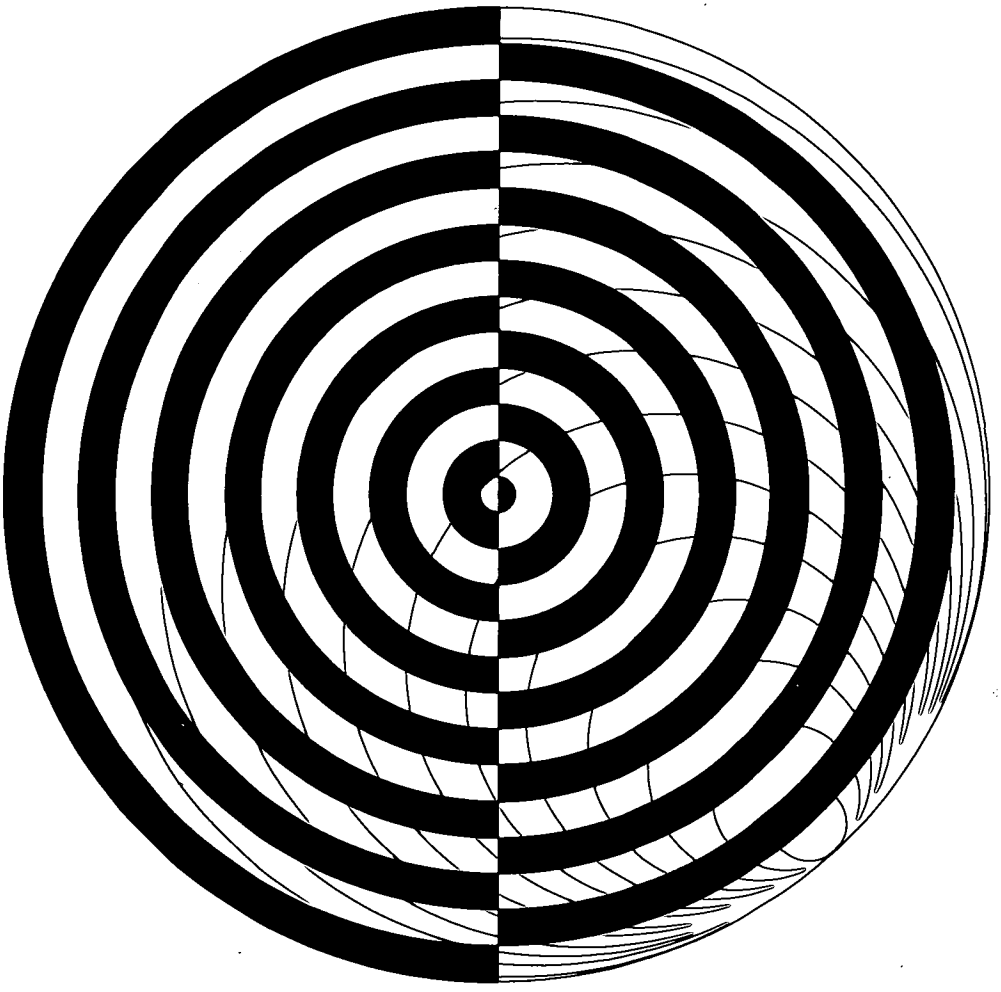
A. E. BENNETT & K. RITCHIE

QUESTIONNAIRES IN MEDICINE

A guide to their design and use

Published for the Nuffield Provincial Hospitals Trust
by the Oxford University Press



BENNETT & RITCHIE
QUESTIONNAIRES IN MEDICINE
OXFORD

A. E. Bennett and K. Ritchie

Questionnaires in medicine

A guide to their design and use

Published for the
Nuffield Provincial Hospitals Trust
by the Oxford University Press
London New York Toronto

Published for the Nuffield Provincial Hospitals Trust
3 Prince Albert Road, London NW1 7SP by
Oxford University Press, Ely House, London W1
Glasgow New York Toronto Melbourne Wellington
Cape Town Salisbury Ibadan Nairobi Dar es Salaam Lusaka Addis Ababa
Bombay Calcutta Madras Karachi Lahore Dacca
Kuala Lumpur Singapore Hong Kong Tokyo

ISBN 0 19 721390 1

A. E. Bennett, MB, FFCM, DIH*
Professor of Clinical Epidemiology and Social Medicine, St George's Hospital Medical School, London

K. Ritchie, BA†
Research Psychologist, The National Health and Medical Research Council Social Psychiatry Unit, Australian National University, Canberra

Formerly *Director and †Research Assistant of the Health Services Evaluation Group, Department of the Regius Professor of Medicine, University of Oxford

Designed by Bernard Crossland
Printed in Great Britain
Burgess & Son (Abingdon) Ltd, Abingdon, Oxfordshire

Contents

Acknowledgements

This work stems from a long-term interest in the theoretical background to questionnaires and their various uses within medicine. I was set on this trail and encouraged thereafter by Walter Holland and Max Wilson and I should like to record my debt to them both.

Many colleagues also deserve our gratitude, none more so than Jessie Garrad for the very many hours of invaluable discussion. We should also like to express thanks to Sir Richard Doll, Marie Johnston, and Rory Williams for reading and criticizing the text, and Christine Lewis for her help at every stage in preparation.

Foreword

There are three aspects to this monograph. First, I think that anyone reading it through should emerge with a clearer understanding of the many factors which influence communication with the patient. Hence, what follows is of wider interest than just to those wishing to use or who are interested in questionnaires. I think it should prove valuable reading for every medical student during training.

Secondly, it touches on the problem of the meaning of symptoms. It raises questions about our understanding of symptoms and symptom clusters and by implication suggests that much of our knowledge is incomplete. As a consequence the diagnostic process remains ill understood and diagnosis remains an art, for we are as yet unable to quantify much of the data required. Only thirdly then do I put the practical considerations of questionnaire design and use. This is not to suggest that the object of the monograph is not achieved but rather to reinforce some of the conceptual thinking underlying it.

A medical questionnaire is concerned with symptoms—perceived changes indicating disease—and communication, a much used and abused word. Whilst communication is difficult to define in a universally acceptable way (there are some twenty different concepts of the word) the word communicate is more easy; moreover its definition is useful for it helps express a purpose. Communicate means share and this is what doctor and patient, researcher and respondent, do concerning their experiences in order to achieve a common description and understanding. The questionnaire is a means of sharing, less personal perhaps and possibly in some situations less complete; in many instances, however, more precise and reliable.

The use of questionnaires in medicine is recent—some twenty years all told—these last ten years to any significant extent. Serious work started from an appreciation of the extent of variability in history-taking, particularly in the investigation of certain diseases, such as chronic bronchitis, that could best be described symptomatically. Since then their use has been encouraged by the development of

computer applications. So it is surprising that little has been written about them in a comprehensive manner. Very few other tools of measurement have been so neglected. Perhaps the resistance is an emotional one. The questionnaire is intrusive and, in part, threatening: it is very demanding. With this challenge let me stress again that what follows is mainly concerned with those things that affect the understanding of doctor and patient, whether using written or verbal systems for communicating.

W. W. HOLLAND

March 1975

Part I

Design and use of questionnaires

1. The principles of questionnaire design

The process of responding

The aim of the questionnaire designer is to *communicate* with potential respondents using the medium of the questionnaire. It is his responsibility to ensure that questions can be fully understood and that the respondent is encouraged to reciprocate in this communication process. For this it is important that the researcher understands something of the nature of the communication process and thus appreciates the task he is setting the respondent.

Responding is not a simple stimulus-response process but a complex procedure by which the respondent selects a small amount of his total information to become the questionnaire data. Each question will alert the respondent in a particular direction causing him to focus on some aspect of his total experience. The relevant experience may be well thought out and organized in his mind, though more likely it will be vague and confused due to the limitations of memory. The effect of memory may well vary according to the nature of the information. For example, it has been demonstrated that some types of medical information are remembered more accurately than others: hospital episodes are better remembered than physicians' visits (Cannell and Marquis, 1967) and physicians' visits are better remembered than acute and chronic conditions (Madow, 1967). Recency and impact are important determinants of retention with regard to medical information: that is, the longer the intervening period between the relevant experience and its recall, the less chance that it will be remembered, and the greater the impact of the experience on the everyday life of the respondent, the better it will be remembered (Ley, 1972*a*). Unfortunately however, the process of forgetting cannot be predicted, for memory is selective and the process by which information is stored or discarded is influenced by incidental emotional factors as well as the continual process of extinction. Furthermore, memory is seldom an all-or-none event. Experience may be remembered in a distorted or incomplete form, confusing events or

magnifying them out of their original proportions. Additionally, the relevant experience may have become associated in time with other experiences so that the question brings to mind a much broader range of ideas than is relevant.

This complex of ideas must then be brought into full awareness—a process complicated by such psychological phenomena as self-analysis, conceptualization of ideas and generalization from specific points. The respondent thus forms a precept of his own experience and the process of producing this may be affected by wishful thinking, a desire to please the research worker and a desire to be fair to oneself and to others; and it will probably be accompanied by a good deal of confusion (Oppenheim, 1968).

The respondent then has to decide what aspects of all this information he is prepared to communicate. He may be reluctant to communicate information which is embarrassing or socially unacceptable. He may have misgivings as to the purpose for which his information will be used and the conclusions that the researcher might draw from it. Working against this censoring procedure will be forces motivating the respondent to answer fully and thoughtfully. Cannell and Kahn (1953) maintain that this positive motivation is derived from two sources, first from a desire to influence his present state, where the researcher is seen as someone able to bring about changes for the respondent's benefit; and secondly from the gratification which the respondent receives from the communication process itself, especially when the respondent sees the situation as an opportunity to express previously suppressed opinion. Criteria which the respondent uses to select information may vary as he goes through the questionnaire in accordance with a changing view of his role. Initially he may not always have a clear idea of what he should do and Cannel *et al.* (1969*a*) have shown wide variations in the views of respondents at different stages in the administration of a questionnaire as to the level of information required.

Then having decided on the information he is prepared to give, the respondent must put it into words. This means he must place the information within the limits of his own vocabulary which may not include the precise terms he would like, so that his ideas are inaccurately expressed or, alternatively, he resorts to clichés and stereotypes. Enough has been said to show that appreciation of the complexity of this response process is vital for the person who would design or use questionnaires.

Questionnaire content

Defining the content of a medical questionnaire must begin with enumeration of those items about which information is sought. They may include personal characteristics, medical history, family medical history, present symptoms, etc. Each item must then be described in terms of its manifestation in the individual. For symptoms this may present considerable difficulty due to the range of individual differences in disease manifestations. As Walshe (1956) puts it, there are no diseases, only the diseased. There may be much disagreement as to the symptom patterns relevant to a particular investigation. For example, Ropes *et al.* (1957) in their attempts to isolate the symptoms of rheumatoid arthritis found that many of the symptoms frequently accepted for diagnostic purposes were non-specific. Another example is provided by Wood and Elwood (1966) who conducted a questionnaire investigation into symptoms of iron deficiency anemia by comparing symptom severity with haemoglobin level. They found that symptoms commonly accepted as indicators of iron deficiency anaemia, such as fatigue, faintness, shortness of breath, anginal pain, and palpitations were not related to haemoglobin level. They point out that these same symptoms occur frequently in people who are not anaemic and that they may be ascribed to other conditions such as the psychoneuroses. It is important, therefore, that symptom patterns should be derived from careful observation of a sample population known to possess the health problem under consideration. Further one should ensure as far as possible that all relevant symptoms are manifest in this sample population and that it is not atypical. For example, Zborowski (1952) found pain response specific to certain cultures.

The content of a questionnaire is generally designed to investigate the minimal amount of an individual's total experience which will provide sufficient information concerning the problem under consideration. This limited selection of a respondent's experience will mean that a good deal of information concerning the individual will be lost. One must consider also the information which the questionnaire will not pick up. Pre-testing of questionnaires by administering to a sample population, where the sample is representative of potential respondents, may indicate where additional information is required. Obviously the greater the number of topics that the questionnaire can cover, the less information will be omitted: but on the other

hand there are the disadvantages that a long questionnaire will be difficult to process and it may also make the respondent reluctant to give full responses so that information loss will then occur through respondent fatigue.

Some researchers maintain that information loss may be reduced if the respondent is permitted to add personal comments concerning information which he feels is important and has not been adequately covered by the questionnaire. Mellner (1970) maintains that the restriction of individual comment causes loss of information to an important degree. He claims on the basis of his various questionnaire studies that a medical history used for patient care cannot be completely standardized and codified. In other words, the questionnaire should contain space for additional comments in narrative by the patient and the physician.

The construction of questionnaire items

The writing of questionnaire items should begin with the selection of a question type. This is important for it will influence the type of data which the questionnaire will yield (that is, quantitative or qualitative) and also determine the nature of the relationship between the respondent and the researcher, ranging from the relative informality of the structured interview to the impersonal method of the self-administered symptom check-list. Broadly speaking two main question types may be defined, the 'open' and the 'closed' question.

Open questions

An open question does not suggest any specific response. The respondent is allowed to answer freely, in his own words, and his response is recorded in full. A closed question is one which requires the respondent to choose his answer from a given, limited selection.

The 'openness' of a question is influenced by its wording and also possibly by the subject matter of the item. For example, 'Have you received any medical treatment in the past year' and 'What medical treatment have you received in the past year' are both open questions if no response categories are suggested but the second is 'more open' in that it may be expected to generate a larger variety of positive answers. They are similar in that they both require the respondent to frame his own responses using his own vocabulary, thus permitting

the expression of individual experience. The open question may, therefore, elicit information and areas of interest not predicted by the investigator.

Because the respondent is given the opportunity to explain his individual problem at length, open questions usually produce more detailed responses, although it is difficult to measure changes in a respondent after an interval of time as consecutive responses may be phrased in different ways and difficult to compare. Another drawback with open questions is that, because responses are recorded in full, administration time is lengthened with the consequent risk of respondent fatigue. Also the long and varied responses are difficult to handle statistically. In order to compare responses to open questions they must be classified in some way and this interpretative and judgemental procedure may be subject to considerable error. The interpretations given to open questions tend to vary amongst different judges. This was demonstrated by Schooler (1956) in a comparison of interpretations given by ten professional coders from the Survey Research Centre of the University of Michigan to open question responses. Schooler found significant inter-coder differences in the quantitative equivalents designated to qualitative data and concluded that inter-coder variability increased with ambiguity or openness of the question. Thus it may be preferable to allow the respondent to interpret his own experience in terms of the given categories of the closed question rather than allow someone else to do it for him, thus risking misinterpretation, inter-coder variation, and the personal biases of the coder. The open question clearly has limitations in a situation aiming at the controlled measurement of medical facts, particularly where it is desirable to combine or compare the results of individual respondents. However, the open question remains an important means of assessing individual experience in detail and its value for pre-design research has been stressed by Parten (1950):

> Only by finding out what aspects of an experience make an impression on people and what words the public use to express it, can intelligent planning proceed.

It is only by research with open questions that closed questions can be designed.

Closed questions

Closed questions may vary considerably in form, their common element being that they require the respondent to consider each

question in terms of a number of given responses. The following examples from current medical questionnaires illustrate the variety of forms which the closed question can take. They represent respectively the 'strict alternative' or 'dichotomous' questions, the 'alternative statement', the 'check-list', and the 'rating scale' types:

	YES	NO
Do you have any pain or discomfort in the chest? (Rose, 1962)	☐	☐

I never worry about anything	☐
I seldom worry about anything	☐
I sometimes get rather worried about things	☐
I tend to worry a great deal	☐
I am always in a state of terrible worry and anxiety	☐

(Ingham, 1965)

Have you ever had:

An injury or operation affecting your chest?	☐
Heart trouble?	☐
Bronchitis?	☐
Pneumonia?	☐
Pleurisy?	☐
Pulmonary tuberculosis?	☐
Bronchial asthma?	☐
Emphysema?	☐
Bronchiectasis?	☐
Other chest trouble?	☐

(MRC, 1966)

Do you suffer from headaches?

seldom or never	☐	some of the time	☐
quite often	☐	almost always	☐

(Rockliff, 1969)

Closed questions may be either dichotomous or multiple choice. The dichotomous question represents a response choice of alternatives such as Yes/No, High/Low. A third category of 'Neither', or 'Undecided', or 'Don't know' is sometimes included, although given no statistical significance unless so many responses fall into this category it is obvious that the questionnaire is not useful. Parten (1950) has suggested, however, that if no specific mention is made of this category in the instructions, then few respondents will fail to make a choice between the opposing alternatives. The dichotomous question type is frequently employed in lengthy questionnaires (such as the Cornell Medical Index, Brodman *et al.*, 1956) and complex question-

naires (such as the multi-level branching general health questionnaire developed by Mayne *et al.*, 1969) because of simplicity of completion, scoring, and statistical evaluation. Furthermore the minimum number of two choices per item maximizes the stability of the response to the test as a whole. The limitations of the dichotomous questions are that a slight misinterpretation of the question or error in recording the response results in a complete reversal of the answer. Also the need to commit oneself to a strongly positive or negative position may lead to information loss and a reluctance on the part of the respondent to complete the questionnaire (Ghiselli, 1939).

The multiple choice question attempts to overcome the limitations of the extremes of classification found in the dichotomous type by permitting a greater variety of responses. The most simple form of multiple choice question, from the point of view of both construction and administration, is the item check-list. This consists of a statement of the problem, followed by a list of possible answers. The respondent is required to check those items relevant to his experience. The main advantage of the check-list is to remind the respondent of a barely remembered experience which he might otherwise have forgotten. In the case of medical questionnaires this may be especially useful where difficult terminology is involved which the respondent may not otherwise recall. For example:

Have you taken any of the following in the past year?
- codeine (except in cough syrup)
- darvon
- penicillin
- demerol
- morphine

(Collen, 1969)

A disadvantage of this type of question is that a respondent might think that an item is familiar when it is not. He might also fail to read an item, and if only positive responses are checked there is no way of detecting such errors.

Another form of multiple choice question is the alternative statement which offers a choice of responses in the form of complete statements. The respondent chooses the one which describes his position best. This type of question is particularly useful for the investigation of experiences which the respondent might find difficult or embarrassing to formulate, as in the case of some psychiatric symptoms. For example:

I feel certain that I look peculiar.
I feel certain that people laugh at me.

The main difficulty associated with this type of question lies in the necessity to ensure that all possible statements are included in the response options and this may lead to the construction of long cumbersome questions. For example:

Is your headache:
always on the right side?
always on the left side?
usually on one side but not always on the same side?
on either side near your temple?
around or near you eyes?
across your forehead?
on top of your head?
at the back of you head or neck?
all over your head?
(Gledhill and Mathews, 1971)

Ingham (1969) described a form of the alternative statement which takes into account the fact that symptoms are not all-or-none events but lie along a continuum of graduated experience. For example, symptoms of fatigue may be considered to lie on a continuum represented by the following statements:

(1) I never get tired.
(2) I only get tired after a very hard day.
(3) I generally feel a little tired by the end of the day.
(4) I always feel very tired by the end of the day.
(5) I feel very tired most of the time.

These statements are then presented to the respondent in all possible combinations of pairs and he is required to indicate the one which he feels is closest to the truth. By presenting only two statements at a time a check may be made on the respondent's consistency in responding and thereby it may be known whether he has really understood the meaning of the statements. This is done by considering the respondent's choices of the various parts as indicating his position on a theoretical scale ranging from 'never tired' to 'always tired'. Suppose that a respondent's position on the scale is about (1); that is, he very rarely gets tired, then each choice he makes must be in accordance with this one point. So, out of a choice of statement (2) and (3) the respondent would choose (2) as being closest to his position. If then from a choice of statements (3) and (4) he chooses (4) then this would be an inconsistency, possibly due to a misunderstanding of the meaning of (4). Ingham (1965) has demonstrated how a table of

consistent and inconsistent responses might be constructed for statement scales of various physical and psychiatric symptoms. For example, given a scale consisting of three statements, *A*, *B*, and *C*, ranging from the least to the most severe, then a table of consistent and inconsistent responses would be constructed thus:

(A, B)	*(A, C)*	*(B, C)*	
A	*A*	*B*	consistent response patterns
B	*A*	*B*	
B	*C*	*B*	
B	*C*	*C*	
A	*C*	*B*	inconsistent response patterns
A	*C*	*C*	
A	*A*	*C*	
B	*A*	*C*	

This method was used by Shapiro (1961) for the development of a questionnaire for psychiatric symptoms and was found to produce acceptable data.

Rating scales

The rating scale is a special form of the multiple choice question in which respondents quantify their response. Rating scales vary in form and the most important type for medical research is the interval scale in which the appropriate gradation of response is defined by the investigator. This gradation is supposed to represent equidistant steps on a cumulative scale. For example:

Do you suffer from headaches . . seldom or never?
some of the time?
quite often?
almost always?

(Rockliff, 1969)

An extension of this is to assign a number sequence to the statements, which may be misleading to the respondent if he thinks the numbers represent some value, but which has the advantage of indicating that there is equality of interval between scale points. For example:

How would you describe your headaches?

0	. .	little or no discomfort
1	. .	mild discomfort
2	. .	moderately painful
3	. .	extremely painful

It has been suggested that negative numbers should be used where the scale is to extend from one extreme to another, such as very pleasant

to very unpleasant, but Guilford (1950) has pointed out that this may be confusing to respondents not familiar with algebra. Some investigators have preferred to use the number sequence alone. For example:

Rate your present health	bad 1 . . . 2 . . . 3 . . . 4 . . . 5 . . . 6 . . . 7 very good
Rate your state of health as it was a year ago	1 . . . 2 . . . 3 . . . 4 . . . 5 . . . 6 . . . 7
What is your normal state of health?	1 . . . 2 . . . 3 . . . 4 . . . 5 . . . 6 . . . 7

Obviously the responses obtained are difficult to interpret and this variation is rather more suited to social and psychological research than medical questionnaires.

The number of points on a rating scale is a problem which has received considerable attention. To have too few categories on a scale will not make use of the discriminative ability of which respondents are capable and which might be required by the research. On the other hand a respondent might not be able to discriminate between a large number of finely graded categories. Conklin (1923) concluded that, for raters without previous experience, the maximum number should be five points for a scale measuring one extreme of a variable, and nine points for a scale measuring both extremes, for example, lethargy (←——→) hyperkinesis. Whether an odd number of categories should be used is debatable. An odd number provides a mid-point which can be regarded as a neutral or 'safe' position. Some researchers think that this is desirable as it allows respondents to avoid committal to a definite position. However, in psychiatric questionnaires where extremes of the scale may be considered as indicating abnormality, the respondent may assume the mid-point to be representative of 'normal' or 'desirable' behaviour. For this reason, many psychiatric questionnaires employ an even number of response options, thus forcing the respondent to place himself to one end of the scale. For example:

Have you recently had the feeling that people were looking at you?	not at all	no more than usual	rather more than usual	much more than usual

(Goldberg and Blackwell, 1970)

Thus there is no right answer to the problem of how many categories to include on a rating scale. It will vary according to the subject matter of the scale and the differentiation required. A last and

perhaps irrelevant consideration for medical questionnaires, but included for completeness, is the need to 'anchor' the scale, an idea described by Hunt and Volkmann (1937). This idea is based on the observation that respondents tend to avoid using the extremes of a scale, thus excluding the terminal categories from their consideration. Hunt and Volkmann demonstrated that if two additional categories, which have no meaning for the researcher, are added one at either end of the scale, then the scores will be more widely dispersed, as the respondents will then use the whole of the original scale.

The difficulty with interval scales is the need to create equal intervals between scale points. This is important if the points are later to be given quantitative values for data analysis. One way of ensuring this empirically is to use a scale which consists of statements whose position on a theoretical interval scale has been determined by some sort of ranking or rating procedure performed by several judges. (A method is described in greater detail by Thurstone and Chave, 1929.) The procedure ensures that the scale contains relatively unambiguous statements, the meaning of which has been agreed so that the risk of respondents misunderstanding their meaning is reduced. The method was used successfully by Dewolfe (1968) in the construction of scale ratings for schizophrenic patients.

Open versus closed questions

Of the relative advantages of open and closed questions a few words should finally be said. It has been noted by Cannel and Kahn (1953) that the closed question is suited to ideal situations in which there is only one frame of reference from which the respondent can reply, and where there is a known range of possible responses which can accurately describe the position of all possible respondents. The ideal is rare and it should be remembered that even experienced medical observers have difficulty in classifying symptoms so that some consideration should be given to whether the lay respondent will be able to recognize the appropriate category for his experience.

Maccoby and Maccoby (1954) have pointed to a fundamental difference between open and closed questions with regard to the demands they make upon the respondent. Open questions require the respondent to recall something whereas closed questions require him to recognize something. From the psychological research dealing with comparisons of recall and recognition memory, it is concluded

that more information will be recognized than will be recalled. This has been confirmed by Belson and Duncan (1962) who found in a comparison of the data yielded by questions of the open and check-list type that the latter gave significantly more information. However, whether such additional information is accurate remains doubtful: as mentioned previously it may be that the respondent thinks certain symptoms or other experiences are familiar, or he may simply check items to shield his ignorance of their meaning. Some researchers have chosen to use a variety of question types within a questionnaire but this may make coding and statistical analysis more difficult. However to offset this, a questionnaire containing different types of questions is less likely to bore respondents and may improve response and the quality of the data. Ultimately the choice of question type will depend largely upon the subject matter of the study. For example, medical surveys and epidemiological studies may require quick coding methods with minimal detail, so the closed question is more suitable: by contrast, social assessments may require a greater appreciation of individual differences so an open question type will be needed.

Question wording

The wording of items may present considerable problems for the questionnaire designer as he must mediate between the needs of two interests. On the one hand are the medical personnel who will make use of the data obtained by the questionnaire for diagnosis or research. They require precise statements of the problem and the symptom entities which comprise it. On the other hand, there is the respondent who requires that items be written in simple non-technical language.

For those concerned with interpreting the data, the use of precise and appropriate medical terminology is important if the results of the questionnaire investigation are to be universally understood. The need for the development of a 'pure' medical terminology has been stressed by Gordon (1966) who complains that jargon and colloquialisms have become respected members of the vocabulary, 'formal' words have gained prominence over plainer rivals, while popularized abbreviations proliferate in areas of fundamental research. In addition there is the problem that words tend to change their meaning over time, particularly when they have been assimilated into everyday usage. Where medical terminology is to be used in item construction it should be standardized as far as possible by

the use of a well-recognized reference such as *Current Medical Terminology*.

Although such precision is obviously desirable in medical research, consideration needs to be given to the respondent. Maccoby and Maccoby (1954) have criticized the all too frequent tendency of researchers to translate their objectives into questions too directly without considering problems of communication. The non-medical orientation of the majority of potential respondents should be taken into account. The importance of this is demonstrated by the work of Boyle (1970) who convincingly illustrated the wide discrepancy between patients' and doctors' interpretations of common medical terms. He found highly significant differences between the definitions given by 234 out-patients and 35 doctors of such terms as 'arthritis' and 'palpitation'. The terms used by Boyle were ones commonly found in general health questionnaires. However, in spite of these findings, many respondents having a particular condition will know it by its proper medical name. This was shown by Young (1972) for conditions such as 'angina' or 'emphysema'. He questioned potential respondents as to the meanings of the terms used in his General Medical Questionnaire and although he found that adequate definitions were rarely given, the questionnaire yielded fairly accurate data. Young concluded that this was partly because respondents experiencing certain conditions learnt the medical terms from their physicians whereas other respondents had not come in contact with them.

An effective way of dealing with the problem of technical terminology is to provide alternative phrasings in parentheses. For example:

Have you had jaundice (yellowing of the eyes or skin) in the past year?
(Collen *et al.*, 1969)

Questions which refer to drugs or compounds should give the medical terms together with a description of their application. For example:

In the past year have you taken any digitalis (heart medicine)?
(Collen *et al.*, 1969)

In phrasing questionnaire items the research should aim at communicating with the lowest rather than the average educational level of the potential population. A useful reference in this regard is a paper by Flesch (1949) describing the development of the Flesch Reading-Difficulty Formula, which evaluated the readability of various texts by considering such factors as length, number of difficult words, and

the number of personal references. By use of this formula, comprehension level can be adjusted to suit various requirements (Ley *et al.*, 1972; Ley, 1973).

Sophisticated terminology is not the only barrier to communication. The use of terms which are vague or ambiguous such as 'occasionally' or 'often' should also be avoided. Such words should be replaced by more precise terms. For example:

Have you, in the past *six months,* had *two or more* bad nose bleeds which were not caused by an injury or a cold?

To question a respondent concerning a vague period of time such as 'When you were very young' creates difficulty and in general it is better to ask about a set period.

Descriptive terms may be particularly misleading. For example, Suchman *et al.* (1958) found that in response to the question:

Do you have trouble with your hearing?

a quarter of the patients who answered yes were judged by the physician as not having any hearing loss. Confusion had arisen among respondents as to what constituted 'trouble'. Collen *et al.* (1969) pointed to the following item as being particularly confusing to respondents:

Have you in the past year been told that you had an abnormal or positive chest X-ray?

Confusion arises because the word 'positive' in customary usage normally indicates a favourable condition. Respondents may then interpret the question as asking whether they had had any X-ray in the past year, normal or abnormal.

Many more examples of bad question wording could be cited from the literature. In an analysis of responses to their general health questionnaire, Anderson and Day (1968) found that 69 per cent of questionnaires contained contradictory responses, due mainly to varying interpretations of a present symptom, as in this example:

Are you presently suffering from a pain in either leg?

In current usage presently means soon or shortly; an older meaning was now, at this time, and this was how it was used in this question, hence the confusion. It is a word to be avoided in the same way that negative phrasings should never be used. Would a positive reply to this question:

You don't think you have lost weight do you?

mean 'Yes: I have lost weight'. or 'Yes: I don't think I have lost weight'?

Sprosty (1964) in an inquiry made into the properties of the well-worded question found that questions containing multiple ideas were of least value. For example:

Have you at any time in the past year had pain or pressure or a tight feeling in the middle of your chest which was brought on by exertion or walking fast uphill, and which left after a few minutes rest?

This question, assumed by researchers to be a good indicator of ischaemic heart disease has been shown to yield poor results (Collen *et al.*, 1969). It should have been broken down into several items as in the following example.

		YES	NO
	Have you ever had any pain or discomfort in your chest?	☐	☐
If NO	Have you ever had any pressure or heaviness in your chest?	☐	☐
If YES	Do you get it when you walk uphill or hurry?	☐	☐
	If you stand still what happens to it?	relieved	☐
		not relieved	☐

(Rose, 1962)

Care should be taken that the wording of a question does not bias the respondent in a particular direction. This may occur if 'leading questions' are used, that is a question which indicates that a certain response is the 'right' one. A respondent may also be biased if 'loaded words' are used. These are words which in their everyday use have gained certain favourable or unfavourable connotations. Many might be reluctant, for example, to give a positive response to the questions:

Have you had any venereal disease in the past year?

or

Have you had an illegitimate child?

Where alternative phrasing is not possible some other method should be designed to lower the impact of such words, for example by placing the question amongst others with less drastic social implications as part of a check-list. For example:

In the past year has a doctor said you had:
Kidney or bladder stones?
Kidney disease such as nephritis?
Kidney or bladder infection?
Venereal disease (such as gonorrhea, syphilis)?
(Collen *et al.*, 1969)

The problem of obtaining accurate answers to embarrassing questions was an important one for Kinsey *et al.* (1953) in their investigations of sexual behaviour. Their technique was to assume that the respondent possessed the low-valued factor and hence place the burden of denial on him. One would then ask 'When did you first have . . .' rather than 'Did you ever have . . .'. This idea has been used by Dlin and Perlman (1971) in the construction of a questionnaire for ileostomy and colostomy patients. The investigators here assumed that all respondents engaged in a wide variety of 'socially unacceptable' behaviour, for example:

	YES	NO
Has there been any change in your self-stimulatory (masturbatory) habits	☐	☐

This technique requires a skilled interviewer or a highly motivating questionnaire as such questions may offend or be threatening to the respondent. Careful thought will usually provide a less dramatic method of making loaded items sound less incriminating. For example, Garrad and Bennett (1971) in their disability study asked respondents 'do you' perform certain tasks, rather than 'can you'.

Particular care should be taken in the construction of questionnaires dealing with psychiatric symptoms. Goldberg and Blackwell (1970) maintain that many patients feel that it is more socially acceptable to be physically rather than emotionally ill and they fear the stigma of being thought of as a 'psychiatric case'. For this reason many questionnaires dealing with psychiatric illness are physiologically orientated. For example:

Do you suffer from severe nervous exhaustion?
Do you usually have great difficulty in falling asleep or staying asleep?
(Brodman *et al.*, 1956)

Slack's computer-based questionnaire system (1971), which measures simultaneously heart-rate and response latency, attempts to deal with items found emotionally disturbing by putting up frames which read 'Relax, you're doing fine'. This is claimed to be an effective (though extravagant) means of controlling the effects of loaded questions.

A check can be made of the effects of question wording by use of the split-ballot technique inaugurated by the American Institute of Public Opinion. This consists of preparing a questionnaire in alternative forms, each being submitted to comparable cross-sections of the population, with some questions being varied in their wording and others being the same to act as controls. Another method is to include in the same questionnaire differently worded versions of the same item and compare the responses for consistency (Anderson and Day, 1968).

Question length

How long should a question be? If it is to be simply constructed then it would seem desirable that it is as short as possible. Oppenheim (1968) maintains that twenty words should be the maximum length. However, it might be argued that it is not the number of words which is important but rather the number of ideas which the question contains.

Consider, for example, the following two questions:

Have you had swelling of both ankles or feet in the morning or later in the day?
Can you tell me if you have ever experienced swelling of both ankles or both feet?
(Collen *et al.*, 1969)

Although both questions are approximately the same length, the former has been shown to produce poor results; the latter is obviously to be preferred in that it attempts to express only one idea.

However, Cannel *et al.* (1969*a*) claim that longer questions may sometimes be preferable in that they can improve the information obtained. They made a comparison of two health questionnaires, one using longer versions of items than the other. For example:

Have you had fainting or blackout spells?

or

How about fainting or blackout spells. Have you had these?

Comparing the results of the two questions with information obtained from physicians, the longer version was found to produce significantly more accurate responses. The researchers concluded that the longer versions had served to increase the general level of motivation of the respondents. The longer questions it should be noted, did

not contain any additional information which could interfere with the main idea.

In conclusion it can be said that the characteristics of a well-worded questionnaire are that it is grammatically simple, avoiding where possible technical terms, adjectives, adverbs, or vague descriptive terms; that leading questions and loaded words are avoided; that additional information where necessary is placed in parentheses to avoid confusion; and that each item contains only one idea.

Question sequence

Parten (1950) and Cannel and Kahn (1953) maintain that the items of a questionnaire should follow a logical sequence resembling as far as possible the sequence which the respondent might be expected to follow. This facilitates the response process for the respondent is not suddenly required to change his frame of reference. However, the investigator must be aware of the effects which might result from placing certain items together. He should remember that the respondent may assimilate the information contained in one item and carry it over to the following item so that it influences subsequent responses. Parten (1950) has pointed out that it is in fact possible to determine the nature of a response by arrangement of items so that the respondent will associate certain items together. This is because the respondent presumes a relationship to exist between certain factors. This might occur in the following example:

112. Are you constantly too tired and exhausted even to eat?
113. Do you suffer from severe nervous exhaustion?
(Brodman *et al.*, 1956)

A respondent might answer 'yes' to 112 on the basis of genuine experience then answer 'yes' to 113 simply because he assumes them to be related.

A common approach to question sequence is to use the technique of the item 'funnel', that is, to ask general questions first followed by successively more detailed and specific questions. This approach closely resembles the personal clinical interview and has the advantage of preventing respondents from answering general questions in terms of preceding specific ones (which is the disadvantage of the 'inverse funnel' technique: that is, moving from specific to general questions). The following questions formed part of a questionnaire for ischaemic heart disease and illustrate the method:

Have you ever had a severe pain across the front of your chest?
Did you see a doctor because of this pain?
What did he say it was?

The disadvantage of this method is that the questionnaire becomes a lengthy and boring one for respondents who answer negatively to the general questions but who must still go on and read through all the specific ones. To overcome this many questionnaires using the funnel technique employ a question 'filter' or 'branching' system; that is, respondents answering 'no' to the general questions are instructed to leave out the following specific ones and proceed to the next general question. For example:

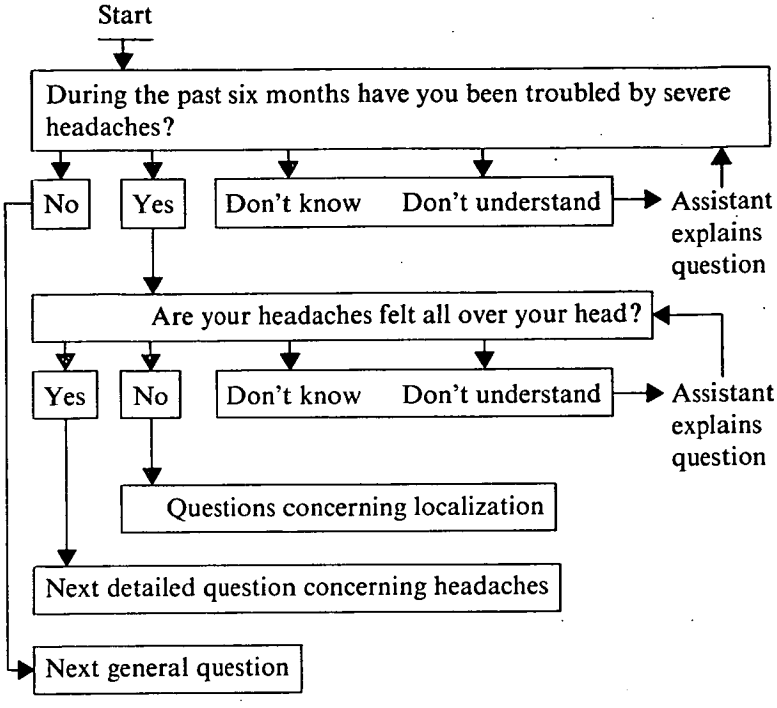
		YES	NO
	A. Do you have a chronic cough, even if just a morning cough?	☐	☐
If YES	*B.* Do you cough only in the morning after arising?	☐	☐
	C. Do you cough off and on all day?	☐	☐
If NO	*D.* Did you ever have a chronic cough?	☐	☐

(Hepper, Hyatt, and Fowler, 1969)

The use of filter questions often leads, however, to the development of cumbersome questionnaires with complex instructions, as illustrated:

If unmarried and over 35, skip to question 40, p. 6.
If unmarried and under 35, answer questions 30–35.
If married, answer questions 30–40.
(Maccoby and Maccoby, 1954)

This difficulty may be overcome by using well-trained interviewers who can move quickly from option to option without confusing the respondent. Alternatively, some researchers have adapted very complex branching questionnaires to computers, which administer the general question by means of a display terminal and the computer puts up detailed questions where relevant or displays the next general question. The computer-administered questionnaire is capable of handling far more complex question sequences, so that several levels of filtering are possible, such that each level is a more detailed inquiry into responses made to preceding levels. This procedure may be demonstrated by the following illustration taken from Kanner (1969) describing the function of his Programmed Medical History:



Items at the beginning of a questionnaire should be as 'neutral' as possible while also appearing relevant to the subject of inquiry. Their aim should be to gain the interest and trust of the respondent. The use of preliminary personal questions such as age, name, sex, occupation, etc., do not adequately fulfil this purpose. This is firstly because these facts may in themselves be embarrassing to the respondent (for example, if he is divorced or has no employment) and secondly because they require the respondent to associate himself by name and other personal details with information the content or implications of which he does not yet know. Such questions should be reserved until the end of the questionnaire. One must always be aware, however, that in the case of the self-administered questionnaire the respondent will be quite free to answer questions out of order, despite instructions to the contrary. In this case it should be ensured that the responses given to the questionnaire are affected as little as possible by the sequence of its items.

Bias

Every step in the construction of a questionnaire may introduce bias. Bias has been defined adequately for our purposes by Kahn and Cannell (1957) as the intrusion of any unplanned or unwanted influence. It may occur firstly through inappropriate wording of questions, for example, the use of leading questions and loaded words as previously described. Alternatively it may result from a failure to state all possible response alternatives as in the following example:

Do you prefer to be examined by a doctor of you own sex?

It would have been more impartial to have asked:

Would you prefer to be examined by a male or female doctor or doesn't it matter which?
(Oppenheim, 1968)

Nor should a question be worded to imply that a positive answer would be foolish as in the following example from a psychiatric questionnaire:

Do you have an unreasonable fear of being in enclosed spaces such as shops, lifts, etc?
(Crown and Crisp, 1966)

The word 'unreasonable' implies the judgement of the investigator and probably should not have been used. Kornhauser (1951) maintains that, if bias due to question wording is to be avoided, the investigator must ask whether the question employs stereotypes, includes prestige-carrying names, employs superlative terms, and finally would the question wording be acceptable to persons with opposite views on the matter.

It must be remembered that the respondent himself is biased in the direction of his own self-interest. This may result in claims that he drinks and smokes less than he really does and that he is more careful of his children's diet that he really is. He may also cheat because of disinterest in the questionnaire or to please the interviewer. If the individual feels that something may be gained by presenting a false image of himself then he may well do so. This problem has been discussed at some length by Whyte (1957) who, in a book which explains how to cheat on questionnaires, upholds the right of an individual to answer falsely:

When an individual is commanded by an organisation to reveal his innermost feelings, he has a duty to himself to give answers that serve his self-interest rather than that of the organisation. In a word he should cheat. . . . Most people instinctively cheat anyway on such tests.

A respondent may be biased by his attitudes towards his doctor, the health service, questionnaires in general, and so on, so the questionnaire designer should concentrate on constructing items which involve such attitudes as little as possible. Cannell *et al.* (1968), however, in their research on household health reporting found that the motives, attitudes, and perceptions of the respondent, as measured the day after questionnaire administration, did not in fact seem to be related to health reporting behaviour. The investigators hypothesized that this was because the household health interview was such a unique experience for the respondent that existing attitudes were no longer relevant. There is also the hypothesis that the individual might be less influenced by personal attitudes in the case of medical questionnaires as he may see his own physical or mental well-being as dependent upon the accurancy of his reporting.

Halo-effect

Another form of bias relating to the respondent is the 'halo-effect'. A positive halo-effect is a tendency on the part of the respondent constantly to over-estimate qualities which he feels are desirable due to an over-all feeling of approval of the person to whom the questions refer. It occurs usually when the respondent is asked to make a judgement concerning others towards whom he is favourable, for example, a mother rating the health status of her child. Thus, a person who is well liked will generally be rated highly on all characteristics, regardless of content. Conversely, a person who is not generally well liked might be rated poorly on all charactristics. This may be termed 'negative halo-effect'.

Rugg (1921) has described this phenomenon of halo-effect in the following words:

We judge our fellows in terms of a general mental attitude towards them and there is, dominating this general mental attitude towards the personality as a whole, a like mental attitude toward particular qualities.

The 'halo-effect' is most likely to occur when a rating scale is used, particularly where the traits measured are vague and not easily confirmed. It may occur when the respondent is rating himself as well as others and it may operate indirectly. For example, a respondent may

continually overstate the effectiveness of certain treatments if prescribed by a doctor whom the respondent really likes. The halo-effect may be counteracted to some extent by careful wording of items, clear definition of the trait to be considered, using follow-up questions and asking the respondent to enlarge upon some of his assertions.

Response set

Another important type of response bias to be considered is the 'response set'. This phenomenon has been described by Cronbach (1950) as the tendency for a respondent persistently to respond in a certain way, irrespective of the content of the question. He may, for example, continually tick 'yes' or always mark the first response alternative given, or he might always respond 'I don't know' whether such a response is correct or not. An interesting example of this was noted by Milne and Williamson (1971) who used a computer to administer questions on ischaemic heart pain, the respondent being required to answer 'no' or 'yes' by means of a red or blue button. For example:

If you stand still what happens to it?
If relieved press *red*, if not relieved press *blue*.

In each case the positive response, red, was indicated first and it always indicated the presence of a symptom. The investigators concluded that the inaccurate responses which they obtained were the result of a response set of consistent red button pressing. Couch and Keniston (1960, 1961) maintain that this indiscriminating tendency stems from a 'central personality syndrome'. The relationship of response set to personality has been discussed also by Eysenck (1970) who claims:

an acquiescent person might tend to reply 'yes' to all questions regardless of their content . . . it is not impossible that the (questionnaire) might measure acquiescence rather than the relevant variable.

Cannel and Marquis (1967) have also discussed the non-acquiescence or 'no' response set. They relate this type of response set particularly to lengthy questionnaires in which it is difficult to main the respondent's full interest and attention throughout. The respondent may, therefore, continually respond 'no' in order to avoid lengthy explanations and so reduce administration time. One might expect this to occur particularly with the complex branching type questionnaire. The existence of a response set may result in the introduction of false

correlations between tests. Vernon (1964) remarked that the common element running through a series of tests may be a response set rather than the presumed trait.

Response sets are more likely to occur when closed questions are used and dichotomous questions are particularly vulnerable. How can response set be avoided? Anastasi (1961) maintains the problem to be one of questionnaire design and claims that the use of multiple choice rather than dichotomous items is a satisfactory measure. Anderson and Day (1968) report that they were able to overcome the problem by the use of dichotomous questions and varying the position where the appropriate response was to be made. For example:

Do you get a pain or discomfort in the chest after having a big meal?	don't know/yes/no
Do you get a pain or discomfort in the chest in bed at night?	no/don't know/yes

The investigators also turned questions around so that a 'yes' response could provide a negative answer, for example 'Is your weight steady' as opposed to 'Have you lost weight'. While such methods may not eliminate the possibility of response set altogether, they allow for its detection should it occur.

Layout

The layout of a questionnaire should be designed with care, as with a self-administered questionnaire the visual impact which the setting out has on the respondent may either arouse his interest or, alternatively, discourage him from co-operating. In the case of the administered questionnaire, layout is also an important consideration as bad design may confuse the interviewer, and result in the omission of items, administration of items out of order, or misreading of instructions.

The questionnaire should begin with an introduction for the respondent, written on a separate page to allow him to read it before glancing at the questions. The introduction should explain fully how the questionnaire should be answered giving an example, and it should also give some explanation of the purpose of the questionnaire. This must be very general or it may introduce bias. The introduction should include the name of the relevant organizing body; the name of a professional organization is usually more acceptable than the names of individuals. Nixon (1954) has recommended that the word

'questionnaire' or 'checklist' should not be used to describe it because of the commercial connotations, but rather the word 'form' or 'instrument'. There should be a section for identifying information, such as name, National Health number, etc., although this should preferably be placed at the end of the questionnaire. Illustrations may emphasize purpose or clarify items in a questionnaire although care must be taken that they do not bias the respondent by suggesting that certain responses are more desirable.

Printing allows the use of different sizes and styles of type to distinguish questions from instructions, etc. Questions should be clarified by the use of lined insets to help guide the interviewer or respondent through the form so that he will not write responses in the wrong places or omit subsidiary questions. The paper on which the questionnaire is written should be strong as it will receive a great deal of handling. Page size should be selected so that the questionnaire folds easily into a standard size envelope should it need to be posted. The paper should be light-coloured to show the print off to the best advantage. Some researchers have recommended the use of coloured paper as being more attractive to respondents. Eastwood (1940) found yellow paper to have the highest percentage of returns in a postal survey, closely followed by pink. The alternation of light colours appeared to elicit more returns when a series of questionnaires was used than did just a single colour. Dark colours were found to be ineffective.

If the data are to be coded for machine processing, it is often an advantage to design the questionnaire accordingly. The simplest method is to provide the appropriate numbered coding boxes alongside each item of data. Then after the questionnaire has been completed, responses and identifying data can be coded into numerical values for key punching on to 80-column punch cards. Alternatively, the layout of the questionnaire may allow direct coding of the data, thereby eliminating this as a separate procedure. This has the advantage of removing a possible source of clerical error: however, the success of either procedure in terms of efficiency and accuracy depends entirely on design.

In conclusion, it must be added that in designing a questionnaire one cannot give too much emphasis to the importance of rigorously pretesting the forms. The objective is a simple one: to achieve the layout that reduces to a minimum errors by interviewers, respondents, coding clerks, or key punch operators.

2. The evaluation of a questionnaire

After having designed a questionnaire appropriate to the needs of the investigator, it is necessary to evaluate its potential usefulness as an instrument of measurement. For this it must be assessed in terms of two criteria: validity and reliability. No questionnaire will be perfectly valid or reliable but the extent to which it is provides the means of predicting its potential usefulness.

Validity

Validity refers to the efficiency with which a measuring instrument measures what it purports to measure, and as such may be seen as a function of three variables:

Relevance: referring to whether the questionnaire obtains the type of information it was designed to obtain.

Completeness: referring to the extent to which a questionnaire has collected all relevant information.

Accuracy: which is a measure of the reliance that can be placed upon the response to each question. It is an indication of the proportion of times that an answer to a question will be correct (Young, 1971).

There are several types of validity which may be discussed with reference to medical questionnaires, each of which is dependent upon the above factors:

(*a*) **Face validity.** This refers to whether a questionnaire looks as though it measures what it is supposed to measure. This is obviously not a very good indicator of potential usefulness although Ley (1972*b*) points out that a questionnaire might be more acceptable to patients if it looks relevant to their problems.

(*b*) **Content validity.** This is the extent to which the questionnaire adequately probes the various aspects of the area it is supposed to measure. This relates to the idea of 'completeness' and is usually tested by reference to clinical experience.

(*c*) **Construct validity.** This refers to whether the questionnaire results are in accord with present theories concerning the relevant areas

of research; for example, do the resultant symptom patterns concur with physiological theories of causality? Construct validity may be assessed by comparison of questionnaire data with the theories expounded in the relevant medical literature.

(*d*) **Predictive validity.** This assumes that a questionnaire has some value as an instrument of prediction—that it, future health developments can be predicted from questionnaire results. This may be assessed by a follow-up study. A questionnaire which was evaluated in terms of its predictive validity was described by Henderson *et al.* (1961). The investigators attempted to predict strokes by identification of transient ischaemic attacks. Although the questionnaire could detect quite accurately the occurrence of attacks, the data obtained were unable to predict strokes. As an instrument of prediction, the questionnaire was, therefore, of low validity (Henderson, 1972).

(*e*) **Concurrent validity.** This is perhaps the most important type of validity in that it refers to the practical usefulness of the questionnaire as an instrument of identification. It is the type usually referred to when investigators discuss the 'validation of questionnaires'. It comes closest to the notion of validity as initially defined in that it refers to the extent to which questionnaire results agree with an independent external criterion (that is, an independent measure of the same variable as that which the questionnaire is investigating). Fairly reliable physical measures may be available for use as criteria, such as chemical and biological tests or X-rays. More usually, however, the investigator has to rely upon less trustworthy criteria, such as the results of a clinical interview, or another questionnaire on the same subject. Examples of the type of criteria used by investigators to validate medical questionnaires may be cited from the literature. Crown and Crisp (1966) used two criteria to establish the validity of their Middlesex Hospital Questionnaire for psychiatric patients. Firstly, they compared the results of sub-tests of the questionnaire with previously made group classifications, and secondly they compared questionnaire results with clinical ratings. Glazunov (1968) assessed the validity of a questionnaire for angina by determining the extent to which answers to the questionnaire were related to physician inquiries and to electrocardiographic examinations. Items in the questionnaire on respiratory symptoms (MRC, 1966) were assessed against a variety of criteria, for example, volume of sputum production.

The use of an external criterion for validity assessment gives rise

to a paradox which has been pointed out by Anastasi (1961): if it is necessary to follow-up subjects or to obtain an independent measure in other ways why not dispense with the questionnaire?

It is, however, only during the initial stages of pretesting questionnaires that the criterion need be used; afterwards the questionnaire will be used alone. Furthermore, several lengthy procedures may be necessary to validate the questionnaire whereas the questionnaire itself should represent a more economical and concise means of measuring the same variable.

The concurrent validity of a questionnaire may be expressed:

(*a*) **In terms of a correlation coefficient,** which is an expression of the relationship between the results obtained by the questionnaire and those obtained by the criterion. This relationship is expressed as a numerical value between -1 and $+1$, where 0 indicates no correlation (or relationship), $+1$ a perfect positive correlation and -1 a perfect negative correlation (that is, questionnaire and criteria are inversely related). A correlation of more than 0·6 is usually considered to represent a useful positive relationship (although the significance of the correlation will depend upon the size of the population tested). Various correlation techniques are available to suit the type of data used; an outline of these is given by Guilford (1950).

(*b*) **In measures of sensitivity and specificity,** which describe the questionnaire's ability to discriminate those who have a particular condition from those who do not. Sensitivity and specificity may be expressed in quantitative terms where sensitivity refers to the ability of a questionnaire to detect all cases of the relevant variable and specificity to its ability to discriminate them from those cases which do not possess the relevant variable. These values may be calculated for a given questionnaire by means of the following formulae:

$$\% \text{ Sensitivity} = \frac{\text{No. of persons with positive result on questionnaire}}{\text{No. positive cases detected by criteria}} \times 100$$

$$\% \text{ Specificity} = \frac{\text{No. of persons with negative result on questionnaire}}{\text{No. negative cases detected by criteria}} \times 100$$

Questionnaires are unlikely to be 100 per cent sensitive or specific, and the limits of acceptability will depend upon the use to which the questionnaire is to be put. It should be noted that sensitivity does not imply specificity: one is often gained at the expense of the other so that the determination of the 'cut-off' point (that is, the point below which all cases are said to be negative and above which all positive)

becomes a difficult problem. The designation of a cut-off point usually involves a degree of error as symptoms are rarely simply present or absent, but lie along a continuum from positive to negative. The errors incurred by the inaccuracy of the cut-off point are termed *false negatives* (when respondents who possess the relevant condition are classified by the questionnaire as not having it) and *false positives* (when respondents not possessing the relevant condition are classified by the questionnaire as having it).

In a research situation such as a disease prevalence study, it is usually better to raise the cut-off point to its uppermost limit to lower the risk of false positives which would exaggerate the findings. For example, Garrad and Bennett (1971) in their questionnaire study of chronic disease and disability define disability in strict terms with a resultant higher proportion of false negatives (8·5 per cent) in identification of area of disability and a low proportion of false positives (0·9 per cent). In a screening situation, however, it is usually better to lower the cut-off point to its lower limit and lower the risk of false negatives: that is, it would be better to assume the respondent has the relevant condition and refute this at a later examination than to ignore borderline cases altogether. For this reason general health questionnaires, such as that used by Gledhill and Matthews (1971) require respondents to answer 'Yes' to questions if in doubt so as to detect all possible cases of a particular condition. Later examination eliminates false positive responses. An experimental analysis of the validity of the 'health questionnaire' was carried out by Suchman *et al.* (1958) by comparing a self-administered health questionnaire with physicians' ratings. They concluded, somewhat sweepingly, that self-rating medical questionnaires have low validity. In reply it must be observed that the investigators used a very crude form of questionnaire and a highly imperfect criterion.

Reliability

In questionnaire construction it is important to ask not only whether the relevant variable is being measured accurately, but also how reliable are the measurements.

This is essentially a measure of the questionnaire's ability to distinguish the extent to which a variable apparently fluctuates as a result of errors of measurement as opposed to real changes in the object of measurement itself. Guilford (1950) has described reliability

in terms of the analogy of the yardstick. If the yardstick is made of rubber and changes its length unpredictably then any measurement will be meaningless: it will give no useful information about the object of measurement or its relationship to other objects.

Basically there are two types of reliability, internal consistency, and repeatability.

(*a*) **Internal consistency.** This refers to the consistency of answers to items in the questionnaire and an aspect dealing with paired comparisons was discussed earlier (see p. 10). More usually its assessment is described by a coefficient of internal consistency which is obtained by correlating one half of the questionnaire responses with the other half: the so-called 'split-half correlation'. This can only be meaningfully applied to questionnaires composed of homogeneous items.

From a theoretical standpoint, Guilford (1950) considers that the measurement of internal consistency comes closest to the central notion of reliability in that it derives from a single administration of the questionnaire, so that no time elapses between measurements during which the object of measurement might undergo change. Because of the variety of items used in medical questionnaires and the differences in the diagnostic values of individual items, measures of the internal consistency of conventional health or symptom questionnaires are difficult to assess.

(*b*) **Repeatability.** This has been defined as the extent to which the questionnaire provides the same results on the same subject on two or more occasions, with the same or another observer, the subject of the test being in the same state of health (Witts, 1959). A 'coefficient of stability' or 'test-retest repeatability' may be calculated. This is determined by readministration of the questionnaire after some period of time and correlation of the results. This would seem to be the most obvious means of assessing reliability. To recall the yardstick analogy, if the stick measures the same object twice and the measurements are the same, then it would be considered reliable. In the case of questionnaires, however, the object of measurement is a characteristic which may be continually changing although the instrument of measurement remains constant, and it may be difficult to tell whether changes in the results on retest are due to poor repeatability of the questionnaire or actual changes in the individual. The repeatability of a questionnaire is an important consideration for medical research in that one is dealing with the variable of 'health' which undergoes

considerable changes. The ability to measure the rate and direction of this change is an important one and can only be achieved if the instrument of measurement is able to give consistent measures unimpeded by irrelevant variables.

Poor repeatability in medical questionnaires may arise as a result of questionnaire design (for example, through the ambiguity of items) or it may be due to factors relating to the respondent. For example, Collen *et al.* (1969) found women to be more consistent than men on retest. Lawton *et al.* (1966) claimed on the basis of a follow-up study by questionnaire of St Louis encephalitis that an average 'normal' person is more consistent in answering health-related questions than emotionally unstable individuals who may give further proof of their instability by the great variation apparent between their answers in the test and retest situations.

The test-retest repeatability of a questionnaire is also a function of the amount of time which is allowed to elapse between test and retest. Collen *et al.* (1969*a*) maintain that this period should be a very short one and allow thirty minutes only for retest of their general health questionnaire. However, a short intertest interval gives rise to difficulties in that the answers given by the respondent may not be the result of his reaction to the question but rather of his having remembered the responses which he gave at the first administration, so that information is carried over from one administration to the other. On the other hand problems may arise if a long interval is allowed due to the multiplicity of changes which may take place in the respondent. For example, Guilford (1950) reported the results of a motor test which had a test-retest correlation of 0·7 after an interval of a year. This relatively low correlation was due mainly to the maturation of motor responses in the respondents and other genuine changes in the variables being measured, and was, therefore, an underestimate of the repeatability of the test.

Several investigators have noted that the data obtained on retest of a medical questionnaire differ significantly from that of the initial administration. With regard to quantity of information, the repeat questionnaire appears to produce less than the original administration. For example, Glazunov (1968) found on re-administration of an angina questionnaire that there was a significant reduction in the amount of positive symptoms recorded to the extent that only half the population originally classified as positive were reclassified as positive on retest. Collen *et al.* (1969) noted a considerable

diminution in the amount of positive symptoms scored on retest and concluded that respondents were more likely to change their second answer when their first answer was 'yes' than when 'no'. This reduction in information may be attributed to a more discriminating view taken by the part of the respondent on retest. However, Cannel *et al.* (1969*a*) maintain that on retest their administered general health questionnaire obtained consistently less valid information than original interviews and concluded that the frequently made assumption that reinterview data are better than first interview data can be called into serious question. They found that use of short questions and reinforcing statements lessened the decrease in quality on retest although not preventing some decline.

The repeatability of a questionnaire may be checked statistically by correlation of test and retest data. Alternatively, repeatability may be assessed by calculating a coefficient of equivalence derived from a comparison of two equivalent versions of a questionnaire administered at essentially the same time. This method overcomes the problem of carry-over of information of the test-retest method. Complications arise, however, firstly due to the difficulty of deriving an equivalent version of the same questionnaire, and secondly with the difficulty of distinguishing whether discrepancies in the results of the questionnaires are due to low repeatability or to differences in the two questionnaires, such as wording, question sequence, and so on.

Item assessment

One might expect a highly valid questionnaire to be highly reliable also, but it should be noted that this is not necessarily the case. For example, a clock which consistently loses ten minutes in every hour is reliable but not valid as a measure of the time of day or of time elapsed. It is possible then that a questionnaire might continually yield inaccurate data in such a way that it is reliable but invalid. Also high validity is usually obtained when a wide variety of information is sampled by heterogeneous items, but as was discussed earlier this makes internal consistency difficult to assess. These points suggest that not only must a questionnaire be evaluated as a whole, but individual items must also be assessed in terms of their relative contribution. Item assessment takes into account the fact that the content of each item is not of equal value as an index of health status in that a positive response to one item in a questionnaire may be far more

significant to the investigator than a positive response to another.

The discrimination of items in terms of their potential significance may be made by, for example:

(*a*) Reference to clinical experience. For example, Young (1972) distinguished 'major' and 'minor' symptoms on the assumption that items recorded in the history of the presenting complaint would be more important than those in the clinical review.

(*b*) Value judgements by panels. For example, the relative importance of individual items may be determined by a series of value judgements given by a group of physicians (Koller, 1967).

(*c*) Concurrent validity of individual items. This method was used by Bennett and Fraser (1972) on their general health screening questionnaire.

(*d*) Reliability of individual items. Collen *et al.* (1969) differentiated the symptoms of their medical history questionnaire by test-retest repeatability for each item.

It should also be taken into account that whereas items may individually be of little value, in combination their value may be greatly increased. The development of the Cornell Medical Index (Brodman *et al.*, 1956) illustrates the need to consider item patterns as well as individual questions. The Cornell Medical Index consists of 195 items, scored 'Yes' and 'No'. These items have been 'grouped' in sections. If a respondent answers 'Yes' to more than 25 items of the questionnaire a serious disorder is to be suspected. However, the researcher must note the distribution of these 'Yes' responses. If they occur mainly in one or two sections then the medical problem is probably a localized physical disorder. If the responses are scattered throughout the questionnaire then some form of emotional disturbance is to be suspected. It is important to consider patterns of response within each section rather than individual items, especially as it has been shown that single items referring to specific disorders (for example, 'Do you suffer from asthma?') have low validity (Abramson, 1966).

Another example of the need to consider 'grouping' of items is given by Bennett and Fraser (1972). The following questions in their general health screening questionnaire are of little value individually, but taken together may provide a useful indicator of emotional disorder.

Question no. 4. Have you found it difficult to concentrate lately?
5. Does your hand often shake when you try to do something?
9. Do you sweat very easily, even on cold days?
14. Do you suffer from palpitations or breathlessness?
15. Do you ever have blackouts, dizzy spells or faints?
23. Do you have difficulty in getting off to sleep?
27. Are you troubled by waking in the early hours and being unable to get off to sleep again?
28. Are you more absent-minded recently than you used to be?

The probability of the respondent being emotionally disturbed increases in proportion to the number of items scored positively.

The discriminative value of questionnaire items may be increased by requiring a set pattern of responses. For example, with the question sequence:

Do you ever have pain or aching in your stomach?
If 'Yes', is it relieved by eating?
is it relieved by drinking milk?
does it often wake you at night?

The respondent must answer 'Yes' to the first question and at least one of the following three questions before he is considered to have symptoms of a possible peptic ulcer (Bennett and Fraser, 1972).

Alternatively, the investigator may assign numerical weights to individual items to indicate their discriminative ability; the higher the individual's total score the greater the likelihood of abnormal health status. This is shown by the questionnaire for detection of registerable blindness developed by Graham *et al.* (1968). This questionnaire consists of eight items, each of which has been designated a score according to its discriminative ability, as in the following table:

Question no.	1	2	3	4	5	6	7	8
Yes	1	–	–	–	–	2	–	–
No	–	–	3	1	2	–	3	1

This coding table implies that a negative reply to question 3, for example ('Can you read the small print in a newspaper?') is a better indicator of visual disability than a negative reply to question 4 ('Can you thread a needle?').

The following cut-off points were then determined for the questionnaire as a whole:

0– 2 = negligible probability of substandard vision
3– 8 = high probability of substandard vision
9–13 = reasonable probability of registerable blindness.

These cut-off points refer to the individual's cumulative score. It can be seen from this that the total score of an individual is dependent

not upon the number of positive responses given, but rather upon the content of the items to which he responded positively. If an interval scale is used, weights may be determined by reference to symptom severity.

The assessment and discrimination of items is important as the validity of the questionnaire as a whole is affected. This was nicely demonstrated by Crown and Crisp (1966) who, in the pilot stage of developing their psychiatric questionnaire determined emotional stability by the number of items scored positively. They found that this method produced a high number of false positives and false negatives. The concurrent validity of the questionnaire was greatly improved when items were weighted so that the total score became a function of the content of positive items.

The scaling of items has been previously discussed in earlier sections dealing with question type (see p. 6) and internal consistency (see p. 32). Guttman scaling or scalogram-analysis (Guttman, 1950) is a refined form of item assessment and a rather sophisticated approach to the concepts of reliability and validity. The technique was originally developed for attitude scaling but it is more widely applicable. It is, for example, appropriate for scaling physical aspects of human performance (Williams *et al.*, 1975), mental performance, and symptom complexes.

The technique allows a group of items to be examined for 'scalability'. The objective is to order the items on a unidimensional, cumulative scale and for this to be achieved a high (usually 0·9 or over) 'coefficient of reproducibility' is required. Given this certain statements are possible. An individual's responses to the items allow him to be assigned to a position on the scale: and further if he shows any improvement or deterioration in the attribute in question his progression can be predicted.

Thus the technique of Guttman scaling is largely concerned with reproducibility which is composed in this instance of both reliability in terms of internal consistency and repeatability and validity. By contrast with most other situations, concurrent validity is a less important consideration and emphasis must be placed on face and content validity: predictive validity is an essential requirement.

3. The administered questionnaire

The administered questionnaire is one which is read out by an interviewer, to which the respondent replies orally and the response is noted down by the interviewer. The administered questionnaire may be broadly classified in terms of whether it is 'standardized' or 'unstandardized' (these two classifications may alternatively be termed 'structured' and 'unstructured', 'formal' or 'informal', or 'passive' and 'active' techniques). The essential difference between these two types is that the standardized questionnaire is based on a series of questions which have been determined in advance of the interview, and which are asked in the same order and with the same wording for every respondent. The interviewer is allowed no initiative with regard to the introduction of additional items or probes which might be relevant to the individual case, nor is he allowed to reword questions to improve communication, or to introduce irrelevant comment for the sake of rapport with particular respondents. His role is strictly limited by the content of the questionnaire.

The unstandardized questionnaire on the other hand allows the interviewer flexibility with regard to questionnaire administration. Thus the questionnaire is in the form of an interview guide comprising the areas which the interviewer must cover. He is at liberty to reword questions to suit individual respondents, to change the order of questions to conform with the subject's spontaneous flow of thought and to introduce supplementary questions or probes when it is felt that further useful information might be found. This type of questionnaire was used by Acheson *et al.* (1969) in their survey of joint disease.

Both styles of interview have their advantages and disadvantages. The standardized interview has the advantage of constituting a standard instrument of measurement which allows for the comparison of information derived from various sources and gives good results on test-retest repeatability. However, as has been discussed, reliability does not necessarily imply validity and the greater adaptability of the

unstandardized procedure might in some circumstances produce more valid information as it allows a more equal level of apparent understanding (Maccoby and Maccoby, 1954). This is because the interviewer is able to reword items and introduce probes to try and ensure that the respondent has correctly understood what information is required. But this flexibility, however, also allows greater opportunity for the interviewer to introduce his own attitudes and opinions thus providing an important source of bias in the data. Furthermore, minor changes in the wording of items have been demonstrated to result in significant changes in response frequency (Cantril, 1944). For most purposes the standardized procedure is to be preferred.

Some researchers have attempted to overcome the problems of these two different approaches by mediating between them and developing a semi-standardized questionnaire which, while maintaining a standardized structure, also includes optional sub-questions to be used at the discretion of the interviewer. An example of this type of questionnaire is the Medical Data Index described by Wynder and Schottenfeld (1969). This questionnaire was originally designed as a fully structured self-administered questionnaire but it was later found desirable to introduce a trained paramedical interviewer to probe items more fully in order to obtain sufficiently detailed information. A further form of the semi-standardized questionnaire is the focused interview described by Merton *et al.* (1956). This type of interview is designed for subjects known to have been involved in the past in a particular situation. This situation will have undergone a content analysis by which hypotheses concerning determining factors of the situation are postulated. An interview guide is then constructed which sets forth the major areas of investigation based on these hypotheses. The interview is then focused on the subjective experiences of respondents involved in these pre-analysed situations in order to determine their definitions of it. This type of interview might be used, for example, in post-treatment follow-up of individual reactions to controversial therapy where therapy has involved an unusual experience followed by a return to normal physical status. For example, a focused interview might be constructed for patients who have recently had abortions in order to reveal the special needs of such patients.

Cannell *et al.* (1967, 1969*a*, *b*) have undertaken a series of investigations concerning the relative efficiency of various forms of administered questionnaires used in household health surveys. Their study

involved a comparison of three administration techniques. The first aimed at stimulating the memory of the respondent with regard to past illnesses by means of a sensitization task (a symptom list) completed before the beginning of the main interview schedule. The second technique attempted to improve recall by rewording or reinforcing it. This reinforcement took the form of statements such as:

We need to know things like that.
We're interested in that.

These statements were made by the interviewer following the respondents' answers to questions. This technique also involved the use of extra words in questions, as for example:

Now I'm going to ask you questions about your health. By asking these questions, the Public Health Service can get a good picture of the nation's health. And to make the information valuable, it's important that you report all your sicknesses, no matter how small or unimportant they may be. Have you ever had: (List of symptoms follows).

Interviewers were also instructed to engage in reinforcing behaviour such as smiling, leaning towards the respondent, and using natural hand gestures. The third technique consisted simply of the standardized interview, identical to the one used in the other two methods but not employing either sensitization or reinforcing methods.

The researchers found on comparing the data by the different methods that the use of a reinforcement technique significantly increased the quantity of material obtained in a health interview: the sensitization procedure used in this study was found to have no effect. An attempt was then made to assess the quality of this additional data by comparison with physician ratings and it was found to have a high validity only when the technique of longer questions was not used. In summary, these experiments with special techniques of administering questionnaires indicate that maximal validity may be obtained by the use of motivating techniques (such as reinforcing statements and encouraging behaviour) in combination with short questions. It thereby stresses the importance of the interviewer–respondent relationship in obtaining valid data in health studies.

Interviewer–respondent interaction

The quality of the relationship which is developed between the interviewer and respondent is important as it may encourage the respondent to talk more honestly and openly of his experience or

alternatively cause him to conceal or distort his experience. This was demonstrated by Kahn and Cannel (1957) for those taking part in a health interview survey using an administered questionnaire. Postal questionnaires were sent to the respondents after the interview to investigate their reactions to it. Their replies indicated that what had impressed most was not the subject of the survey or the type of questions asked, but rather the nature of the relationship which had been established with the interviewer. This relationship may be seen as the outcome of an interaction between various factors. Cannell *et al.* (1968) have suggested that attitudes, expectations, appearances, and behaviour of both the interviewer and respondent can influence the accuracy of the collected data, and of these, overt behaviour in terms of encouraging gestures, smiling, etc., as previously mentioned, is the most important.

The interaction of interviewer and respondent may have negative as well as positive effects. On the positive side, as well as those effects already described, from the interviewer's point of view a good relationship with the respondent may permit her to judge the quality of the responses given. Kinsey *et al.* (1953) comment:

> It has been asked how it is possible for an interviewer to know whether people are telling the truth, when they are boasting, or otherwise distorting the record. The experienced interviewer knows when she has established a sufficient rapport to obtain an honest record.

On the negative side, the relationship between interviewer and respondent may be detrimental to the quality of the data obtained. For example, Verplanck (1955) and Greenspoon (1955) have demonstrated that 'operant conditioning' of verbal behaviour may occur as a result of such a relationship. That is to say, if the interviewer reacts encouragingly to certain words by saying for example 'mhmm' or 'good', then those words will be used more often. Such procedures have been shown to increase, for example, the number of plural nouns, hostile words, and statements of opinion which the respondent uses. Furthermore, the respondent is apparently quite unaware of such influences. Hallauer (1972) demonstrated how such verbal conditioning could affect reported illness behaviour in the context of a health survey. These findings point to the danger of a 'manufactured relationship' (Rogers, 1961) in which the interviewer is the dominant element in the interview and the respondent's role is essentially a passive one. When the interviewer is in such a dominant position as this the respondent may exhibit dependent responses such as

asking for help, advice, or praise (Maccoby and Maccoby, 1954). The interviewer must then deflect such demands without seeming to reject the respondent and hence place a barrier to further communication. Another problem which may arise from the interviewer–respondent relationship is an attempt on the part of the respondent to preserve that relationship by presenting as favourable an image of himself as possible. In this case he may deny any experience which he feels the interviewer might find unacceptable.

What are the characteristics of a good interviewer–respondent relationship? Rogers (1942), in his discussion of the importance of this relationship for psychotherapy, mentions several desirable characteristics which might be relevant to questionnaire administration. He maintains the need for warmth and responsiveness, interest in and acceptance of the respondent, permissiveness of expression, and freedom from pressure and coercion. Rogers stresses that the interview should be a reciprocal relationship in which the 'psychological fields' of both interviewer and respondent are constantly being modified, avoiding the type of 'manufactured relationships' discussed above. Kinsey *et al.* (1953) maintain that the interviewer–respondent relationship is most under strain at the beginning of the interview so at this time the interviewer should concentrate on putting the respondent at ease. This should be done by discussion of the subject of the interview rather than by personal questions.

The selection and training of interviewers

One of the greatest problems with the administered questionnaire is that it introduces another variable into the question–response situation, that of the interviewer. The use of an interviewer in questionnaire administration allows for an increase in flexibility. The interviewer can modify his approach and change his tone of voice to suit the attitude of the respondent. The interviewer may serve to give the questionnaire a more personal aspect and improve response rates by the establishment of an encouraging, response-inducing relationship, that is rapport: 'that elusive motivating force that will keep the respondent interested and responsive to the end of the interview' (Kahn and Cannell, 1957). On the other hand it must be remembered that flexibility is gained at the expense of administrative standardization. The researcher may not be able to tell whether diverse findings are attributable to individual differences among respondents or whether

they are the result of a particular interviewer style. This is the problem of interviewer variation. The problem first arose when researchers observed that answers reported by one interviewer differed significantly from those reported by another interviewer, although the populations interviewed were similar.

Differences in results obtained by interviewers may be attributed to:

1. Variations in the personal characteristics of the interviewers.
2. Deviations by the interviewer from the standardized procedure.

The first problem must be dealt with by careful selection of interviewers, and the second by careful training of interviewers.

Personal characteristics of interviewers

Personal factors such as sex, age, and race of interviewers have been shown to have an important effect on the data collected.

The influence of the sex of the interviewer has been demonstrated by a National Opinion Research Centre interview survey conducted in 1947 which was concerned with sexual behaviour (Hyman *et al.*, 1954). It was found that unsure subjects sympathized to a significant degree with problems relating to the interviewer's sex rather than their own. Sheatsley (1950, 1951*a*, *b*), analysing data obtained from interviews conducted by the National Opinion Research Centre, also concluded that the sex of the interviewer had had a significant effect; the results obtained when interviewer and respondent were of the same sex differed from those obtained when interviewer and respondent were not of the same sex. The researchers concluded that this was because the communication of personal data was facilitated when interviewer and respondent were of the same sex.

These studies identify the tendency for respondents to formulate their opinions in accordance with the sex of the interviewer. It would thus seem advisable to use interviewers of the respondents' own sex where possible, particularly when the information sought is of a very personal nature. This suggestion is supported by the results of a survey conducted by the then Ministry of Health in Great Britain (Hyman *et al.*, 1954) in which all the interviewers were women. The aim of the survey was to collect evidence for a campaign against venereal disease. The interviewers described 14 per cent of male respondents as being embarrassed, shy, or nervous, whereas only 8 per cent of female respondents were described as acting in this way.

However, the influence of the interviewer's sex on questionnaire data should be examined by pretesting and not simply assumed. For example, researchers concerned with industrial relations have frequently assumed that women would be less successful than men as interviewers in this area. The Survey Research Centre of the University of Michigan has demonstrated that this is not in fact the case. They found that women were generally more successful than men for obtaining personal information (Kahn and Cannell, 1957). The reason for the success of female interviewers in what is traditionally considered a male environment was presumably that the respondents did not associate them with the power hierarchy and, therefore, felt able to talk more openly.

The appropriate age for an interviewer is another factor to be considered. From their research into this question Hyman *et al.* (1954) concluded that the quantity and validity of interview data is greatest when interviewer and respondent are approximately the same age. This would seem appropriate in that to some extent persons of the same generation would be more familiar with each others' experience and percepts. However, such a situation is not always feasible as, for example, in the interviewing of children and elderly persons. As Brown (1937) explains:

> Investigators who are too young often lack tact and the ability to approach people, while those who are too old are often not willing to follow instructions carefully, or not aggressive enough to obtain an interview.

An age range for interviewers of 25 to 45 years seems desirable.

The general appearance of the interviewer is a potentially important factor. Many of the decisions made during everyday life concerning other persons are based upon immediate impression. A respondent may, therefore, regard as untrustworthy an interviewer who appears irritable, impatient, or condescending. A glamorous face may be as great a barrier to communication as a sinister frown. Parten (1950) has said the following concerning the appearance of interviewers:

> Interviewers should be clean and neat in appearance . . . and not wear badly spotted clothes. It is well to avoid appearing so neat that the housewife refuses admission to her disorderly home. Similarly, it is best not to appear too prosperous.

Brown (1937) has produced evidence to suggest that the plainly attired interviewer tends to be the most successful one.

However, both of these references are over twenty years old and the last decade has seen considerable changes of style and appearance

in younger generations. Furthermore, many of these changes are not liked by older people who consider that they reflect standards of behaviour unacceptable in professional workers. By contrast, the younger generations tend to reject the conventionally attired young person considering them to be different and, therefore, lacking understanding. These changes emphasize the need for approximating the ages of interviewers and respondents and for training young interviewers in the need for dressing and appearing in a less trendy or generation-characteristic way when interviewing older people.

The interviewer's social status may affect the responses to an administered questionnaire as class differences and 'social distance' may influence the communication process. This has been shown by Back *et al.* (1950) and Riesman and Glaser (1948) who found that lower-class respondents were more pleased to be asked to discuss their problems and opinions than upper-class respondents who tended to feel that they had little to gain by talking to an interviewer. Also Katz (1960) found that interviewers from working-class backgrounds consistently obtained more radical social opinions from respondents than interviewers from the middle class. However, class differences may give rise to other reactions. A lower-class respondent may resent the authority assumed by the interviewer or may distort his responses if he feels the interviewer is in a position of influence. He may be reluctant to reveal his socially unfortunate circumstances to an interviewer perceived as middle class. As an example, in one social survey some 10 per cent of the families interviewed failed to report that they were receiving benefits (Neely, 1938).

Ideally it would be best to choose interviewers of the same social class as the respondents but a main problem here is how to determine an individual's social class. It may be indicated by factors such as dress and appearance, but as commented earlier there are now large generational differences. Kinsey *et al.* (1953) have stressed that speech in particular is an important indicator of class membership. They maintain that both accent and the use of slang or jargon are telling and may require special understanding on the part of the interviewer. Where there is also a racial difference between interviewer and respondent, the problem is particularly acute. In the National Centre for Health Statistics survey of hospitalization it was found that, in a variety of procedures conducted by white interviewers, the rate of under-reporting was consistently higher for white than for non-white respondents. Where racial differences exist there

is a tendency for respondents to tailor their replies in accordance with the social background of the interviewer. Again in a study of negroes using both negro and non-negro interviewers it was found that the latter obtained significantly less information on racial discrimination. Another study demonstrated that increasingly less hostile opinions concerning Jews were expressed as interviewers with more definite Jewish characteristics were used (Hyman *et al.*, 1954).

Interviewing persons from another society involves many additional problems. In the first place the interviewer must gain the acceptance of that society if he is to learn personal information. This is greatly facilitated if he can be introduced by a person already trusted by potential respondents. He must then be prepared to submit his own social values to those of the society which he is investigating. Maccoby and Maccoby (1954) report the extreme example of a field-worker in an Afghanistan village who had to wrestle with the village wrestling champion and take high hurdles on horseback before he was able to interview the villagers. Further, the interviewer must be prepared to discard the presuppositions of his own society. For example, he should not take for granted holidays and feast days, patterns of child-rearing and education, and other customary features of family life. In some societies commonplace symptoms such as cough are considered socially unacceptable and Maccoby and Maccoby (1954) have also pointed out that whereas in western society it is customary for the interviewer to look directly at the respondent while asking a question, in other societies this may be considered impolite.

The attitude of the interviewer towards the questionnaire study itself may be an important consideration. An interviewer who is pessimistic from the beginning as to the success of his interviews may well be unable to produce sufficient data. Maccoby and Maccoby (1954) maintain that interviewers who are confident and expect to get respondent co-operation usually have the fewest refusals. Similarly, with regard to interpretation of questionnaire responses, the interviewer who has little interest in his work may not make the necessary effort to understand. Quinn (1950) found after examining a series of psychotherapeutic interviews that 'understanding' of the respondent's meaning is essentially an attitude of 'desiring to understand'. Attitudes of interviewers towards their work might be examined by means of a work satisfaction index such as the SRA Employment Inventory or the Job Descriptive Index. These and other means of measuring attitudes to work are described by Smith *et al.* (1969).

Interest in the influence of interviewer variation on data collection has led many investigators to study the relationship between interviewers and the type of responses they tend to elicit. It has been found that in many cases that interviewers tend to elicit replies similar to their own views. This was nicely demonstrated by Rice (1929) who found in a study of the desitute that respondents questioned by an interviewer with prohibitionist sympathies tended to blame alcohol for their circumstances, whereas respondents interviewed by an interviewer with social sympathies tended to blame industrial conditions.

Thus, it may be concluded that variations in interviewer attitude can influence questionnaire response. Sheatsley (1951*c*) has pointed to several disqualifying personal characteristics of interviewers. These include intense partisanship, a long background of undesirable experiences (such as selling), and any tendency to regard interviews in general as therapeutic devices.

Standardization of interview procedure

Failure by an interviewer to adhere to the standardized administration procedure has been shown to constitute a common source of error in studies using administered questionnaires. Fairbairn *et al.* (1959) found considerable variation in the responses given to a respiratory symptoms questionnaire. They came to the conclusion that this was almost always a result of failure by the interviewer to keep to the rules.

A prime source of interviewer error arises when the interviewer introduces subsidiary questions or probes to clarify the respondent's initial replies. Where the interviewer is encouraged to initiate probing at his own discretion, considerable variations may occur. Feldman *et al.* (1951) showed large variations in the number of answers obtained by different interviewers due to the fact that some probed more than others and so gained more information. Probes formulated spontaneously by the interviewer may also be leading or loaded, or stress the interviewer's attitude to the subject so influencing the response obtained. It is possible, even easy, to change the meaning of an item without actually changing the wording by the stress given to different words. Oppenheim (1968) exemplifies this by the question 'Why do you say that?' The question 'Why do you say *that*?' constitutes a challenge to the respondent's reply, whereas 'Why do *you* say that?' is a more personal inquiry.

Wyatt and Campbell (1950) maintain that interviewers' expectations of the frequency distributions of the responses will affect the results obtained. This has been demonstrated by Fairbairn *et al.* (1959) to be relevant to medical questionnaires. In comparing responses to a respiratory symptom questionnaire, they found that doctors tended to report significantly more symptoms than did lay interviewers. They further commented that health visitors or lay interviewers starting with fewer fixed beliefs about the answers they will get seem to be able to be trained to report more consistently than doctors. Not only may the personal characteristics and behaviour of the interviewer influence both the quantity and quality of data obtained as was previously discussed, but to this must be added the consideration that interviewers' perceptions may be biased by their background or experience.

Interviewer variation has also been demonstrated to occur as a result of faulty coding of responses. In a study aimed to assess the effect of interviewer error in the recording of data, Guest (1947) recorded interviews conducted by 15 college students all questioning the same respondent. Comparing the recordings with the data collected by the interviewers he found 32 errors of commission (where something was recorded which was not said or said differently) and 60 errors of omission. Additionally, 41 cases were found in which probes had not been differentiated from the regular questions in the write-up. Stewart and Flowerman (1949) examined recorded interviews and found many instances where items had not been asked at all although responses had been recorded by the interviewers!

Interviewer variation cannot be totally controlled: even if the same interviewer is used he may vary his technique between respondents. It can, however, be reduced by training. The aim of an interviewer is to attain uniformity in the asking of questions and recording of answers. In consequence, the training of an interviewer is orientated towards efficiency in following instructions (Moser, 1968). The means that should be employed are:

1. Supervised practice interviews where observed faults can be pointed out and corrected. A tape-recorder is particularly useful for this purpose as the trainee can then listen to recordings of his own performance in terms of criticisms which have been made.

2. Interpreting and coding responses from recorded interviews. These codings can then be compared to those of more experienced interviewers and any differences identified and discussed.

3. The need for strict standardization should be explained. The interviewer must be impressed with the importance of asking each question exactly as it is worded on the form and understand that even a slight change of wording might change the meaning of an item. It should be emphasized that the questionnaire functions as a whole with each item dependent upon the others, so there must be no errors or omissions.

4. Interviewers should be warned of any foreseen difficulties. A handbook should be written which covers problems such as the treatment of ambiguous responses and appropriate probing methods. An example of this is the Medical Research Council (1966) handbook for their questionnaire on respiratory symptoms. This handbook gives an account of the research leading to the development of the questionnaire and general instructions for its use. It discusses in detail any problems which might arise with interpreting replies to individual items. For example:

> Some subjects admit to bringing up phlegm without admitting to cough. This should be accepted without changing the replies to the question about cough. A claim that phlegm is coughed from the chest but swallowed counts as a positive reply.

Interviewer drift

The initial training of interviewers is not sufficient to ensure a standardized procedure because over time efficiency may wane and the interviewer may cease to uphold his initial standards. The more times an interviewer uses a given questionnaire, the more remote the training period becomes and the more he will forget his briefing. This process of subtle and imperceptible change is known as interviewer drift and several hypotheses may be postulated as its cause.

1. It may result from forgetting. Thus the interviewer may lose his frame of reference or objective criterion against which to judge his performance. He can no longer compare his own behaviour and judgements with those of others, as was the case during training.

2. The initial frame of reference may cease to be rewarding. This may occur when the interviewer is no longer being encouraged in appropriate behaviour by the favourable comments of a superior. Alternatively over the relevant time the interviewer might become committed to a new point of view and this might influence his involvement or lead to the adoption of different standards.

Obviously there is a need for some type of check to be made on the interviewer's progress if drift is to be controlled. The following methods may be used:

1. Observation of the interviewer from time to time in the process of administering the questionnaire, for example, by placing a tape-recorder in the interview situation. For a true sample of interview behaviour this would have to be done without the knowledge of either interviewer or respondent. This raises many difficulties, especially where respondents are interviewed in their own homes. There are also ethical considerations, especially if the respondent has been given an assurance of anonymity.

2. Serial analysis of completed questionnaires. This is the method used by the National Opinion Research Centre. Their procedure is that each interviewer's work is rated regularly, and upon the completion of each assignment the interviewer receives a personal letter from the central office in which errors of procedure, in so far as they can be detected from examination of the completed schedules, are pointed out to him. For example, marked or unusual patterns in the responses, changes in patterns of responses, the repetition of particular words or phrases in free-answer replies, indications that suggest probes are used, deviant behaviour as revealed by comments on the interviewer's report form, and the like faults are noted and called to the attention of the interviewers (Hyman *et al.*, 1954).

3. Retraining sessions which check knowledge of administrative procedures and provide opportunity for fresh interviewer comparisons. These permit a frame of reference to be re-established by which an individual can judge his own performance. A retraining session may also serve to revive the interest of the interviewer in the project itself, and so help to reduce errors due to the interviewer's lack of interest or boredom with the task. The main difficulty with this method is that of arranging training sessions, particularly where the research is limited by factors of time and finance. This problem was overcome by the Medical Research Council for their questionnaire on respiratory symptoms by preparing recordings of a series of test interviews. These were designed to illustrate difficulties arising during field surveys in the interpretation of answers to the standard questionnaire. They allow the interviewer to compare his own ratings with those of other interviewers and provide a frame of reference by which he can continually check his own performance (MRC, 1966).

4. Reinterviewing of respondents by different interviewers and

comparing results. This provides a more objective assessment of interviewer drift which can be expressed statistically in terms of a correlation coefficient. Apart from the time-consuming nature of this method, the main difficulty lies in the problem of differentiating to what extent disagreement in the findings of the two observers is due to the effect of the interviewers and to what extent it is due to factors relating to the repeat use of the questionnaire.

The questionnaire form should be constructed in such a way that some of the questions are logically interdependent. This provides a potential check on interviewer inattention or false reporting of answers. The method is most effective if interviewers are required to write out some of the respondents' replies in full. According to Parten (1950) this helps to prevent the 'faking' of responses as few people are ingenious enough to invent comments which are as varied as those encountered in the real interview situation. However, it is not a perfect control, for Fothergill and Willcock (1950) in their study of interviews made with dummy subjects, found that only 12 per cent of observers' errors could be detected from the data on the questionnaire form.

In conclusion it may be noted that each of the methods suggested introduces an element of day-to-day supervision of working practice. This is the essential basis for any method designed to detect and control interviewer drift. The knowledge that our performance is being continually monitored is enough to make most of us try to keep on the straight and narrow path demanded by the standardized questionnaire.

4. The self-administered questionnaire

The self-administered questionnaire, as its name suggests, is completed by the respondent without the assistance (or interference) of an interviewer. The questions are preset and the respondent records his answer on the questionnaire form or on a specially prepared answer sheet, or by some other means such as operating a computer. The feasibility of doing this with regard to medical factors has been explored. Suchman *et al.* (1958) were able to show that self-ratings of both general and specific health problems were significantly related to the ratings of a physician.

Many studies which have used self-administered questionnaires have found that they are able to produce a greater amount of information than the administered type. For example, in a comparison of a self-administered medical questionnaire with the traditional medical history obtained by interview, Young (1972) found that the former obtained about three times as many symptoms. He divided the relevant symptoms into 'significant' and 'non-significant' categories and showed that the self-administered questionnaire collected twice as many significant symptoms. Again, in a comparison of a self-administered questionnaire with an interview, Stouffer *et al.* (1950) showed that nearly identical information was obtained by both methods. However, the self-administered questionnaire was found to be superior as respondents had included more free comments on the forms than had been recorded by the interviewers. Further investigations have been made into the quality of self-administered questionnaire data by Metzner and Mann (1952) and Kahn (1952). These investigations found questionnaire data to be comparable to interview data and in the instances where different answers were recorded, the questionnaire data were judged to be superior.

As a means of data collection, therefore, the self-administered questionnaire would seem to be comparable to, and in some instances superior to traditional case-history taking or the administered questionnaire. It is a particularly useful instrument in cases where large

amounts of factual data are desired, for example, the medical history. A medical history questionnaire which has good validity is that developed by Mayne *et al.* (1969): the preliminary section of this questionnaire has a sensitivity of 90·1 per cent and a specificity of 96·2 per cent.

The self-administered questionnaire has several advantages. First, it ensures standardization of measurement in that all subjects are asked precisely the same questions. Closed questions are often used to refine the method still further and this strict standardization of format improves test–retest repeatability. Collen *et al.* (1969), in their testing of a self-administered medical questionnaire, found that on retest patients changed their answers to less than 6 per cent of the questions in 95 per cent of the cases. Eliminating the interviewer, apart from removing a source of bias, saves considerable training costs and, further, the self-administered questionnaire may be used in the clinical situation to improve efficiency. For example, Hall (1972) describes a brief out-patient questionnaire which aims at reducing the necessary routine questioning in subsequent physician interviews. Less tangible advantages are that the respondent can proceed at his own speed and that, by studying the questions and saying the words to himself, he had the opportunity for both visual and auditory recognition of terms and phrases. This might be of particular importance with questions including technical terms or with checklists as are often found in medical questionnaires.

However, several serious disadvantages of self-administered questionnaires must also be considered. Firstly, they can only be used where simple and straightforward questions can be formulated and understood with the aid of written instructions. They are, therefore, inappropriate for persons of low intelligence or poor reading ability: many elderly people find self-administered questionnaires very taxing, and often impossible if visual acuity is beginning to fade. The usual sort of questionnaire form requires respondents to be able to express themselves in writing and to be able to write, again something that may be a difficult task for an elderly person or for one handicapped by arthritis. Although self-administered questionnaires achieve exact replication of questions, the sequence may vary: in spite of instructions to the contrary, questions may be answered in any order and this may influence the answers. Further, the answers obtained by a self-administered questionnaire must be accepted as final: there is no opportunity for uncertain responses to be probed: there is

no opportunity for observational data, such as emotional state, to be noted: and lastly, forms completed at home may be the work of many hands, not necessarily including the intended respondent!

Self-administered questionnaires have been developed in a variety of ways.

1. The paper and pencil questionnaire, consisting of a standardized printed form, has undergone considerable development in the last decade. Its most popular use is in surveys or for the collection of past history and symptom data prior to a clinical consultation. Examples of this type of questionnaire are the Cornell Medical Index, as described by Brodman *et al.* (1956) and the Self-Administered Medical Questionnaire developed by Collen *et al.* (1969). This type of questionnaire may also be adapted to specialized research, for example the Hypothyroid Questionnaire (Barker and Bishop, 1969) and the Cardiovascular Questionnaire (Rose, 1962).

2. Forms processed by computer using the 'mark-sense' procedure. This involves the use of specially printed forms which are marked in a particular manner to record specific information. Through the use of a mark-sensing reproducer, these marks are automatically converted in holes in a punch card corresponding to the relative positions of the pencil marks, the punched holes assuming a numerical value equivalent to the original mark. A modification of this method with specially printed IBM cards is used by Mayne *et al.* (1969) in their self-administered symptom history. The mark-sense technique requires subjects to use a considerable amount of skill if errors in the reproducing process are to be avoided.

3. Items printed on prepunched cards which can be 'posted' directly into a computer via 'Yes' and 'No' slots. This saves some time and also questions may be worded more completely as there is less limitation of space on a card than on a form. Martin *et al.* (1969), however, discovered several disadvantages with the card method. They noted that subjects tended to fumble with and drop the cards; that they frequently shuffled through them and, therefore, administered them out of order; and that for future research there seemed little possibility of incorporating a branching system into the card procedure.

4. The computer-administered questionnaire is perhaps the most sophisticated form of self-administered questionnaire developed so far. The equipment includes a visual display screen on which projected photographs of questions appear, the subjects responding by pressing appropriate buttons. The number of buttons (and therefore,

the number of response categories) must be limited to prevent the method from becoming too complex. However, this limitation introduces the danger of a response set (Milne and Williamson, 1971) and further refinements include the use of typewriter-style keyboards and 'touch sensitive' screens. Here it is necessary to add that such sophisticated technology does not alter any of the fundamentals underlying the design of questionnaires.

Use of self-administered questionnaires

Self-administered questionnaires have been used for a wide variety of medical purposes. Their potential for both clinical practice and research is demonstrated by the number and diversity of published studies which have used them as a basic instrument of measurement. For example, Dunn and Hawkes (1966) describe a self-administered questionnaire in use at the Memorial Hospital for Cancer and Allied Diseases which provides a review of symptoms and historical background information as a guide for the examining physician. Fahy *et al.* (1969) using self-administered questionnaires to investigate post-anaesthetic symptoms of discharged patients, found they were able to obtain much valuable information not revealed in a previous interview. Sydiaha *et al.* (1968) gained useful information concerning the process of psychiatric patients after discharge by using two questionnaires, one being sent to the patient and the other to a parent or guardian.

The ability to post the questionnaire to the respondent has made it particularly attractive to researchers. By this means a questionnaire can be sent to respondents unable to leave home or anxious about visiting a hospital or physician. It may also prove useful as a means of following up patients who need to have regular medical examinations. For example, Barker and Bishop (1969) used postal questionnaires to follow up thyroid patients. They found the data obtained by questionnaire compared favourably with that obtained by interview, and commented that in fact the postal questionnaire may be a superior measure in that two subjects who had confessed to feeling unwell in the questionnaire, failed to report this at the interview. However, a possible disadvantage of the postal questionnaire is the complete loss of personal contact with the respondent, although a well-planned introductory letter may compensate to some extent for this loss of rapport. The problem of maintaining the subject's interest

is discussed by Reeder (1963) and Skipper and Ellison (1966). They maintain that some form of personal contact is desirable and will increase co-operation and suggest that this may be achieved by personal delivery of the questionnaire form or by use of the telephone.

It is, however, in the field of epidemiological research that the self-administered questionnaire is best seen as when used in surveys it provides a relatively cheap and simple method of investigating a large sample of respondents. Delivered by post it allows studies of local populations, national samples, or migrants to other countries. It may be used to identify subsamples of the population for further study, to measure the frequency of particular symptoms or conditions in a cross-sectional study, to measure the frequency of a characteristic as the basis for a prospective study or to obtain data in follow-up studies. For example, with the use of postal questionnaires Adler *et al.* (1973) undertook the first stage of a two-stage prevalence study of cardio-respiratory disease; Waters (1973) investigated the presence of headache in a random sample of 2,000 people in Wales; Bennett *et al.* (1970) examined the association of sugar consumption and cigarette smoking in a random sample of men in a London borough; Doll and Hill (1964) determined the smoking habits of British doctors as the basis for a prospective study of the association of cigarette smoking and mortality; Rickels *et al.* (1968) followed up patients lapsing from drug trials.

The main problem in the use of postal questionnaires for surveys is the very poor response rates they tend to produce; 40–60 per cent is a common rate of return. This may mean that insufficient data will be obtained for any proper conclusion to be drawn as a high non-response rate may lead to a bias in the results because the returns are not representative of the whole population. A good example of bias due to non-response is given by Burgess and Tierney (1970) in their study of physicians' smoking habits. Their initial finding that physicians tended to be non-smokers was reversed when it was later revealed that eight out of ten non-respondents were smokers.

Goode and Hatt (1952) have pointed out that unreturned postal questionnaires bias the sample in a predictable direction but to an unknown degree: the direction of bias being usually towards those respondents who are in the higher social and economic strata and towards those who are better educated. Durant and Maas (1956) maintain that the response rates are highest when the questionnaire deals with a subject of particular interest to the respondents. A medical

questionnaire which is designed to investigate, for example, personal handicaps of respondents might be expected to give reasonably good returns. Graham *et al.* (1968) used a postal questionnaire to establish the prevalence of registerable blindness and had a response rate of 92·5 per cent.

Two methods of controlling response bias are available: correction and prevention. A correction can be made by extending the sample, selecting replacements comparable to the non-respondents, until sufficient replies are received to compensate for the initial non-response. Alternatively a hypothesis can be formulated as to the reason for non-response. To test this the return date of every questionnaire must be noted for Oppenheim (1968) has pointed out that those who make late returns are similar to non-respondents. Response bias might then be judged by comparing respondents with non-respondents with regard to different factors such as age and economic status, and by comparing the answers given in early and late returns.

Perhaps a better means of controlling the effects of response bias is by prevention: by constructing the questionnaire in a manner which will give most returns. To this purpose, a number of studies have been carried out which have aimed at isolating those factors which induce response in postal questionnaire surveys. The use of coloured paper was discussed earlier in relation to questionnaire layout and printing. Cartwright and Ward (1968), in a study of general practitioners' responses to postal questionnaires, showed that amongst other factors potentially influencing response were questionnaire length, whether envelopes were typed or handwritten, number and denomination of the stamps used, tone of accompanying letter, layout and method of recording, space allowed for qualifying statements, and the time of year in which the questionnaire was sent. Scott (1961) from his experience with the British Government Social Survey, comments that a higher response rate may be obtained by using stamped, handwritten reply envelopes as opposed to printed machine-stamped envelopes which may be associated with advertising. He found also that a printed letterhead, or the day of the week on which the questionnaire is sent, makes no difference. King (1970) found that questionnaire anonymity did not necessarily improve returns.

It should be noted that an effective means of reducing non-response is to follow up non-respondents with another copy of the questionnaire. The effectiveness of this procedure is demonstrated by the results of a questionnaire mailed to respondents as part of the St

Thomas's Health Survey in Lambeth (Holland, 1968). Questionnaires were posted to 2,247 persons. After ten days non-respondents were sent a further copy of the questionnaire and a letter asking for it to be completed on the supposition that the original copy had been lost. After another ten days this was repeated for respondents who had still not replied. This procedure improved the percentage of returns as follows: first mailing: 60·2 per cent; second mailing: 79·3 percent; third mailing: 86·4 per cent. The response rate was thus raised from an unacceptably low level of 60·2 per cent to a reasonably high level of 86·4 per cent, leaving only a small minority of the questionnaires to be recovered by household visits.

Administered versus self-administered questionnaires

Previously it was thought opportune to include some comments on the topic of open versus closed questions. It was concluded that the two question types differed fundamentally and, furthermore, each has its uses. So it is with administered and self-administered questionnaires. The administered questionnaire is in many instances easier to design but more difficult and expensive to use. Its use requires interviewer training and supervision but given these it introduces certain safeguards. The self-administered questionnaire is easy and cheap to use, but more difficult to design. In its most common paper and pencil form it must be simple; complicated instructions and complex branching questions are unsuitable. As a result it is generally agreed that the formulation of content and questions is one of the most difficult tasks facing a researcher.

Thus, as the questionnaire types differ so do their potential uses and the constraints which surround them. Most often if the situation of use is clearly defined, the objectives clearly stated, the constraints in terms of time, money, and numbers of people involved clearly identified, then the choice of instrument becomes obvious.

5. User attitudes to medical questionnaires

In his discussion of the possibility of using an automated questionnaire as a means of collecting data concerning a patient's medical history, Kanner (1969) points to five aspects of a questionnaire which he claims determine its potential usefulness: the quality of the medical data obtained, the time it saves the researcher or other medical staff, its economic feasibility, its acceptace by respondents, and its usefulness to the researcher. In most questionnaire studies very little consideration is given to these last two factors. This is a serious failing on the part of questionnaire designers for if the method is to be a useful one then it must have the support of its users.

The respondent

Respondent acceptance is important in that its chief expression will be in terms of motivation to complete the questionnaire and a negative attitude may well result in low motivation and carelessness thus enhancing the number of false answers or unanswered questions (Mellner, 1970). There are several factors which may motivate a respondent to complete a questionnaire and certain studies have been directed at isolating these factors. Pomeroy (1963) found that respondents were less reluctant to give information if assured of its importance, of the confidentiality of the information given and a lack of moral judgement on the part of the researcher with regard to his responses. He concluded that poorly trained interviewers and faulty interviewing creates more reluctance on the part of respondents than any other factor.

Budd *et al.* (1969) maintain there is evidence to suggest that respondents have a favourable response to the initial presentation of a questionnaire but that interest decreases in inverse proportion to its length. They quote Slack and Haessler (personal communication) as suggesting thirty minutes to an hour as the limit before respondent saturation.

Cannell *et al.* (1968) found that the interpersonal behaviour which takes place between an interviewer and respondent was an incentive to health reporting. This hypothesis, that personal contact with the respondent increases motivation, is further supported by the work of Bueker (1969) who attempted to identify sources of respondent resistance to answering questionnaires. She concluded that questionnaires were more acceptable when they were handed out personally to the respondents.

Although it is obviously desirable to motivate a respondent to complete a questionnaire, the researcher should avoid 'over-motivating' him as this may also lead to inaccurate results. For example, Stout *et al.* (1969) administered the Cornell Medical Index to 603 respondents seeking public assistance for physical and emotional disability. Their results contained an extremely high number of false positives, demonstrating that the level of motivation of their respondents (due to their need for aid) led to inaccurate results. Also, Heron (1956) found that respondents gave significantly less 'neurotic' responses to a questionnaire on emotional adjustment when they thought it was for job selection than when they thought it was for research purposes.

When a medical questionnaire is used as a survey instrument user attitudes are usually determined in terms of the response rate. For example, the New Haven survey of joint diseases (Acheson *et al.*, 1969) used a semi-structured administered questionnaire which was claimed to have a high acceptance rate on the basis of a 92 per cent co-operation rate. Glasunov (1968) claims his postal questionnaire for angina to be highly acceptable to respondents in that response rates for the various social groups tested ranged between 76·6 per cent and 82·8 per cent. It should be remembered, however, that a respondent may complete a questionnaire although he may disapprove of it but is too embarrassed to refuse; and if he does not reply it does not always mean that he finds the questionnaire unsatisfactory.

The use of medical questionnaires in a clinical setting raises additional problems with regard to respondent acceptance as in this case it is his personal interests which are at stake. The patient must have confidence in the methods of medical science so that he will voluntarily seek its services, otherwise medicine ceases to be a public service and becomes a science interesting for its own sake only. As Hulka *et al.* (1970) comment:

Only recently has the idea become accepted that the consumers of medical care as well as the providers should be consulted as to their attitudes, concerns and satisfactions with the medical care they receive. Patient satisfaction . . . has been suggested as one criterion for the measurement of quality of care.

One method which has been used to assess respondent acceptance of medical questionnaires in a clinical context has been simply to ask the respondent his opinion. For example, Haeger *et al.* (1965) report that, in a sample of 215 respondents given a general medical questionnaire, 6 per cent of men and 4 per cent of women considered it troublesome or hard to fill in. Kanner used a computer-based medical history questionnaire for collecting case-histories and he records that respondents repeatedly expressed their approval of the method (Kanner, 1969). However, this sort of loose approach to assessing respondent acceptance and satisfaction is not acceptable and if respondent attitude is to be properly examined then a validated attitude questionnaire must be used.

Apparently the only attempt yet made to design a questionnaire for measurement of patient acceptance of a specific medical questionnaire has been that of Grossman (1971). Two hundred and fifty respondents were asked a series of questions, such as:

DID THE QUESTIONS ALLOW YOU TO GIVE ACCURATE ANSWERS?

	Positive responses %
1. All of the time	23
2. Most of the time	65
3. About half of the time	11
4. Seldom	0·5
5. Hardly ever	0·5

The results of this questionnaire study indicated that in addition to feeling they were able to give accurate answers, respondents were not bored, found the questions related to their individual medical problems, and felt that they were able to give complete histories. With regard to whether they would want to be reinterviewed by the computer their reactions were mixed:

WOULD YOU LIKE TO TALK TO THE MACHINE DURING FUTURE VISITS?

	Positive responses %
1. All of the time	22
2. Most of the time	32
3. About half of the time	23
4. Seldom	10
5. Hardly ever	13

This last question is especially interesting in that it points to the possibility that what a respondent may accept once on an experimental basis, he may not accept as a permanent method. This possibility throws doubt upon all methods used to assess respondent acceptance of medical questionnaires. That is, in asking a respondent his opinion of a medical questionnaire is one asking his opinion of it as a diagnostic instrument or an amusing method of research? Furthermore, is his response to such an inquiry an attitude upon which he might act or is it a gratuitous opinion given to satisfy the research worker? These questions stress the need for the development of a reliable measure of respondent acceptance of medical questionnaires. If the respondent claims to have confidence in medical questionnaires then he must support this by appropriate behaviour; that is, he must choose them in preference to other forms of assessment when his own health is at stake. No such investigation appears to have been made so far.

The researcher

Some consideration must now be given to the other user of medical questionnaires: the researcher or questionnaire designer. In the use of survey questionnaires he may be an epidemiologist or behavioural scientist; if the questionnaire is for a clinical purpose he may be a physician or surgeon collecting case-history data. For the former, acceptability of a questionnaire will depend primarily on its validity, in particular its sensitivity and its relative value in comparison with other available survey methods. Satisfaction with the questionnaire will depend primarily upon its efficiency as a measuring instrument.

In the clinical situation, while the efficiency of the questionnaire remains an important consideration, other factors are operant. For example, Hall (1972) maintains on the basis of personal inquiry that although his self-administered medical history questionnaire was acceptable to patients there was some reluctance on the part of physicians to use it. They rejected the idea as too mechanistic, and erosive of the mystique of anamnesis. Mayne *et al.* (1972) in a study of the acceptability and usefulness to patients of their paper-and-pencil Patient Inventory Questionnaire (PIQ) found that although the data which the questionnaire collected was accurate, physicians judged it of limited use. The questionnaire did not indicate the urgency of the medical problem and moreover the method was time-consuming.

Although initially the medical staff were enthusiastic about the technique, after months of use they complained that it accumulated too much trivial information which increased physician time and work, and that the data were not helpful in making medical decisions. Thus earlier comments by Budd and his colleagues on the growing acceptance of general medical history questionnaires in out-patient clinics proved optimistic.

Perhaps it should be concluded that the stage of romance with questionnaires, when they excited us with their promises, is now over. Realistically, their use as a research tool is now well established: their use in clinical practice has been most successful in a few limited areas. However, an indirect benefit of some of the work to date has been an increase in our knowledge of symptom manifestation and meaning. From this may ultimately stem a wider use of questionnaires in clinical medicine.

Part II

Questionnaires relevant to medicine

Introductory note

A considerable number of medical questionnaires have been used in the past but few have been validated and documented. This section consists of brief notes on those of general interest. References are given should a more detailed account of a particular questionnaire be desired.

We have excluded questionnaires which are eclectic or derivatives of other well-known questionnaires. Examples of psychiatric questionnaires have been included on the basis that they are mainly concerned with physiological correlates of emotional disturbance.

Obviously the more complete such a catalogue the more useful it is. We should be pleased to learn of other questionnaires that have been developed and documented to the point where they are suitable for wider use.

6. General health questionnaires

Automated Medical History

This questionnaire forms part of a system for the automated collection of a patient's medical history. It consists of 187 multiple choice items with a branching system so that the patient may answer as few as 90 items. The patient answers by denoting the number of the appropriate response and the computer produces a printed summary.

The questionnaire was validated by comparison with the data collected at physicians' interviews. Information collected by the questionnaire agreed closely with that obtained at interview with an average error of two false positive and three false negative responses for each administration. The questionnaire was found to produce more information than the interview. A 16-item rating scale has been devised to assess the attitudes of respondents who used the questionnaire. In general, attitudes were favourable although many respondents expressed some reluctance at the idea of 'talking' to the machine during further visits.

GROSSMAN, J. H., BARNETT, G. O., MCGUIRE, M. T., and SWEDLOW, D. B. (1971). 'Evaluation of computer-acquired patient histories', *JAMA*, **215**, 1286.

Cornell Medical Index

The Cornell Medical Index (CMI) was developed to provide a quick and reliable means of obtaining facts about a patient's medical history without expenditure of the physician's time. It is a self-administered questionnaire which takes ten to thirty minutes to complete; it may also be administered by an interviewer. The questionnaire consists of 195 questions which the respondent answers by reference to Yes/No categories. The items are of four kinds (symptomatology, past history, family history, and behaviour) and compose eighteen groups designated on the questionnaire form by a letter of the alphabet:

A	eyes and ears	J	illness frequency
B	respiratory system	K	miscellaneous diseases
C	cardiovascular system	L	habits
D	digestive tract	M	inadequacy
E	musculoskeletal system	N	depression
F	skin	O	anxiety
G	nervous system	P	sensibility
H	genito-urinary system	Q	anger
I	fatigability	R	tension

The questionnaire was validated by comparison with information recorded by physicians in the case-records of 179 hospital patients. It was found that 95 per cent of the data appeared in both the case-records and the questionnaires. An average of 6·4 items per patient appeared in the case-records and not on the questionnaire: however only 1·4 of these were items of critical data about the present illnesses. An average of 176·1 items appeared on each questionnaire and not in the case-records: however the large majority of these were negatives and only 27·7 were positives and thereby considered to have diagnostic potential or significance.

Further experiments were designed to determine the usefulness of the CMI as a diagnostic instrument. It was found that physicians were able to identify almost all (94 per cent) of the diagnostic areas in which disease was found and in 87 per cent of these areas they could infer the specific medical diseases present. However, the subjective interpretations of a medical observer were still necessary for this judgement to be made as the CMI alone is of little value as an indicator of specific somatic disorders even when a weighted system of scoring is used.

As an indicator of the presence and degree of emotional disturbance the CMI has been demonstrated to be more successful. In a comparison of CMI responses given by 'normal' and 'abnormal' respondents, it was demonstrated that a high probability of emotional disturbance could be inferred by the number and type of questions marked in the 'psychiatric' subsections (M–R) and also if the respondent made thirty or more 'yes' responses throughout the questionnaire, if he answered three or more items 'yes' and 'no', omitted answers to six or more questions or wrote in several remarks on the questionnaire.

Many studies have been made using the CMI in a large variety of

settings. The following references refer only to the original design and validation.

BRODMAN, K., ERDMANN, A. J. JR, LORGE, I., and WOLFF, H. G. (1949). 'The Cornell Medical Index: An adjunct to medical interview', *JAMA*, **140**, 530.
—— —— —— —— (1951). 'The Cornell Medical Index-Health Questionnaire. II. As a diagnostic instrument', ibid. **145**, 152.
—— —— and WOLFF, H. G. (1956). *Cornell Medical Index; Health Questionnaire* (manual) (New York: Cornell University Medical College).

The CMI has also been translated into Hebrew and has been validated for a Jerusalem population.

ABRAMSON, J. H. (1966). 'The Cornell Medical Index as an epidemiological tool', *Am. J. Publ. Hlth*, **56**, 2.
—— TEREPOLSKY, L., BROOK, J. G., and KARK, S. L. (1965). 'The Cornell Medical Index as a health measure in epidemiological studies', *Br. J. prev. soc. Med.* **19**, 103.

Disability Schedule (see p. 92)

This schedule was designed and validated for the performance of prevalence studies of disability. The authors put forward a concept of disability based on appreciation of inability to perform essential activities of daily living. The schedule contains sections for the description of functional ability in the areas of mobility, self-care, domestic duties, and occupation. Comparison of data obtained by trained interviewers using the schedule with data contained in the clinical record showed levels of agreement between 90 and 100 per cent. For a series of assessments repeated after an interval of one year agreement was found in 80 per cent but in only 2 per cent were the changes inexplicable. A short simple screening questionnaire for disability is described in one of the references.

GARRAD, J., and BENNETT, A. E. (1971). 'A validated interview schedule for use in population surveys of chronic disease and disability', *Br. J. prev. soc. Med.* **25**, 97.
BENNETT, A. E., GARRAD, J., and HALIL, T. (1970). 'Chronic disease and disability in the community: a prevalence study', *Br. med. J.* **3**, 762.

General Medical Questionary

This self-administered questionnaire was developed for use as a screening device for out-patients in the belief that it could elicit a great deal of information which out-patient practitioners did not have the time to discover. It consists of 172 questions for male, and

185 for females, covering two main areas: preliminary (including personal information, past medical history, family and social history) and clinical (concerned with presenting symptoms). The questionnaire does not incorporate a branching system but a card-sort technique is used to ensure all items are answered and to facilitate the process of response analysis. A subject takes an average of 23 minutes to complete the questionnaire.

Assessment of the questionnaire was made by administering it to 69 patients, male and female, from a general medical unit and then comparing responses with the results of interviews. The interview consisted of asking the patient all the items on the questionnaire and eliciting further details where 'yes' responses were given. A comparison of results revealed that the preliminary section of the questionnaire had a sensitivity of 90·1 and specificity of 96·2 and the clinical section a sensitivity of 77·3 and specificity of 99·5. As the questionnaire is desired as a screening device it is desirable that the negative responses should be highly accurate, particularly in the clinical section. The specificity values indicate that the questionnaire meets this requirement. The use of card-sorting as an administration method has frequently been criticized and it was found that 43 per cent of subjects made errors in self-administration. These were mainly due to not putting the cards face down in the box as instructed, putting them out of order and putting them in the wrong box.

Further assessment of the questionnaire results was then made by comparison with the patients' medical records. It was found that the questionnaire obtained more items of information, both positive and negative, than were recorded although 17 per cent of items found in the records were not found by questionnaire. All the information collected by the questionnaire does not have the same value and in any one case it was found impossible to determine what was the most important part of the history. Much of the information collected appeared trivial and to assess this further an attempt was made to divide the symptoms into 'significant' and 'non-significant'. On this basis it was estimated that the questionnaire still collected twice as many significant symptoms.

YOUNG, D. W. (1972). 'Evaluation of a questionary', *Meth. Inform. Med.* **11**, 15.
—— (1972). 'Comparison of information collected by a questionary with that in the patient's hospital record', ibid. **11**, 20.

Health Questionnaire for Secondary School Students

This questionnaire has been developed as a means of rapidly and economically assessing the health status of American secondary school students. It is essentially a screening device rather than a diagnostic instrument and has been demonstrated to show no bias towards race, sex, age, socio-economic group, or IQ. It consists of 182 general health questions each of equal weight with a dichotomous response system.

The reliability of the questionnaire was assessed by readministration of the questionnaire to 1,166 secondary school students after a two-week interval. This gave a correlation coefficient of 0·90. An assessment of the validity of the questionnaire was made by comparing the responses of 388 secondary school students with ratings from physicians' examinations. It was found that questionnaire scores were able to differentiate, at a statistically significant level, those children with and those without health problems. Of the nineteen health subsections examined by the questionnaire, thirteen were found to differentiate significantly the healthy from the non-healthy respondents in their respective areas: however, six subsections (nose, throat/neck, heart/blood, lungs, male genito/sexual and gastro-intestinal areas) could not. The questionnaire was also able to distinguish the priority of need although 14 per cent of the subjects classified as having health problems to be seen with priority were classified by the physician as having no problem. This low specificity restricts the usefulness of the questionnaire in a survey or diagnostic context although the apparent high sensitivity recommends it as a screening instrument. It is possible that the introduction of a weighting system into the scoring of the questionnaire would improve specificity.

ALEXIOU, N. G., and WIENER, G. (1968). 'Reliability of a self-administered health questionnaire for secondary school students (adolescents)', *Am. J. Publ. Hlth*, **58**, 1439.

—— —— SILVERMAN, M., and MILTON, T. (1969). 'Validity studies of a self-administered health questionnaire for secondary school students', ibid. **59**, 1400.

Medical Questionnaire

This self-administered questionnaire aims to obtain from respondents an extensive case-history. It consists of 531 questions which cover major systems, general health, previous illness, operations, drugs, and

family and social history. In order to assist the respondent through the large number of items, a branching technique has been incorporated so that a minimum of 93 items may be answered. The difficulties of complicated procedures and confusing layouts normally associated with the use of branching techniques in self-administered questionnaires are minimized by placing different sections on different coloured paper in order to facilitate transitions from question to question. The position and type of response categories are varied to prevent response sets and contradictory questions are included as a means of detecting carelessness and false responses.

The questionnaire was evaluated by a study of 70 male subjects aged between 13 and 82. Of these 80 per cent were found capable of completing the questionnaire. These histories were compared with physicians' standard notes and the former were found to produce a far greater amount of information, both positive and negative (no assessment seems to have been made of the value of this information). A mean of six errors was made in completing each questionary: on only three occasions were contradictory answers given. Respondents generally expressed favourable attitudes towards the questionary (only one was displeased with the technique). It was found that 'yes' responses tended to cluster within sections suggesting that the branching system was successful in directing respondents to further useful information.

ANDERSON, J., and DAY, J. L. (1968). 'New self-administered medical questionary', *Br. med. J.* **4**, 636.

Out-patient Medical Questionnaire

This questionnaire was designed for use in an out-patient department to reduce the time spent on routine questioning and identification of primary symptoms. It consists of six questions relating to personal and family medical history and thirty-five questions relating to common symptomatology. The questionnaire was found to be acceptable to patients and able to distinguish efficiently the four main conditions which occurred at a general hospital (peptic ulcer, diabetes, neurosis, and rheumatism). Further validation studies are presently being carried out (personal communication).

HALL, G. H. (1972). 'Experiences with outpatient medical questionnaires', *Br. med. J.* **1**, 42.

Patient Inventory Questionnaire

This general medical questionnaire was developed at the Mayo Clinic. It consists of a large number of case-history and symptom items with an individualized branching potential. A computer is used to process results recorded on mark-sense forms. The medical content of the questionnaire has been drawn from ten currently used medical history questionnaires, repetitious material being eliminated and all questions being recorded in a common style. The selected questions were compared with a list of diagnoses which had been recorded twenty or more times in one year at the Mayo Clinic to ensure each was represented by an item in the questionnaire.

The questionnaire is organized on three levels.

1. The first level is a broad screen presented as a symptom review checklist and case-history.

2. The second asks for greater detail, concerning frequency, location and severity of those symptoms which have been checked in the first level.

3. The third level demands still greater detail involving past signs, symptoms, and treatment.

The computer prints out all positive information and the qualifying details from levels 2 and 3.

The reliability of the questionnaire was assessed by retest of 101 patients after several hours. The first administration produced an average number of 14·2 health problems per patient, and the second produced an average of 13·5. The validity of the questionnaire was checked by comparing responses with patients' medical records (N = 500). Thirty-one patients (6·3 per cent) omitted the chief complaint and all but two of these were picked up in the section concerned with present health problems. The remaining two omitted their complaint through faults of memory: one subject being senile and the other schizophrenic. Correct recording of the chief complaint was made by 93·8 per cent of all subjects tested, and 99 per cent of all summary sheets were found to contain useful medical information.

A later evaluation was made using 300 patients attending the Mayo Clinic. The data from their summary sheets was found to be accurate but not useful for determining urgency of medical problems. Nor was the questionnaire able to usefully discriminate functional versus organic nature of the problem, physician times needed for consultation, appropriate medical specialist required, or specific laboratory

tests which were needed to clarify the diagnosis. The general opinion of the clinicians who used the questionnaire was one of initial enthusiasm, but later there were complaints that it collected too much trivial information which required physician time and effort to discriminate. The researchers hope to overcome this by the development of a more discriminating branching technique.

Some of the items on the questionnaire itself would seem to suffer from ambiguity, which is perhaps why the use of a clerical assistant was found to improve accuracy. Care should be taken with these items and also others which lack sufficient explanation. The questionnaire is commendable, however, in that it appears to have been developed and evaluated relatively efficiently and critically, although some further refinements of the technique would seem to be necessary.

MAYNE, J. G., MARTIN, M. J., MORROW, G. W. JR, TURNER, R. M., and HISEY, B. L. (1969). 'A health questionnaire based on paper-and-pencil medium, individualised and produced by computer. I. Technique', *JAMA*, **208**, 2060.

MARTIN, M. J., MAYNE, J. G., TAYLOR, W. F., and SWENSON, M. N. (1969). 'A health questionnaire based on paper-and-pencil medium and produced by computer. II. Testing and evaluation', ibid. **208**, 2064.

MAYNE, J. G., MARTIN, M. J., TAYLOR, W. F., O'BRIEN, P. C., and FLEMING, P. J. (1972). 'A health questionnaire based on paper-and-pencil medium, individualised and produced by computer. III. Usefulness and acceptability to physicians', *Ann. Intern. Med.* **76**, 923.

Programmed Medical History

The programmed medical history is a lengthy, complex, symptom questionnaire, proceeding system by system to obtain an exhaustive case-history. It consists of 52 pages containing 935 questions, but employs a branching technique so as few as 221 questions may be answered by a healthy patient. It is self-administered although help may be sought from a clerical assistant or relative of the patient.

A criticism to be made of this questionnaire is the lack of evidence of its validity. An attempt was made to assess the validity of the PMH-s (a screening form of the PMH) but did not reach a conclusion. A comparison was made of a few questionnaire results with those obtained by senior medical students using conventional means. The results were termed 'comparable' but were not elaborated in statistical terms. It is claimed that patients are able to respond as well on questionnaire forms as in a formal interview and occasionally additional information is brought to light.

KANNER, I. F. (1969). 'Programmed medical history-taking with or without computer', *JAMA*, **207**, 317.

Screening Questionnaire (see p. 91)

This is a short self-administered questionnaire designed as part of a long-term care programme for patients with chronic diseases. It is suitable for use in a screening programme or for routine health examinations. It consists of thirty-five items dealing with cardio-respiratory symptoms, alimentary and urinary symptoms, miscellaneous conditions, and psychiatric status. Questions are answered Yes/No and the significance of the responses given are determined by reference to response frequency and clustering. During pretesting the questionnaire was administered to 2,000 respondents. It was found acceptable to respondents, easy to complete and valid when compared with clinical records and clinical interviews.

BENNETT, A. E., and FRASER, I. G. P. (1972). 'Impact of a screening programme in general practice: a randomized control trial', *Internat. J. Epidemiol.* **1**, 55.

Self-administered Symptom History

The self-administered symptom history has been developed to overcome the practical problems associated with obtaining medical histories in hospital (such as illegibility, use of physician time and lack of formalization). It is a computer-based questionnaire method which requires patients to record their own medical histories either before or after their arrival at hospital if necessary with the help of friends or relatives. The questionnaire is presented in two stages. The primary questionnaire consists of 300 questions in five sections (present health, symptoms review, major complaint, past medical history, social and family history) which is administered to all patients. A computer program uses the responses to these questions to develop a branching technique by searching a library file of some 4,000 secondary questions. Those appropriate to the patient's primary responses are then printed out for him to answer.

The researchers claim that the self-administered symptom history has efficiently overcome the problems associated with traditional case-histories and is of diagnostic value. However, more formal assessments of validity and repeatability would appear to be needed.

GLEDHILL, V. X., and MATHEWS, J. D. (1971). 'Acquisition and storage of clinical data by computer', *Br. J. Hosp. Med. Equip. Supp.*

—— (1972). 'Medical decision making by computer', *Australian Computer J.* **4**, 12.

—— MATHEWS, J. D., and MACKAY, I. R. (1972). 'Computer aided diagnosis: a study of bronchitis', *Meth. Inform. Med.* **11**, 228.

7. Questionnaires for specific conditions

Cardiovascular Questionnaire

This questionnaire was developed for the detection of the presence or absence of angina, intermittent claudication, and history of myocardial infarction. It consists of 61 questions, administered by an interviewer, incorporating a simple branching system. Detailed instructions are provided for the interviewer concerning the interpretation of responses.

The questionnaire has been validated by comparison with physicians' interviews for 57 respondents with chest pain. One hundred per cent agreement was found in the negative cases and 94·4 per cent in the positive cases. The small advantage in sensitivity found in the physicians' judgements was gained by taking fifteen times as long as the questionnaire and proceeding in an unrepeatable manner. Predictive validity was assessed by correlation of questionnaire diagnoses with subsequent mortality experiences and good correlations were found. Repeatability was assessed by retesting 1,136 respondents with a one-year interval. On this basis it was found that over-all prevalence rates remained quite stable, changing only from 3·3 to 3·4 per cent. However, for individual respondents repeatability was very low as in the subsequent year only 40 per cent of the 43 men originally classified as positive were reclassified positive. The researchers consider that this may be due in large part to variability in the natural history of angina.

A self-administered version of the questionnaire consisting of only 23 items has also been developed. The sensitivity of this instrument may be better than the interviewer-administered version (personal communication).

ROSE, G. A. (1962). 'The diagnosis of ischaemic heart pain and intermittent claudication in field surveys', *Bull. Wld Hlth Org.* **27**, 645.

—— (1965). 'Cardiovascular diagnosis by questionnaire', *Millbank Meml Fund Q.* **43**, no. 2. 32.

—— (1968). 'Variability of angina: some implications for epidemiology', *Br. J. prev. soc. Med.* **22**, 12.

ROSE, G. A. (1971). 'Predicting coronary heart disease from minor symptoms and electrocardiographic findings', *Br. J. prev. soc. Med.* **25**, 160.

Headache Questionnaire

This questionnaire was developed for a survey of migraine prevalence. It is a self-administered questionnaire and consists of 39 items, 18 referring to type and severity of headache and associated symptoms, and 21 items concerning general health. Severity and type of headache are classified by reference to symptom frequency and symptom clustering.

The questionnaire was validated by comparison of responses with neurologists' clinical diagnoses. This produced very favourable results. Further validity studies using this criterion are at present being carried out (personal communication). An attempt to validate the questionnaire was made by comparison of questionnaire responses with familial prevalence of migraine. However, this was unsuccessful due to the inadequacy of the criteria used. The researchers concluded that family history could not be included in the definition of migraine. An attempt to use the treatment response to ergotamine tartrate as a diagnostic criterion for validation was also unsuccessful. The questionnaire has been found to be highly reliable in terms of test–retest stability. It has been shown that if the questionnaire is administered similar prevalence rates are obtained as with the self-administered version.

WATERS, W. E. (1971). 'Migraine: intelligence, social class and familial prevalence', *Br. med. J.* **2**, 77.

—— and O'CONNOR, P. J. (1971). 'Epidemiology of headache and migraine in women', *J. Neurol. Neurosurg. Psychiat.* **34**, 148.

Ileostomy and Colostomy Questionnaire

This questionnaire consists of 85 questions designed for follow-up study of physical, social, and emotional adjustment of persons who have undergone ileostomy and colostomy. It consists of subsections on medical history, physiological data, sexual history, miscellaneous facts, and reactions. The questionnaire is designed for adults and validation studies have been carried out on persons over the age of 50. Validity was assessed by comparing responses to the questionnaire with clinical data and information obtained from other studies of ileostomy and colostomy. However, no statistical analyses of the

validity study have been published although the investigators claim it to produce very frank responses.

DLIN, B. M., PERLMAN, A., and RINGOLD, E. (1969). 'Psychosexual response to ileostomy and colostomy', *Am. J. Psychiat.* **126**, 374.
—— —— (1971). 'Emotional response to ileostomy and colostomy in patients over the age of fifty', *Geriatrics*, **26**, 113.

Medical Research Council Questionnaire on Respiratory Symptoms
(see p. 89)

This questionnaire was designed for studies of the prevalence of chronic bronchitis and related disorders. The questionnaire is administered by an interviewer, consists of 48 items and takes approximately three to five minutes to complete. Detailed instructions are given for interviewers regarding the interpretation of responses, and a series of recorded interviews has been prepared which can be used by the interviewer during training and as a check to interviewer drift.

The questionnaire responses of 192 male and 192 female respondents aged 40–59 have been compared with objective measurements of sputum volume, ventilatory capacity and inspection of sickness absence records. It was found that the questions on dyspnoea compared badly with tests of ventilatory capacity and so provided a poor indication of lung function. On the other hand questions concerning illness compared well with sickness absence records showing no marked tendency to exaggeration or under reporting of past illness. Questions concerning sputum production (which was considered by the investigators to be the most significant indicator of chronic bronchitis) compared closely with measurement of sputum volumes. While the latter was considered a more accurate measure, it was felt to be a more troublesome and distasteful procedure.

The questionnaire has been translated into many languages including Japanese.

'Medical Research Council's Committee on the aetiology of chronic bronchitis', *Br. med. J.* **2** (1960), 1665.
Medical Research Council's Committee on research into chronic bronchitis (1966). *Questionnaire on respiratory symptoms* (London: MRC).
Medical Research Council's Committee on research into chronic bronchitis (1966). *Instructions for the use of the questionnaire on respiratory symptoms* (London: MRC).
HOLLAND, W. W., ASHFORD, J. R., COLLEY, J. R. T., MORGAN, D. C., and PEARSON, N. J. A. (1966). 'A comparison of two respiratory symptoms questionnaires', *Br. J. prev. soc. Med.* **20**, 76.

The ECCS questionnaire is derived largely from the Medical Research Council questionnaire. It was developed by a special working group of the European Coal and Steel Community with a view to using it as a standardized instrument of investigation. To this end it has been translated into four languages.

It was found on testing that consistent answers were obtained to questions asked on two occasions, agreement ranging from 80 to 90 per cent. Sixty-two per cent of the disagreement results from faulty interviewer technique, 16 per cent to inadequate formulation of the question, and 21 per cent to changed responses.

BRILLE, D., VAN DER LENDE, R., SANNA-RANDACCIO, F., and SMIDT, U. (1972). *Commentaires relatifs au questionnaire de la CECA*. Collection d'hygiene et de medicine du travail no. 14 (Luxemburg).

VAN DER LENDE, R., and ORIE, N. G. M. (1972). 'The MRC-ECCS questionnaire on respiratory symptoms (use in epidemiology)', *Scand. J. resp. Dis.* **53**, 218.

Menstrual Distress Questionnaire

This questionnaire was developed to provide a standardized technique for studying menstrual symptomatology at the menstrual, premenstrual, and inter-menstrual phases. It is a rating scale and consists of 47 symptoms which were found by factor analysis to represent eight factors (pain, concentration, behavioural change, autonomic reactions, water retention, negative affect, arousal, and control).

The data obtained by questionnaire have been demonstrated to be stable over time, scores not being affected by the menstrual phase at the time of responding. Concurrent validity was determined by comparison of MDQ results with reviews of previous studies and the conclusions seemed consistent. Non-relevant symptoms included in the questionnaire also act as a check on the validity of responses.

MOOS, R. H. (1968). 'The development of a menstrual distress questionnaire', *Psychosomatic Medicine*, **30**, 853.

—— KOPELL, B. S., MELGES, F. T., YALOM, I. D., LUNDE, D. T., CLAYTON, R. B., and HAMBURG, D. A. (1969). 'Fluctuations in symptoms and moods during the menstrual cycle', *J. Psychosom. Res.* **13**, 37.

Peptic Ulcer Questionnaire

This questionnaire was designed for prevalence surveys of peptic ulcer. It is an administered questionnaire consisting of six items dealing with symptomatology. The questionnaire was validated by administration to 201 patients in Jerusalem prior to X-ray. The results of the

X-ray were then compared with the responses given at interview and it was found that the questionnaire was able to differentiate persons who had a peptic ulcer from those who had other gastro-intestinal pathology or were free of pathology. The questionnaire was estimated to have a sensitivity of 82 per cent and specificity of 56 per cent. The respondents were also asked eleven questions from the gastro-intestinal section of the Cornell Medical Index and these were found to produce less accurate results than the peptic ulcer questionnaire.

The researchers have found significant differences in the pattern of responses from different ethnic groups. In particular, they noted a significantly lower sensitivity and higher specificity for European respondents than for respondents from Asia or the Middle East.

DUNN, P. J., and COBB, S. (1962). 'Frequency of peptic ulcer among executives, craftsmen and foremen', *J. occ. Med.* **4**, 343.

EPSTEIN, L. M. (1969). 'Validity of a questionnaire for diagnosis of peptic ulcer in an ethnically heterogeneous population', *J. chron. Dis.* **22**, 49.

Structured Interview for Joint Diseases

This questionnaire forms the basis for a five-minute semi-structured interview for the investigation of the symptoms and visible stigmata of joint disease. It also aims to assess the impact of these diseases on respondents' life-style and use of medical resources.

The structured interview was considered by the investigators to provide a satisfactory instrument for survey work as it proved simple and quick to administer using non-professional staff. Interview technique was standardized by the use of tape-recorded interviews. The questionnaire was validated against X-rays and shown to be an effective discriminator of joint disease. It was found to have a high acceptance rate (92 per cent) and to be more acceptable to respondents than the clinical examination and X-ray proceedings which followed it.

ACHESON, R. M., COLLART, A. B., GREENBERG, R. H., and CLEMETT, A. R. (1968). 'New Haven survey of joint diseases: photographs and other variables in screening for arthritis of the hands', *Am. J. Epidemiol.* **90**, 224.

—— CHAN, Y-K., and PAYNE, M. (1969). 'New Haven survey of joint diseases: the interrelationships between morning stiffness, nocturnal pain and swelling of the joints', *J. chron. Dis.* **21**, 533.

—— —— and CLEMETT, A. R. (1970). 'New Haven survey of joint diseases: XII. Distribution and symptoms of osteoarthritis in the hands with references to handedness', *Ann. rheum. Dis.* **29**, 275.

Thyroid Follow-Up Questionnaire (see p. 94)

This questionnaire was designed as a means of following up patients, who have been treated for thyrotoxicosis by radioactive iodine therapy, in order to detect whether they develop symptoms of hypothyroidism. It is a postal questionnaire consisting of nine questions. The questionnaire was validated by a comparison of questionnaire results with clinical assessments. It was shown that the discrimination of symptoms compared favourably with that of the clinical interview. Return rates for the questionnaire were very high.

BARKER, D. J. P., and BISHOP, J. M. (1969). 'Computer-based screening system for patients at risk of hypothyroidism', *Lancet*, **ii**, 835.

Transient Ischaemic Attacks Questionnaire

This administered questionnaire covers ten symptoms of transient ischaemic attacks, each symptom being dealt with by a series of detailed questions administered according to a simple branching technique. The questionnaire is scored by a weighting system developed by reference to severity and frequency of symptoms. The development of the questionnaire took place in three stages, the first involved the estimation of symptom frequencies and significance in clinical practice, the second involved the identification of needed additional information about each positive symptom and the third involved pretesting the phrasing and discriminating value of the questionnaire.

Evaluation of the questionnaire was carried out by interviewing 336 out-patients in their homes using the questionnaire. Of these 91 were given a standardized clinical examination by two neurologists, who classified them as having 'definite', 'possible', or 'no transient attack' on the basis of the examination. This statement by each neurologist was accepted as the absolute criterion of the presence of transient ischaemic attacks. It should be added that neurological and cardiovascular findings did little to influence the opinions of the neurologists and their decisions were made largely on the basis of the case-history; this suggests that the subject lends itself to questionnaire analysis.

Comparison of questionnaire responses with the findings of the clinical examinations demonstrated that for subjects classified as 'definite' the questionnaire had a sensitivity of 83 per cent and a specificity of 65 per cent. It was found, however, that the frequency of

'possible' diagnoses decreased as the study progressed and each neurologist gained more experience. The value of the examination as a criterion would, therefore, seem to be a function of clinical practice. To check this, all examination reports were reread by one of the neurologists and this resulted in a change in respondent allocation, the most significant difference being a reduction in the number of 'possible' diagnoses. On the basis of this review the questionnaire's sensitivity and specificity for definite attacks were 92 and 66 per cent respectively.

The researchers hope to improve the sensitivity of the questionnaire, thus enhancing its value as a screening instrument, by improving the weighting system and method of analysing symptom response. The clinical value of the questionnaire was assessed by a two-year follow-up study of 'definite' diagnoses to see if any deaths had occurred which could be attributable to the initial diagnoses. None of the subjects had died from conditions associated with transient ischaemic attacks. This has caused the researchers to express some reservation concerning the potential usefulness of such a questionnaire programme, although the samples used in the follow-up study were too small to yield significant conclusions.

HENDERSON, M., CUCINELL, E., and NELSON, E. (1969). 'Identification of transient ischaemic attacks by questionnaire', *Trans. Am. Neurol. Ass.* **93**, 218.

8. Psychiatric questionnaires

Cornell Word Form

Cornell Word Form (CWF) was designed as a means of making rapid psychiatric assessment of large numbers of persons. It provides a descriptive sketch of the individual's adaptative and adjustive mechanisms in a manner not apparent to the respondent. The investigators claim it to be particularly useful in situations where strong motivation might interfere with response accuracy. The questionnaire is designed to be an adjunct to and not a replacement for the interview by providing insight into how the respondent deals with real-life situations.

The questionnaire is self-administered and consists of 60 items. The respondent is presented with a list of words: for each word he must choose a word from two others which he thinks goes with it. About half of the items refer to symptoms of maladjustment associated with body dysfunctions, and the rest to behavioural, attitudinal, and emotional disturbances. Reliability was assessed by test–retest after an interval of a week with 100 respondents. This yielded a correlation coefficient of +0·81. Validity was assessed by administering the questionnaire to 100 'normal' respondents and 100 respondents who were emotionally disturbed. A comparison of the results revealed that 29 items were able to differentiate the two groups at a statistically significant level, while the remaining 51 were able to give qualitative cues which could contribute to clinical insight.

MITTLEMAN, B., and BRODMAN, K. (1946). 'The Cornell Indices and Cornell Word Form: construction and standardisation', *Ann. N.Y. Acad. Sci.* **46**, 575.

WEIDER, A., and WESCHLER, D. (1946). 'The Cornell Indices and Cornell Word Form: No. 2. Results', ibid. **46**, 579.

—— MITTELMAN, B., WESCHLER, D., and WOLFF, H. G. (1955). 'Further developments of the Cornell Word Form', *Psychiat. Q.* **29**, 588.

Middlesex Hospital Questionnaire

The Middlesex Hospital Questionnaire (MHQ) is a self-rating questionnaire for detection of symptoms of psychoneurosis. It consists of

48 items divided into six sub-tests of free-floating anxiety, phobic anxiety, obsessive-compulsive traits and symptoms, and hysterical traits and symptoms. Validation was carried out by comparing sub-test results of normal respondents with psychiatric hospital patients and also by correlation of sub-tests with clinical ratings. All sub-tests were found to be valid on at least one of the criteria used. Comparison of results of the sub-tests by correlation revealed that they measured distinct and separate factors.

CROWN, S., and CRISP, A. H. (1966). 'A short clinical diagnostic self-rating scale for psychoneurotic patients. The Middlesex Hospital Questionnaire', *Br. J. Psychiat.* **112**, 917.

—— (1974). 'The Middlesex Hospital Questionnaire (MHQ) in clinical research: a review', in Pichot, P. (ed.), *Psychological Measurement* (Basle-Kager).

Self-rating Questionnaire for Depression

This is a self-administered questionnaire which consists of 18 items, 12 of which are orientated symptomatically positive, that is, the more affirmative the response the greater the abnormality. The remaining 6 items are negatively orientated and scored separately, and they act as controls for validation and detection of acquiescence set. Validation was carried out by assessing the significance of the difference between scores obtained by 'depressed' respondents and 'non-depressed' respondents. A further assessment was made by correlation of questionnaire scores with follow-up clinical assessments. On the basis of these criteria the questionnaire appears to be an effective discriminator of depression, although the researchers have pointed to certain weaknesses which arise from the use of a rating-scale. Given these limitations, however, the researchers consider the questionnaire to be a useful instrument of measurement and it is at present in use in both an office and clinical setting (personal communication).

ROCKLIFF, B. W. (1969). 'A brief self-rating questionnaire for depression (SRQ-D)', *Psychosomatics*, **10**, 236.

General Health Questionnaire (see p. 94)

Belying its title, this questionnaire was designed as a means of assessing the severity of psychiatric disturbance in both a clinical and research context. It consists of 60 items dealing with recent symptoms which aim at the assessment of present mental state rather than personality traits or prediction of future behaviour. The questionnaire

begins with overtly physical symptoms and proceeds to more obviously psychiatric items. The questionnaire has been validated for a general practice setting by administration to 200 respondents followed by an independent assessment by a psychiatrist. On this basis it was found that 91·5 per cent of respondents were correctly classified as 'well' or 'ill' by the questionnaire and correlation of clinical assessment of severity of disturbance with questionnaire scores yielded a coefficient of +0·80. The questionnaire was estimated to have a sensitivity of 95·1 per cent and a specificity of 87·8 per cent, indicating that it may be a useful screening instrument.

GOLDBERG, D. P. (1969). 'The assessment of the severity of non-psychotic psychiatric illness by means of a questionnaire', DM thesis, Oxford University.
—— and BLACKWELL, B. (1970). 'Psychiatric illness in general practice: a detailed study using a new method of case identification', *Br. med. J.* **2**, 439.

Depression Inventory (Beck)

This questionnaire is composed of 21 categories of symptoms and attitudes, as follows:

A	Sadness	L	Social withdrawal
B	Pessimism	M	Indecisiveness
C	Sense of failure	N	Body image change
D	Dissatisfaction	O	Work retardation
E	Guilt	P	Insomnia
F	Expectation of punishment	Q	Fatigability
H	Self-accusation	R	Anorexia
I	Suicidal ideas	S	Weight loss
J	Crying	T	Somatic preoccupation
K	Irritability	U	Loss of libido

It will immediately be apparent that most categories deal with emotional disturbance rather than physiological correlates. However, the questionnaire is interesting for it is administered by an interviewer rather than being self-administered which is more usual for this type of instrument. Within each category the respondent is asked to pick out one statement to describe his present feelings. A depression score is obtained which is the sum of the weighted responses to all categories. The questionnaire has been extensively tested and reliability and validity shown to be of a high order.

BECK, A. T. (1967). *Depression: Clinical, Experimental and Theoretical Aspects* (London: Staples Press).

Part III

Questionnaire examples

Introductory note

The following extracts from five questionnaires have been selected to demonstrate various aspects of questionnaire style and content. Thus, they provide illustrative models but it is necessary to add that they, like most other measurement techniques, are not beyond reproach or improvement. We are grateful to the authors for their kind permission to publish these examples.

Questionnaire on Respiratory Symptoms

(MRC, 1966; see p. 79)

Part of the first page of this questionnaire showing the items on cough and phlegm production. Note the interpolation of instructions and the emphasis given to the word **usually** by printing this in heavy type. The numbering of questions is not continuous as questions 2, 4, 7, and 9, which were present in the 1960 version, have been omitted. This part of the questionnaire is precoded so that answers are assigned a numerical value for recording on punched cards. Column numbers are shown in heavy type down the right-hand margin.

Use the actual wording of each question. Put X in appropriate square after each question. When in doubt record 'No'.

PREAMBLE I am going to ask you some questions, mainly about your chest. I should like you to answer 'YES' or 'NO' whenever possible.

COUGH

1. Do you **usually** cough first thing in the morning [on getting up*] in the winter?
   Count a cough with first smoke or on first going out of doors. Exclude clearing throat or a single cough.
   1 ☐ Yes 2 ☐ No **1**

3. Do you **usually** cough during the day—or at night— in the winter?
   Ignore an occasional cough.
   1 ☐ Yes 2 ☐ No **2**

   If 'No' to both questions 1 and 3, go to question 6.
   If 'Yes' to either question 1 or 3:

5. Do you cough like **this** on most days [or nights*] for as much as three months each year? 1 ☐ Yes 2 ☐ No 9 ☐ N.A. **3**

PHLEGM

6. Do you **usually** bring up any phlegm from your chest first thing in the morning [on getting up*] in the winter? 1 ☐ Yes 2 ☐ No **4**
Count phlegm with the first smoke or on first going out of doors. Exclude phlegm from the nose. Count swallowed phlegm.

8. Do you **usually** bring up any phlegm from your chest during the day—or at night—in the winter? 1 ☐ Yes 2 ☐ No **5**
Accept twice or more.
If 'No' to both questions 6 and 8, go to question 12a.
If 'Yes' to either question 6 or 8:

10. Do you bring up phlegm like **this** on most days [or nights*] for as much as three months each year? 1 ☐ Yes 2 ☐ No 9 ☐ N.A. **6**

* *For subjects who work by night.*

Questionnaire on Respiratory Symptoms

(MRC, 1966; see p. 79)

A later section of this questionnaire showing the structuring of questions designed to elicit present or past smoking habit. This section is not precoded.

TOBACCO SMOKING

35a. Do you smoke? ☐ Yes ☐ No
Record 'Yes' if regular smoker (as defined in question 35b) up to one month ago.
If 'No' to question 35a, ask question 35b.
If 'Yes' to question 35a:

Do you inhale the smoke? ☐ Yes ☐ No

Would you say you inhale the smoke slightly (S), moderately (M), deeply (D)? ☐ S ☐ M ☐ D

How old were you when you started smoking regularly? ..... years old

How many manufactured cigarettes do you usually smoke per day? ..... per working day
..... at weekends

How much tobacco (oz/g) do you usually smoke per week in hand-rolled cigarettes? .....

How much pipe tobacco (oz/g) do you usually smoke per week? .....

How many cigars do you usually smoke per week? .....
Specify large (L) or small (S).

35b. Have you ever smoked as much as one cigarette a day [or one ounce of tobacco a month] for as long as a year? ☐ Yes ☐ No

If 'No' to question 35b, go to question 38.
If 'Yes' to question 35b:

How old were you when you started smoking regularly? ..... years old

How old were you when you last gave up smoking? ..... years old

How many manufactured cigarettes per day were you smoking before you gave up? ..... per working day
..... at weekends

How much tobacco (oz/g) per week were you smoking in hand-rolled cigarettes before you gave up? .....

How much pipe tobacco (oz/g) per week were you smoking before you gave up? .....

How many cigars per week were you smoking before you gave up? .....

Specify large (L) or small (S).

Screening Questionnaire

(Bennett and Fraser, 1972; see p. 76)

This is the first of two pages of questions many of which are derived from other questionnaires. Most are of the dichotomous type with the positive answers first. This favours a response set. However, simple reversion of some responses, or rephrasing of some questions to reverse the responses, were rejected during development in favour of clarity in a short instrument designed for screening purposes.

INSTRUCTIONS ON ANSWERING

These questions are asking about your present and recent health and well-being. Please circle the answer which is right, thus: (YES)/NO. Try to answer all the questions on these pages. At the end there is a space to add any additional symptoms that are now troubling you.

1. Do you have difficulty in reading small print? (with spectacles if you have them) YES / NO
2. Do you have difficulty in seeing distant objects? (with spectacles if you have them) YES / NO
3. Do you have difficulty in hearing and understanding most things people say, without seeing their face or lips? YES / NO
4. Have you found it difficult to concentrate recently? YES / NO

5. Does you hand often shake when you try to do something? YES / NO
6. Are you ever troubled by shortness of breath when hurrying on the level or walking up a slight hill? YES / NO
7. Do you get short of breath when walking with other people at an ordinary pace on level ground? YES / NO
8. How many flights of stairs can you climb without stopping? 0 / 1 / 2 / 3 / 4
9. Do you sweat very easily, even on cold days? YES / NO
10. Are you troubled with a frequent or persistent cough? YES / NO
11. Do you usually cough first thing in the morning in the winter? YES / NO
    If 'YES', Do you cough like this on most mornings for as much as three months each winter? YES / NO
12. Do you usually bring up any phlegm from your chest first thing in the morning in the winter? YES / NO
    If 'YES', Do you bring up phlegm like this on most mornings for as much as three months each winter? YES / NO
13. Have you ever had attacks of shortness of breath with wheezing? YES / NO
14. Do you suffer from palpitations or breathlessness? YES / NO
15. Do you ever have blackouts, dizzy spells or faints? YES / NO
16. Have you ever had any pain or discomfort in your chest? YES / NO
    If 'YES', Do you always get it when you walk uphill or hurry? YES / NO
    Does it pass off if you stop still or slow down? YES / NO
17. Have you ever had a severe pain across the front of your chest lasting for half-an-hour or more? YES / NO
18. Do you ever have severe pain in either leg on walking? YES / NO
    If 'YES', Does the pain ever wear off while you are still walking? YES / NO

Disability Schedule

(Garrad and Bennett, 1971; see p. 70)

This is the first of two pages of a questionnaire designed to identify and score disability. Each sub-section is introduced by a lead question asking in most instances 'Do you . . . ?' rather than 'Can you . . . ?' or 'Could you . . . ?'. Interviewers are trained to probe to identify the appropriate level of performance. When a respondent's answer falls between two defined levels of performance, the higher level is recorded so that disability is underestimated rather than overestimated.

SECTION I

MOBILITY

Walking Do you walk outdoors in the street (with crutch or stick if used)?

If 'Yes': one mile or more	☐	If 'No':		and:	
¼ mile	☐	Between rooms	☒	Unaccompanied	☐
100 yds.	☒	Within room	☒	Accompanied	☒
10 yds.	☒	Unable to walk	☒	Acc. & support.	☒

Stairs Do you walk up stairs?		Do you walk down stairs?	
To 1st floor or above	☐	From 1 floor to another	☐
5–8 steps or stairs	☒	5–8 steps or stairs	☒
2–4 steps or stairs	☒	2–4 steps or stairs	☒
1 step	☒	1 step	☒
mount stairs other than by walking	☒	goes down stairs other than by walking	☒
unable to mount stairs	☒	unable to descend stairs	☒
Unaccompanied	☐	Unaccompanied	☐
Accompanied	☒	Accompanied	☒
Acc. & support	☒	Acc. & support	☒
No need to mount stairs	☐	No need to descend stairs	☐

Transfer

	Yes	No		Yes	No
Do you need help to get into bed?	☒	☐	Do you need help to sit down in a chair?	☒	☐
Do you need help to get out of bed?	☒	☐	Do you need help to stand up from a chair?	☒	☐
Bedfast	☒		Not applicable	☐	

Travel

Do you drive yourself in a car?		Do you travel by bus or train?			
		If 'Yes':		If 'No':	
Normal (unadapt.)	☐	Whenever necessary	☐	Unable to use bus and train	☒
Adapted	☐	Only out of rush hour	☒	Unable to use bus, train and car	☒
Invacar	☒	and:			
Self-propelled vehicle (outdoors)	☒	Unaccompanied	☐	Does not travel by choice	☐
Does not drive	☐	Accompanied	☒	Uses private transport by choice	☐

Thyroid Follow-Up Questionnaire
(Barker and Bishop, 1969; see p. 82)
A short, simple questionnaire for a very specific purpose. The clarity of content and layout contribute greatly to its successful use and high return rate. The method of writing in answers to dichotomous questions can only be used successfully with such a small number of questions.

PLEASE ANSWER YES OR NO TO THE FOLLOWING QUESTIONS.

(1) Do you feel as well as you did a year ago? Answer ☐
(2) Do you now feel the cold more than ever before, so that you cannot get properly warm? Answer ☐
(3) Is your appetite as good as it was a year ago? Answer ☐
(4) Do you feel less energetic than usual? Answer ☐
(5) Do you think you have put on weight in the last year? Answer ☐
(6) Have you, or any of your family or friends, noticed that your voice has recently become huskier or weaker? Answer ☐
(7) Are you getting any fuller in the face? Answer ☐
(8) Has the skin of your arms or legs become more dry or rough during the past year? Answer ☐
(9) Has your hair recently become unruly or more difficult to manage? Answer ☐

General Health Questionnaire
(Goldberg, 1969; see p. 85)
This is the first of four pages containing a total of sixty questions. The instructions are clear and the respondent is not asked to start by recording his name and other personal details. The questions are all in the form of rating scales with four points requiring the respondent to place himself to one end of the scale. Scoring the questionnaire is very simple and takes only a few seconds and a score of twelve or more is considered to indicate a high probability of psychiatric illness.

GENERAL HEALTH QUESTIONNAIRE

Please read this carefully:
We should like to know if you have had any medical complaints, and how your health has been in general, *over the past few weeks.* Please answer ALL the questions on the following pages simply by underlining the answer which you think most nearly applies to you. Remember that we want to know about present and recent complaints, not those that you had in the past.

It is important that you try to answer ALL the questions.

Thank you very much for your co-operation.

HAVE YOU RECENTLY:

1. been feeling perfectly well and in good health?
   Better than usual / Same as usual / Worse than usual / Much worse than usual
2. been feeling in need of a good tonic?
   Not at all / No more than usual / Rather more than usual / Much more than usual
3. been feeling run down and out of sorts?
   Not at all / No more than usual / Rather more than usual / Much more than usual
4. felt that you are ill?
   Not at all / No more than usual / Rather more than usual / Much more than usual
5. been getting any pains in your head?
   Not at all / No more than usual / Rather more than usual / Much more than usual
6. been getting a feeling of tightness or pressure in your head?
   Not at all / No more than usual / Rather more than usual / Much more than usual
7. been able to concentrate on whatever you're doing?
   Better than usual / Same as usual / Less than usual / Much less than usual
8. been afraid that you were going to collapse in a public place?
   Not at all / No more than usual / Rather more than usual / Much more than usual
9. been having hot or cold spells?
   Not at all / No more than usual / Rather more than usual / Much more than usual
10. been perspiring (sweating) a lot?
    Not at all / No more than usual / Rather more than usual / Much more than usual

References

ABRAMSON, J. H. (1966). 'The Cornell Medical Index as an epidemiological tool', *Am. J. Publ. Hlth*, **56**, 2.

ACHESON, R. M., CHAN, Y-K., and PAYNE, M. (1969) 'New Haven survey of joint diseases: the interrelationships between morning stiffness, nocturnal pain and swelling of the joints', *J. chron. Dis.* **21**, 533.

ADLER, M. W., ILIFFE, J. L., HOLLAND, W. W., and KASAP, H. S. (1973). 'Assessment of medical care needs of individuals with chronic cardio-respiratory disease using subjective measures', *Internat. J. Epidemiol.* **2**, 73.

ANASTASI, A. (1961). *Psychological Testing* (New York: Macmillan).

ANDERSON, J., and DAY, J. L. (1968). 'New self-administered medical questionary', *Br. med. J.* **4**, 636.

BACK, K., FESTINGER, L., HYMOVITCH, B., KELLEY, H., SCHACHLER, S., and THIBAUT, J. (1950). 'The methodology of studying rumor transmission', *Hum. Relat.* **3**, 307.

BARKER, D. J. P., and BISHOP, J. M. (1969). 'Computer-based screening system for patients at risk of hypothyroidism', *Lancet*, **ii**, 835.

BELSON, W. A., and DUNCAN, J. A. (1962). 'A comparison of the checklist and the open response questioning systems', *Appl. Stats*, **2**, 120.

BENNETT, A. E., DOLL, R., and HOWELL, R. W. (1970). 'Sugar consumption and cigarette smoking', *Lancet*, **i**, 1011.

—— and FRASER, I. G. P. (1972). 'Impact of a screening programme in general practice: a randomized controlled trial', *Internat. J. Epidemiol.* **1**, 55.

BOYLE, C. M. (1970). 'Difference between patients' and doctors' interpretation of some common medical terms', *Br. med. J.* **2**, 286.

BRODMAN, K., ERDMANN, A. J. JR, and WOLFF, H. G. (1956). *Cornell Medical Index; Health Questionnaire* (manual) (New York: Cornell University Medical College).

BROWN, L. (1937). *Market Research and Analysis* (New York: Ronald Press).

BUDD, M., BLEICH, H., SHERMAN, H., and REIFFEN, B. (1969). *Survey of Automated Medical History Acquisition and Administering Devices*. Lincoln Laboratory, MIT Project Report ACP-4.

BUEKER, K. (1969). 'Reactions to a questionnaire survey', *J. Psychiat. Nurs.* **7**, 215.

BURGESS, A. M., and TIERNEY, J. T. (1970). 'Bias due to non-response in a mail survey of Rhode Island physicians' smoking habits', *New Engl. J. Med.* **282**, 908.

CANNELL, C. F., and KAHN, R. L. (1953). 'The collection of data by interviewing', in Festinger, L., and Katz, D. (eds), *Research Methods in the Behavioural Sciences* (New York: Dryden).

CANNELL, C. F., FOWLER, F. J., and MARQUIS, K. H. (1968). 'The influence of interviewer and respondent psychological and behavioural variables on the reporting in household interviews', *Vital Health Statistics*, series 2, no. 26.

—— and MARQUIS, K. H. (1967). *Effect of Some Experimental Interviewing Techniques on Reporting in the Health Interview Survey* (Ann Arbor: Survey Research Center, Michigan University).

—— —— and LAURENT, A. (1969*a*). *An Experimental Study of the Effects of Reinforcement, Question Length and Re-interviews on Reporting Selected Chronic Conditions in Household Interviews* (Ann Arbor: Survey Research Center, Michigan University).

—— —— —— (1969*b*). *Effects of an Extensive Questionnaire and a Diary Procedure on Health Reporting* (Ann Arbor: Survey Research Center, Michigan University).

CANTRIL, H. (1944). *Gauging Public Opinion* (Princeton NJ: Princeton University Press).

CARTWRIGHT, A., and WARD, A. W. M. (1968). 'Variations in general practitioners' response to postal questionnaires', *Br. J. prev. soc. Med.* **22**, 212.

COLLEN, M. F., CUTLER, J. L., SIEGELAUB, A. B., and CELLA, R. L. (1969). 'Reliability of a self-administered medical questionnaire', *Arch. intern. med.* **123**, 664.

CONKLIN, E. S. (1923). *The Scale of Values Method for Studies in Genetic Psychology* (Oregon: Univ. Orl. Press).

COUCH, A., and KENISTON, K. (1960). 'Yeasayers and naysayers: agreeing response set as a personality variable', *J. abnorm. Psychol.* **60**, 151.

—— —— (1961). 'Agreeing response set and social desirability', ibid. **62**, 175.

CRONBACH, L. J. (1950). 'Further evidence on response sets and test design', *Educ. Psych. Meas.* **10**, 3.

CROWN, S., and CRISP, A. H. (1966). 'A short clinical diagnostic self-rating scale for psychoneurotic patients. The Middlesex Hospital Questionnaire (MHQ)', *Br. J. Psychiat.* **112**, 917.

DEWOLFE, A. S. (1968). 'Self-reports and case histories of schizophrenic patients: reliability and validity of Phillips scale ratings', *J. clin. Psychol.* **24**, 415.

DLIN, B. M., and PERLMANN, A. (1971). 'Emotional response to ileostomy and colostomy in patients over the age of 50', *Geriatrics*, **26**, 113.

DOLL, R., and HILL, A. B. (1964). 'Mortality in relation to smoking: Ten years observations of British doctors', *Br. med. J.* **1**, 1399.

DUNN, J. P., and HAWKES, R. (1966). 'Comparison of respondents and non-respondents in a periodic health examination program to a mailed questionnaire', *Am. J. Publ. Hlth*, **56**, 230.

DURANT, I., and MAAS, I. (1956). 'Who doesn't answer? A symposium on the problem of bias through non-returns in questionnaire surveys and follow-up conducted by post', *Bull. Br. Psychol. Soc.* **29**, 33.

EASTWOOD, R. P. (1940). *Sales Control by Quantitative Methods* (New York: Columbia University Press).

EYSENCK, H. J. (1970). *The Structure of Human Personality* (London: Methuen).

FAHY, A., WATSON, B. G., and MARSHALL, M. (1969). 'Post-anaesthetic follow-up by questionnaire: a research tool', *Br. J. Anaesth.* **41**, 439.

FAIRBAIRN, A. S., WOOD, C. H., and FLETCHER, A. M. (1959). 'Variability in answer to a questionnaire on respiratory symptoms', *Br. J. prev. soc. Med.* **13**, 175.

FELDMANN, J. J., HYMAN, H., and HART, C. W. (1951). 'Interviewer effects on the quality of survey data, *Publ. Opin. Q.* **15**, 734.

FLESCH, R. (1949). *The Art of Readable Writing* (New York: Harper).

FOTHERGILL, J. E., and WILLCOCK, H. D. (1956). 'Interviewers and interviewing', in Edwards, F. (ed.), *Readings in Market Research* (London: British Market Research Bureau).

GARRAD, J., and BENNETT, A. E. (1971). 'A validated interview schedule for use in population surveys of chronic disease and disability', *Br. J. prev. soc. Med.* **25**, 97.

GHISELLI, E. E. (1939). 'All or none versus graded response questionnaires', *J. appl. Psychol.* **23**, 405.

GLASUNOV, I. S. (1968). 'Detection of angina pectoris of effort in a population using a postal questionnaire', *Cor Vasa*, **10**, 183.

GLEDHILL, V. X., and MATHEWS, J. D. (1971). 'Acquisition and storage of clinical data by computer. I. The self-administered symptom history', *Br. J. hosp. Med., Equip. Supp.*

GOLDBERG, D. P., and BLACKWELL, B. (1970). 'Psychiatric illness in general practice: a detailed study using a new method of case identification', *Br. med. J.* **2**, 439.

GOODE, W. J., and HATT, P. K. (1952). *Methods in Social Research* (Chicago: Free Press).

GORDON, B. L. (1966). 'Preciseness and specificity for medical records', *JAMA*, **197**, 569.

GRAHAM, P. A., WALLACE, J., WELSBY, E., and GRACE, J. H. (1968). Evaluating postal detection of registrable blindness. *Br. J. prev. soc. Med.* **22**, 238.

GREENSPOON, J. (1955). 'The reinforcing effect of two spoken sounds on the frequency of two responses', *Am. J. Psychol.* **68**, 409.

GROSSMAN, J. H., BARNETT, O. G., MCGUIRE, M. T., and SWEDLOW, D. B. (1971). 'Evaluation of computer-acquired patient histories', *JAMA*, **215**, 1286.

GUEST, L. (1947). 'A study of interviewer competence', *Internat. J. Opin. Att. Res.* **1**, 17.

GUILFORD, J. R. (1950). *Fundamental Statistics in Psychology and Education* (New York: McGraw-Hill).

GUTTMAN, L. (1950). 'The basis for scalogram analysis', in Stouffer, S. A. (ed.), *Measurement and Prediction* (Princeton NJ: Princeton University Press).

HAEGER, K., ENGKIRST, O., DANIELSSON, T., HALL, P., and MELLNER, A. (1965). *Datamaskinell journalyoring och 'driftsknotrell' vid en kirurgisk klinik. Preliminara erfarenheter* ('Computerized medical records and production control in a surgical department. Preliminary results') *Lakartan*, **62**, 4239.

HALL, G. H. (1972). 'Experience with outpatient medical questionnaires', *Br. med. J.* **1**, 42.

HALLAUER, D. S. (1972). 'Illness behaviour—an experimental investigation', *J. Chron. Dis.* **25**, 599.

HENDERSON, M. (1972). Personal communication.

HENDERSON, M., CUCINELL, E., and NELSON, E. (1961). 'Identification of transient ischaemic attacks by questionnaire', *Trans. Am. neurol. Ass.* **93**, 218.

HEPPER, N. G., HYATT, R. E., and FOWLER, W. S. (1969). 'Detection of chronic obstructive lung disease', *Arch. Environ. Health*, **19**, 806.

HERON, A. (1956). 'The effects of real-life motivation on questionnaire response', *J. appl. Psychol.* **40**, 65.

HOLLAND, W. W. (1968). Unpublished data.

HULKA, B. S., ZYZANSKI, S. J., CASSEL, J. C., and THOMPSON, S. J. (1970). 'Scale for the measurement of attitudes towards physicians and primary medical care', *Medical Care*, **8**, 429.

HUNT, W. A., and VOLKMANN, J. (1937). 'The anchoring of an effective scale', *Am. J. Psychol.* **49**, 88.

HYMAN, H. H., COBB, W. J., FELDMAN, J. J., and HART, C. W. (1954). *Interviewing in Social Research* (Chicago: University of Chicago Press).

INGHAM, J. G. (1965). 'A method for observing symptoms and attitude', *Br. J. soc. Psychol.* **4**, 131.

—— (1969). 'Quantitative evaluation of subjective symptoms', *Proc. R. soc. Med.* **62**, 492.

KAHN, R. L. (1952). 'The comparison of two methods of collecting data for social research: the fixed-alternative questionnaire and the open-ended interview', Unpublished doctoral thesis, Michigan University.

—— and CANNELL, C. F. (1957). *The Dynamics of Interviewing* (New York: Wiley).

KANNER, I. F. (1969). 'Programmed medical history-taking with or without computer', *JAMA*, **207**, 317.

KATZ, D. (1960). 'The functional approach to the study of attitudes', *Publ. Opin. Q.* **24**, 63.

KING, F. W. (1970). 'Psychology in action: anonymous versus identifiable questionnaires in drug usage surveys', *American Psychologist*, **25**, 982.

KINSEY, A. C., POMEROY, W. B., and MARTIN, C. E. (1953). *Sexual Behaviour in the Human Male* (Philadelphia: Saunders).

KOLLER, S. (1967). 'Mathematisch-statistische Grundlagen der Diaknostik', *Klinische Wochenschrift*, **45**, 1065.

KORNHAUSER, A. (1951). 'Questionnaire and interview schedules', in Jahoda, M., Deutsch, M., and Cook, S. W. (eds), *Research Methods in Social Relations*, vol. ii (New York: Dryden).

LAWTON, A. H., SEABURY, C., BRANCH, L. C., AZAR, G. J., and BOND, J. O. (1966). 'Follow-up studies of St. Louis encephalitis in Florida', *Southern Med. J.* **59**, 1409.

LEY, P. (1972*a*). 'Primacy, rated importance, and the recall of medical statements', *J. Hlth & Soc. Behav.* **13**, 311.

—— (1972*b*). *Quantiative Aspects of Psychological Assessment: An Introduction* (London: Duckworth)'

—— GOLDMAN, M., BRADSHAW, P. W., KINCEY, J. A., and WALKER, C. M. (1972). 'The comprehensibility of some X-ray leaflets', *J. Inst. Hlth Educ.* **10**, 47.

—— (1973). 'The measurement of comprehensibility', ibid. **11**, 17.

MACCOBY, E. E., and MACCOBY, N. (1954). 'The interview: a tool of social science', in Lindzey, G. (ed.), *Handbook of Social Psychology* (Mass.: Addison Wesley).

MADOW, W. G. (1967). 'Interview data on chronic conditions compared with information derived from medical records', *Vital Health Statistics*, series 2, no. 23.

MARTIN, M. J., MAYNE, J. G., TAYLOR, W. F., and SWENSON, M. N. (1969). 'A health questionnaire based on paper + pencil medium, individualised and produced by computer. II. Testing and evaluation', *JAMA*, **208**, 2064.

MAYNE, J. G., MARTIN, M. J., MORROW, G. W. JR, TURNER, R. M., and HISEY, B. L. (1969). 'A health questionnaire based on paper + pencil medium, individualized and produced by computer. I. Technique', ibid. **208**, 2060.

—— —— TAYLOR, W. F., O'BRIEN, P. C., and FLEMING, P. J. (1972). 'A health questionnaire based on paper + pencil medium, individualised and produced by computer. III. Usefulness and acceptability to physicians', *Ann. Intern. Med.* **76**, 923.

MEDICAL RESEARCH COUNCIL'S COMMITTEE ON RESEARCH INTO CHRONIC BRONCHITIS (1966). *Questionnaire on Respiratory Symptoms* (London: MRC).

—— (1966). *Instructions for the Use of the Questionnaire on Respiratory Symptoms* (London: MRC).

MELLNER, C. (1970). 'The self-administered medical history', *Acta chir. scand.* **40**, 611.

MERTON, R. K., FISKE, M., and KENDALL, P. L. (1956). *The Focused Interview* (Chicago: Free Press).

METZNER, H., and MANN, F. (1952). 'A limited comparison of two methods of data collection: the fixed alternative questionnaire and the open-ended interview', *Am. sociol. Rev.* **17**, 486.

MILNE, J. S., and WILLIAMSON, J. (1971). 'Comparison of teaching machine with an observer in detection of angina pectoris by questionnaire', *Br. J. prev. soc. Med.* **25**, 105.

MOSER, C. A. (1968). *Survey Methods in Social Investigation* (London: Heinemann).

NEELY, T. E. (1938). 'A study of error in the interview', unpublished doctoral thesis, New York University.

NIXON, J. E. (1954). 'The mechanics of questionnaire construction', *J. educ. Res.* **47**, 481.

OPPENHEIM, A. N. (1968). *Questionnaire Design and Attitude Measurement* (London: Heinemann).

PARTEN, M. B. (1950). *Surveys, Polls and Samples* (New York: Harper).

POMEROY, W. B. (1963). 'The reluctant respondent', *Publ. Opin. Q.* **27**, 286.

QUINN, R. D. (1950). 'Psychotherapists' expressions as an index to the quality of early therapeutic relationships', unpublished doctoral thesis, Chicago University.

REEDER, L. G. (1963). 'Mailed questionnaire in longitudinal health studies: the problem of maximizing and maintaining response', *Publ. Opin. Q.* **27**, 294.

RICE, S. A. (1929). 'Contagious bias in the interview', *Am. J. Sociol.* **3**, 420.

RICKELS, K., ANDERSON, J., and HOWARD, K. (1968). 'Dropout contact by mail: a useful research tool', *Dis. nerv. Syst.* **29**, 545.

RIESMAN, D., and GLASER, N. (1948). 'The meaning of opinion', *Publ. Opin. Q.* **12**, 633.

ROCKLIFF, B. W. (1969). 'A brief self-rating questionnaire for depression (SRQ-D)', *Psychosomatics*, **10**, 236.

ROGERS, C. R. (1942). *Counselling and Psychotherapy* (Mass.: Houghton-Mifflin).

—— (1961). *On Becoming a Person* (Mass.: Houghton-Mifflin).

ROPES, M. W., BENNETT, G. A., GOBB, S., and JESSAR, R. A. (1957). 'Proposed diagnostic criteria for rheumatoid arthritis', *Ann. rheum. Dis.* **16**, 118.

ROSE, G. A. (1962). 'The diagnosis of ischaemic heart pain and intermittent claudication in field surveys', *Bull. Wld Hlth Org.* **27**, 645.

RUGG, H. O. (1921). 'Is the rating of human character practicable', *J. educ. Psychol.* **12**, 425.

SCOTT, E. (1961). 'Research on mail surveys', *Jl R. statist. Soc.* **24**, 143.

SCHOOLER, K. K. (1956). 'A study of error and bias in coding responses to open end questions', *Dissertation Abstr.* **16**, 2542.

SHAPIRO, M. B. (1961). 'A method of measuring psychological changes specific to the individual psychiatric patient', *Br. J. med. Psychol.* **34**, 151.

SHEATSLEY, P. B. (1950). 'An analysis of interviewer characteristics and their relationship to performance. I', *Internat. J. Opin. Att. Res.* **4**, 473.

—— (1951*a*). 'An analysis of interviewer characteristics and their relationship to performance. II', ibid. **5**, 79.

—— (1951*b*). 'An analysis of interviewer characteristics and their relationship to performance. III', ibid. **5**, 191.

—— (1951*c*). 'The art of interviewing and a guide to interviewer training and selection', in Jahoda, M., Deutsch, H., and Cook S. W. (eds), *Research Methods in Social Relations*, vol. ii (New York: Dryden).

SKIPPER, J. K., and ELLISON, M. D. (1966). 'Personal contact as a technique for increasing questionnaire returns from hospitalized patients after discharge', *J. Hlth & Hum. Behav.* **7**, 211.

SLACK, W. (1971). 'Computer-based interviewing system dealing with nonverbal behaviour as well as keyboard responses', *Science*, **171**, 84.

SMITH, P. C., KENDALL, K. M., and HULIN, C. I. (1969). *The Measurement of Satisfaction in Work and Retirement* (Chicago: McNally).

SPROSTY, P. J. (1964). 'The use of questions in the diagnostic problem solving process', in Jacquez, J. H. (ed.), *The Diagnostic Process* (Ann Arbor: Michigan University Press).

STEWART, N., and FLOWERMAN, S. H. (1949). 'An investigation of two different methods for evaluation of interviewer job performance', Paper read at AAPOR.

STOUFFER, S. A. (ed.) (1950). *Measurement and Prediction* (Princeton NJ: Princeton University Press).

STOUT, C., WIGHT, M. A., and BRUHN, J. G. (1969). 'The Cornell Medical Index in disability evaluation', *Br. J. prev. soc. Med.* **23**, 251.

SUCHMAN, E. A., PHILLIPS, B. S., and STREIB, G. F. (1958). 'An analysis of the validity of health questionnaires', *Social Forces*, **36**, 223.

SYDIAHA, D., STEWART, A., and LAFARE, H. G. (1968). 'A study of mail questionnaire technique with discharged psychiatric patients', *Internat. J. soc. Psychiat.* **14**, 135.

THURSTONE, L. L., and CHAVE, E. J. (1929). *The Measurement of Attitude* (Chicago: University of Chicago Press).

VERNON, P. E. (1964). *Personality Assessment* (London: Methuen).

VERPLANCK, W. S. (1955). 'The control of the content of conversation: reinforcement of statements of opinion', *J. abnorm. soc. Psychol.* **51**, 668.

WALSHE, F. M. R. (1956). 'The nature and dimensions of nosography in modern medicine', *Lancet*, **ii**, 1059.

WATERS, W. E. (1973). 'The epidemiological enigma of migraine', *Internat. J. Epidemiol.* **2**, 189.

WYATT, D. F., and CAMPBELL, D. T. (1950). 'A study of interviewer bias as related to interviewer's expectations and own opinions', *Internat. J. Opin. Att. Res.* **4**, 77.

WHYTE, W. H. (1957). *The Organization Man* (London: Jonathan Cape).

WILLIAMS, R., JOHNSTON, M., WILLIS, L., and BENNETT, A. E. (1975). 'Disability: a model and a measurement technique', *Br. J. prev. soc. Med.* (in press).

WITTS, L. J. (ed.) (1959). *Medical Surveys and Clinical Trials* (London: Oxford Medical Publications).

WOOD, M. M., and ELWOOD, P. C. (1966). 'Symptoms of iron deficiency anaemia: a community survey', *Br. J. prev. soc. Med.* **20**, 117.

WYNDER, E. L., and SCHOTTENFELD, D. (1969). 'A standard medical history form', *Cancer*, **24**, 850.

YOUNG, D. W. (1971). 'Assessment of questions and questionaries', *Meth. Inform. Med.* **10**, 222.

—— (1972). 'Comparison of information collected by a questionary with that in the patient's hospital record', ibid. **11**, 20.

—— (1972). 'Evaluation of a questionary', ibid. **11**, 15.

ZBOROWSKI, M. (1952). 'Cultural components in response to pain', *J. Social Issues*, **8**, 16.

Additional references

BASS, B. M., and WURSTER, C. R. (1956). 'Using mark-sense for ratings and personal data collection', *J. appl. Psychol.* **40**, 269.

BENNETT, E. M., BLOMQUIST, R. L., and GOLDSTEIN, A. C. (1954). 'Response stability in limited response questioning', *Publ. Opin. Q.* **18**, 218.

BENNETT, G. K. (1938). 'The relative efficiency of fine and coarse weighting of questionnaire items', *Psychol. Bull.* **35**, 642.

BOTWINICK, J. (1969). 'Disinclination to venture response versus cautiousness in responding; age differences', *J. gen. Psychol.* **115**, 55.

COCHRANE, A. L., CHAPMAN, P. J., and OLDHAM, P. D. (1951). 'Observer's errors in taking medical histories', *Lancet*, **i**, 1007.

EDWARDS, A. C. (1957). *The Social Desirability Variable in Personality Assessment and Research* (New York: Dryden).

GRAY, P. G. (1955). 'The memory factor in social surveys', *J. Am. Statist. Ass.* **50**, 344.

GRUNTZIG, A., and GALLA, J. (1970). 'The results of a questionnaire in comparison to physician's diagnosis', *Method. Inform. Med.* **9**, 121.

GUETZKOW, H. (1950). 'Utilizing and categorizing problems in coding qualitative material', *J. clin. Psychol.* **6**, 47.

GULLAHORN, J. E., and GULLAHORN, J. T. (1963). 'An investigation of three factors in response to mailed questionnaires', *Publ. Opin. Q.* **27**, 294.

HERSHBERG, P. I. (1969). 'Medical diagnosis: the role of a brief open-ended medical history questionnaire', *J. med. Educ.* **44**, 293.

—— ENGLEBARDT, G., HARRISON, R., ROCKART, J. F., and MCGANDY, R. B. (1971). 'The medical history question as a health screening test', *Arch. intern. Med.* **127**, 266.

HOLLAND, W. W. (ed.) (1970). *Data Handling in Epidemiology* (London: Oxford Medical Publications).

KAPLAN, F. (1964). *The Conduct of Inquiry* (California: Chandler).

KAPLAN, S., and COLE, P. (1970). 'Factors affecting response to postal questionnaires', *Br. J. prev. soc. Med.* **24**, 245.

KENNEDY, F., CLEARY, J. J., ROY, A. D., and KAY, A. W. (1968). 'Switch: a system producing a full hospital case history on computer', *Lancet*, **ii**, 1230.

LEWIS, B. N., and COOK, J. A. (1968). 'Some comments and recommendations concerning inter-coder reliability', *Psychology Reports*, **22**, 1213.

MARLOWE, D., and CROWNE, D. P. (1964). *The Approval Motive* (New York: Wiley).

MILNE, J. S., HOPE, K., and WILLIAMSON, J. (1970). 'Variability in replies to a questionnaire on symptoms of physical illness', *J. chron. dis.* **22**, 805.

MONSON, R. R. (1970). 'Response to questionnaires', *New Engl. J. Med.* **282**, 1377.

MOUNT, G. E. (1947). 'An investigation of the influence of preceding items on subsequent items in a questionnaire', *Am. Psychologist*, **2**, 416.

NEWCOMB, T. (1931). 'An experiment designed to test the validity of a rating technique', *J. educ. Psychol.* **22**, 279.

OLDHAM, P. D. (1968). 'Observer error in medicine', *Proc. R. Soc. Med.* **61**, 447.

PAYNE, S. L. (1951). *The Art of Asking Questions* (Princeton, NJ: Princeton University Press).

PEAK, H. (1953). 'Problems of objective observation', in Festinger, L., and Katz, D. (eds), *Research Methods in the Behavioural Sciences* (New York: Dryden).

ROBINS, L. N. (1963). 'The reluctant respondent', *Publ. Opin. Q.* **27**, 276.

ROGOT, E., and GOLDBERG, I. D. (1966). 'A proposed index for measuring agreement in test re-test studies', *J. chron. Dis.* **19**, 991.

ROSS, H. L. (1963). 'The inaccessible respondent: a note on privacy in city and country', *Publ. Opin. Q.* **27**, 269.

SCADDING, J. G. (1959). 'Principles of definition in medicine with special reference to chronic bronchitis and emphysema', *Lancet*, **i**, 323.

SCHILLING, R. S. F., HUGHES, J. P. W., and DINGWALL-FORDYCE, I. (1955). 'Disagreement between observers in an epidemiological study of respiratory disease', *Br. med. J.* **1**, 65.

SHUTTLEWORTH, F. K. (1941). 'Sampling errors included in incomplete returns to mail questionnaires', *J. appl. Psychol.* **25**, 588.

SMITH, H. L., and HYMAN, H. (1950). 'The biasing effect of interviewer expectations on survey results', *Publ. Opin. Q.* **14**, 491.

SUCHMAN, E. A., and MCCANDLESS, B. (1940). 'Who answers questionnaires?', *J. appl. Psychol,* **24,** 758.

SYMONDS, P. M. (1924). 'On the loss of reliability in ratings due to coarseness of the scale', *J. exp. Psychol.* **7,** 456.

—— (1925). 'Notes on rating'. *J. appl. Psychol.* **9,** 188.

THORNDIKE, E. L. (1920). 'A constant error in psychological ratings', ibid. **4,** 25.

TUCKER, L. R. (1946). 'Maximum validity of a test with equivalent items', *Psychometrics,* **11,** 1.

UHRBROCK, R. S. (1932). 'Rating tendencies of personally selected judges', *J. educ. Psychol.* **23,** 594.

WILDE, G. J. S., and DE WIT, O. E. (1970). 'Self-report and error choice inter-individual differences in the operation of the error-choice principle and their validity in personality and questionnaire tests', *Br. J. Psychol.* **61,** 2.

WYLER, A. R., MASUDA, M., and HOLMES, T. H. (1970). 'The seriousness of illness rating scale: reproducibility', *J. psychosom. res.* **14,** 59.

Index

A selection of publications by the Oxford University Press for the Nuffield Provincial Hospitals Trust

Screening in Medical Care. Thomas McKeown, A. L. Cochrane, *et al. 1968. Demy 8vo. 184pp. £1.75 net.*

In the Beginning. *Studies of the maternity services.* Edited by Gordon McLachlan and R. F. A. Shegog. *1970. Demy 8vo. 192pp. £1.75 net.*

Psychiatric Epidemiology. *Proceedings of the International Symposium held at Aberdeen University 22–5 July 1969.* Edited by E. H. Hare and J. K. Wing, *1970. Demy 8vo. 396pp. £4 net.*

Medical History and Medical Care. *A symposium of perspectives.* Edited by Gordon McLachlan and Thomas McKeown. *1971. Demy 8vo. 256pp. £3 net.*

Patient, Doctor, Society. *A symposium of introspections.* Edited by Gordon McLachlan. *1972. Demy 8vo. 168pp. £2.50 net.*

Evaluating a Community Psychiatric Service. *The Camberwell Register 1964–71.* Edited by J. K. Wing and Anthea M. Hailey. *1972. Demy 8vo. 468pp. £6.50 net.*

Problems and Progress in Medical Care, Ninth Series. *The Future—and Present Indicatives.* Edited by Gordon McLachlan. *1973. Demy 8vo. 206pp. £2.50 net.*

Positions, Movements, and Directions in Health Services Research. *Papers and proceedings of a meeting held at Hertford College, Oxford, 1974.* Edited by Gordon McLachlan. *1974. Demy 8vo. 182pp. £2.50 net.*

Measuring for Management. *Quantitative methods in health service management.* J. R. Ashford, *et al.,* based on research at the University of Exeter. *1975. Demy 8vo. 198pp. £2.75 net.*

Bridging in Health. *Reports of studies on health services for children.* Edited by Gordon McLachlan. *1975. Demy 8vo. 228pp. £2.60 net.*

Specialized Futures. *Essays in honour of Sir George Godber, GCB.* Edited by Gordon McLachlan. *1975. Demy 8vo. 310pp. £6 net.*

Probes for Health. *Essays from the Health Services Research Centre, Birmingham.* Edited by Gordon McLachlan. *1975. Demy 8vo. 188pp. £3 net.*

Prelude to Harmony on a Community Theme. *Health care insurance policies in the Six and Britain.* Jozef Van Langendonck, English text edited and introduced by Gordon Forsyth. *1975. Demy 8vo. 318 pp. £8 net.*

Questionnaires in medicine

A guide to their design and use

A.E. BENNETT & K. RITCHIE

This book is primarily aimed at those who would like to design and use a questionnaire as part of their research, or to aid them in their everyday clinical practice. It goes behind the deceptive simplicity of questionnaires and underlines their importance as a measurement technique. It draws extensively on past medical and psychological research and provides for the first time a short, comprehensive manual.

The book is divided into three parts. The first deals with principles of questionnaire design and use; the second documents questionnaires available; and the third illustrates some model questionnaires. The text contains many examples and there is an extensive list of references.

On a more theoretical level, Part I deals with many of the fundamentals underlying doctor–patient communication. As such it should be interesting and helpful to a wide audience.

ISBN 0 19 721390 1

£3 net