstanding of the intent of the survey questions. They also provided a mechanism to consult with interviewers and to listen to their suggestions and recommendations [41].

The following texts will provide further information on focus groups, in particular, providing details on their conduct and the analysis of their results:

- Greenbaum TL. The Handbook for Focus Group Research. [42]
- Morgan DL. Focus groups. [68]
- Morgan DL, Krueger RA. The Focus Group Kit. [69]
- Stewart DW, Shamdasani PN. Focus group research: exploration and discovery. [95]

Cognitive Testing

Survey design in developed countries has incorporated the use of cognitive testing since the 1980s. With cognitive testing, survey researchers attempt to understand how respondents have interpreted the questions in relation to the identification of potential problems with perception, sensitivity, comprehension, memory, context, format, vocabulary, reference periods, judgment and response categories [51, 107]. Essentially cognitive testing is conducted in a laboratory where volunteers are interviewed by trained cognitive interviewers. The respondents are asked to give details of their understanding of the questions (or sets of questions) along with the processes they used to derive their answers.

Questions may be difficult to answer if they are difficult to understand, if response requires detailed memory recall, or if response categories fail to cover the range of respondent experience [35]. In addition, the time required by respondents to consider their response is examined. For example, some questions may require extensive consideration before the respondent is ready and able to provide their answer. Cognitive interviewing seeks to confirm suspected problems, discover unknown problems, and engineer solutions to problems that are identified [70]. It can be considered that cognitive interviewing trades off strength in numbers for intensity of focus where the effect of probing brings out problems in the questions that are otherwise not directly observable and pays attention to the thought processes that respondents use to answer survey questions [35, 105].

In 1999, the ABS established a cognitive laboratory for the development of survey questionnaires, particularly for new questions. In the laboratory trained cognitive interviewers conduct the interviews while cognitive psychologists attempt to learn just what respondents are thinking when they are attempting to answer questions. Currently there is no set of standards for cognitive interviewing that have been developed, evaluated and systematically implemented. Thus cognitive interviewing approaches have evolved independently across organisations and reflect the preferences of the researchers in them [14].

The three most common procedures used are:

1. “think-aloud” interviews, where the respondent is encouraged to include in their response to a question what aspects of their experience lead them to a particular response;
2. asking probe or follow-up questions after each question, or short series of questions; and
3. going through the questions twice, first under regular interview conditions followed by a discussion with the respondent about the responses they provided and the thought processes they used to come up with their response [36].

For the think-aloud interviews, respondents are asked to report their thought processes while they are answering the question, or immediately after they have answered the question. The administration of this procedure can vary widely between survey researchers and the way they analyse the findings, based on the researcher’s impressions [22].

These last two procedures focus predominantly on the respondents’ understanding of the questions and the meanings of their responses, rather than the cognitive processes involved [14]. However, there have also been developments in cognitive testing with the inclusion of hypothetical vignettes and card sorting type tasks. In the use of hypothetical vignettes there is a testing of how consistently respondents classify various marginal or ambiguous activities related to the question topic [40, 65]. This is also important for respondents who are reticent or less literate, being able to draw out difficulties that might not be expressed either in focus groups or in the three procedures presented above. For example, Martin & Polivka (1995) report on an investigation of the questions relating to work [65]. In this investigation respondents in a field test were asked to react to different imaginary work situations such as

- “Sam spent 2 hours last week painting a friend’s house and was given 20 dollars.
Would you report him as working for pay (or profit) last week?”
- “Last week, Sarah cleaned and painted the back room of her house in preparation for setting up an antique shop there.
Would you report her as working for pay (or profit) last week?”

There have been concerns that the results of cognitive testing need to be accepted by survey sponsors. In a test of draft BRFSS “quality of life” measures that respondents found difficult to answer, Beatty et al. (1996) reported that the survey sponsors were concerned that given the interviewing methodologies used in the cognitive testing, reported difficulties subjects had in providing codeable responses might have been as a result of the open cognitive procedures. In addition, the sponsors required evidence of the magnitude of the problems identified in the draft questions [13]. In an attempt to overcome this, Rothgeb et al. (2000) recommends that survey sponsors be given opportunities to observe the testing process [81].

Although cognitive testing is predominantly conducted in a laboratory setting, the incorporation of telephone interviewing is possible and can reveal issues with questions that are specific to the mode of collection. For CATI questions it is essential to integrate the delivery of the questions over the telephone into the laboratory setting [12, 70, 83]. As an example, in the ABS cognitive laboratory in Canberra, telephone interviews from adjacent rooms can incorporate CATI interviews.

Another benefit of conducting the CASIC testing in a laboratory setting is that interviewer and respondent problems can be identified [61]. Identification of major wording changes and the failure to probe are two key areas where interviewers can incorrectly conduct the interview, while respondents can:

- interrupt questions;
- provide multiple answers;
- answer outside of the response frame;
- be uncertain of their answers;
- qualify their answers
- seek a definition or have the question repeated; and
- digress or distract the interview process.

Cognitive testing provides a rich source of information in the development of questions, particularly to ensure that respondents comprehend questions the way they are intended to be understood. Dillman (2000) concludes that through the use of cognitive interviewing, questions are greatly improved and this stage of question development has become an indispensable step prior to field testing [32].

Focus Groups

Used to assess the question and answering process

- what are the underlying assumptions in the question?
- do the questions cover the topic?
- do the respondents understand what the questions are covering?

Cognitive Testing

Assessment of interviewing in a laboratory by trained cognitive interviewers

- in 'think aloud' interviews respondents report their thought processes while answering questions
- use of probing and follow-up questions
- going through questions a second time to understand the thought processes used by the respondent while answering questions
- hypothetical vignettes can be used to test situations that can be seen as likely problem areas in a topic.

Question Design

Fowler (1998) presents five basic characteristics of questions and answers that are fundamental to a good self report measurement process, namely:

1. questions need to be consistently understood;
2. questions need to be consistently administered or communicated to respondents;
3. what constitutes an adequate answer should be consistently communicated;
4. unless measuring knowledge is the goal of the question, all respondents should have access to the information needed to answer the question accurately; and
5. respondents must be willing to provide the answers called for in the question [36].

This document does not cover the detailed procedures (such as question wording) required for the design of questions. However, Fowler (1998) also provides some general principles that apply to all types of questions [36].

Principle 1: The strength of survey research is asking people about their first-hand experiences – what they have done, their current situations, their feelings and perceptions.

- Principle 1a: Beware of asking about information that is only acquired second-hand
- Principle 1b: Beware of hypothetical questions
- Principle 1c: Beware of asking about perceptions of causality
- Principle 1d: Beware of asking about solutions to complex problems

Principle 2: Questions should be asked one at a time

- Principle 2a: Avoid asking two questions at once
- Principle 2b: Avoid questions that impose unwarranted assumptions
- Principle 2c: Beware of questions that include hidden contingencies

Principle 3: A survey question should be worded so that all respondents are answering the same question

- Principle 3a: To the extent possible, choose the words in questions so that all respondents understand their meaning and all respondents have the same sense of what the meaning is
- Principle 3b: To the extent that words or terms must be used that have meanings that are likely not to be shared, provide definitions to all respondents
- Principle 3c: The time period referred to by a question should be unambiguous
- Principle 3d: If what is to be covered is too complex to be included in a single question, ask multiple questions

Principle 4: If a survey is to be interviewer administered, wording of the questions must constitute a complete and adequate script such that when the interviewer reads the question as worded, the respondent will be fully prepared to answer the question

- Principle 4a: If definitions are to be given, give them before the question itself is asked
- Principle 4b: A question should end with the question itself. If there are response alternatives, arrange the question so that they constitute the final part

Principle 5: All respondents should understand the kind of answer that constitutes an adequate answer to a question

- Principle 5a: Avoid questions that begin with adverbs: how, when, where, why, to what extent. Such questions do not specify the terms of an adequate answer
- Principle 5b: Specify the number of responses to be given to questions for which more than one answer is possible

Principle 6: Survey instruments should be designed so that the tasks of reading questions, following instructions, and recording answers are as easy as possible for interviewers and respondents

One important difference between computer assisted interviewing and paper-based surveys, is that there is no questionnaire in the usual sense in CASIC surveys. The 'questionnaire' is a computer program which creates a unique questionnaire for each interview that contains only those questions that apply to the interview. To ensure that the presented construction and relationship of questions throughout the interview are in the desired order, flow charts of the different question responses are of great assistance [50, 53]. In addition, flow charts can assist in ensuring that errors in the preparation of the 'questionnaire' are reduced with questions having the appropriate links and skips in the interview.

Another key advantage of computer assisted interviewing is the ability to provide additional information for the interviewer to answer additional anticipated queries with the preparation of set introduction/selection procedures and the provision of fallback statements for use by the interviewers [47, 60]. Auxiliary information can be placed on the screen as part of the question or as a special area reserved for instructions. Unlike printed questionnaires this additional support for the interviewer is not limited by the layout or space in the 'questionnaire'.

Once draft questions have been prepared, Dillman (2000) recommends that each of the following eight questions needs to be addressed with regard to each question and that answers to these questions assist in the diagnosis of problems [32]:

1. Does the question require an answer?
2. To what extent do survey recipients already have an accurate, ready-made answer for the question they are being asked to report?
3. Can people accurately recall and report past behaviours?
4. Is the respondent willing to reveal the requested information?
5. Will the respondent feel motivated to answer each question?
6. Is the respondent's understanding of response categories likely to be influenced by more than words?
7. Is survey information being collected by more than one mode?
8. Is changing a question acceptable to the survey sponsor?

In addition, respondents sometimes do just enough to satisfy the survey request, but no more. This has been termed 'satisficing', giving minimal acceptable answers, rather than optimal answers. To avoid many of the undesirable effects of 'satisficing', such as response order effects and acquiescence bias, question design must maximise respondent motivation, minimise task difficulty and minimise response effects. The following approaches are recommended to encourage respondents to provide optimal answers [57, 58].

Maximise Respondent Motivation:

- Describe the purpose and value of the study;
- Provide instructions to think carefully;
- Obtain commitment to think carefully;
- Give booster instructions to think carefully;
- Include random probes (why do you say that?); and
- Keep questionnaires short and place important questions early.

Minimise Task Difficulty:

- Minimise number of words in questions;
- Maximise familiarity of words;
- Minimise use of words with multiple definitions;
- Focus on salient, current or very recent events;
- Ask only about one object in each question;
- Ask only about a single evaluative dimension in each question;
- Ask for absolute, not relative, judgements;
- Decompose questions whenever possible;
- Keep pace slow; and minimise distraction.

Minimise Response Effects:

- Offer responses in balanced or random order;
- Avoid blocks of ratings on the same scale; and
- Do not offer no opinion response options.

There are a large number of references that outline details of question design. The following will provide greater detail and direction:

- Belson WA. The design and understanding of survey questions. [15]
- Oppenheim AN. Questionnaire Design and Attitude Measurement. [76]
- Payne SL. The art of asking questions. [78]
- Schuman H, Presser S. Questions and Answers in Attitude Surveys. Experiments on Question Form, Wording, and Context. [86]

Validity and Reliability

As an integral part of the testing and design stages, there needs to be specific attention to the validity and reliability of the questions in the CASIC environment [2]. Simplistically, validity is the ability to actually measure what it is intended to measure while reliability is the consistency of the measurement. Detailed definitions of validity and reliability are provided as Attachment B.

Validity

Validity refers to the extent to which a question (or set of questions) measures what it is intended to measure and does not measure what it is not intended to measure [94]. In addition, because health measures can be used for different purposes, their validity needs to be evaluated separately for each purpose. There are a number of different aspects of the validity of questions with the three broad types being content, criterion and construct validity. Specific definitions are provided in Attachment B.

Content validity (namely, are all relevant concepts represented in the question or set of questions) is essential during the construction and development of questions and is addressed with key informants and/or a reference group established to assist in the evaluation of alternative questions both before and after cognitive testing. In addition, if changes are recommended following field testing, again there is a need to examine whether the questions adequately represent the topic concept.

Criterion validity should be included as an integral component of the field testing component. Every effort to identify a “true” value or “gold standard” should be incorporated in the testing process with measures of sensitivity and specificity determined for the questions against the “true” value (or criterion indicator). This external validation is often difficult to obtain and is predominantly achieved with questions that are of a factual nature [96]. For example, in the testing of questions on child immunisation, an examination of immunisation records from respondents can provide criterion validation of the responses.

As measures of criterion validity both sensitivity (the proportion that the question picks up with the criterion) and specificity (the proportion that the question does not report the value that do not have the criterion) need to be computed and reported following field testing. For example, again with immunisation, sensitivity would be reported as the proportion of immunised children who were reported as being immunised from the study, while specificity would be the proportion of non-immunised children who were reported as not being immunised from the study. The higher the sensitivity and specificity the greater the criterion validity.

As an example of the comparison of self reporting and a ‘true’ value, in a South Australian study on hearing impairment, Wilson et al. (1999) reported that 54% of those who reported that they had hearing impairment were so classified from

measurement by standard audiological methods (sensitivity = 0.54), while of those who reported that they did not have a hearing impairment, 17% were found to have a hearing impairment from audiological assessment (specificity = 0.83). From these findings, the authors state that “this means that on self-report there is a considerable misclassification problem ... (and) self-reported hearing disability cannot be considered valid for planning purposes” [108].

Construct validity (where the questions have a hypothesized association with other measures being collected) should be documented both in the results of the field testing and when the questions are used as part of the health survey.

Reliability

Reliability of a question is the consistency and stability of the question from one survey to another. When the question is repeated and gives identical or very similar results, then it is said to be reliable [102].

Test-retest and internal consistency (see Attachment B for definitions) are the key reliability measures in question development. Questions with low reliability are ones in which the answers respondents give vary widely as a function of when the questions are asked, who asks them and the fact that the particular questions chosen from a set of items seem to be asking the same thing, but are not [1].

Test-Retest Reliability:

- As part of the field testing process the same questions need to be asked of the same respondent at different points in time;
- For a sub-sample of the survey respondents (in the order of 5%) as part of the evaluation of the data collection; and
- Computation of reliability coefficients (Pearson correlation coefficient for interval-level data, Spearman rank order coefficient for ordinal-level variables and Kappa coefficient for nominal variables) to examine the association between the answers at two different points in time.

It should be noted that there are two important issues that impact on test-retest results, namely answering at one occasion can influence measurements on subsequent occasions and many responses change all the time. Both of these phenomena may occur in the test-retest period, thus lowering the estimate of reliability [37].

Internal Consistency Reliability:

- As part of the evaluation of the field testing process, the inter-correlation among a number of different questions that are supposed to reflect the same concept needs to be computed and reported;
- As part of the analysis of the data collection evaluation of the inter-correlation among a number of different questions that reflect the same concept needs to be computed and reported; and
- Internal consistency (corrected item-total correlation, split-half coefficients, Cronbach's alpha coefficients) needs to be computed and reported.

The aim of CASIC question development is to standardise the delivery of the question by the interviewer to the respondent. Therefore, inter-rater reliability does not have a significant impact on CASIC health survey questions, particularly those that are developed along the systematic lines outlined in this document.

As an example of the reporting and value of measuring question reliability, Starr et al. (1999) report on the re-interviewing of 4% of respondents in a CATI health survey of mental health in South Australia in 1997. They found moderate to almost perfect reliability for health risk factor questions and the highest reliability for demographic variables. Table 1 below presents selected reliability estimations on demographic, health risk factors and co-morbidity questions [91].

In Table 1 it can be seen that age and sex were reported with perfect reliability, while the BMI (body mass index), calculated from reported weight (in kilograms) divided by the square of reported height (in metres), has a high level of reliability, along with smoking status. These variables could be expected to remain with little change between the two interviews. Reported exercise showed moderate reliability, however the questions involved could be expected to have high variation in response between the two interviews given the high variability in reported exercise patterns over time and in their study the re-interviews were conducted over a month after the initial interview.

Table 1: Selected reliability estimates on demographic, health risk factors and co-morbidity questions

Variables	n	Response at first interview (%)	Response at second interview (%)	Reliability value ^a	95% CI
Age	102			1.00 ^b	-
18-34 years		20.6	20.6		
35-54 years		27.5	27.5		
≥55 years		52.0	52.0		
Gender	102			1.00 ^b	-
Male		34.3	34.3		
Female		65.7	65.7		
BMI (calculated)	97			0.89 ^b	0.84-0.92
up to 20		10.9	11.3		
>20-25		45.5	43.3		
>25-30		29.7	29.9		
>30		13.9	15.5		
Exercise					
walked for fitness	102	68.6	58.8	0.54 ^c	0.37-0.70
increase in heart rate	102	52.0	46.1	0.45 ^c	0.28-0.62
vigorous	102	22.6	20.6	0.48 ^c	0.27-0.69
Smoking status	102			0.92 ^d	0.86-0.98
non-smoker		52.0	50.0		
ex-smoker		25.5	30.4		
smoker		22.5	19.6		
Told by doctor have ...					
Diabetes	102	5.9	5.9	1.00 ^c	-
Arthritis	102	28.4	25.5	0.73 ^c	0.58-0.88
Heart attack/angina	102	6.9	6.9	0.85 ^c	0.64-1.00
Stroke	102	4.9	5.9	0.71 ^c	0.40-1.00
Cancer	102	2.9	2.9	0.31 ^c	-0.18-0.81

Notes: (a) Reliability measures
 (b) Intra-class correlation coefficient
 I Kappa
 (d) Weighted Kappa

Source: Starr et al. (1999, Table 2, page 530)

Content Validity

- are all relevant concepts represented in the questions?

Criterion Validity

- how do the questions compare to a “true” value or “gold standard”?

Construct Validity

- how do the questions match to other related measures?

Reliability

- are responses to the questions consistent and stable from one survey to another?

Field Testing

Testing the questions in the field is a vital stage in the development and assessment of survey questions, along with the development and testing of associated procedures and instructions [75]. In the past, field testing has been criticised as being handled too casually with procedures relying on interviewers' reporting on whether the question 'worked'. These reports are often based on a few interviews by a small number of interviewers who often disagree on which question 'worked' [27, 96].

The qualitative testing outlined earlier is predominantly conducted in laboratory settings with respondents who are self-selected by volunteering to participate and with interviewers who are often survey researchers. It can be anticipated that there will be important differences in these settings that will have major effects on the development of survey questions [106].

No matter how much development work has been done, the question (or set of questions) has to be tested under field conditions. In particular, the questions have to be tested in the mode they will be used; specifically CATI/CAPI questions require field testing in a CATI/CAPI survey environment.

Field testing can be interpreted to achieve different things to different people. It can cover an evaluation of procedures by collection from a small sample of respondents, the identification of any production mistakes and learning whether respondents understand the questions [32, 73]. Observations of field testing can provide valuable insight into the effectiveness of changes to questions during the development process.

The ABS User's Guide to Sample Surveys provides direction to the background for survey testing, stating that testing is used to:

- a. assess the suitability of the chosen sampling methodology;
- b. estimate sampling error and variability in the target population;
- c. estimate likely response rates;
- d. identify weaknesses in sample framework, questionnaire design, and the method of data collection;
- e. assess field work procedures, and processing procedures; and
- f. estimate costs [7].

Within the context of the testing of questions, points d and e are the most important, although there may be aspects of the other points that need to be kept in mind. The guide also highlights that there are three main types of testing – skirmishing, pilot testing and dress rehearsals.

Skirmishing is the process of informally testing the questions and can be a part of the discussions with key informants and/or a reference group. This can also be part of an iterative process with the cognitive testing procedures. A skirmish of a questionnaire provides a valuable process to eliminate a range of more obvious errors or incorrect assumptions that may have been overlooked in the initial design and phrasing of questions. Skirmishing also results in the efficient use of

later testing, which can concentrate on the refinement of questions rather than being distracted by more obvious errors that surface through skirmishing. However, skirmishing does not provide information on the time to ask questions, nor does it identify problems from non-expert respondents.

Pilot testing involves formally testing the questions with a small sample of respondents in the same way that the final survey will be conducted. Pilot testing is used to:

- a. identify problems with the questions;
- b. provide validation, reliability and usability assessment of the questions;
- c. assess the adequacy of instructions to interviewers and supervisors; and
- d. ascertain the time required to ask the questions.

Pilot testing should also be directed to as wide a population as possible and with a range of interviewers. For criterion validity (see Attachment B for definition), if possible the pilot test should incorporate a “true” value for the measure(s) being gathered. For example, if CAPI questions on immunisation are being tested, then as part of the field test, and at the conclusion of the interview, the interviewer would seek to examine immunisation records.

As part of the pilot testing, debriefing of the interviewers is required to identify possible problems and anomalies with the questions [27, 34, 101]. This can incorporate self-administered debriefing questionnaires and participation in focus group discussions with other interviewers. As part of interviewer debriefing, the following questions need to be addressed by the interviewers about each question:

1. Did you have any difficulty reading the question exactly as worded?
2. Does the question contain words or concepts that respondents do not understand?
3. Do respondents have difficulty retrieving information or providing an answer to the question?

From the pilot testing, it is also possible to identify whether the questions are generating high item non-response, as well as providing estimates of the likely overall survey response rate [32].

Debriefing of respondents with a post interview can also include vignettes and follow-up questions to help understand why certain questions may be posing conceptual problems to respondents [34]. Respondent debriefing can also be used to evaluate other tasks which are required in the interview, such as their understanding of the questions and the records or recall that they used to answer the questions [101].

A **dress rehearsal** is a trial run of a survey where the chosen sampling methodology is used to select a small sample from the target population. Dress rehearsals are used to conduct a final test of the questions before moving to the final survey. It is in the dress rehearsal that a final check of the procedures and instructions can be tested and evaluated. This is important to ensure that there are no problems with the preparation of the CASIC software programming and question sequencing.

The pretesting policy of the US Census Bureau highlights that [101]:

“Changes in survey procedures, including the questions asked, may affect the continuity of time series data. Seemingly minor changes in question wording or sequence sometimes can affect survey responses in important and unexpected ways. When a time series measure may be affected by a questionnaire or procedural change, the Census Bureau recommends that an experimental field test be conducted to evaluate the effects of the proposed changes on survey estimates.”

It is therefore important, particularly for the collection of health surveillance data where changes over time are a key component, that as part of the evaluation of questions, there is measurement of the impact of changes to existing questions. If existing questions are to be replaced, then there is a need to provide estimates of how the replacement questions will affect the ongoing measures.

In conclusion, in the development of the question (or set of questions) it is essential to include skirmishing and at least one pilot test in field testing. In addition, the pilot test should incorporate a test-retest component to establish the question reliability. Where substantial revisions are required as a result of the pilot test then a subsequent pilot test should be conducted. When the question (or set of questions) is incorporated into a final survey, then a dress rehearsal should be conducted, even at the expense of reducing the sample size in the resulting survey.

Usability Testing

Usability is the concept of putting the question into the survey environment away from the artificial setting of question development, and includes issues such as the ability of the interviewer to ask the questions as intended and the ability of the respondent to answer the questions. Through usability testing other work practices can be considered, such as in CAPI the interviewer's ability to physically hold the computer and enter the responses in a poorly lit doorway.

There is a need to evaluate the 'usability' of the CASIC instrument to ascertain how easy or difficult it is for both interviewers and respondents to interact and to determine the accuracy of responses. Usability testing focuses on the interviewer's interaction with the CASIC system and survey instrument, shifting the focus from system feasibility and functionality to design of instruments from the user's perspective, and increases the importance of usability testing [45].

A number of additional design issues are introduced with CASIC procedures, particularly the many ways in which the technology used affects the interviewers and respondents. The main focus of usability testing is directed toward the CATI and CAPI software. However, there are major question design issues, particularly related to the interviewer/respondent interaction that can be resolved through usability testing of questions. In the CASIC environment the interview flow is segmented, where the interviewer reads the question, gets an answer, enters the response, presses [enter] and moves to the next question, mostly on a new screen. With this segmented interaction it is sometimes difficult for interviewers to maintain consistent performance [39]. It should be noted that this is unlikely to be a major problem with comprehensive interviewer training and with familiarisation of both the computer system and survey itself.

The US Census Bureau has found that usability testing identified new problems associated with the switch from paper-based to computer-based instruments [64, 71]. Usability testing of questions can be achieved by the behaviour coding of respondent/interviewer interactions or by the use of specific usability laboratory testing where the respondent/interviewer interaction is video-taped and reviewed.

Behaviour coding involves coding how the interviewer asked the question and how the respondent reacted. With the video-taping of interviews becoming a popular medium for question evaluation and interviewer training, video-taping also provides valuable insight into the usability (or delivery) of the questions [109]. Video-taped usability tests serve as a visual record of the kinds of difficulties experienced by interviewers and respondents. They also provide supporting evidence that is sometimes necessary to convince the survey sponsors, researchers and programmers that changes to the questions and instructions are necessary [44].

Usability problems arise, in part, because the design of the user interface imposes both cognitive and physical demands that interviewers and respondents are sometimes unable or unwilling to meet [71]. With the use of usability testing

of CASIC questions, increased confidence can be gained that the delivery and reception of the questions will be in accordance with the original aims of the questions. In addition, question problems that are identified through usability testing can provide valuable input into what definitions, items and summaries are required in interviewer 'help screens' [48].

Reporting

The wide dissemination of the details of the question development process is required to ensure that extensive value can be gained from the resources used and advances that have been achieved. Not only does this support the input from the key informants but it facilitates the wide use of comparable questions across a number of surveys.

An example of the publication of the question development process is the ABS report into the development of the ethnicity question for the population census. This report presents documentation of the concept to be measured, requirements of the topic, detailed results of the five pilot tests that were conducted including the variations in wording that were tried, other issues including the particular response to the concept and questions by the indigenous population. In addition, the report included a review of the literature and a listing of the persons and organisations consulted in the process [5].

It could be expected that a report covering similar material, including reporting on the qualitative testing, validity and reliability findings, should be publicly available. In addition, the combining of sets of questions into a comprehensive manual with support instructions and explanations for interviewers should be prepared and widely circulated.

There would also be considerable value in including the developed questions in the National Health Data Dictionary (NHDD) [9]. Although the evaluation processes required for inclusion in the NHDD are additional to the stages presented in this overview of question development principles and practices, there would be even broader utilisation of standard (comparable) questions, procedures and instructions across data collections once they were an integral part of the NHDD. This would assist in the comparison of findings from different survey collections.

Interviewers and Supervisors

One of the strengths of CATI is the centralised supervision and support for interviewers [16]. With regard to the development of questions there is also a need to provide interviewers and supervisors with additional material related to the questions [109].

The training of interviewers and supervisors is an area of survey quality that has received scant attention compared to other survey areas, such as sampling errors. It would appear to be counter-productive to strive for as many interviews as possible in an endeavour to minimise sampling error, if consistent interviewer error was not being addressed.

When there are a small number of interviewers, each having a large workload then poor performance from a single interviewer can have a major effect on the results of the collection. Improvements in ongoing supervision is particularly important to monitor each interviewer [16].

So that respondents can clearly and consistently understand questions, the thorough development and wording of survey questions is of major importance for successful standardisation of interviewing [11, 23]. With a greater emphasis on question development there is generally a reduction in the amount of probing required by interviewers. However, there is a need for interviewer training to cover the concepts that are inherent in the questions so that interviewers can assess the completeness of respondents' answers and provide "non-directive probing that is based on a sound understanding of the question's objectives rather than the interviewers' own speculation of what constitutes a complete answer" [11].

When examining the quality of responses to questions used across a number of surveys, Cannell & Oksenberg (1988) reported that in the worse situation nearly a quarter of interviewers made major wording changes and over half used unacceptable probes [20]. It is therefore necessary to support the question (or set of questions) with additional material that can be included in interviewer training. In addition, interviewers with less familiarity with computing technology require additional training to raise their confidence and precision in the CASIC environment [80].

Resources and techniques for interviewer training include:

1. written resource materials;
2. lectures and demonstrations;
3. home study and written exercises;
4. role playing and practice interviews; and
5. coding of interviewer behaviour [1].

Interviewers must be trained to distinguish between “deviation” (when they respond to a query with an equivalent meaning to the original question) from “error” (when they respond with a different meaning to the original question) [1]. For example:

Original: “What is the highest grade or year of school you have completed?”

Deviation: “How many years of school have you finished?”

Error: “How many years were you in school?”

It is important that questions are supported with instructions for interviewers, both as a part of the CASIC software and as part of interviewer training. Interviewers need material, such as instructions and definitions, that they can draw upon to correctly respond to respondent queries or misunderstandings [43, 84]. For more complicated questions both scripted and para-phrased additional wording clarification increases the accuracy of responses, but also increases the length of interviews [85].

The requirement to develop uniform questions that can be utilised across many surveys needs to be mindful of the different positions that are held by the interviewer and the respondent. Will all interviewers perceive and deliver the question in a similar fashion? Will the majority of respondents perceive what they are being asked in a similar way?

The aim of an ideal CASIC health survey is to establish a partnership between the interviewer and the respondent and not to see those being interviewed as a homogeneous mass or as objects to be manipulated. If the question (or set of questions) fails with key groups, these may well be the groups that require the greatest policy attention and/or experience the greatest barriers to accessing health services.

It is also beneficial to provide interviewers with a thorough orientation of the question (or set of questions) as well as seeking their involvement in the survey development process. “Interviewers jump at the opportunity to help refine a questionnaire” [70]. Similarly, supervisors need support material in their training and at hand to respond to queries on the questions from interviewers and respondents. This is also important, as supervisors are generally involved in training the interviewing staff.

Although interviewer training is generally accepted as being essential either when the interviewer is first employed or at the introduction of a new data collection, there can be substantial value in scheduling regular training. As part of this training, interviewer debriefing can be incorporated to monitor the progress of survey questions and highlight difficulties both the interviewers and their respondents might be experiencing [59].

The ability to monitor interviews is one of the major features of CATI surveys, aiding in the improvement of costs, productivity and efficiency [25]. A large number of CATI systems provide the facility to have online monitoring of interviews with the system able to reproduce any interviewer’s screen at the supervisor’s terminal where audio monitoring can also be linked [74]. Under

such settings, not only can the interviewer be monitored but also the performance of the question, in particular how well it is delivered and received by respondents, can be assessed. “When appropriately used, (monitoring) permits reinforcement of training guidelines, presumably leading to a reduction in interviewer-induced error” [25]. However, this does require a CATI facility with a centralised bank of telephones where the supervisor’s telephone is able to monitor all interviewers’ lines [60].

Additional Issues

Although some of the additional issues presented below do not relate directly to the development of questions, the issues do impact on the conduct and delivery of questions in the data collections.

Ethical and Related Issues

Ethical issues are fundamental in social research and are not just an add-on that is only addressed in the planning stages of a project. The design of questions needs to assume that participation is voluntary and that informed consent safeguards the rights of respondents. There have been concerns that a 'hit and run' model of research ignores the relationships which might be built between researcher and researched in discussing sensitive or shared concerns [49]. Accordingly, if there are important ethical issues associated with a question (or set of questions), appropriate additional material needs to be prepared and included in interviewers' instructions.

Voluntary informed consent implies that as part of the introduction to the survey the respondent is advised of the questions to be asked and given opportunities to ask for more information about the questions [88]. To assist in responding to these inquiries interviewers need relevant information both in training and readily at hand.

Concerns are being expressed regarding the intrusion of information collection, particularly via telephone. These relate mainly to the unethical practices of market research and tele-marketing, such as 'push-polls' (political propaganda disguised as legitimate polling, but using biased question wording solely to expose respondents to a highly partisan viewpoint), 'FRUGing' (fund-raising under the guise of surveying), and 'SUGing' (selling under the guise of surveying). However, CATI health surveys could fall under legislation that limit future interviewing, with respondents being contacted when they have expressed their consent and are not contacted without their consent. Such a development would lead CASIC health surveys into the area of survey 'panels' [60].

The establishment of panels opens up a broader range of collection issues and biases, although the impact on question development may not be significant. Panels are being established by market research and are being used to conduct an extensive range of population and product surveillance. The largest of these panels is the Harris Interactive (see www.gsbc.com), which claims to currently have over seven million respondents who have joined up to their online panel.

The interviewing of children and adolescents raises a number of issues such as consent of their parents/guardians plus their own assent to participate. In addition, a number of key factors relevant to interviewing children include:

1. they have limited psychological, as well as legal, capacity to give informed consent;
2. they may be cognitively, socially, and emotionally immature;

3. there are external constraints on their self-determination and independent decision making;
4. they have unequal power in relation to authorities, such as parents, teachers, and researchers;
5. their parents and certain institutions, as well as the youngsters themselves, have an interest in their participation; and
6. national priorities for research on children and adolescents include research on drug use, the problems of runaways, pregnancy among teenagers, and other sensitive topics, compounding the ethical and legal surrounding research on minors [88].

However, it would seem patronising to exclude children and adolescents from answering a range of questions on their own behalf, provided the questions have been appropriately designed and consent is obtained (as discussed above).

Core Questions

One issue associated with the development of sets of questions for a topic is the determination of which questions should be considered as core questions as part of a core component of an ongoing series of surveys and which questions can be considered as expanded/additional questions. To a major extent this depends on the priority given to the topic and the requirement to monitor this topic. This determination would be made by the body responsible for the survey following consultation with appropriate reference groups.

As an example, in the BRFSS diabetes is included as one of their 12 core sections. The question asks “Have you ever been told by a doctor that you have diabetes?”

In the BRFSS optional module on diabetes, a series of questions related to the management of their diabetes plus the age of onset of diabetes are included. In Australia, it would not be surprising if the same distinction would be made. Diabetes is one of the national health priority areas and the monitoring of the prevalence of diabetes would be incorporated in CASIC health surveys that are measuring chronic disease. However, questions on further aspects of diabetes would be included occasionally or when these additional questions match the focus of the survey.

Survey Coverage

Coverage of CASIC surveys is a major concern, particularly when the characteristics of those not covered by the population being studied can be expected to be substantially different from those covered by the survey methodology. In particular, it has been shown that telephone ownership remains lower in households with young adult males, households with a non-English speaking background and those with low household income who live in public housing [4, 8, 92].

There are also key groups who are generally excluded from CASIC health surveys, namely those who are:

- homeless;
- live in sparsely settled areas;
- in the armed forces,
- in long stay hospital, nursing homes, hostels;
- in prison/detention; and
- in other long-term communal accommodation.

It could be expected that for most of these groups, and for particular health issues, there could be substantial and important differences. For example, for questions related to sexually transmitted diseases or hepatitis B the majority of the homeless and those in prison would have different and important experiences to the rest of the population. However, neither group would be included in CASIC health surveys although this situation may partly change with the Australian Institute of Health and Welfare and some State health departments investigating prisoner health.

Survey Introduction

Although there has been little research conducted on the nature and extent of the survey introduction in the CASIC environment, Meegama & Blair (1999) observe that “the goal of an effective introduction appears to be not only to provide minimum information, but the right kind of information – eliciting cooperation of the potential respondent and ... enhancing rapport during the interview” [67]. They recommend the tailoring of the introduction to gain cooperation by following a verbatim brief introduction with additional information that might be used, depending on the respondent’s level of cooperation. The additional information covers:

- I’m not selling anything/this is not a sales call
- Length of survey
- Confidentiality/legitimacy
- Importance of study
- Topics in survey
- Sponsors name

A similar strategy is advocated by Snijkers et al. (1999) where they found that successful interviewers (with regard to response rates) tailored their introductions to suit the respondent and, in the case of personal interviews, the situation at hand. In their study in the Netherlands they found that amongst the most important tactics to obtain participation, interviewers need to know what they are talking about [90]. Accordingly, this supports the need for the inclusion of additional background information on the question (or set of questions) for both interviewer training and as part of the CASIC help screens and definitions that are readily accessible at the beginning of the interview.

Order Effect

There are major differences between the different modes of data collection, particularly in relation to the cognition and communicative processes assumed to

underlie the process of question answering [87]. The mode of data collection affects the:

- emergence of question order and context effects;
- emergence of response order effects;
- validity of retrospective reports; and
- degree of socially desirable responding.

Care is required to ensure that the question (or set of questions) is placed in a logical position in the data collection and the order of response categories in a question follows a logical sequence. The questions (or set of questions) are part of a context with other questions, as well as part of a continuous flow of questions [38]. Accordingly, the context in which the question (or set of questions) is placed can affect an individual's response, with the cognitive cues required to respond to prior questions influencing answers to subsequent questions [98]. In addition, there is also the response order effect where the order of response categories presented for a question influences the selection made by the respondent [52].

One particular concern arises when expanded questions are included in a survey, and what their subsequent impact might be on responses to later questions. It will require care from those developing the survey to balance what might be considered to be a logical sequencing of questions with an untoward influence on questions to follow. If doubt remains, the expanded questions should be included after core questions, with appropriate statements to explain to the respondent that some additional questions are being asked related to earlier questions.

Re-interviewing and Monitoring

As part of an ongoing assessment of the questions, there should be a component of re-interviews as part of the final survey. The BRFSS incorporates a 5% re-interview of the completed interviews (see <http://www.cdc.gov/nccdphp/brfss>). As mentioned earlier this is required for the establishment of reliability of the questions. It could be expected that some questions might need to be further developed and the incorporation of a re-interview component would provide an early warning whether the questions were in need of further development, for example with a particular population.

In addition to randomly selected re-interviews, in 1997 the US Census Bureau introduced a 'focused re-interview' process as part of the National Health Interview Survey to specifically identify those interviewers who falsify data in the survey. The focused re-interview is selected for those interviewers where patterns of unusual answers for several different variables can be identified [46].

In large or ongoing surveys, question development and testing should be seen as a continuous and integrated process rather than as a set of independent activities [106]. For example, the use of re-interviews to monitor the reliability of questions should be incorporated as an integral component of each round of data collection.

References

- 1 Aday LA. Designing and conducting health surveys. 2nd Edition. San Francisco: Jossey-Bass, 1996.
- 2 Akkerboom H, Dehue F. The Dutch model of data collection development for official surveys. *International Journal of Public Opinion Research* 1997; 9: 126-145.
- 3 Akkerboom H, Luiten A. Selecting pretesting tools according to a model of questionnaire development, with illustrations concerning patient satisfaction with medical care. *Proceedings of the Section on Survey Research Methods, American Statistical Association* 1996; 911-916.
- 4 Allister J, Brennan D, Davies M, Hoskin G, Roberts-Thomson K, Spencer J, Slade G, Stewart J, Szuster F. A research database on dental care in Australia. Adelaide: The University of Adelaide, 1993.
- 5 Australian Bureau of Statistics. The measurement of ethnicity in the Australian census of population and housing. Catalogue No. 2172.0. Canberra: ABS, 1984.
- 6 Australian Bureau of Statistics. 1995 national health survey data reference package. Canberra: ABS, 1996.
- 7 Australian Bureau of Statistics. An introduction to sample surveys: a user's guide. Catalogue No. 1299.0. Canberra: ABS, 1999.
- 8 Australian Bureau of Statistics. Population survey monitor. Unpublished tables, 2000.
- 9 Australian Institute of Health and Welfare. National Health Data Dictionary. Version 9. Catalogue No. HWI 24. Canberra: AIHW, 2000.
- 10 Baker T. (ed.) Key performance indicators manual. South Melbourne: Pearson Professional, 1995.
- 11 Beatty P. Understanding the standardized/non-standardized interviewing controversy. *Journal of Official Statistics* 1995; 11: 147-160.
- 12 Beatty P, Schechter S. An examination of mode effects in cognitive laboratory research. *Proceedings of the Section on Survey Research Methods, American Statistical Association* 1994; 1275-1280.
- 13 Beatty P, Schechter S, Whitaker K. Evaluating subjective health questions: cognitive and methodological investigations. *Proceedings of the Section on Survey Research Methods, American Statistical Association* 1996; 956-961.
- 14 Beatty P, Schechter S, Whitaker K. Variation in cognitive interviewer behavior – extent and consequences. *Proceedings of the Section on Survey Research Methods, American Statistical Association* 1997; 1064-1068.

- 15 Belson WA. The design and understanding of survey questions. London: Gower, 1981.
- 16 Bennett DJ, Steel D. An evaluation of a large-scale CATI household survey using random digit dialing. *Australian and New Zealand Journal of Statistics* 2000; 42: 255-270.
- 17 Bickman L, Rog DJ. (eds.) *Handbook of Applied Social Research Methods*. Thousand Oaks: Sage, 1998
- 18 Biemer PP. Nonresponse bias and measurement bias in a comparison of face to face and telephone interviewing. *Journal of Official Statistics* 2001; 17: 295-320.
- 19 Bischooping K, Dykema J. Toward a social psychological programme for improving focus group methods of developing questionnaires. *Journal of Official Statistics* 1999; 15: 495-516.
- 20 Cannell C, Oksenberg L. Observation of behavior in telephone interviews. In: Groves RM, Biemer PP, Lyberg LE, Massey JT, Nicholls WL II, Waksberg J. (eds.) *Telephone Survey Methodology*. NY: John Wiley & Sons, 1988.
- 21 Computer Assisted Telephone Interview Technical Reference Group. Report of the Module Development Project. National Public Health Partnership, 2000.
- 22 Conrad F, Blair J. From impressions to data: increasing the objectivity of cognitive interviews. *Proceedings of the Section on Survey Research Methods, American Statistical Association* 1996; 1-9.
- 23 Conrad FG, Schober MF. Clarifying question meaning in a household telephone survey. *Public Opinion Q* 2000; 64: 1-28.
- 24 Couper MP, Baker RP, Bethlehem J, Clark CZF, Martin J, Nicholls WL II, O'Reilly JM. (eds.) *Computer Assisted Survey Information Collection*. NY: John Wiley & Sons, 1998.
- 25 Couper MP, Holland L, Groves RM. Developing systematic procedures for monitoring in a centralized telephone facility. *Journal of Official Statistics* 1992; 8: 63-76.
- 26 Couper MP, Nicholls WL II. The history and development of computer assisted survey information collection methods. In: Couper MP, Baker RP, Bethlehem J, Clark CZF, Martin J, Nicholls WL II, O'Reilly JM. (eds.) *Computer Assisted Survey Information Collection*. NY: John Wiley & Sons, 1998.
- 27 Czaja R. Questionnaire pretesting comes of age. *Marketing Bulletin* 1998; 9: 52-66.
- 28 DeLeeuw ED. The effect of computer-assisted interviewing on data quality: a review of the evidence. Cologne: 5th Social Science Methodology Conference, 2000.

- 29 DeMaio TJ, Rothgeb J, Hess J. Improving survey quality through pretesting. *Proceedings of the Section on Survey Research Methods, American Statistical Association* 1998; 50-58.
- 30 Deming WE. On errors in surveys. *American Sociological Review* 1944; 9: 359-369.
- 31 Desvousges WH, Frey JH. Integrating focus groups and surveys: examples from environmental risk studies. *Journal of Official Statistics* 1989; 5: 349-363.
- 32 Dillman DA. *Mail and Internet Surveys. The Tailored Design Method*. 2nd Edition. NY: John Wiley & Sons, 2000.
- 33 Eisenhower DL. Design-oriented focus groups and cognitive laboratories: a comparison. *Proceedings of the Section on Survey Research Methods, American Statistical Association* 1994; 1374-9.
- 34 Esposito JL, Campanelli PC, Rothgeb J, Polivka AE. Determining which questions are best: methodologies for evaluating survey questions. *Proceedings of the Section on Survey Research Methods, American Statistical Association* 1991; 46-55.
- 35 Forsyth BH, Hubbard ML. A method for identifying cognitive properties of survey items. *Proceedings of the Section on Survey Research Methods, American Statistical Association* 1992; 470-475.
- 36 Fowler FJ Jr. Design and evaluation of survey questions. In: Bickman L, Rog DJ. (eds.) *Handbook of Applied Social Research Methods*. Thousand Oaks: Sage, 1998: 343-374.
- 37 Frankfort-Nachmias C, Nachmias D. *Research methods in the social sciences*. 4th Edition. London: St Martin's press, 1992.
- 38 Frey JH. *Survey research by telephone*. 2nd Edition. Newbury Park: Sage, 1989.
- 39 Fuchs M. Screen design and question order in a CAI instrument effects on interviewers and respondents. *Proceedings of the Section on Survey Research Methods, American Statistical Association* 1999; 1013-1018.
- 40 Gerber ER. Hidden assumptions: the use of vignettes in cognitive interviewing. *Proceedings of the Section on Survey Research Methods, American Statistical Association* 1994; 1269-1274.
- 41 Gower AR, Bélanger B, Williams M-J. Using focus groups with respondents and interviewers to evaluate the questionnaire and interviewing procedures after the survey has taken place. *Proceedings of the Section on Survey Research Methods, American Statistical Association* 1998; 404-409.
- 42 Greenbaum TL. *The Handbook for Focus Group Research*. 2nd Edition. Thousand Oaks: Sage Publications, 1998.

- 43 Hansen SE. The impact of design on interaction in the computer assisted survey interview. Proceedings of the Section on Survey Research Methods, American Statistical Association 1999; 1019-1024.
- 44 Hansen SE, Couper MP, Fuchs M. Usability evaluation of the NHIS CAPI instrument. Proceedings of the Section on Survey Research Methods, American Statistical Association 1998; 928-933.
- 45 Hansen SE, Fuchs M, Couper MP. CAI instrument usability testing. Proceedings of the Section on Survey Research Methods, American Statistical Association 1997; 1023-1028.
- 46 Hilton JS, Bushery JM. Automating the focused reinterview. Proceedings of the Section on Survey Research Methods, American Statistical Association 1998; 421-424.
- 47 House CC. Questionnaire design with computer assisted telephone interviewing. Journal of Official Statistics 1985; 1: 209-219.
- 48 House CC, Nicholls WL II. Questionnaire design for CATI: design objectives and methods. In: Groves RM, Biemer PP, Lyberg LE, Massey JT, Nicholls WL II, Waksberg J. (eds.) Telephone Survey Methodology. NY: John Wiley & Sons, 1988.
- 49 Humphries B. Unsettling ethics in social research. 5th Social Science Methodology Conference, Cologne, 2000.
- 50 Jabine TB. Flow charts: a tool for developing and understanding survey questionnaires. Journal of Official Statistics 1985; 1: 189-207.
- 51 Jobe JB, Mingay DJ. Cognition and survey measurement: history and overview. Applied Cognitive Psychology 1991; 5: 175-92.
- 52 Johnson T, O'Rourke D, Severns E. Effects of question context and response order on attitude questions. Proceedings of the Section on Survey Research Methods, American Statistical Association 1998; 857-860.
- 53 Katz I, Stinson L, Conrad F. Questionnaire designers versus instrument authors: bottlenecks in the development of computer-administered questionnaires. Proceedings of the Section on Survey Research Methods, American Statistical Association 1997; 1029-1034.
- 54 Kerr C, Taylor R, Heard G. (eds.) Handbook of Public Health Methods. Sydney: McGraw-Hill, 1998.
- 55 Kinsey SH, Jewell DM. A systematic approach to instrument development in CAI. In: Couper MP, Baker RP, Bethlehem J, Clark CZF, Martin J, Nicholls WL II, O'Reilly JM. (eds.) Computer Assisted Survey Information Collection. NY: John Wiley & Sons, 1998.
- 56 Khurshid A, Sahai H. A bibliography on telephone survey methodology. Journal of Official Statistics 1995; 11: 325-367.

- 57 Krosnick J. Survey research. *Annual Review of Psychology* 1999; 50: 537-567
- 58 Krosnick J. The threat of satisficing in surveys: the shortcuts respondents take in answering questions. *Survey Methods Newsletter*. London: National Centre for Social Research 2000; 20 (1): 4-8.
- 59 Lavrakas PJ. Telephone survey methods. Sampling, selection, and supervision. Newbury Park: Sage, 1987.
- 60 Lavrakas PJ. Methods for sampling and interviewing in telephone surveys. In: Bickman L, Rog DJ. (eds.) *Handbook of Applied Social Research Methods*. Thousand Oaks: Sage, 1998: 429-472.
- 61 Lepkowski JM, Couper MP, Hansen SE, Landers W, McGonagle KA, Schlegel J. CAPI instrument evaluation: behavior coding, trace files, and usability methods. *Proceedings of the Section on Survey Research Methods, American Statistical Association* 1998; 917-922.
- 62 Lessler JT, Kalsbeek WD. *Nonsampling Error in Surveys*. NY: John Wiley & Sons, 1992.
- 63 Lindstrom K, Dallner M, Elo A-L, Gamberale F, Knardahl S, Skogstad A, Orhede E (eds.) *Review of psychological and social factors at work and suggestions for the general Nordic questionnaire*. Copenhagen: Nordic Council of Ministers, 1997.
- 64 Marquis K, Nichols E, Tedesco H. Human-computer interface usability in a survey organization: getting started at the Census Bureau. *Proceedings of the Section on Survey Research Methods, American Statistical Association* 1998; 416-420.
- 65 Martin E, Polivka AE. Diagnostics for redesigning survey questionnaires. *Measuring work in the current population survey*. *Public Opinion Quarterly* 1995; 59: 547-567.
- 66 McDowell I, Newell C. *Measuring Health. A Guide to Rating Scales and Questionnaires*. 2nd Edition. NY: Oxford University Press, 1996
- 67 Meegama N, Blair J. The effects of telephone introductions on cooperation: an experimental comparison. *Proceedings of the Section on Survey Research Methods, American Statistical Association* 1999; 394-397.
- 68 Morgan DL. Focus groups. *Annual Review of Sociology* 1996; 22: 129-152.
- 69 Morgan DL, Krueger RA. *The Focus Group Kit*. Thousand Oaks: Sage Publications, 1998.
- 70 Mullin PA. Feasibility of using CAPI/CATI interviewers to conduct cognitive interviews during survey field tests. *Proceedings of the Section on Survey Research Methods, American Statistical Association* 1999; 38-43.

- 71 Murphy E, Marquis K, Hoffman R, Saner L, Tedesco H, Harris C, Roske-Hofstrand R. Improving electronic data collection and dissemination through usability testing. Paper SM00/01, US Bureau of the Census, 2000.
- 72 National Public Health Partnership. Performance indicator frameworks for population health. Report on workshop held on 16 March 2000. Melbourne: NPHP, 2000.
- 73 Nelson DD. Informal testing as a means of questionnaire development. *Journal of Official Statistics* 1985; 1: 179-188.
- 74 Nicholls WL II. Computer-assisted telephone interviewing: a general introduction. In: Groves RM, Biemer PP, Lyberg LE, Massey JT, Nicholls WL II, Waksberg J. (eds.) *Telephone Survey Methodology*. NY: John Wiley & Sons, 1988.
- 75 Oksenberg L, Cannell C, Kalton G. New strategies for pretesting survey questions. *Journal of Official Statistics* 1991; 7: 349-365.
- 76 Oppenheim AN. *Questionnaire Design and Attitude Measurement*. London, Heinemann, 1966.
- 77 Organisation for Economic Co-operation and Development. *The environmental goods and services industry. Manual for data collection and analysis*. Paris: OECD, 1999.
- 78 Payne SL. *The art of asking questions*. Princeton: Princeton University Press, 1951.
- 79 Platek R. Some important issues in questionnaire development. *Journal of Official Statistics* 1985; 1: 119-136.
- 80 Potter DEB, Sperry S, Edwards B, Dulaney R. Can information from and about interviewers predict their CAPI behavior? *Proceedings of the Section on Survey Research Methods, American Statistical Association* 1998; 987-991.
- 81 Rothgeb JM, Loomis LS, Hess JC. Challenges in gaining acceptance of the results of cognitive methodological research. Cologne: 5th Social Science Methodology Conference, 2000.
- 82 Schaeffer NC. A decade of questions. *Journal of Official Statistics* 1995; 11: 79-92.
- 83 Schechter S, Blair J, van de Hey J. Conducting cognitive interviews to test self-administered and telephone surveys: which methods should we use? *Proceedings of the Section on Survey Research Methods, American Statistical Association* 1996; 10-17.
- 84 Schober MF, Conrad FG. Response accuracy when interviewers stray from standardization. *Proceedings of the Section on Survey Research Methods, American Statistical Association* 1998; 940-945.

- 85 Schober MF, Conrad FG, Fricker SS. When and how should survey interviewers clarify question meaning? Proceedings of the Section on Survey Research Methods, American Statistical Association 1999; 985-990.
- 86 Schuman H, Presser S. Questions and Answers in Attitude Surveys. Experiments on Question Form, Wording, and Context. NY: Academic Press, 1981.
- 87 Schwarz N, Strack F, Hippler H-J, Bishop G. The impact of administration mode on response effects in survey measurement. *Applied Cognitive Psychology* 1991; 5: 193-212.
- 88 Sieber JE. Planning ethically responsible research. In: Bickman L, Rog DJ. (eds.) *Handbook of Applied Social Research Methods*. Thousand Oaks: Sage, 1998: 127-156.
- 89 Sirken M, Herrmann D. Relationships between cognitive psychology and survey research. Proceedings of the Section on Survey Research Methods, American Statistical Association 1996; 245-249.
- 90 Snijkers G, Hox J & DeLeeuw ED. Interviewers' tactics for fighting survey nonresponse. *Journal of Official Statistics* 1999; 15: 185-198.
- 91 Starr GJ, Dal Grande E, Taylor AW, Wilson DH. Reliability of self-reported behavioural health risk factors in a South Australian telephone survey. *Australian and New Zealand Journal of Public Health* 1999; 23: 528-530.
- 92 Steel D, Boal P. Accessibility by telephone in Australia: implications for telephone surveys. *Journal of Official Statistics* 1988; 4: 285-297.
- 93 Steel D, Vella J, Harrington P. Quality issues in telephone surveys: coverage, non-response and quota sampling. *Australian Journal of Statistics* 1996; 38: 15-34.
- 94 Stewart AL, Hays RD, Ware JE Jr. Methods of validating MOS health measures. In: Stewart AL, Ware JE Jr. (eds.) *Measuring functioning and well-being*. Durham: Duke University Press, 1992.
- 95 Stewart DW, Shamdasani PN. Focus group research: exploration and discovery. In: Bickman L, Rog DJ. (eds.) *Handbook of Applied Social Research Methods*. Thousand Oaks: Sage, 1998: 505-526.
- 96 Sykes W, Morton-Williams J. Evaluating survey questions. *Journal of Official Statistics* 1987; 3: 191-207.
- 97 Tanur JM, Fienberg SE. Cognitive aspects of surveys: yesterday, today, and tomorrow. *Journal of Official Statistics* 1992; 8: 5-17.
- 98 Todorov A. Context effects in national health surveys. *Public Opinion Quarterly* 2000; 64: 65-76.
- 99 Trewin D. How do we reduce non-sampling errors? *Journal of Official Statistics* 1987; 3: 343-347.

- 100 Trewin D, Lee G. International comparisons of telephone coverage. In: Groves RM, Biemer PP, Lyberg LE, Massey JT, Nicholls WL II, Waksberg J. (eds.) Telephone Survey Methodology. NY: John Wiley & Sons, 1988.
- 101 United States Census Bureau. Pretesting policy and options: demographic surveys at the Census Bureau. Washington, DC: US Department of Commerce, 1998.
- 102 Vogt WP. Dictionary of Statistical and Methodology. 2nd Edition. Thousand Oaks: Sage, 1999.
- 103 Weeks MF. Computer-assisted survey information collection: a review of CASIC methods and their implications for survey operations. Journal of Official Statistics 1992; 8: 445-465.
- 104 Willis GB, Royston P, Bercini D. The use of verbal report methods in the development and testing of survey questionnaires. Applied Cognitive Psychology 1991; 5: 251-267.
- 105 Willis GB, Schechter S, Whitaker K. A comparison of cognitive interviewing, expert review, and behavior coding: what do they tell us? Proceedings of the Section on Survey Research Methods, American Statistical Association 1999; 28-37.
- 106 Willis GB, Trunzo D, Stussman BJ. The use of novel pretesting methods in the development of survey questionnaires. Proceedings of the Section on Survey Research Methods, American Statistical Association 1992; 824-828.
- 107 Wilson BF, Peterson LS. Using the NCHS Cognitive Lab to help design Cycle VI of the National Survey of Family Growth. Proceedings of the Section on Survey Research Methods, American Statistical Association 1999; 997-1002.
- 108 Wilson DH, Walsh PG, Sanchez L, Davis AC, Taylor AW, Tucker G, Meagher I. The epidemiology of hearing impairment in an Australian adult population. International Journal of Epidemiology 1999; 28: 247-252.
- 109 Wojcik MS, Hunt E. Training field interviewers to use computers: past, present, and future trends. In: Couper MP, Baker RP, Bethlehem J, Clark CZF, Martin J, Nicholls WL II, O'Reilly JM. (eds.) Computer Assisted Survey Information Collection. NY: John Wiley & Sons, 1998.

Abbreviations

ABS	Australian Bureau of Statistics
BMI	Body Mass Index
BRFSS	Behavioral Risk Factor Surveillance Survey
CAPI	Computer Assisted Personal Interviewing
CASIC	Computer Assisted Survey Information Collection
CASS	Centre for Applied Social Survey
CATI	Computer Assisted Telephone Interviewing
CATI TRG	CATI Technical Reference Group
CCHS	Canadian Community Health Survey
CI	Confidence Interval
DALE	Disability Adjusted Life Expectancy
DHAC	Department of Health and Aged Care
FRUGing	Fund-raising under the guise of surveying
NHDD	National Health Data Dictionary
NHS	National Health Survey
NPHIWG	National Public Health Information Working Group
NPHP	National Public Health Partnership
PHIDU	Public Health Information Development Unit
SUGing	Selling under the guise of surveying
UK	United Kingdom
US	United States

Glossary

<i>Bias</i>	Anything in the conduct of a survey that produces systematic error.
<i>Cognitive testing</i>	Interviewing in a laboratory setting to understand how respondents have interpreted questions in relation to the identification of potential problems with perception, sensitivity, comprehension, memory, context, format, vocabulary, reference periods, judgment and response categories. Essentially cognitive testing is conducted in a laboratory where volunteers are interviewed by trained cognitive interviewers. The respondents are asked to give details of their understanding of the question along with the processes they used to derive their answer.
<i>Construct validity</i>	The hypothesized association between the question and a measure of the same concept or a different concept is confirmed with other measures being collected.
<i>Content validity</i>	The subjective evaluation of the researcher that the question will measure the concept/topic.
<i>Criterion validity</i>	The ability of question to predict or agree with the “true” value or “gold standard” for the topic being measured.
<i>Debriefing</i>	Discussions with interviewers and respondents to identify possible problems and anomalies with the question.
<i>Dress rehearsal</i>	A trial run of a survey where the chosen sampling methodology is used to select a small sample from the target population. Dress rehearsals are used to conduct a final test of the questions before moving to the final survey. It enables a final check of the procedures and instructions can be tested and evaluated.
<i>Field testing</i>	Testing the questions in the mode of collection. Particularly, questions are tested in the CASIC environment to assess of survey questions, along with the development and testing of associated procedures and instructions.

<i>Focus groups</i>	In question development, they consist of approximately a dozen people who engage in an intensive discussion on the question topic. The discussion is lead by a moderator and respondents are encouraged to express their opinions on the topic at hand and to discuss the underlying assumptions related to the question.
<i>Internal consistency</i>	The inter-correlation among a number of different questions that are supposed to reflect the same concept.
<i>Key informants</i>	Individuals who have demonstrated extensive knowledge in the topic, such as researchers, academics, policy makers and representatives of special interest groups. They also include expert subject matter groups.
<i>Non-response</i>	The part of a survey, those individuals who do not respond to the interview. Usually referred to as a percentage of the total selected sample.
<i>On-line range</i>	The recognition of valid responses during interviews. These can result from answers that are inconsistent with earlier responses (such as the receipt of an aged pension when the respondent is below the pension age) or from incorrect entry into the computer by the interviewer.
<i>Pilot test</i>	Formally testing the questions with a small sample of respondents in the same way that the final survey will be conducted. Pilot testing is usually directed to as wide a population as possible and with a range of interviewers.
<i>Pretest</i>	All aspects of question testing including both the qualitative and quantitative development and testing of questions.
<i>Probes</i>	Follow-up questions or inquiries to expand on an answer provided in an interview.
<i>Recall bias</i>	Consistent under- or over-reporting of items, such as the level of alcohol consumption or the frequency of health preventive behaviour.
<i>Reliability</i>	The consistency and stability of the question from one survey to another. When the question is repeated and gives identical or very similar results, then it is said to be reliable.

<i>Routing</i>	The sequencing of a respondent through an interview.
<i>Satisficing</i>	The provision of minimal acceptable answers, rather than optimal answers, in an interview.
<i>Sensitivity</i>	In criterion validity, the proportion that the question picks up with the criterion.
<i>Skirmish</i>	The process of informally testing questions. It can be a part of discussions with key informants and/or a reference group.
<i>Specificity</i>	In criterion validity, the proportion that the question does not report the value that doesn't have the criterion.
<i>Surveillance</i>	The measurement of the question over time.
<i>Test-retest reliability</i>	The ability of the same question to yield consistent results at different times. Test-retest reliability can be computed using correlation coefficients between answers a respondent provides to the same question at two different points in time. The closer the value of the coefficient is to +1.00, the more stable or consistent the question
<i>Usability testing</i>	Putting the question into the survey environment away from the artificial setting of question development, and includes issues such as the ability of the interviewer to ask the questions as intended and the ability of the respondent to answer the questions.
<i>Validity</i>	The extent to which a question measures what it is intended to measure and does not measure what it is not intended to measure
<i>Vignettes</i>	Brief narratives containing elements of social situations and actions relevant to the topic under examination.

Attachment A: Stage One Template Example for the Collation of CASIC Questions

Objective	Question Used	Source	Use of the information	Expected Information Outcome	Comments	Any results – basic demographic results
To identify if marital status may predispose people to health problems.	AA2 WHAT IS YOUR CURRENT MARITAL STATUS? (Tick one only) MARRIED DE FACTO SEPARATED BUT NOT DIVORCED DIVORCED WIDOWED NEVER MARRIED	WA Health Survey 1995	To see if marital status category predisposes people to health problems.	Certain life style behaviors were expected to be significantly related to marital status.	The results from this question were quite valuable in some of the lifestyle and health status interpretation	WA 1995 Married (50.9%) Defacto (4.2%) Separated (3.4%) Divorced (5.6%) Widowed (8%) Never Married (28%)

Source: CATI Technical Reference Group, 2000

Attachment B: Validity and Reliability

The following definitions have essentially been drawn from Aday (1996: 50-63) and Frankfort-Nachmias & Nachmias (1992: 158-67) [1, 37].

Content Validity: relies on judgments about whether the questions chosen are representative of the concepts they are intended to reflect. How good are these questions in measuring the theoretical domain they are presumed to represent?

The extent to which the questions adequately represent the concept.

The predominant type of content validity is often referred to as **face validity** where the question/variable content rests on the subjective evaluation of the researcher that the question(s) will measure the concept/topic.

Criterion Validity: is the extent to which the question(s) predict or agree with the “true” value or “gold standard” for the topic being measured.

The extent to which the questions predict or agree with criterion indicators.

The two main types of criterion validity are

predictive - can the criterion be predicted by the question(s) or variable), and
concurrent - does the criterion currently correspond to the question(s) or variable.

Criterion validity is generally quantified through correlation between the question and the future or concurrent criterion source value. It can also be quantified by examining the sensitivity and specificity of the questions or variable with the criterion.

	Criterion: Yes	Criterion: No
Variable: Yes	a	b
Variable: No	c	d

$$\text{Sensitivity} = a \div (a+c) \quad \text{Specificity} = d \div (b+d)$$

Sensitivity – the proportion that the question picks up with the criterion

Specificity – the proportion that the question does not report the value that doesn't have the criterion.

The higher the sensitivity and specificity of the questions, the higher the criterion validity.

Construct Validity: refers to the relationship of the questions with other questions being measured in the same collection. Again correlation of these questions can be used to quantify construct validity.

Construct validity tests whether a hypothesized association between the questions and a measure of the same concept (convergent validity) or a different concept (discriminant validity) is confirmed.

The reliability of questions refers to the stability and equivalence of measures of the same concept over time or across methods of collecting the data. Stability of questions refers to the consistency of answers to the same question when asked at different points in time if there have been no changes in relation to the topic being measured, while the equivalence of different data gathering methods refers to the consistency of answers when the same questions are asked by different interviewers.

Test-retest Reliability: refers to whether the same question yields consistent results at different times.

Test-retest reliability can be computed using correlation coefficients between answers a respondent provides to the same question at two different points in time. The closer the value of the coefficient is to +1.00, the more stable or consistent the question.

Inter-rater Reliability: refers to whether different people collecting data on the same questions tend to get comparable answers.

Correlation or kappa coefficients can be used to measure the strength of agreement between two data gatherers. However, in personal interview surveys consideration in advance should be given to what might cause random variation in interviewers' performance and standardised training and supervision procedures put in place to reduce these sources of variation.

Internal Consistency Reliability: refers to whether different questions that are assumed to collect the same underlying concept are correlated. Internal consistency reliability is used primarily in the construction of summary scales.

Corrected item-total correlation, split-half and alpha reliability coefficients are the main procedures for estimating the internal consistency among a number of different questions that are supposed to reflect the same concept.

Corrected item-total correlation is the correlation of a data item with the sum of all other items (in a scale). A correlation of 0.4 or higher has been used as a cutoff for identifying items that are internally consistent.

Split-half reliability reflects the correspondence between answers to two subsets of questions when an original set of questions about a topic is split in half and a correlation coefficient is computed between scores from the two halves. The higher the correlation, the more reliable the set of questions.

Coefficient alpha reliability is similar to the split-half approach except that it is based on all possible ways of splitting and comparing the set of questions. Cronbach's alpha is used when there are multilevel response categories, while Kuder-Richardson's alpha is used when there are dichotomous categories. The higher the value of alpha, the more reliable the set of questions. For group-level data the lowest level of acceptable internal consistency reliability is 0.7, while for individual-level analysis the lowest level is 0.9.