WHO/MNH/PSF/94.8
English only
Distr. General

A USER'S GUIDE TO THE SELF REPORTING QUESTIONNAIRE (SRQ)

DIVISION OF MENTAL HEALTH

WORLD HEALTH ORGANIZATION

GENEVA

WHO/MNH/PSF/94.8
English only
Distr. General

A USER'S GUIDE TO THE SELF REPORTING QUESTIONNAIRE (SRQ)

COMPILED BY M. BEUSENBERG AND J. ORLEY

The Self Reporting Questionnaire (SRQ) has been developed by WHO as an instrument designed to screen for psychiatric disturbance, especially in developing countries. The SRQ consists of 20 questions which have to be answered by yes or no. It may be used either as a self-administered or as an interviewer administered questionnaire. Various additional questions have been used with the SRQ-20, to screen for psychotic disorder and substance abuse.

This document will provide the reader with some insight into the background and applications of the instrument. Statistical data have been reviewed and special attention is given to issues of reliability and validity.

It is apparent from the review that no global, generally applicable cut off score can be recommended for the SRQ, and each study should determine its own. The score used will depend upon the language used, the method of administration, the population answering it, as well as the needs of the research design (for instance: high sensitivity, high specificity or optimum predictive value).

This document has been written for health professionals and those who are planning to conduct research which includes the use of a screening instrument for mental disorders. The SRQ can also be used within general medical practice to rapidly identify those likely to be suffering from a mental disorder and who might therefore benefit from more detailed assessment and treatment for this.

DIVISION OF MENTAL HEALTH
WORLD HEALTH ORGANIZATION
GENEVA
1994

Further copies of this document may be obtained from

Division of Mental Health
World Health Organization
1211 Geneva 27
Switzerland

PREFACE AND ACKNOWLEDGEMENTS

The need for a user's guide for the Self Reporting Questionnaire (SRQ) became apparent after the realization that an increasing number of investigators are using this screening test around the world. Potential researchers have found it hard to obtain adequate information on the SRQ. Until now, no English document has been available for potential users of this screening instrument, although a Spanish psychiatric manual for primary care workers, which includes a section on the SRQ, has been available since 1983 (Climent and De Arango, 1983). This latter provides information on the background of the SRQ, recommendations on its use, a description of each question and an individual glossary and pictorial representations of each item for respondents with a limited education.

The SRQ-20 is an instrument with 20 items which question respondents about symptoms and problems likely to be present in those with neurotic disorder. Versions exist with additional questions concerning psychiatric disorders, substance abuse and epilepsy. Results from all available studies that have used the SRQ-20 (up to 1994) are reviewed in this present document and should provide the reader with some insight into the applications of the instrument. Furthermore, investigators who are planning to conduct research, using a screening instrument for mental disorders are likely to be interested in the psychometrics of the instrument. Since the SRQ is an instrument with proven reliability and validity, this document will be of value to them.

In addition to English, the SRQ has also been used in Afrikaans, Amharic, Arabic*, Bahasa Malaysia, Bengali, Filipino, French*, Italian, Hindi*, Kiswahili, Njanja Lusaka, Portuguese*, Shona, Siswati, Somali*, South Sotho and Spanish*. The versions marked with an asterisk are available upon request, together with further information, from the Director, Division of Mental Health, World Health Organization, 1211 Geneva, Switzerland.

The authors are very much indebted to Dr Mohan Isaac and Dr Willem Kuyken of the WHO Division of Mental Health who gave their advice and lent their expertise. They would also like to thank the following people for their comments on a draft version of this document: Professor Carlos Climent, Professor David Goldberg, Dr Trudy Harpham, Dr Lourdes Ignacio, Dr Jair Mari, Dr Ahmad Mohit, Dr Charles Parry, Dr Naotaka Shinfuku and Dr Mick Power.

M. Beusenberg and J. Orley

Geneva, 1994

CONTENTS

LIST OF TABLES AND FIGURES

INTRODUCTION

Research has shown that mental disorders are common among general medical patients but are not often identified, treated, or referred. There has been increasing interest in ways to improve the diagnosis and treatment of mental disorders in these patients.

It is currently estimated that at least 500 million people in the world suffer from mental disorders, and that only a small proportion of them receive appropriate care. In many developing countries, in particular, trained staff are few, and specialized psychiatric facilities are limited to large urban centres. It has become apparent therefore that the development of human resources is vital, so that mental health care can be extended and made accessible to a greater number of people.

Evidence of widespread and disabling psychiatric morbidity and the lack of care has led the World Health Organization to allocate a high priority to psychiatric provision at primary health care level. Within the framework of a collaborative study on Strategies for Extending Mental Health Care, coordinated by WHO, the Self Reporting Questionnaire (SRQ) was developed as an instrument which was designed to screen for psychiatric disturbance in primary health care settings, especially in developing countries (Harding et al., 1980).

Use of the SRQ as a screening instrument, or more precisely, as a case-finding instrument, has not however been restricted to primary health care settings. Applications of the SRQ vary from research among elderly people in South Africa to research among relatives of people with schizophrenia in a psychiatric clinic in Malaysia.

In recent years the SRQ has been used in some thirty studies from which its psychometric properties can be assessed. The results of a review of these is presented here. The historical background of the SRQ and instructions on its use are also included in this document. Special attention is paid to issues of reliability and validity.

Originally, the SRQ consisted of 25 questions, 20 related to neurotic symptoms, 4 concerning psychotic symptoms and one asking about fits (convulsions)[1]. This is referred to in the text as the SRQ-25. In some studies, just the "neurotic" and "psychotic" items are used, and this is referred to as the SRQ-24. This document concentrates on the SRQ-20 consisting only of the "neurotic" items, for the following reasons.

a) few patients with functional psychoses come spontaneously to primary health facilities for help;
b) reaching psychotic patients usually calls for more active case finding by primary health workers in the community;
c) the need for "psychotic items" to detect psychoses is questionable (often, patients are easily recognized as being psychotic, and, in most situations, psychotic patients are not aware of their condition, hence, a questionnaire might be inappropriate).
d) the psychometric properties of these questions (e.g. their sensitivity and specificity) have not been assessed.

A Spanish version (Climent and De Arango, 1983) also includes five items exploring harmful alcohol use. The World Health Organization has also produced a 10 item Alcohol Use Disorders Identification Test - AUDIT (Babor et al., 1989). This latter could be used together with the SRQ if required.

1 The 4 psychotic items are:
1. Do you feel that somebody has been trying to harm you in some way?
2. Are you a much more important person than most people think?
3. Have you noticed any interference or anything else unusual with your thinking?
4. Do you ever hear voices without knowing where they come from or which other people cannot hear?

and the fifth item is:
5. Have you ever had any fits, convulsions or falls to the ground, with movements of the arms and legs, biting of the tongue or loss of consciousness?

SRQ-20

A copy of the English version of the Self Reporting Questionnaire-20 is shown below.

1.	Do you often have headaches?	yes/no
2.	Is your appetite poor?	yes/no
3.	Do you sleep badly?	yes/no
4.	Are you easily frightened?	yes/no
5.	Do your hands shake?	yes/no
6.	Do you feel nervous, tense or worried?	yes/no
7.	Is your digestion poor?	yes/no
8.	Do you have trouble thinking clearly?	yes/no
9.	Do you feel unhappy?	yes/no
10.	Do you cry more than usual?	yes/no
11.	Do you find it difficult to enjoy your daily activities?	yes/no
12.	Do you find it difficult to make decisions?	yes/no
13.	Is your daily work suffering?	yes/no
14.	Are you unable to play a useful part in life?	yes/no
15.	Have you lost interest in things?	yes/no
16.	Do you feel that you are a worthless person?	yes/no
17.	Has the thought of ending your life been on your mind?	yes/no
18.	Do you feel tired all the time?	yes/no
19.	Do you have uncomfortable feelings in your stomach?	yes/no
20.	Are you easily tired?	yes/no

BACKGROUND OF THE SRQ

The WHO collaborative study on strategies for extending mental health care

The SRQ was developed as part of a collaborative study coordinated by WHO, on strategies for extending mental health care. This study started in 1975 with teams composed of psychiatrists, public health workers and others in Colombia, India, Senegal and Sudan. Later, three more teams from Brazil, Egypt and the Philippines joined the original four. Some areas were rural, some semi-rural and some semi-urban. The work was done in areas with populations varying from 30 000 to 75 000. The areas selected had primary health care facilities but no mental health services. This lack of mental health care was one of the key criteria in selecting the study areas and corresponds to the realities of the vast majority of communities in developing countries.

The technical stimulus for the study came from specific recommendations of a WHO Expert Committee that met in 1974 (World Health Organization, 1975). The Committee strongly endorsed a policy of decentralization and integration of services and addressed itself to the urgent problem of adequate coverage of the population advocating the provision of basic mental health care by primary health workers. Since such workers have only a limited training and are expected to cope with many pressing health problems, the need to limit and define the scope of mental health care provided at this level was stressed, so that only simple and circumscribed tasks should be included in their work.

One of the objectives of the collaborative study was to evaluate the effectiveness of alternative and low-cost methods of mental health care introduced into basic health services. In cooperation with local health authorities and local communities, baseline data were collected by screening adults and children attending primary health facilities to detect mental disorders and by means of an assessment of the attitude and knowledge of local health workers concerning mental health and mental disorders (Climent et al., 1980).

These baseline observations showed a consistent tendency of health workers to miss cases of mental disorders among patients seen in their routine work. Approximately two cases out of three were missed among adults. It was clear that health workers tended to overlook psychological symptoms when physical symptoms were also present. It is possible that their lack of knowledge about treatment methods for psychiatric cases discouraged diagnosis. It has therefore been one of the explicit aims of training programmes to increase health workers' sensitivity to psychological symptoms, to improve diagnostic accuracy and to provide the knowledge and skills for treatment.

The study design called for the development and testing of new research tools. In particular, a reliable and valid technique of case detection was needed. Furthermore, the technique had to be suitable for use by workers with limited training, often working in remote health facilities.

The process of designing such instruments is time-consuming, and it was impossible for each team to develop new instruments based on a detailed study of local symptom patterns, symptom frequency, and their association with mental disorders. On the other hand, making use of an instrument developed in an industrialized country and imposing it on culturally diverse settings would seriously limit the validity of the results. This constraint provided the impetus for developing a screening instrument applicable in different cultural settings: the SRQ.

Development of the SRQ

The items of the SRQ were selected by the participants in the WHO Study, by comparing items in four screening instruments for psychiatric morbidity, each designed for a particular setting, but together representing a wide variety of cultural backgrounds:

1) The Patient Self-report Symptom Form (PASSR), an instrument developed and tested in Cali, Colombia (Climent & Plutchik, 1980);

2) The Post Graduate Institute Health Questionnaire N2 developed by Wig and his colleagues in Chandigarh after they found the Cornell Medical Index to be inappropriate for the Indian setting (Verma & Wig, 1977);

3) The General Health Questionnaire (GHQ) developed originally by Goldberg in England (Goldberg, 1972) but subsequently validated in the United States, Australia, Jamaica, Germany and many other settings;

4) The symptom items on the shortened version of the Present State Examination (PSE) (Wing et al., 1974). The full version of the PSE has been adapted and tested in a wide range of cultural settings.

This comparison produced a preliminary list of 32 items by pooling those which were either identical or very similar in meaning. These were reviewed by the investigators in the WHO collaborative study. Of these 32 items, 20 were finally selected by agreement of representatives from each of the four first-phase study areas (Colombia, India, Senegal and Sudan) as indicative of the presence of non-psychotic mental disorders. Additional selection criteria were ease of translation and comparability of meaning between each geographic region; both criteria facilitated transcultural comparisons (Harding et al., 1980).

The four "psychotic" items were based on items in Foulds' Symptom Sign Inventory (Foulds and Hope, 1968) shown to be most effective in detecting such illnesses.

USE OF THE SRQ

The SRQ was originally designed as a self-administered scale, but was also found to be suitable as an interviewer administered questionnaire. Because of the low literacy rate in some developing countries, respondents may have to have their questionnaires read to them. In order to keep the results comparable within any one study, it is imperative to use the SRQ either as a self-administered questionnaire for all respondents or as an interviewer administered questionnaire for all, and not to mix the two in a single study. When questions are read out to respondents, it presents the researcher with the problem of the influence of subjective elements stemming from the different backgrounds of the research workers. This can interfere with the way items are read, the way they are answered by the patients and on how the answers are finally scored. This chapter offers some guidelines which might be helpful in formatting a common interview pattern.

In this chapter a distinction is made between the use of the SRQ as a screening instrument and the use of the SRQ as a training instrument.

Scoring

Each of the 20 items is scored 0 or 1. A score of 1 indicates that the symptom was present during the past month, a score of 0 indicates that the symptom was absent. The maximum score is therefore 20.

SRQ as a screening instrument

In order to standardize the way in which the questionnaire is answered, it is essential that respondents receive the same instructions. Therefore, the following instructions should be added to the SRQ:

> *"Please read the entire introduction before you fill in the questionnaire. It is very important that everyone taking the questionnaire follows the same instructions.*
>
> *The following questions are related to certain pains and problems, that may have bothered you the last 30 days. If you think the question applies to you and you had the described problem in the last 30 days, answer YES.*
>
> *On the other hand, if the question does not apply to you and you did not have the problem in the last 30 days, answer NO.*
>
> *Please do not discuss the questions with anyone while answering the questionnaire.*

If you are unsure about how to answer a question, please give the best answer you can.

We would like to reassure that the answers you are going to provide here are confidential.

It is very important that if a decision is taken that the SRQ will be interviewer administered in a study, that all interviewers follow exactly the same procedure. (It should be repeated here that it is preferable that self-administration and interviewer administration should not be mixed in any one study.) In order to ensure standard interviewer administration it is preferable to keep this to a very simple procedure. This can for instance be done by getting the interviewers to introduce themselves in a standard way. Insist on being alone with the subject, give a standard introduction of why the questions are being asked and then read an adapted version of the introduction as set out for self-administration (but without the first paragraph). The subject can be told not to discuss questions with others during the administration and that the interviewer is not allowed to discuss the items with the subject during the session, but could for instance go over it again with the subject after all questions have been asked. This second round should not be used to change any ratings, but only to allow subjects to feel that they are being listened to, and discourage them from insisting on discussion during the first round. The subjects could be told that they really must try to answer Yes or No, even if it is approximate and that if they fail to give either answer, the interviewer will repeat the question once more before moving on to the next item. The only additional intervention that could be allowed to the interviewer after the first asking may be to remind the subject, in a standard way, that the item can be discussed once all the items have been answered. The question is then repeated. The above standard procedure is only a suggestion. What is important is that a standard procedure is followed for all subjects and by all interviewers, if there are more than one. The procedure should be piloted for acceptability before the SRQ is used in a study. The above sort of procedure may sound very "impersonal" but something like that is necessary if reliability is to be maintained.

SRQ as a training instrument

Screening of patients in general health clinics in the participating countries within the collaborative study showed that a significant proportion of mental symptoms were reported by patients but were not being picked-up by the health workers (Climent et al., 1980; Harding et al., 1980; Ladrido-Ignacio et al., 1983).

The involvement of primary health workers in mental health care has made clear the need for simple and reliable instruments for their training, so that they can improve their diagnostic skills. In that respect, the SRQ is useful as a training device for several reasons.

First of all, it is a simple and objective scale, easy to evaluate, which covers many important areas of psychopathology.

Secondly, the questions of the instrument are written in a simple language, which is easy to understand. Only a brief training is necessary to use it properly.

Thirdly, its use does not represent a great difficulty either for the trainee or for the trainer, making evaluation of the student's progress an easier task. Due to highly selected items and their specificity, the instrument does not allow major doubts about each of the symptoms.

The fact that questions have been selected on the basis of their frequency, makes its use more relevant for the student since some positive items are found in many medical patients. Finally, the flexibility of the scale permits individual considerations regarding training. As the primary health worker matures in content acquisition and ability to relate to patients, the instrument allows different degrees of knowledge that could be advanced during supervision. In the same way that there are some primary health workers capable of handling only concrete items such as "headaches", others are able to comprehend more complex items such as "hallucinations", which appear in additional questions to the SRQ-20 to screen for psychotic disorders (SRQ-24).

The above-mentioned reasons facilitate the extension of knowledge of psychopathology beyond the item description depending on progress, abilities and motivation of the primary health worker.

An important aim for a training programme is to make primary health workers capable of identifying important psychological symptoms and to make them aware of the relevance of mental disorders in their everyday work.

This should make them more receptive of patients' problems, because the training programme broadens their conception of illness to its emotional components. A simple training manual based on the SRQ items has been published in Spanish (Climent and de Arango, 1983).

THE VALIDITY OF THE SRQ

The validity of a test refers to how well a test measures what it intends to measure. The SRQ has been tested for validity in a series of studies carried out between 1978 and 1993. These studies are analyzed in the section "Review of research with the SRQ". In this section different facets of validity are dealt with individually:

- Face validity
- Content validity
- Criterion validity
- Construct validity

Face validity

The term face validity is a technical description of the judgement that the items are meaningful and relevant. Face validity simply indicates whether, on the face of it, the instrument appears to be assessing the desired qualities. The criterion here represents a subjective judgement based on a review of the measure itself by one or more experts; empirical approaches are rarely used. As is explained in the "Development of the SRQ" on page 4, a team of experts selected the SRQ items from different questionnaires. The SRQ items have a high face validity, that is, they do appear to be assessing symptoms and problems that would be expected to be present in those with mental disorders.

There may be some circumstances in which high face validity is not desirable. Certain questions could be considered offensive by some respondents. Questions like "Do you feel that you are a worthless person?" or "Has the thought of ending your life been on your mind?" might be answered with "no" because respondents would feel ashamed of a positive answer. But of course, the use of this kind of direct, personal question in screening tests is not limited to the SRQ. On the other hand, there are situations in which patients may want to appear worse than they actually are in order to ensure that they receive help. This is easy to do if the items have face validity and much harder if they do not. In this case, it is worth taking into account research done in Ethiopia by Kortmann and Ten Horn (1988), who suggest that some of the high scorers on the SRQ used the questionnaire to emphasise their request for a medical certificate, regardless of the severity of their illness. The respondents who asked for a medical certificate, belonged significantly more often to the group of high scorers than those who did not ask for it. De Jong (1987) wrote a descriptive account of the use of the SRQ in Guinea-Bissau, in which he also stresses the fact that respondents may "misuse" the SRQ e.g. to report subjective feelings of being ill or to complain about situational difficulties.

Another way of examining face validity is to ask the target population what they think the instrument is supposed to measure. No research done so far has ascertained the face validity of the SRQ in this manner.

Content validity

Content validity is closely related to face validity, consisting of a judgement of whether the instrument samples all the relevant or important content or domains (Steiner and Norman, 1992). These two forms of validity consist of a judgement by experts as to whether the scale appears appropriate for the intended purpose. This approach to validation is called 'validation by assumption', meaning the instrument measures what it is supposed to because an expert says it does. So, the theme of content validity is whether the instrument is representative of the universe of questions that could have been asked.

The analysis is a practical or pragmatic approach, which in psychopathology refers to clinical knowledge. The problem here is that the SRQ-20 was designed to cover "mental distress" and more specifically neurotic disorders; there is no agreed upon definition for this concept. It should be stressed that the SRQ was not intended to be either a substitute for, or equivalent to, a clinical diagnosis. It was explicitly developed as the first research instrument in a two-stage detection procedure. The first stage refers to the use of a screening test so that probable cases will be detected, and in the second stage a clinical diagnosis is made. Thus, the SRQ can do no more than express the likelihood of mental disorder. The specific nature of the disturbance has to be determined by a "second stage" diagnostic instrument and/or a clinical interview. Nevertheless, the questions of the SRQ might be expected to tap into diagnostic categories, as defined for instance, in ICD-10 or DSM-III-R.

Inherent in designing a screening test, is the limitation in the ability to assess every mental disorder as described in the ICD-10. Consequently, content validity with respect to covering all the "mental disorders" is fairly low. The scale has insufficient items to cover the domain under investigation. The risk is that some people with mental disorders are not detected by the SRQ. This issue of misclassification is treated in the paragraph on criterion validity. On the other hand, the content validity related to covering "neurotic disorders" is high. Table 1 sets out some mental disorders which are commonly encountered in community settings and in general medical clinics (Goldberg and Huxley, 1992). All these disorders can be screened by using the SRQ. At first sight, there is no question in the SRQ related to obsessive-compulsive disorders, but since individuals with obsessive-compulsive disorders often have depressive and anxiety symptoms, this form of disorder may be detected by the instrument.

Table 1. Diagnostic classification of some common mental disorders, showing current international diagnostic codes (ICD-10)

Depression:	
F32	Depressive episode
F33	Recurrent depressive disorder
F34.1	Dysthymia
Anxiety-related disorders:	
F40	Phobic anxiety disorders
F41.0	Panic disorder
F41.1	Generalized anxiety disorder
F41.2	Mixed anxiety-depressive disorder
F42	Obsessive-compulsive disorder
F43.2	Adjustment disorder
Somatoform disorders:	
F45.0	Somatization disorder
F45.1	Undifferentiated somatoform disorder
Other neurotic disorders:	
F48.0	Neurasthenia

Another interpretation of the concept of content validity is given by Kortmann (1987, 1990) and Kortmann and Ten Horn (1988) who examined this in an Amharic version of the SRQ. They asked their respondents to explain their yes-answers, and to give as many examples as possible to support their information. These explanations were used to assess whether or not the yes-answers validate the concept psychiatrists have in mind when asking each particular question. Do their yes-answers reflect the presence of the psychopathological phenomena which the question intended to uncover? If a question does not have the same meaning for the respondent as it has for the investigator, problems in transcultural communication may emerge. Obviously, since this investigation involved only the Amharic version from Ethiopia, one should be cautious about making generalizations to other cultural settings. If other investigators want to use the SRQ in a particular cultural setting for which there is no information available, it is advisable to do a preliminary study in a small number of patients (a pilot study) in order to evaluate that the instrument is clinically meaningful when used in a target population. The following table summarizes the problems with the content validity of this version according to the investigators.

Table 2. Analysis of yes-answers to the SRQ-20 (Amharic version) by 110 Ethiopian subjects (Kortmann & Ten Horn, 1988)

SRQ-items	Number of yes-answers	Percentage of invalid yes-answers due to			Total (%)
		Language	Motives	Concept	
1. Headache	45	--	2	27	29
2. Lack of appetite	39	3	3	46	51
3. Sleeping problems	51	--	2	29	31
4. Being frightened	51	2	8	33	43
5. Shaking hands	29	7	--	28	34
6. Feeling nervous	45	4	13	9	27
7. Poor digestion	32	--	3	31	34
8. Not thinking clearly	30	3	10	10	23
9. Being unhappy	39	10	5	28	44
10. Crying more than normally	21	5	--	33	38
11. Not enjoying activities	49	8	8	35	51
12. Difficulty with decision-making	35	14	11	26	51
13. Work suffering	28	4	--	14	18
14. Not feeling life is useful	27	--	15	4	18
15. Loss of interest in life	59	7	13	30	51
16. Feeling worthless	32	9	31	9	50
17. Thinking of ending life	19	5	10	16	32
18. Always feeling tired	33	3	9	18	30
19. Stomach problems	50	14	6	10	30
20. Easily tiring	37	5	3	8	16
Total 1-20	755	6	8	23	37

There are three stages at which a yes-answer may become invalid:

1) The respondent may give a yes-answer without having clearly understood the language of the question. Such questions had to be repeated one or more times or needed further explanation before an answer was obtained.

2) The respondent had other motives for saying yes than those pointing to the recognition and confirmation of the phenomena that was being asked for. The investigators assigned a yes-answer to this category if the respondent stuck persistently to his yes, but appeared to be unable to give any further details or examples that made it likely that he answered from his experience. This kind of invalid answer seemed to be based upon the wish to benefit from the advantages of being in the sick role, such as receiving somewhat lighter work or being transferred in order to be reunited with other family members.

3) The investigator, trained in western psychiatry, and the Ethiopian respondent, did not understand one another properly, due to differences of conceptualization of the question.

Criterion validity

The criterion validity of the SRQ is assessed by determining the relationship of the scale with some other measure of the construct under study, ideally, a "gold standard" which has been used and accepted in the field.

In practice, the most appropriate criterion is often a clinical assessment by a trained investigator using a structured research interview of known reliability (Goldberg and Williams, 1988). The underlying assumption is that the psychiatric assessment in the second stage is correct.

A traditional way in which the relationship between the testing instrument, the SRQ, and the criterion is assessed, is by measuring the correlation. The higher the correlation between the scores, the stronger the evidence favouring criterion validity. In this context it is worthwhile mentioning three studies in which the validity has been assessed by measuring the correlation.

1) Araya et al. (1992) calculated the correlation coefficients between the total score on the Spanish translation of the Revised Clinical Interview Schedule (CIS-R; Lewis et al., 1992) and the score on the SRQ (Spanish version) and found a figure of $r = + 0.69$ ($n = 163$).

2) Mari and Williams (1985), using a Portuguese version of the SRQ in Brazil compared the SRQ score with the weighted scores on the CIS (Goldberg et al., 1970). They calculated a correlation coefficient of + 0.74 (n = 260). Since coefficients of more than 0.60 are usually regarded as highly acceptable, these results suggest the SRQ has good criterion validity.

3) El-Rufaie and Absood (1994) calculated a correlation coefficient between the total weighted CIS score and the "best" six SRQ questions as determined by stepwise regression analysis. The results indicate a model in which 60% of the variation in CIS score is explained by variations in these six questions (n = 217).

Araya et al. (1992) also reported a kappa of 0.51 (standard error [SE] = 0.07) when they assessed the SRQ-20 against the CIS-R using a score of 12 or more on CIS-R as the criteria for caseness. Cohen's kappa is a measure of association which corrects for chance agreement. It has a value of 1 if agreement is perfect, of zero if agreement is no better than that due to chance, and a negative value if agreement is less than chance. Agreement is usually considered satisfactory when the kappa value is between 0.40 and 0.75. In the same study they found a kappa for the GHQ-12 of 0.48 (SE = 0.07). Cohen's kappa is usually used for assessing inter-rater reliability and will be discussed again in the section on reliability.

Sensitivity and specificity

A common approach in measuring criterion validity in the case of screening tests, is the use of validity indices like sensitivity and specificity. The measures calculated are designed to indicate how well the results from the SRQ match the results from the criterion instrument, a psychiatric assessment interview.

After administering the SRQ and a criterion interview to a group of respondents, the respondents can be allocated to one of the cells in Table 3. With the dichotomous scores obtained from the screening test as well as the diagnosis from the criterion instrument, the criterion-related validity coefficients (Table 4) can be calculated.

Table 3. 2 x 2 Decision matrix

Screening instrument results	Truth (criterion instrument results)		
	disorder present	disorder absent	total
positive (above cut-off)	A (true positives)	B (false positives)	A + B
negative (below cut-off)	C (false negatives)	D (true negatives)	C + D
total	A + C	B + D	

Table 4. Validity coefficients

Prevalence independent indices of validity
Sensitivity = A / (A+ C)
Specificity = D / (B + D)

Prevalence dependent indices of validity
Positive Predictive Value (PPV) = A / (A + B)
Negative Predictive Value (NPV) = D / (C + D)
Overall Misclassification Rate (OMR) = (B + C) / (A + B + C + D)

Sensitivity is a measure of the screening instrument's ability to detect the true cases of mental disorder identified by the criterion instrument (i.e. the probability of the test correctly identifying a case). Specificity is a measure of the screening instrument's ability to identify the true non-cases identified by the criterion instrument (i.e. the probability of correctly identifying non-cases). When the sensitivity is 100%, it means that the screening test detects all cases. When the screening test excludes all non-cases, it has a specificity of 100%. In theory, both sensitivity and specificity could reach a 100% level. But in practice, this is seldom the case. In fact, the sensitivity and specificity will vary with the cut-off value used in a study to distinguish a probable case from a probable non-case. The cut-off point between

what is normal and what is abnormal is usually arbitrary. The choice of a particular cut-off value involves a trade-off between sensitivity and specificity. Increasing sensitivity inevitably causes some decrease in specificity. Conversely, increasing specificity decreases sensitivity. Decisions on the appropriate criteria for a screening test depend on the consequences of identifying false negatives and false positives. The following table summarizes the results from different studies in which sensitivity and specificity figures of the SRQ have been calculated.

Table 5. Sensitivity and specificity figures of the SRQ-20

Author(s) & Year	Cut-off	Sensitivity	Specificity
Harding et al. (1980)	5-11	73% - 83%	72% - 85%
Dhadphale et al. (1982)	7/8	89.7%	95.2%
Mari and Williams (1985)	7/8	83%	80%
Sen et al. (1987)	11/12	79%	75%
Kortmann and Ten Horn (1988)[1] "psychiatric" group "somatic" group "control" group	8/9	 77% 63% 0%	 44% 68% 100%
Deshpande et al. (1989)	8/9	62.9%	62%
Aldana et al. (1990)	3/4	70%	70%
Penayo et al. (1990)	7/8	81%	58%
Salleh (1990)	5/6	84.8%	83.7%
Araya et al. (1992)	9/10	74%	73%
Carta et al. (1993)[2]	7/8	90%	70%
El-Rufaie and Absood (1994)	5/6	78.3%	75.2%

[1] Figures for the total sample are not available. The study was conducted in a outpatient department of a hospital. Patients who are appointed to one of the somatic clinics constitute the "somatic" group (n = 40); patients referred to the psychiatric clinic constitute the "psychiatric" group (n = 30). A sample from the Addis Ababa (Ethiopia) community is called the "control" group (n = 40).
[2] Based on the SRQ-24.

It can be seen from Table 5 that sensitivity figures range from 62.9% to 90%, whereas specificity figures range between 44% and 95.2%. The variability of the validity indices highlights the fact that a screening instrument needs to be validated in a variety of settings with different populations (Chan & Chan, 1983). The acceptability of the results depends on the purposes of the studies. Some investigators prefer to be over-inclusive, that is, accept the risk of having more false positives by lowering the cut-off score; it is preferable to include false positives rather than exclude false negatives (Carta et al., 1993). Indeed, in epidemiological studies it might be extremely important not to lose probable cases during the first stage of the study, since the majority of those classified as non-cases by the screening test are not followed up. False positives will be excluded in the second stage of the study, and the disadvantage of having too many is that it adds extra work to the second stage. False negatives however are likely to be lost to the study altogether. In most of the studies reported in scientific journals, a proportion of the probable non-cases is followed up in order to calculate figures such as sensitivity and specificity. Normally the cut-off score is calculated in a small preliminary study. If however the translation (version) of the instrument has been used in previous studies on similar populations using the same method of administration, then a cut-off score obtained from that study can be used.

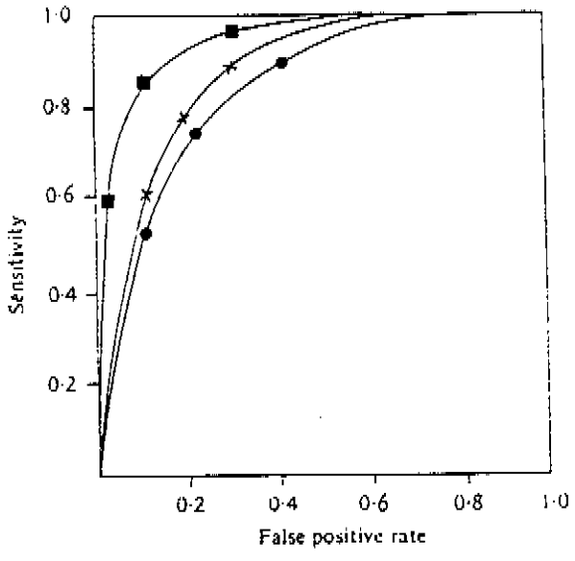
Weighting techniques

At this point it is important to call attention to the use of weighting techniques. Usually, in selecting respondents for the second stage, only a randomly selected proportion of those with low scores on the SRQ will be further examined. On the other hand, all the high scorers should preferably be followed up. It is to be expected that there will be very few psychiatric cases in the group of respondents who scored under the cut-off score. Since case ascertainment is a lengthy and costly procedure, it makes sense to make such a pre-selection. The proportion of those followed up can be increased with increasing screening scores (e.g. *0-1*: 5%, *2-3*: 10%, *4-5*: 20% $\geq$ *6*: 100%). The consequence of using subsamples, however, is that findings have to be weighted to give valid information on the total population studied and on the sensitivity and specificity of the instrument. Failing to do so will lead to inaccurate validity indices and prevalence figures. In Appendix I, a weighting technique is described.

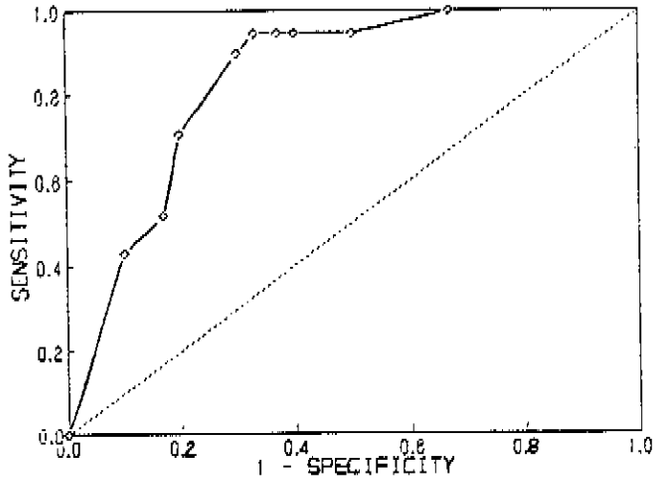
ROC curve

As awareness grew that the showing of sensitivity and specificity figures for a single cut-off score was probably not the most sophisticated way of presenting results, a graphical way was found to be useful: a Relative Operating Characteristics (ROC) curve. ROC analyses for the SRQ against a criterion have been used by Mari and Williams (1985), Araya, Wynn and Lewis (1992), Carta et al. (1993) and El-Rufaie and Absood (1994). Mari and Williams, and Carta et al. actually presented graphs, shown in Figure 1. El-Rufaie and Absood presented graphs not for the whole SRQ 20, but for the six best questions selected by regression analysis and another for the eight most sensitive questions.

Figure 1. ROC curves of the SRQ



ROC curve of the SRQ-20 for each of the primary care clinics.
● Brasilandia; x Servidor; ■ Barra Funda
Mari and Williams (1985)



ROC curve of the SRQ-24.
Carta et al. (1993)

A ROC curve is obtained by plotting sensitivity against the false positive rate (the complement of specificity) for all possible cut-off points of the screening instrument. Thus, a ROC curve can be described as summarizing the possible set of 2 x 2 decision matrices, that results when the cut-off is varied from the largest to the smallest possible value. Mari and Williams (1985) concluded that the ROC curve was superior to the conventional method of presenting validity data.

Goldberg and Williams (1988) sum the advantages of using a ROC analysis compared with giving a single value for sensitivity and specificity at one cut-off point:

- assessment of the discriminating ability of the instrument across the total spectrum of morbidity;
- comprehensive comparative assessment of the performance of two or more screening tests (see the paragraph on construct validity);
- assessment of the effect of varying the threshold score (i.e., raising the cut-off point increases sensitivity at the expense of specificity).

The ascending diagonal of the graph is called the "chance line". Points on this line express the fact that the ability of the SRQ to discriminate between "cases" and "non-cases" is no better than chance. In those situations the sensitivity will be equal to the false positive rate. The greater the height of the curve above this chance line, the better the discrimination between "cases" and "non-cases". Points below the diagonal reflect worse results than could be obtained by chance alone, that is, a low discriminating ability. The investigators who applied a ROC analysis to the SRQ, used a computer program (ROCFIT, Metz et al. 1987), which fits a curve by a maximum likelihood technique. This ROCFIT program makes it possible to calculate the area under the curve, which is regarded as a useful index of the discriminating ability of the screening test. This area can take values from 0.5, implying an ability to discriminate no better than chance to 1.0 for a test with a perfect discrimination.

Looking at the ROC curves made by Mari and Williams, it can be seen that the values were highest for Barra Funda, 0.958, intermediate for Servidor, 0.876, and lowest for Brasilandia, 0.854. The area for the total sample was 0.90. Carta et al. did not give an exact figure but mentioned that the area delimited by the curve and the diagonal is reasonably wide. Araya, Wynn and Lewis calculated a figure of 0.85 (SE = 0.03) for the ROC area .

In addition to the mentioned prevalence-independent indices of validity, there are also prevalence-dependent indices (see Table 4). The latter makes it hard to compare figures from different studies, because prevalence figures of mental disorders from numerous studies revealed significant differences between the figures. They were not only dependent on the culture in which the study took place, but also on the setting (community, hospital or primary health settings).

Predictive values

Predictive values describe the frequency with which the test results represent correct identification of the mental health status among those screened. The Positive Predictive Value (PPV) reflects the probability that a high scorer on the SRQ will be confirmed to be a true case in the subsequent diagnostic assessment. It is hard to achieve a high PPV when mental disorders are rare within the population. The PPV decreases as prevalence increases.The Negative Predictive Value (NPV) reflects the probability that a low scorer on the SRQ will be confirmed to be a true non-case in the subsequent diagnostic assessment. The NPV

increases as prevalence decreases. Because predictive values vary with the prevalence of mental disorders in the screened population, they are useful in deciding whether to use a particular test in a given population. In Table 6 below, it can be seen that in two studies, by Mari and Williams (1985, 1987) and Salleh (1990), the prevalence of mental disorders is mentioned. A prevalence figure calculated by Penayo et al. (1990) in a validation study is not included. They identified 99 cases in a subsample of 171 respondents taken from the general population, which means that the prevalence would reach the unlikely level of 58%. Since those respondents with medium to high SRQ scores were deliberately selected for a clinical interview, it is not considered to be a valid estimation of the prevalence. The paper by El-Rufaie and Absood (1994) does not give a prevalence rate, only reporting on validity data for the SRQ.

Table 6. Predictive values of the SRQ-20

Author(s) & Year	Prevalence	Cut-off	PPV	NPV
Mari and Williams (1985, 1987)	47-56%	7/8	81 % total sample	82 % total sample
Penayo et al. (1990)	not mentioned	7/8 9/10 11/12	73 % 77 % 78 %	69 % 67 % 55 %
Salleh (1990)	23%	5/6	83.9 %	84.7%
El-Rufaie and Absood (1994)	not given	5/6	54.7%	90.1%

Overall misclassification rate

The overall misclassification rate (OMR) is the proportion of respondents who are not correctly assessed. It gives some indication of the performance of the test but one might wrongly assume that false positives and false negatives are of equal weight. It is in fact more interesting to know how many respondents are false positives and how many false negatives. It is common practice to keep the false negatives as low as possible. Even so, in the "Review of research with the SRQ" the OMRs for the SRQ are reported in the papers by Harding (1980), Mari and Williams (1985), Kortmann and Ten Horn (1988), Salleh (1990), Araya et al. (1992) and El-Rufaie and Absood (1994). The figures vary from 15.7% to 33%.

Construct validity

Criterion validity and construct validity both require that the test being assessed be compared against a yardstick which is independent of the test itself. Construct validity differs from criterion validity. There is no direct measure of the hypothetical variable underlying the test for construct validity, whereas an independent and direct measure can be taken in criterion related validity. With regard to the criterion validity, the mental health status of a respondent is also measured independently from the SRQ, by using a standardised psychiatric assessment instrument. Evidence for construct validity, however, is generally circumstantial, e.g. the pattern of symptoms or scores seems to conform to the hypothetical construct of mental illness / minor psychiatric disorder.

The construct validity refers to how well a scale measures the proposed underlying factors or dimensions, in this case of mental illness. Two of the most commonly used types of construct validation are convergent validity and factorial validity. Convergent validity focuses on the extent to which several measures of the same concept correlate with each other. For example, a high correlation between the SRQ and the GHQ is to be expected. Through factor analysis it can be empirically demonstrated whether discernible underlying theoretical constructs exist.

Convergent validity

Chan and Chan (1983) focused in their study on the reliability and validity of the General Health Questionnaire (GHQ-30) in a Chinese context. They evaluated the validity of the GHQ against the SRQ[2]. The results, based on 224 subjects who yielded complete data on the GHQ and SRQ, indicated that the GHQ correlated 0.54 ($P < 0.001$) with the SRQ-20. This is a moderately high correlation. Besides looking at the total score, Chan and Chan also examined the ability of the GHQ and the SRQ in assigning subjects to the "case" or "non-case" categories. As an index of agreement they employed the proportion of agreement or concordance. For details, see Table 9B: "Review of research with the SRQ". High concordance, when a large number of subjects were declared by both instruments to be "cases" or by both to be "non-cases", would indicate convergent validity. It can be seen from the Table that reasonable concordance rates of above 75% were achieved.

Araya et al. (1992) validated the SRQ-20 and the GHQ-12 simultaneously against the criterion of the Revised Clinical Interview Schedule (CIS-R) in a primary care clinic in Santiago, Chile. The differences between the GHQ and the SRQ were small. The SRQ-20 turned out to be slightly more specific than the GHQ (77% versus 73%) which in turn was somewhat more sensitive (76% versus 74%). The area under the ROC curve was slightly larger for the SRQ-20 than for the GHQ score, but this difference was not statistically significant ($Z = 1.40$, $P = 0.15$).

[2] The SRQ contains 5 items taken from the GHQ. Therefore, the results presented here should be interpreted cautiously.

Mari and Williams (1985) concluded that both questionnaires (SRQ-20 and GHQ-12) showed similar results. These psychiatric screening instruments were assessed against the CIS in three primary care settings in Brazil. The results for the total sample will be mentioned here. The product-moment correlation between the two set of scores was +0.72. The validity coefficients for the SRQ and the GHQ were respectively: sensitivity 83% versus 85% and specificity 80% versus 79%. Mari and Williams (1985) concluded further after comparing the two ROC curves that the two instruments were similar in performance; the SRQ was doing slightly, but not significantly, better ($Z = 0.92$). The area under the curves was 0.900 for the SRQ-20 and 0.872 for the GHQ-12.

In an attempt to estimate psychiatric morbidity, Rahim and Cederblad (1989) randomly sampled 204 adults of a suburban area of Khartoum, Sudan. Those who scored high on the Life Events Scale (Theorell, 1982; Cederblad and Höök, 1980), showed significantly more psychiatric symptoms (65.4%) than those with medium or low life events scores (33.8% and 34.5% respectively). High scores on the SRQ were associated with a high life events score ($P<0.05$). Subjects who reported marital discord and divorce had a significantly higher rate of psychiatric morbidity (37.5%) than those who were married without conflicts or divorce (10.5%). High scores on the SRQ were associated with marital discord or divorce ($P < 0.01$). Those with a high score on the Loneliness Scale (Russell et al., 1980) had a 54.3% frequency of psychiatric symptoms and those with minimum loneliness score had only 20.5%. Again, a significant association was found between high scores on the SRQ and a high loneliness index ($P < 0.001$). These results add evidence for a high convergent validity.

Factorial validity

In the remainder of this section, three studies in which a factor analysis has been performed on the SRQ will be discussed. All three used a principal components technique with varimax rotation.

Sen (1987) published the results of an analysis of depressive phenomena in India. The Indian version of the SRQ was administered to 202 attenders at primary health care clinics. Table 7 shows the results of the principal components analysis. This procedure isolated seven factors. Items with factor loadings less than 0.5 were disregarded. The author concluded that only factor II represents uniquely depressive symptoms. Although the labelling of factors can sometimes be a subjective process, it was referred to as a "severe depression" category. Other depressive symptoms were found in other factors. Extrapolating along the lines of reasoning of the author, it was concluded that no proof was found for depression as a distinct diagnostic category. Factor I, which explains 21.8% of the variance, corresponds with anxiety-depression / dysphoria. Factor V and VI together, represent somatization. Factor VII covers neurasthenia. The other factors are less clinically meaningful. The author did not however, publish the eigenvalues of the components (factors). This factor analysis is therefore of concern because of the lack of eigen values and because factors containing only two items such as Factors V, VI and VII, are not normally considered as factors.

Table 7. Factor structure of the SRQ-20 in India (Sen, 1987)

Seven factor structure of the SRQ-20 (figures in parentheses indicate the percentage of variance explained)		
Item No.	**Item**	**Loadings**
	Factor I (22%)	
9	Do you feel unhappy?	.75
6	Do you feel nervous, tense or worried?	.68
14	Are you unable to play a useful part in life?	.51
	Factor II (8%)	
15	Have you lost interest in things?	.77
10	Do you cry more than usual?	.62
17	Has the thought of ending your life been in your mind?	.51
	Factor III (7%)	
4	Are you easily frightened?	.74
1	Do you often have headaches?	.56
8	Do you have trouble thinking clearly?	.51
	Factor IV (6%)	
11	Do you find it difficult to enjoy your daily activities?	.75
8	Do you have trouble thinking clearly?	.51
12	Do you find it difficult to make decisions?	.51
	Factor V (6%)	
7	Is your digestion poor?	.82
19	Do you have uncomfortable feelings in your stomach?	.81
	Factor VI (6%)	
2	Is your appetite poor?	.68
3	Do you sleep badly?	.68
	Factor VII (7%)	
20	Are you easily tired?	.81
18	Do you feel tired all the time?	.74

Iacoponi and Mari (1989) applied a factor analysis (principal components analysis with varimax rotation of SRQ-20) to 1182 completed Portuguese questionnaires from Brazil and extracted four factors with an eigenvalue greater than 1. Together they accounted for 41% of the total variance (see Table 8). Only those questionnaire items loading equal or more than 0.40 are included. Factor I which has been labelled by the investigators as "Decreased energy", consists of both anxiety and depressive items. The somatic items load primarily on Factor II. The investigators stressed that the somatic items were discriminated clearly from the other items. One might argue that the distribution of the items on factor III and IV is not as straightforward as the investigators hoped to achieve. They labelled factor III as "Depressive/ Anxious mood" and factor IV as "Depressive thoughts". However, item 6 (Do you feel nervous, tense or worried?) does not refer exclusively to a mood or a thought. "Feeling worried" is clearly referring to depressive thoughts (factor IV). Though factor III was not loaded highly by this question, it still led to some interpretation problems. Once again, the factor structure supports the notion that "mental illness" is not unidimensional.

Table 8. Factor structure the SRQ-20 in Brazil (Iacoponi and Mari, 1989)

Four factor structure of the SRQ-20		
Item No.	**Item**	**Loadings**
	Factor I - Decreased Energy - Variance, 22.3%; Eigenvalue, 4.46	
20	Are you easily tired?	.649
18	Do you feel tired all the time?	.623
12	Do you find it difficult to make decisions?	.573
13	Is your daily work suffering?	.501
8	Do you have trouble thinking clearly?	.493
11	Do you find it difficult to enjoy your daily activities?	.419
	Factor II - Somatic Symptoms - Variance, 7.3%; Eigenvalue, 1.47	
19	Do you have uncomfortable feelings in your stomach?	.765
7	Is your digestion poor?	.713
2	Is your appetite poor?	.533
1	Do you often have headaches?	.410

Factor III - Depressive Mood - Variance, 5.9%; Eigenvalue, 1.17		
10	Do you cry more than usual?	.714
9	Do you feel unhappy?	.681
6	Do you feel nervous, tense or worried?	.437
Factor IV - Depressive Thoughts - Variance, 5.4%; Eigenvalue, 1.08		
16	Do you feel you are a worthless person?	.678
14	Are you unable to play a useful part in life?	.632
17	Has the thought of ending your life been in your mind?	.567
15	Have you lost interest in things?	.438

In Ethiopia, Tafari et al. (1991) described the four factor structure of the Amharic version of the SRQ-24 without going into statistical details. There were 2000 respondents who filled in the SRQ in this community survey. The main finding was the fact that there was a clear discrimination between the neurotic scale (items 1-20) and the psychotic scale (items 21-24). Within the neurotic scale three meaningful factors were derived:

Factor I - Cognitive items

item 8. Do you have trouble thinking clearly?
item 12. Do you find it difficult to make decisions?
item 13. Is your daily working suffering?

Factor II - Anxiety and depression

item 4. Are you easily frightened?
item 9. Do you feel unhappy?
item 10. Do you cry more than usual?
item 16. Do you feel that you are a worthless person?

Factor III - Somatic symptoms

item 1. Do you often have headaches?
item 2. Is your appetite poor?
item 7. Is your digestion poor?
item 3. Do you sleep badly?

Overall correspondence is not found among the factor structures of the SRQ. So, at this stage, it is too early to propose the use of factors as sub-scales. The likelihood of cultural differences in constellations of symptoms might discourage researchers from pursuing this possibility. An obvious explanation of differences in the factor structures is found in the setting in which the study took place. The first two studies used the SRQ to measure psychiatric morbidity in primary care settings, while the third study used the SRQ in a community survey.

RELIABILITY OF THE SRQ

Although validity and reliability are usually discussed as independent concepts, validity is in fact dependent on reliability. Assessment of reliability amounts to determining that the instrument is measuring something in a reproducible and consistent fashion. Conceptually, reliability reflects how free the screening test is from measurement error. The term "measurement error" refers to fluctuations of individuals' test scores from what their actual "true scores" would be if the test were a perfect instrument and scores were not influenced by irrelevant, chance factors (Tulsky, 1990). So, the question that needs to be answered is: "Does the test produce a score that is relatively uninfluenced by chance events?"

There are various ways to assess reliability. Before the SRQ can be accepted as an interviewer administered screening instrument, it is essential that it is shown to provide consistent results between interviewers (inter-rater reliability). Furthermore, measuring consistency between items on a scale, the so-called "internal consistency", is a second method of assessing reliability. Both methods are going to be dealt with in this section. Another method for demonstrating reliability is the measurement of test-retest reliability, which gives an indication of consistency across time. To date, there is no research addressing the test-retest reliability of the SRQ. Because the SRQ measures the mental health status of people, which is variable over time, traditional test-retest estimates of reliability would have interpretation problems. It is hard to distinguish between true change (the condition of the respondent has improved or got worse), and unreliability (the source of variance is the test itself). Evidence of consistency over different time periods would have been useful however, despite the difficulties related to its measurement.

There are few reported data on the reliability of the SRQ. Only four articles have been found which mention reliability, with researchers mainly focusing on validity indices like sensitivity and specificity. Since reliability is a necessary condition for validity, more data on reliability is required.

Inter-rater reliability

An inter-rater reliability coefficient of the SRQ can logically only be calculated if the SRQ is interviewer-administered. Because of the high illiteracy rate in the developing world, the SRQ was interviewer-administered in a majority of the studies.

Iacoponi and Mari (1989) assessed the inter-rater reliability of the SRQ-20 in primary care settings in Sao Paulo, Brazil; 27 patients were involved in the inter-rater study and were interviewed in simultaneous scoring sessions. The Clinical Interview Schedule was employed as the criterion. The reliability coefficients (inter-rater) were derived from the scores of the four interviewers: one research psychiatrist and three research assistants. This was done in two ways. (i) The SRQ score was taken as a continuous variable with scores ranging from 0 to 14. They calculated the intra-class correlation via ANOVA (analysis of variance). The

correlation found was 0.978 (F = 174.98). (ii) When the final SRQ score was analyzed as a case/non-case dichotomy (i.e. those scoring below versus those scoring above a 7/8 cut off score), it yielded an intra-class correlation of 0.963 (χ^2 = 198.32; *df* = 26). Both coefficients are close to unity, indicating that the agreement between observers was very high. Yet, the authors cautioned the readers that the coefficients might have been lower if more respondents had been included in the inter-rater study, since with a larger sample it is more likely that the full spectrum of minor mental disorders would have been covered.

Another reliability study was carried out by Kortmann and Ten Horn (1988) in Ethiopia. The SRQ was submitted to 40 patients attending a psychiatric clinic, 30 a somatic clinic and 40 subjects not attending a clinic (control group). This study has been more fully described on pp 13-15 of this document. The researchers examined the reliability between four independent raters and the investigator, in determining that a yes-answer to any of the items was valid. The pairwise interrater congruency between the four raters varied between 66% and 78%. The congruency between the opinions of the five raters (including the investigator) simultaneously reached a kappa of 0.48.

The investigators also computed a kappa value on the agreement between four independent raters on the validity of yes-answers given on individual items of the SRQ. This reliability study is unorthodox because it does not provide information on agreement of the capacity of the SRQ in selecting cases and non-cases. For that reason the results are not included in this section. However, the results are illustrated in Table 9B.

Internal consistency

One might argue that the construct of mental illness is not unidimensional, which could result in low internal consistency reliability of the SRQ, since this latter is a measure of the homogeneity of the test. On the other hand, a satisfactory coefficient might be obtained if the underlying dimensions are strongly connected, and if it is expected that there is an underlying general factor.

There are two investigations published in which the internal consistency of the Portuguese version of the SRQ is measured. Iacoponi and Mari (1989) used the Kuder Richardson (KR-20) formula, which is appropriate for scales composed of dichotomous items. When the variance of the total composite (four interviewers interviewing 1182 respondents) was considered together, a very satisfactory coefficient of 0.81 was obtained.

Tafari et al. (1991) computed the item-total correlations in research in Ethiopia, using a translation of the SRQ into Amharic on 2000 respondents. For the individual items, correlations of 0.40 and above were found. Exception to the rule was item 5 (hands shake), 0.33. These correlations demonstrate a highly acceptable degree of internal consistency and at the same time add evidence to the idea that there is an underlying general factor of mental illness as stated above.

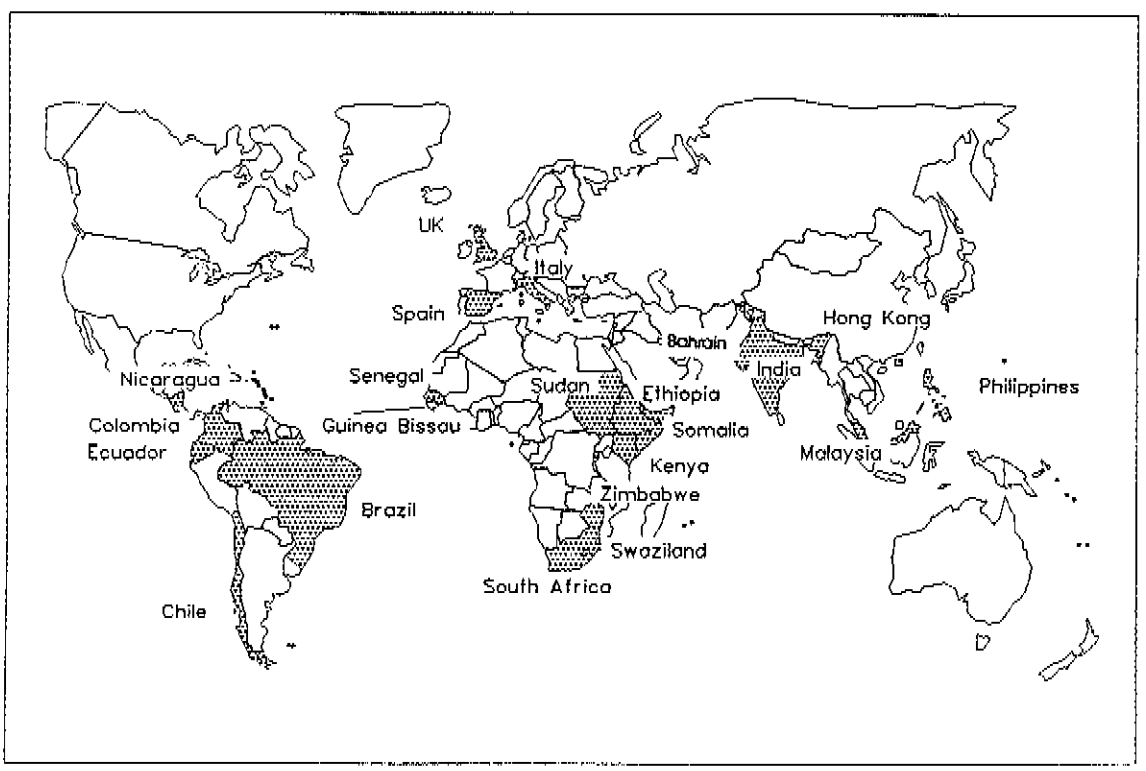
RESEARCH REVIEW

This section analyses the results of the research with the SRQ to date. Only those papers that were published in scientific journals, have been selected for this review. For reasons of clarity, a table format is chosen with the following headings:

1) Author(s) and Year
2) Country and City
3) Purpose of the study
4) Sample and Setting
5) Number of SRQs
6) Cut-off scores
7) Instrument used for the second stage
8) Findings
9) Validity assessment
10) Reliability assessment
11) Comments

Headings 1 through 7 are described in Table 9A and the rest is described in Table 9B. Added together the two tables read as one. The heading "Author(s) and year" is common to both tables. The contents of the table are meant to be self-explanatory. To give an impression of the use of the SRQ throughout the world, the following figure is inserted.

Figure 2. Research with the SRQ throughout the world



The designations employed and the presentation of material on this map do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country or concerning the delimitation of its frontiers or boundaries.

Table 9A. Review of research with the SRQ

1)	Author(s) and Year
2)	Country and City
3)	Purpose of the study
4)	Sample and Setting
5)	Number of SRQs
6)	Cut-off scores
7)	Instrument used for the second stage*
8)	Findings
9)	Validity assessment
10)	Reliability assessment
11)	Comments

* In the majority of cases where an instrument was used in the second stage, this was either the Present State Examination - PSE (Wing, Cooper and Sartorius, 1974) or the Standardised Psychiatric Interview - SPI, also referred to as the Clinical Interview Schedule - CIS (Goldberg et al., 1970). The Revised Clinical Interview Schedule - CIS-R, has also been used (Lewis et al., 1992)

The way, however, that the instruments were used in the second stage has been very variable, and an indication of this will be given in the following table in column 7 on second stage instrument and method. In most cases the structured interview was used as the basis for making a clinical judgement. In some cases the papers report that this judgement was based on ICD or DSM-III or -IIIR criteria, usually following a structured interview, but occasionally without this. Thus where there was a second stage, this was carried out with either **1.** The PSE, **2.** The SPI/CIS or CIS-R, **3.** An unspecified structured interview, **4.** A presumed unstructured psychiatric assessment, since no structured interview is mentioned.

Any of these four latter could be followed by the making of a clinical judgement/diagnosis by a psychiatrist on the basis of the interview, sometimes using ICD or DSM criteria.

There are five studies in which the structured interview was actually scored (rather than being used as the basis for making a clinical judgement) and this score taken as the criterion against which the SRQ was judged (PSE = 1, CIS = 3, CIS-R = 1) (Araya, 1992; Carta, 1993; El-Rufaie and Absood, 1994; Mari and Williams, 1985; Sen, 1987). In the case of the PSE, ID-Catego level 5 and above was used as a case/non-case cut-off. In the case of the CIS and CIS-R, the score was sometimes used to produce a case/non-case dichotomy (e.g. weighted score of 20 or above = case), and sometimes the score was also treated as a continuous variable and a correlation against the SRQ score was calculated.

Author(s) & Year	Country & City	Purpose of the study	Sample & Setting	Number of SRQs	Cut-off score(s)	Instrument and method used for second stage
Harding et al. (1980)	Colombia, Union de Vivienda Popular	- Measurement of the frequency of mental disorders among patients presenting at the primary level of health care in developing countries; - The main part of the research is based on assessing diagnostic skills of health workers.	- 444 primary health care attenders; - adults; - male : female ratio =1 : 6.9.	444	10/11	- 'Potential' cases: shortened version of the Present State Examination (PSE); - 10% of the 'non-cases': shortened version of the PSE. Diagnostic assessment and formulation on basis of PSE ratings (clinical judgement). Inter-centre reliability checks were made on diagnoses by the research psychiatrists.
	India, Raipur Rani		- 361 primary health care attenders; - adults; - male : female ratio =1 : 2.0.	361	5/6	
	Sudan, Shagara Jebel Awlia		- 360 primary health care attenders; - adults; - male : female ratio =1 : 1.8.	360	3/4	
	Philippines, Sampaloc		- 459 primary health care attenders; - adults; - male : female ratio =1 : 3.5.	459	6/7	
Dhadphale et al. (1982)	Kenya, Kisii	Assessment of the frequency and diagnosis of mental disorders in patients.	- 186 patients attending hospital outpatients clinics; - age: 18-55; - largely rural population.	186	not mentioned, but according to the doctorate thesis by Dhadphale (1984), the cut-off point was set at 7/8	Standardized Psychiatric Interview (SPI) and clinical judgement (ICD-8 criteria).
Diop et al. (1982)	Senegal, Niakhar	Ascertainment of the pattern of mental disorders among patients presenting at primary health facilities.	- 933 patients attending primary health facilities; - largely rural population.	933	7/8	none

Author(s) & Year	Country & City	Purpose of the study	Sample & Setting	Number of SRQs	Cut-off score(s)	Instrument and method used for second stage
Chan and Chan (1983)	Hong Kong	- Examination of the structure and the psychometric properties of the 30-item GHQ in a Chinese community sample; - Validation of the GHQ against multiple criteria, like the SRQ-24.	- 255 first-year undergraduates attending the Chinese University of Hong Kong; - age: 17-25; - different socioeconomic backgrounds.	224	moving cut-off point	none
Dhadphale et al. (1983)	Kenya, Nyanza-Kisumu	Assessment of the frequency and diagnosis of mental disorders in patients.	- 200 patients attending hospital out-patient clinics; - age: 18-55; - semi-urban population.	200	not mentioned (see Dhad-phale et al., 1982)	Standardized Psychiatric Interview (SPI) and clinical judgement (ICD-8 criteria)
Krishnaswamy and Kyaw (1985)	Malaysia, Kuala Lumpur	Assessment of the applicability of the SRQ in establishing an appropriate cut-off point sensitive enough to pick up psychiatric cases in Malaysia (pilot study).	- 100 respondents: ▪ 50 psychiatric outpatients with neurotic disorders (mean age = 32.38); ▪ 50 medical staff as controls (mean age = 32.42).	100	not clear (see validity assessment)	not mentioned
Mari and Williams (1985)	Brazil, Sao Paulo	- Comparison of the validity of two psychiatric screening questionnaires (GHQ-12 and SRQ-20) in Brazil; - Investigation of the feasibility and applicability of Relative Operating Characteristic (ROC) analysis in the evaluation of psychiatric screening instruments.	- 875 patients attending primary health facilities: ▪ 291 in the Brasilianda sample (health centre); ▪ 349 in the Servidor sample (health centre); ▪ 235 in the Barra Funda sample (out-patient clinic); - age ≥ 16.	875	7/8 (GHQ-12 : 3/4)	Clinical Interview Schedule (CIS), (n = 260) with use of the total weighted score as the criterion for caseness / non-caseness and also treated as a continuous variable.

Author(s) & Year	Country & City	Purpose of the study	Sample & Setting	Number of SRQs	Cut-off score(s)	Instrument and method used for second stage
De Jong et al. (1986)	Guinea-Bissau	- Development and evaluation of low-cost methods for mental health care; - Determination of treatment priorities within the framework of a general health service in Guinea-Bissau.	- 252 primary health care attenders; - age ≥ 16; - rural population.	251	7/8	All 'potential cases' and 10% of the 'non-potential cases': - PSE (shortened version), administered by the research psychiatrist; - Diagnostic Assessment Form (DAF) completed by the research psychiatrist on the basis of the : a) PSE ratings; b) Social Unit Rating (SUR); c) Disability Assessment Schedule (DAS).
Hall and Williams (1987)	Zimbabwe, Bindura	Ascertainment of the extent of hidden psychiatric morbidity in patients presenting at an out-patient facility.	- 448 patients attending an out-patient department at a hospital; - age ≥ 16; - diverse patient population; - questions were posed orally.	448	7/8	An in-depth psychiatric interview based on the Present State Examination (not further explained).
Kortmann (1987)	Ethiopia, Addis Ababa	Illustration of a number of communication problems in the area of transcultural psychiatry, by making use of research findings: "What meanings do Ethiopians attach to each of the SRQ questions?"	- 110 respondents: ▪ 40 from a Psychiatric Outpatient Clinic, the "psychiatric group"; ▪ 30 from various somatic outpatient clinics, the "somatic group"; ▪ 40 from the population of residents of Addis Ababa, the "healthy group"; - questions were posed orally; - no information about age.	110	not mentioned	none

Author(s) & Year	Country & City	Purpose of the study	Sample & Setting	Number of SRQs	Cut-off score(s)	Instrument and method used for second stage
Lima et al. (1987)	Colombia, Armero	- Assessment of the mental health consequences of a natural disaster (volcanic eruption) in a developing country; - Identification of specific personal or environmental factors associated with an increased risk for developing emotional disorders.	- 200 respondents (victims) from two shelters and two camps; - age ≥ 18; - male : female ratio=1 : 0.9; - questions were posed orally.	200	7/8	none Unspecified semi-structured interview and symptom checklist to obtain diagnoses according to DSM-III criteria.
Sen (1987) See also Sen, Wilkinson and mari (1987) and Sen and Williams (1987)	India, Calcutta	Analysis of the nature of depressive phenomena in primary health care utilising multivariate statistical techniques.	- 202 respondents: - 24% males and 76% females; - non- and semi-literate people belonging to the lowest socio-economic classes; - age ≥ 15; - urban population.	202	7/8 and 11/12 (see Comments)	A random sample of 114 subjects were interviewed with the CIS and the Hamilton Rating Scale for Depression. Cut off scores (unspecified) for CIS and for Hamilton, (20 or more) as caseness criteria.
Kortmann and Ten Horn (1988) (see Kortmann, 1987)	Ethiopia, Addis Ababa	Measurement of the content validity and the criterion validity of the SRQ in Ethiopia.	(see Kortmann, 1987)	110	different cut-off points	- The psychiatric group: an extensive semi-structured psychiatric interview, using a self-developed checklist, resulting in a DSM-III diagnosis; - The two other groups: a shorter semi-structured interview to determine whether they were assessed as psychiatrically ill (without detailed classification into diagnostic groups).

Author(s) & Year	Country & City	Purpose of the study	Sample & Setting	Number of SRQs	Cut-off score(s)	Instrument and method used for second stage
Deshpande et al. (1989)	India, New Delhi	Estimation of the prevalence of psychiatric symptoms in a general medical ward.	- 350 patients.	326	10/11	- Patients scoring above the cut-off score: PSE (n = 68); - A random sample of the patients scoring below the cut-off score: PSE (n = 90). Caseness determined by clinical judgement.
Iacoponi and Mari (1989)	Brazil, São Paulo	Measurement of the following psychometric properties of the SRQ-20: a) the coefficients of internal consistency; b) the inter-rater reliability coefficients derived from the scores of four interviewers when rating the same patients; c) the factor structure of the SRQ.	- 1288 patients from 40 primary health clinics in São Paulo; - age ≥ 16; - questions were posed orally.	1182	7/8	none
Rahim and Cederblad (1989)	Sudan, Khartoum	Estimation of the psychiatric morbidity in the young adult population of a suburban area of Khartoum, Sudan.	- 204 respondents: ▪ 104 subjects represent the indigenous population; ▪ 100 subjects represent the newcomers; - age: 22-35; - suburban population; - questions were posed orally.	174	not mentioned	Psychiatric diagnoses were made according to DSM-III categories (conditions were judged to be mild, moderate or severe).
Aldana et al. (1990)	Spain, Valencia	- Validation of the SRQ-20 as an assessment instrument for detection of psychiatric problems in the general population: - Secondary prevention: testing the feasibility of using the SRQ in the periodic health evaluation of a sample of the general population.	- 2855 employers working for the local government of Valencia ("ayuntamiento de Valencia"); - age: 18-65; - male: female ratio= 1:0.36.	2849	3/4	Psychiatric interview (not standardized) (n=335) with caseness by clinical judgement.

Author(s) & Year	Country & City	Purpose of the study	Sample & Setting	Number of SRQs	Cut-off score(s)	Instrument and method used for second stage
Kortmann (1990) (see Kortmann, 1987)	Ethiopia, Addis Ababa	Answering the following questions: ▪ What is the content validity of the yes-answers to each of the questions of the SRQ in Ethiopia? ▪ What is the criterion validity of the SRQ as a whole in predicting psychiatric 'caseness' in Ethiopia?	(see Kortmann, 1987) - age > 18.	110	- clinic attenders 8/9; - non-attenders 4/5.	- Psychiatric group: extensive self-designed, semi-structured psychiatric interview, based upon DSM-III; - Other two groups: shorter interview (not further specified) to determine whether they were to be classified as psychiatric cases.
Penayo et al. (1990)	Nicaragua, Ciudad Sandino and Subtiava	- Validation of the SRQ-20 as a screening instrument to identify probable psychiatric cases in the general population in Nicaragua; - Assessment of the prevalence and identification of mental disorders in primary health care.	- 1357 respondents: ▪ 576 respondents from the general population in Subtiava; ▪ 781 patients from primary health care centers in Ciudad Sandino and Subtiava; - age (only given for the subsample, n = 171): 16-71.	1357	9/10	- General population: PSE (n = 171); - Primary health care patients: no second stage (n = 781). Caseness in general population sample probably determined by clinical judgement based on PSE.

Author(s) & Year	Country & City	Purpose of the study	Sample & Setting	Number of SRQs	Cut-off score(s)	Instrument and method used for second stage
Salleh (1990)	Malaysia	- Assessment of the applicability of the SRQ as a screening instrument in a high risk population; - To detect the frequency of emotional disturbance in the first degree and non-first degree relatives of schizophrenics who were staying together with them and/or were actively involved in their care.	**First stage** ▪ psychiatric group: 178 new cases attending the psychiatric clinic of the University Hospital (USM) (139 neurotic and 39 schizophrenic); ▪ normal group: 178 respondents (medical staff, fourth year medical students and relatives of the patients); - age: 15-65; **Second stage** ▪ study group: - 160 first degree relatives; - 104 non-first degree relatives; - age: 18-65.	620	5/6 (based on stage 1)	The diagnoses were based on ICD-9 and PSE was used as an adjunct to clinical judgment: - psychiatric group: all patients; - normal group: not mentioned; - study group: 101 suspected cases (38%) were further interviewed.
Upadhyaya et al. (1990)	United Kingdom, Manchester	- Examination of the validity of the SRQ as a screening instrument in the immigrant population from a developing country; - Testing of the hypothesis that the Asian immigrant mothers with a high total score on the SRQ would score predominantly on the somatic items, whereas the local Caucasian mothers would do mostly on the psychological items.	- 149 mothers attending a well-baby clinic in a multi-ethnic area; - age: 16-38; - illiterate respondents had the items read out to them in their native language; - urban population.	149	4/5	38 women with a score above 5 and a random sample (n = 8) of low scorers: psychiatric interview by a psychiatrist with a psychiatric assessment schedule (total: 46). Caseness determined by clinical judgement.
Bester et al. (1991)	South Africa, Orange Free State	Assessment of psychiatric disorders amongst the black elderly population in the Orange Free State.	- 400 respondents; - age ≥ 65; - questions were posed orally.	375	8/9	none

Author(s) & Year	Country & City	Purpose of the study	Sample & Setting	Number of SRQs	Cut-off score(s)	Instrument and method used for second stage
Freeman et al. (1991)	South Africa, South Eastern Transvaal	- Establishment of the extent to which patients attending clinics and outpatient departments suffer from mental health problems; - Assessment of the diagnostic skills (mental disorders) of the health workers in clinics and outpatient departments.	- 363 patients from six clinics in Kangwane; - age ≥ 16; - male : female ratio =1 : 1.9; - questions were posed orally; - patients came from rural, peri-rural and more urban areas.	363	9/10 (SRQ-20) + 1 (SRQ-5: the psychotic items)	- Patients who screened positively on the SRQ-25: PSE; - Low scorers: no second stage. Caseness determined by clinical judgement following examination of completed PSEs.
Kigamwa (1991)	Kenya	Providing data on psychiatric morbidity and referral for psychiatric evaluation of medical in-patients at Kenyatta National Hospital.	- 200 medical in-patients; - no information about age; - questions were posed orally.	200	8/9	- High scorers (n = 44) on the SRQ: modified version of the SPI; - Low scorers: no second stage. An overall severity rating was calculated (range 0-3) on a modified SPI. Author reports that of these, 32 obtained scores 1-2 (mild to moderate) and 12, scores 2-3 (moderate to severe). SPI score not used to determine caseness, merely to confirm it in SRQ high scores.
Reichenheim and Harpham (1991)	Brazil, Rio de Janeiro	Assessment of the prevalence of mental disorders of mothers rearing small children in a low-income community in Rio de Janeiro.	- 480 randomly selected mothers with under-fives; - the research took place in a squatter settlement; - the SRQ-20 was applied to the mother, the adoptive mother or the co-rearer (if the mother was unable to respond); - questions were posed orally; - response rate 95.8%.	460	7/8	none

Author(s) & Year	Country & City	Purpose of the study	Sample & Setting	Number of SRQs	Cut-off score(s)	Instrument and method used for second stage
Tafari et al. (1991)	Ethiopia, Kembata/ Hadiya district	- Assessment of the prevalence of mental illness; - Identification of risk factors (stress and demographic variables) associated with mental illness.	- 2000 respondents; - age: 15-55; - rural adult population; - questions were posed orally.	2000	10/11	none
Araya et al. (1992)	Chile, Santiago	- Comparison of the sensitivity and specificity of the GHQ-12 and the SRQ-20; - Examination of the influence of sociodemographic factors on misclassification; - Examination of the prevalence of individual symptoms obtained with the questionnaires.	- 170 primary care attenders; - poor urban population; - illiterate patients had their questions read out.	163	9/10	All subjects: Revised Clinical Interview Schedule (CIS-R). Caseness defined as those scoring 12 or more as total for CIS-R.
Guinness (1992)	Swaziland	Estimation of the prevalence and the expression of psychiatric morbidity in school populations using the SRQ-24 and a open question (" Do you have any problems with your health at school").	- 2040 students from the top three classes of a cross selection of secondary schools, remote rural schools, periurban and elite (51 classes were surveyed); - age: 14-24 years;	- 705 students (1985); - 276 students (1990).	11/12	none
Lima et al. (1992)	Ecuador, Imabura area	- Assessment of the psychosocial consequences of disasters in developing countries; - Exploration of the role of the primary care sector in providing mental health services to disaster victims in developing countries.	- 150 adult patients attending 10 health clinics in an area recently devastated by an earthquake; - questions were posed orally.	150	7/8	37 of the 60 'potential' cases: semi-structured psychiatric interview, to which a standardized symptom checklist was appended to obtain diagnoses in accordance with DSM-III.

Author(s) & Year	Country & City	Purpose of the study	Sample & Setting	Number of SRQs	Cut-off score(s)	Instrument and method used for second stage
Carta et al. (1993)	Italy, southern Sardinia	- Assessment of the degree of accuracy of the Italian version of the SRQ-24 applied by GPs to their patients; - Analysis of convenient cut-off points according to ROC analysis.	- 106 patients who consulted a GP; - mean age is 48 years; - the SRQ was self-administered; - overall level of education is low; - the study took place in an industrialized country.	102	7/8	- Italian version of the PSE. Cases were classified as those classed at level 5 or above according to the PSE-ID-Catego procedure.
Thom, Zwi and Reinach (1993)	South Africa, Johannes-burg, Soweto	- Assessment of the prevalence of psychiatric disorders in a primary care clinic in Soweto; - Determination of how effective the clinic staff were in detecting such disorders.	- 299 patients presenting at the Zola Community Health Care Centre; - age: 16-60; - male : female ratio =1 : 2.8; - the Zola clinic is situated in a socio-economically deprived area; - questions were posed orally.	299	6/7	- Patients scoring above the cut-off : screening version of the PSE with the subsequent use of DSM-III criteria to determine caseness - 5% (11 patients) of the low scorers: screening version of the PSE; - patients scoring positive on any psychotic/fits item (SRQ-25): screening version of the PSE.

Author(s) & Year	Country & City	Purpose of the study	Sample & Setting	Number of SRQs	Cut-off score(s)	Instrument and method used for second stage
El-Rufaie and Absood (1994)	United Arab Emirates, Al Ain	- Examination of the validity of the Arabic version of the SRQ-20 as a screening test for mental disorders in a primary care setting; - Assessment of the feasibility of the two stage method used for a wider research project in the same setting.	- 224 randomly selected patient - response rate (96.9%) - age: 16-80 - the population of Al Ain is multi-ethnic in origin	217	5/6	All patients (217) were interviewed by a psychiatrist using the CIS. Cases defined by fulfilling at least 2 out of 3 criteria: (i) CIS score ≥ 20 (ii) Score ≥ 2 on 5 point scale of functional impairment (iii) Psychiatric diagnosis (ICD-9) The latter two however appear to depend on psychiatrist's clinical judgement.

Table 9B. Review of research with the SRQ

1)	Author(s) and Year
2)	Country and City
3)	Purpose of the study
4)	Sample and Setting
5)	Number of SRQs
6)	Cut-off scores
7)	Instrument used for the second stage
8)	Findings
9)	Validity assessment
10)	Reliability assessment
11)	Comments

Author(s) & Year	Findings	Validity assessment	Reliability assessment	Comments
Harding et al. (1980)		sensitivity 73% - 83% specificity 72% - 85% misclassification rate 18% - 24%		- The number of low scorers interviewed in the second stage was very small, so the findings must be treated with caution.
Dhadphale et al. (1982)	- PM (psychiatric morbidity)-rate = 25.8 %	Because sensitivity and specificity figures in the articles of Dhadphale (1982 and 1983) are inconsistent, corrected specificity and sensitivity figures have been derived from the raw data reported in Dhadphale (1984): n= 130 (65 patients, 65 matched controls) patients: sensitivity = 93.4% specificity = 75% controls: sensitivity = 57.1% specificity = 96.6% total: sensitivity = 89.7% specificity = 95.2%	not mentioned	- Psychiatrists in the second stage were aware of the SRQ data, thereby interfering with their diagnosis, which should have been merely based on using the SPI; - The authors concluded that the SRQ was easy to use, reliable, sensitive as well as specific, when positive cases were reviewed by a psychiatrist using the SPI.
Diop et al. (1982)	- It was found that 16% reported more than 7 symptoms commonly associated with psychiatric illness on the SRQ-24; - 'psychotic' and 'psychological' symptoms were much more likely to be associated by health workers with a positive diagnosis of mental disorder than psycho-physiological or somatic symptoms.	not mentioned	not mentioned	- The investigation was carried out in two parts dealing with children and adults respectively; - The relationship between frequency of positive responses to SRQ-items and the diagnosis of psychiatric problems by health workers has been calculated.

Author(s) & Year	Findings	Validity assessment	Reliability assessment	Comments
Chan and Chan (1983)	- The results, based on 224 subjects who yielded complete data on the GHQ and SRQ, indicated that the GHQ correlated 0.49 (P <0.001) with the total SRQ, 0.54 (P <0.001) with the non-psychotic items of the SRQ, and non-significantly (0.06) with the psychotic items of the SRQ; - Indication of convergent validity: two-way classification of potential case-non-case by the GHQ and the SRQ-20 (N = 224): GHQ score SRQ score: ≥5 ≤4 | ≥6 ≤5 | ≥7 ≤6 ≥5: 36 17 | 32 21 | 30 23 ≤4: 37 134 | 20 151 | 17 154 Conc.*: 76% | 82% | 82% ≥6: 33 10 | 30 13 | 29 14 ≤5: 40 141 | 22 159 | 18 163 Conc.*: 78% | 84% | 86% ≥7: 26 7 | 23 10 | 22 11 ≤6: 47 144 | 29 162 | 25 166 Conc.*: 76% | 82% | 84% * Concordance is the proportion of agreement.	see Findings	none	
Dhadphale et al. (1983)	- PM (psychiatric morbidity)-rate = 30.5 %	see Dhadphale et al., 1982	not mentioned	

Author(s) & Year	Findings	Validity assessment	Reliability assessment	Comments
Krishnaswamy and Kyaw (1985)	- The mean SRQ scores for the two groups is as follows: • 11.98 for the neurotic group (s.d. = 4.9); • 1.7 for the normal controls (s.d. = 2.2); The difference is significant (t= 68.5, Df = 98, P< 0.005); - Neurotic patients responded positively and significantly more than the normal population on all of the items. Out of these items, item nos. 2, 3, 4, 6, 8, 9, 11, 12, 18, 19 and 20 were scored positively by more than 50% of the neurotic patients (which constitutes the proposed shortened version of the SRQ).	100% of the neurotic patients scored 3 or more than 3 and 90% of them scored 6 or more than 6. Whereas in the control group 74% scored less than 3 and only 6% scored more than 6. In this study a cut-off score of 2/3 is overinclusive and would label 26% of the normals as potential psychiatric cases. If 6 is taken as a cut-off point, only 6% of the normals would be labelled as potential psychiatric cases. But 10% of the psychiatric cases would be missed in any further epidemiological survey.	not mentioned	- The questionnaires were distributed in both English and Malay translation and in this study all those not conversant with two languages were excluded; - There were no backtranslations made; - Proposed shortened version of the SRQ (10 most sensitive items, see Findings): 1. Is your appetite poor? 2. Do you sleep badly? 3. Are you easily frightened? 4. Do you feel nervous, tense or worried? 5. Do you have trouble thinking clearly? 6. Do you feel unhappy? 7. Do you find it difficult to enjoy your daily activities? 8. Do you find it difficult to make decisions? 9. Do you feel tired all the time? 10. Do you have uncomfortable feelings in your stomach? - Item number 20 was omitted because it was similar to item number 18.

Author(s) & Year	Findings	Validity assessment		Reliability assessment	Comments
Mari and Williams (1985)	The validity coefficients for the GHQ-12 and the SRQ-20 were approximately the same, indicating that both instruments were equally effective in identifying psycho-emotional symptoms in primary care clinics in São Paulo; - The 'optimum' cut-off point (the best compromise between high sensitivity and low false positive rate) derived from the ROC curve, was found to be 7/8 for the SRQ-20; - Both questionnaires showed a similar ability to identify minor psychiatric morbidity when assessed against the criterion of the Clinical Interview Schedule; - The validity coefficients were not affected by socio-demographic variables (age, sex, marital status, race, educational level, and family income). - The estimated true prevalence in the Brasilandia sample was 56%, in the Servidor sample 50% and in the Barra Funda sample 47% (Mari et al., 1987).	<u>SRQ-20</u> total sample: sensitivity = 83% specificity = 80% Brasilandia sample: sensitivity = 84% specificity = 66% Servidor sample: sensitivity = 81% specificity = 84% Barra Funda sample: sensitivity = 85% specificity = 90% Negative predictive value for the total sample = 82% Positive predictive value for the total sample = 81% Overall misclassification rate for the total sample = 19%	<u>GHQ-12</u> total sample: sensitivity = 85% specificity = 79% Brasilandia sample: sensitivity = 82% specificity = 69% Servidor sample: sensitivity = 84% specificity = 81% Barra Funda sample: sensitivity = 89% specificity = 84% Negative predictive value for the total sample = 82% Positive predictive value for the total sample = 82% Overall misclassification rate for the total sample = 18%	not mentioned	- The ROC curve is a representation of the ability of the screening instrument to discriminate between 'cases' and 'non-cases' across the total spectrum of morbidity; - This thorough research is the first serious attempt to assess the psychometric qualities of the SRQ-20 (in relation to the GHQ-12).

Author(s) & Year	Findings	Validity assessment	Reliability assessment	Comments
Mari and Williams (1985) continued		Caseness was determined for the above by using the CIS author's recommended cut-off of the total weighted score, - Product-moment correlation (between GHQ-12 and SRQ-20) : • Total sample: .72; • Brasilandia sample: .68; • Servidor sample: .70; • Barra Funda sample: .78. - Correlations between the questionnaire scores and the total weighted scores from the CIS, considering each as continuous variables: -Total interviewed subsample (n=260): .70 (GHQ-12) .74 (SRQ-20) -Brasilandia (n=80): .58 (GHQ-12) .70 (SRQ-20) - Servidor (n=100): .72 (GHQ-12) .76 (SRQ-20) - Barra Funda (n=80): .76 (GHQ-12) .78 (SRQ-20)		

Author(s) & Year	Findings	Validity assessment	Reliability assessment	Comments
De Jong et al. (1986)	35 potential cases: - 18 scored above the cut-off scores on the non-psychotic items; - 4 scored more than one psychotic item; - 13 were added (8 with a positive epilepsy score only, and 5 who were clinically suspected of being mentally ill); -The minimum estimate of psychiatric morbidity is 12%: 8.4% identified at follow-up and 3.6% on the basis of half the frequency of psychiatric disorder found in the 10% sample.	Sensitivity and specificity figures are based on the comparison of the diagnostic rating of the health worker and that of the research worker, which does not reveal any information on the validity of the SRQ.	not mentioned	- The researchers added two questions to the SRQ-24, one to detect epileptic or other fits and one to detect alcoholism; - Comparison between data from this research with results from other research is problematic because of the unconventional 'constitution' of potential cases.
Hall and Williams (1987)	- 10.5% scored above the cut-off score.	not mentioned	not mentioned	
Kortmann (1987)	-In all, 37% of the yes-answers to the non-psychotic items (1-20) and 68% of the yes-answers to the psychotic items (21-24) were shown not to be a valid indication of the psychopathological phenomena which the question was seeking to uncover.	see Findings	- 4 psychiatrically trained co-judges carried out the same assessment as the investigator of the elucidations of the yes-answers of the psychiatric group. They judged 35% of the yes-answers of the non-psychotic items to be invalid (as compared to 32% according to the investigators' rating of that group); - The simultaneous agreement among the 5 independent judges (items 1-24) lay between 66% and 78% (pairwise interrater congruency).	- An elucidation of the yes-answers on SRQ-items was required in order to judge whether or not the yes-answer was a valid contribution for a Western psychiatric diagnosis; - In this research the investigators distinguished three types of problems: those which could be traced back to language, to motivation and to conceptual difficulties; - According to the author it is questionable whether Western psychiatry will be capable of contributing substantially to finding appropriate therapeutic answers to the psychoneurotic suffering of people from another culture.

Author(s) & Year	Findings	Validity assessment	Reliability assessment	Comments
Lima et al. (1987)	- The overall prevalence of emotional and psychiatric problems was 56%; - Various aspects of the disaster experience (for instance, having been injured) that could be thought as being closely related to the victims' mental health were not significantly associated with increased emotional distress; - The most frequently reported symptoms were feeling nervous, tense or worried, being easily frightened and having headaches; - The strongest predictors of SRQ positivity were the symptoms of feeling unable to play a useful part in life, feeling tired all the time, and having problems in thinking clearly.	not mentioned	not mentioned	
Sen (1987) See also Sen, Wilkinson and Mari (1987 and Sen and Williams (1987)	- Out of 48 males, 29% were judged depressed by the depression items of the CIS and 10% by the Hamilton. Likewise, out of 154 females, 47% were judged depressed by the CIS depression scale and 27% by the Hamilton; - The only three items of the SRQ which are statistically significant in contributing to the explained variance of the Hamilton total score do not include any depressive or even anxiety items. All three items (13, 8 and 14) were relatively non-specific; - The results of a principal components analysis with varimax rotation of the SRQ-20 are reported in Table 7.	cut-off point = 11/12 sensitivity = 79% specificity = 75% PPV = 76% NPV = 78% OMR = 23% (Sen, Wilkinson and Mari, 1987) It is not clear whether caseness was or was not determined from the total weighted score on the ICS. A Hamilton Rating Scale for Depression of 20 or over was used as a criterion for depressive caseness.	not mentioned	- Initially, the cut-off point was taken as the median of the cut-off scores determined by Harding et al. (1980) and by Mari and Williams (1986). However, the optimum cut-off point for this study , based on a trade-off of between sensitivity and specificity, was 11/12; -The Screening for Depression Questionnaire (SDQ-9) was constructed in Bengali for the purpose of this study. It consists of nine questions designed to elicit symptoms commonly elicited in almost any variety of depression by means of simple yes/no-answers. Seven of the nine questions were taken from the SRQ.

Author(s) & Year	Findings	Validity assessment	Reliability assessment	Comments
Kortmann and Ten Horn (1988)	- In the group of subjects who came to the psychiatric clinic, the investigators found 31 psychiatric cases (75%), in the somatic group 8 (25%) and in the control group 5 cases (12%); - Respondents who asked for a medical certificate belonged significantly more often to the group of high scores on the SRQ than did those who did not ask for it.	- problems concerning the content validity are discussed in Table 2 - criterion validity (SRQ-20) Psychiatric group: cut-off point = 8/9 sensitivity = 77% specificity = 44% misclassification = 30% Somatic group: cut-off point = 8/9 sensitivity = 63% specificity = 68% misclassification = 33% Control group: cut-off point = 8/9 sensitivity = 0% specificity = 100% misclassification = 12%	- Related to the validity of the yes-answers in the psychiatric group, the congruency in opinion of the five raters simultaneously reached a kappa value of 0.48; - The simultaneous agreement of four independent raters in assigning visitors to the psychiatric clinic to the diagnostic category of the non-psychotic disorders reached a kappa value of 0.58; - The simultaneous agreement upon assignment to the category of psychiatric cases in the somatic group reached a kappa value of 0.44 and in the control group, 0.60.	- This study demonstrates that merely achieving semantic equivalence of the items (through the procedure of translation and independent translating back to the original language of the questions) does not ensure that questions are understood in the same way by Ethiopian respondents as by Western-trained psychiatrists; - To measure the illness behaviour of the Ethiopians, and especially to estimate whether they were attempting to benefit from being ill, their request for a medical certificate was used as a parameter.

Author(s) & Year	Findings	Validity assessment			Reliability assessment	Comments
Deshpande et al. (1989)	- The prevalence of definite psychiatric disorders was 34.3% (weighted figures, using prevalence data from patients administered the PSE). An additional 14.8% had significant psychiatric symptoms that could not be diagnosed using ICD-9. Thus 49% of the sample had distressing psychiatric symptoms; - A cut-off point of 11 selected on the basis of a pilot study, was found to be unsatisfactory. A lower cut-off score (9 or lower) may be better; - Female patients had a significantly higher mean score than the male patients on the SRQ.	SRQ score (n=326)	sens. (%)	spec. (%)	not mentioned	- Twenty four patients (21 with delirium, 3 with mental retardation), who could not answer the SRQ because they were too ill, were assigned a psychiatric diagnosis on the Mental State Examination (MSE) alone; - According to the investigators, some questions appeared to be unsuited for general medical patients, for example those on sleep, appetite or work, as such questions are usually answered in the affirmative and increase the SRQ score.
		0	100	0		
		1	100	7.6		
		2	100	13.9		
		3	100	24.6		
		4	96.1	33.8		
		5	91.0	39.9		
		6	91.8	47.8		
		7	77.4	51.1		
		8	66.6	54.4		
		9	*62.9*	*62.0*		
		10	51.5	66.9		
		11	*46.6*	*74.1*		
		12	39.0	78.6		
		13	28.2	84.5		
		14	24.2	88.2		
		15	24.8	93.5		
		16	21.7	97.3		
		17	12.8	98.1		
		18	7.0	99.5		
		19	1.2	100		
		20	1.2	100		

Author(s) & Year	Findings	Validity assessment	Reliability assessment	Comments
Iacoponi and Mari (1989)	The results of the factor analysis are reported in Table 8.	not mentioned	Internal consistency (KR20/ n= 1182) = 0.81 Interrater reliability a) total SRQ-20 score as case/non-case dichotomy → intra-class correlation 0.963; b) total SRQ-20 score as a continuous variable (score from 0 to 14) → intra-class correlation (via Anova) 0.978. (four interviewers)	- In some of the questionnaire items (items 11, 13, 14 and 16), understanding was noticeably limited, particularly for the less educated items, however, it did not cause a decrease in the mean inter-item correlation; - Only 27 patients were involved in the inter-rater study; - The SRQ-20 achieved very satisfactory psychometric standards.
Rahim and Cederblad (1989)	- 40.3% of the studied adults had at least one psychiatric symptom; - In 23.7% the complaints were mild, sub-clinical, but in the remaining 16.6% they were moderate to severe symptoms that warranted a clinical diagnosis.	The SRQ score had a strong, positive correlation with the psychiatric assessment (r = .69, P<0.001). Correlations of SRQ score with other measures used are also quoted e.g. - life events score - loneliness index - marital discord or divorce	none	Besides the SRQ, the following psychometric batteries were used: • the Eysenck Personality Inventory (EPI); • Raven's Progressive Matrices Test; • Goodenough's Draw-a-man Test; • Well-Being Scale; • Life Events Scale; • Daily Activities Scale; • Loneliness Scale; • Interview Schedule for Social Integration (ISSI); • The Rahim Anxiety-Depression Scale.
Aldana et al. (1990)	The SRQ was found to be an useful tool in the detection of psychiatric cases in the general community.	cut-off point= 3/4 sensitivity = 70% specificity = 70% PPV = 35% NPV = 96%	none	Diagnosed assessment was based on an unstructured psychiatric interview. Since only 335 respondents had had a second stage assessment, the risk of false negatives is considerable.

Author(s) & Year	Findings	Validity assessment	Reliability assessment	Comments
Kortmann (1990)	The SRQ seems to be a useful instrument for epidemiological case finding in the community. Its feasibility for use in an outpatient clinic is somewhat less.	- problems concerning the content validity are discussed in Table 2 - criterion validity (SRQ-24) Psychiatric group: cut-off point = 8/9 sensitivity = 90% specificity = 22% misclassification = 25% Somatic group: cut-off point = 8/9 sensitivity = 75% specificity = 55% misclassification = 40% Non-attenders: cut-off point = 4/5 sensitivity = 100% specificity = 71% misclassification = 17%	see Kortmann and Ten Horn (1988)	The criterion validity of the SRQ may increase by improving the content validity of the questions.

Author(s) & Year	Findings	Validity assessment	Reliability assessment	Comments
Penayo et al. (1990)	- General population study: in the interviewed subsample (n = 171), 58% were identified as cases according to clinical interviews (true positives); - Applying a cut-off score of 9/10 to the general population (n = 576), 133 people (23%) were identified as probable cases according to SRQ-20; - Among primary health care patients, 364 of 781 patients (46%) were identified as probable cases according to the SRQ-20; - The six most frequent symptoms among probable cases (prevalent items) were item 6 (feeling nervous), item 9 (sadness), item 20 (easily tired), item 1 (headache), item 4 (being frightened) and item 18 (always feeling tired).	- Not assessed for the primary health care population; - Sensitivity, specificity and predictive power (PP) of SRQ-20 in identifying clinical cases in the general population (%): Cut-off 7/8: sensitivity = 81 specificity = 58 Positive PV = 73 Negative PV = 69 Cut-off 9/10: sensitivity = 75 specificity = 69 Positive PV = 77 Negative PV = 67 Cut-off 11/12: sensitivity = 55 specificity = 79 Positive PV = 78 Negative PV = 55	not mentioned	- The validation of the SRQ-20 versus clinical interviews should ideally have been made in the primary health care population instead of the general population; - The investigators stress the fact that the SRQ-20 is a useful and valid instrument for identifying probable cases in a general population.

Author(s) & Year	Findings	Validity assessment	Reliability assessment	Comments
Salleh (1990)	- In the study group 62 subjects (23%) were confirmed to have emotional disturbance and fitted into one of the ICD-9 diagnoses (prevalence of psychiatric morbidity); - 1% of the 264 relatives of schizophrenics had latent schizophrenia; - Question numbers 7, 10, 13, 14, 15, 16 and 17 were scored positively less than 50% by the patients (psychiatric group); - The six most prevalent items among the psychiatric group were item 18 (always feeling tired), item 1 (headache), item 6 (feeling nervous), item 8 (trouble thinking clearly), item 3 (sleeping badly) and item 20 (easily tired); - The six most prevalent items among the normal group were item 18 (always feeling tired), item 20 (easily tired), item 12 (difficult making decision), item 19 (uncomfortable stomach) and item 6 (feeling unhappy).	<u>SRQ-20</u> (It is not clear if the validity indices are based on the results of the normal group and the psychiatric group (n = 356, stage 1) or just on the results of the psychiatric group (n = 178)): sensitivity = 84.8% specificity = 83.7% positive predictive value = 83.9% negative predictive value = 84.7% misclassification rate = 15.7%	not mentioned	The results of this study suggest that the SRQ-20 in conjunction with ICD-9 is suitable for use in two-stage screening of a high risk group of population.

Author(s) & Year	Findings	Validity assessment	Reliability assessment	Comments

Upadhaya et al. (1990)	- The threshold score on the SRQ that most closely responded with being a psychiatric case was found to be 7/8; - The 46 psychiatric interviews yielded 21 women who had definitely had psychiatric illness according to DSM-III criteria; - There was no difference between the white and the Asian mothers in the psychological score (total of 12) and the somatic score (total of 8). That is, there is no predominance of somatic versus psychological symptoms. - The results of the discriminant function analysis of the SRQ items (cumulative ability of a positive score to discriminate between cases and non-cases (%)): White item 15: Lost interest 48 item 6: Nervous 70 item 20/18*: Tired 75 item 7: Digestion 80 item 10: Crying 84 item 3: Sleep 89 item 1: Headache 92 item 8: Thinking 94 item 11: Enjoy act 95 Asians item 9: Unhappy 41 item 20/18*: Tired 61 item 3: Sleep 72 item 15: Lost interest 78 item 1: Headache 84 item 4: Frightened 87 item 13: Daily work 90 item 14: Life 93 item 5: Shake 94 (*= It is not clear from the article if the "tired"-item is referring to item 18	SRQ scores --------- < 5 5 6 7 8 9 10 11 ≥12 Total	Cases ------- 0 0 0 1 4 3 3 2 8 21	Non-cases ------- 8 4 6 4 0 1 1 1 0 25	not mentioned	

Author(s) & Year	Findings	Validity assessment	Reliability assessment	Comments
Bester et al. (1991)	- the estimation of the psychiatric non-psychotic morbidity based on scores on the SRQ-20, was 39% (males: 27%, females: 48%). However, no second stage instrument has been used.	not mentioned	not mentioned	A shortcoming of the SRQ when working with the elderly is its lack of ability to detect dementia.
Freeman et al. (1991)	- 14.3% (52 patients) scored positively on the SRQ-25; - Of these 52 patients, 7 people were not followed up; - After using the PSE, 30 patients out of 45 were given a positive psychiatric diagnosis: 8.3%; - The 8.3% is a conservative estimate since the sample consisted of patients with only minor ailments; - 9 patients were suffering from depression, 6 from anxiety, 7 from a mixed depression/anxiety syndrome, 2 from psychosis and 6 from a variety of other disorders; - Less than one third of the mental health problems identified by the researchers were identified by the clinic staff.	not mentioned	not mentioned	The low prevalence is probably a result of the high cut-off score. No argument is given for the use of this high cut-off score. Furthermore, no sensitivity and specificity coefficients have been calculated.
Kigamwa (1991)	PM-rate = 22% (44 cases): • depressive illness: 26 patients; • anxiety states: 8 patients; • psychosis: 6 patients; • alcohol dependence: 4 patients.	not mentioned	not mentioned	

Author(s) & Year	Findings	Validity assessment	Reliability assessment	Comments
Reichenheim and Harpham (1991)	- The proportion of probable cases of mental disorder was 36%; - The items that most frequently contributed to high scores were, in decreasing order, tenseness, sadness, frequent headaches, frequent frights, and difficulties in thinking and decision making; - The lower the household income, the worse the physical housing conditions and the less educated the mother, the higher was the chance of mental disturbance;	not mentioned (cut-off is taken from other studies, for example Mari 1986)	not mentioned	
Tafari et al. (1991)	- Prevalence rate for neurotic mental illness was 11.2%. However, this figure was based on the SRQ-20 score instead of on a diagnostic interview; - Results from the factor analysis: • factor 1: cognitive items from the neurotic scale (e.g. trouble thinking clearly, indecisive, work suffering); • factor 2: items which tap anxiety and depression from the neurotic scale (e.g. easily frightened, unhappy, cry often, worthless); • factor 3: somatic manifestations (e.g. headache, poor appetite and digestion, sleep badly); • factor 4: psychotic items.	not mentioned	The internal consistency of the 2 scales is supported by item-total correlations. The neurotic item-total correlations were .45 and above, except for headache (.45), cry often (.40) and hands shake (.33). The psychotic items correlated poorly with the neurotic total.	As a result of previous use of the scale in Ethiopia (Kortmann) and the pretesting in this research, certain items of the SRQ were clarified by the interviewers.

Author(s) & Year	Findings	Validity assessment	Reliability assessment	Comments
Araya et al. (1992)	- Both instruments performed equally well in identifying cases of minor psychiatric morbidity in this representative sample of primary care attenders in Santiago, Chile. The small difference between the questionnaires in their ability to detect cases of minor psychiatric disorder was not statistically significant; - The estimated prevalence of minor psychiatric disorder was around 50% in this sample; - Misclassification was affected by sociodemographic variables, particularly education; - The six most prevalent items of the SRQ-20 were: Do you feel tense, nervous or worried? 90.2% Do you feel unhappy? 67.5% Do you often have headaches? 64.4% Are you easily tired? 61.3% Do you have uncomfortable feelings in your stomach? 60.7% Do you find it difficult to enjoy your daily activities? 58.3%	'Cases' were defined as those scoring 12 or more as a total score on the CIS-R.. SRQ-20 sensitivity = 74% specificity = 77% negative predictive value = 73% positive predictive value = 78% overall misclassification rate = 25% <u>GHQ-12</u> sensitivity = 73% specificity = 76% negative predictive value = 73% positive predictive value = 76% overall misclassification rate = 26% Agreement coefficients for SRQ-20 and GHQ-12 against CIS-R as criterion: ■ Kappa (s.e.): .51 (0.07) (SRQ-20) .48 (0.07) (GHQ-12) ■ Index of Agreement on positives (IAp): .62 (SRQ-20) .61 (GHQ-12) s.e. = standard error Correlation coefficients: - between CIS-R and SRQ score = 0.69 - between CIS-R and GHQ score = 0.67	not mentioned	ROC analysis was used to estimate the optimal threshold score and to compare the ability of these two questionnaires to discriminate between "cases" and "non-cases" using the CIS-R as the criterion.

Author(s) & Year	Findings	Validity assessment	Reliability assessment	Comments
Guinness (1992)	- 1985 (English version of the SRQ, n= 705): The distribution of symptom prevalence showed higher rates in rural schools (33% scored above the cut off score) than periurban (27%), and very low in elite (10%); - 1990 (Siswati translation, n = 276): rates were consistent with the earlier findings in rural and elite schools; - Significant differences between responses to English and Siswati versions of the SRQ were found for items 4 (easily frightened), 8 (trouble thinking), 9 (feeling unhappy), 10 (crying often), 11 (difficult enjoying things), 14 (useful part in life), and item 15 (losing interest in things); - When the SRQ is using the same idiom as the cultural expression, then volunteered symptoms (open question) are significantly related to SRQ items. This is strongest with somatic symptoms, weak but consistent with cognitive symptoms, but inconsistent with affective; - Prevalence of the brain fag syndrome in African adolescents in terms of definite psychiatric disorder remains to be established.	The cultural validity was tested by comparing the open-question volunteered symptoms with the SRQ items using χ^2 association and kappa correlation. For more detailed information, see the article.	none	- Only symptoms elicited by the open question that were possibly related to anxiety ("brain fag symptoms") were counted. Students with three or more symptoms were arbitrarily defined as brain fag; - SRQ-24: a cut off score of 11 was determined as indicating possible psychiatric disorder (this equates with three symptoms by the open question); - The article examines culture (language, culture specific constellation of symptoms) related issues which are not further explained here; - The authors stresses the invalidity of the psychotic items (items 21-24); - The chief limitation of this study was that it was not possible to use the two stage process to establish psychiatric morbidity rather than simple symptom prevalence rates.
Lima et al. (1992)	- The findings of this study reveal a significant amount of psychiatric morbidity among adult patients attending primary care clinics three months after a major disaster; - The most frequent diagnoses were post-traumatic stress disorder and major depression.	Of the 37 potential cases (based on the score on the SRQ-24 + 4 questions on alcohol-abuse) who were given a psychiatric interview, the investigators found 29 (78%) of the subjects were given a formal psychiatric diagnosis and 8 (22%) were not.	not mentioned	- As there is no control group, it is not possible to state from these data that the observed prevalence of psychiatric disorders in this sample differed significantly from the frequency seen in unaffected communities; - Because of the importance of alcohol abuse among the study population, the SRQ-24 has been supplemented with four questions on alcohol.

Author(s) & Year	Findings	Validity assessment			Reliability assessment	Comments
Carta et al. (1993)	- The prevalence of mental disorders, determined using the PSE, was found to be 37%; 23% suffered from anxiety disorders and 14% from depression; - Mental disorders were more prevalent among women than men (49% versus 23%); - The ROC trend is mentioned in the section on criterion validity.	Cut-off SRQ-24 ---------- 2/3 3/4 4/5 5/6 6/7 7/8 8/9 9/10 10/11	sens. (%) ------- 100 95 95 95 95 90 71 52 43	spec. (%) ------- 33 50 60 63 67 70 80 83 90	The interviewers had reached "good" (not further specified) inter-rater reliability in the course of a previous epidemiological study.	- According to the investigators, the simplicity of the SRQ-24 is one of the most encouraging findings of the study and probably one of the main reasons for the satisfactory validity results obtained; - The study seems to demonstrate the usefulness of the SRQ in industrialized countries, particularly in areas where the overall level of education is rather low, or where the prevalent daily use of a dialect or a second language would preferably indicate use of a simple and easily comprehensible screening test, such as the SRQ-24.
Thom, Zwi and Reinach (1993)	- The total percentage of psychiatric disorders in the sample was 14.38% (43/299). This is an underestimation because patients who did not have a PSE were considered to be normal; - The male/ female ratio in those who scored positively on the SRQ was 1 : 5.18; - Of the cases diagnosed by using the SRQ and the PSE, 93% were missed by the clinic staff.	The cut-off score was determined by means of tests for the highest sensitivity and specificity for various cut-off points on a pilot group of people with known psychiatric disorders and normal controls.			not mentioned	The study encountered some methodological problems: 1) How to differentiate between a "case" of psychiatric disorder and a normal response to stress; 2) The use of non-psychiatrically trained health care workers in epidemiological psychiatric research; 3) The extremely low prevalence of substance abuse-related psychiatric disorders and reported substance abuse in the study population; 4) Most patients had a concurrent physical illness.

Author(s) & Year	Findings	Validity assessment	Reliability assessment	Comments
El.Rufaie and Absood (1994)	- The study furnished evidence that the Arabic version of the SRQ-20 was a valid instrument for detecting minor psychiatric morbidity in primary health care settings: - The most sensitive questions were questions 1, 3, 4, 6, 9, 10, 18 and 20; - Several subsets of questions were tested but the total of all 20 questions were superior to all of them; - The authors investigated the statistics of each question, the results of which are not presented here.	Content validity: reasonably strong, biserial correlation values between each item and the total score of SRQ were found, Criterion validity (SRQ-20): cut-off point = 5/6 sensitivity = 78.3% specificity = 75.2% PPV = 54.7% NPV = 90.1% misclassification= 24.0%	not mentioned	- The results of statistical analysis indicates that the misclassification rate was significantly lower in males than females and the specificity is significantly higher in males. It also suggest that the rate of negative prediction is higher in males than females. There were no significant differences between the illiterates and the literate. - The authors feel that, instead of the yes/no response, three or four alternative responses would be more appropriate for most SRQ-20 items.

Appendix I. Weighting technique

Weighting technique

As mentioned before, it is common practice to include a sample of the low scorers on the SRQ in the second stage; All the high scorers on the SRQ are usually followed up. This is an example of stratified sampling. In the table below, uppercase letters are used for the first stage and lowercase letters are used for the second stage. For the first stage, only information on "A + B" and "C + D" is initially available. These are the results from the SRQ. In the second stage the case ascertainment takes place. Since not all the low scorers on the SRQ are followed up, "C" (false negatives in the first stage) and "D" (true negatives in the first stage) have to be estimated. They can only be estimated by using the results in "c" and "d".

Screening instrument results	FIRST STAGE SAMPLE		
	disorder present	disorder absent	
positive	A	B	A + B
negative	C	D	C + D
total	A + C	B + D	

SECOND STAGE SAMPLE		
disorder present	disorder absent	
a	b	a + b
c	d	c + d
a + c	b + d	

A simple example is used to illustrate a weighting technique. As can be seen from the table below, the total sample size is 800. All the high scorers are followed up in this example, but only half the low scorers.

Screening instrument results	FIRST STAGE SAMPLE		
	disorder present	disorder absent	
positive	120	80	200
negative	C	D	600
total	A + C	B + D	

SECOND STAGE SAMPLE		
disorder present	disorder absent	
120	80	200
50	250	300
170	330	

The only complete data available are of the high scorers on the SRQ. Calculating the prevalence, sensitivity and specificity based on the data in the second stage, would inevitably lead to a distorted image. It is therefore important to estimate "C" and "D" of the first stage. The proposed weighting technique is as follows:

Estimation of "C" (false negatives): c x (C + D) /(c + d) → 50 x (600/300) = 100
Estimation of "D" (true negatives): d x (C + D) /(c + d) → 250 x (600/300) = 500

Now, the data of the first stage are complete. It is possible to calculate the estimation of sensitivity and specificity by making use of the equations in Table 4. In this example the estimated sensitivity is 0.545 (without weighting 0.71), the estimated specificity is 0.86 (without weighting 0.76) and the estimated prevalence (A + C / (A + B + C + D)) is 27.5% (without weighting 34%).

REFERENCES

References

Aldana, L.L., Moreno, L.R., Carabantes, A.D. and Moscardo, e I.B. (1990). Validacion del SRQ en los examenes de salud mental en la poblacion general. *Actas Luso-Esp. Neurol. Psiquiatr.*, **18**(5), 286-289 (in Spanish)

Araya, R. I., Wynn, R. and Lewis, G. (1992). Comparison of two self-administered psychiatric questionnaires (GHQ-12 and SRQ-20) in primary care in Chile. *Social Psychiatry and Psychiatry Epidemiology*, **27**, 168-173.

Babor, T.F., de la Fuente, J.R., Saunders, J., Grant, M. (1989). The Alcohol Use Disorders Identification Test: Guidelines for use in primary health care. Unpublished document WHO/MNH/DAT/89.4. WHO, Geneva.

Bech, P., Malt U. F., Dencker, S. J., Ahlfors, U. G., Elgen, K., Lewander, T., Lundell, A., Simpson, G. M. and Lingjærde (Eds.) (1993). Scales for assessment of Diagnosis and severity of mental disorders. Supplementum 372, *Acta Psychiatrica Scandinavica*, **87**.

Beiser, M. and Fleming, J. A. E. (1986). Measuring psychiatric disorder among Southeast Asian refugees. *Psychological Medicine*, **16**, 627-639.

Bester, F. C. J., Weich, D. J. V. and Gagiano, C. A. (1991). Mental disorders of elderly blacks in the Orange Free State. *The Journal of Age Related Disorders*, **3**, 13-16.

Carta, M. G., Carpiniello, B., Cicone, V., Sannais, C., Paulis, M. and Rudas, N. (1993). Standardization of a psychiatric screening test for use by general practitioners in Sardinia, preliminary results. *Acta Psychiatrica Scandinavica*, **87**, 342-344.

Cederblad, M. and Höök, B. (1980). Daghemsvard för trearingar: en trvärvetenskaplig, experimentell studie - X. Sammanfattning. *Reports from the Laboratory for Clinical Research on Stress, Karolinska Institute, Stockholm,* no. 121 (in Swedish).

Chan, D. W. (1985). The Chinese version of the General Health Questionnaire: does language make a difference? *Psychological Medicine*, **15**, 147-155.

Chan, D. W. and Chan, T. S. C. (1983). Reliability, validity and the structure of the General Health Questionnaire in a Chinese context. *Psychological Medicine*, **13** (2), 363-371.

Climent, C. E. and De Arango, M. V. (1983). *Manual de Psiquiatria para Trabajadores de Atencion Primaria. [Psychiatric Manual for Primary Care Workers.]* Washington, Organización Panamericana de la Salud. [Pan American Health Education Foundation (PAHEF).] (Serie PALTEX para Técnicos Medios y Auxiliares, No. 1).

Climent, C. E., Diop, B. S. M., Harding, T. W., Ibrahim, H. H. A., Ladrido-Ignacio, L. and Wig, N. N. (1980). Mental health in primary health care. *WHO Chronicle*, **34**, 231-236.

Climent, C. E., Harding, T. W., Ibrahim, H. H., Ignacio, L. L. and Wig, N. (1989). El cuestionario de sintomas para la deteccion de problemas en adultos. [the symptom questionnaire for the detection of mental problems in adults]. *Acta Psiquiatrica Psicológica América Latina*, **35** (3-4), 124-131.

Climent, C. E. and Plutchik, R. (1979). Confiabilidad y validez de un cuestionario de autoreportaje de síntomas de enfermedad mental (PASSR). [Reliability and validity of a self reporting questionnaire of symptoms of mental illness]. *Revista Colombiana de Psiquiatría*, Vol. VIII, **3**, 321-330.

Deshpande, S. N., Sundaram, K. R. and Wig, N. N. (1989). Psychiatric disorders among medical in-patients in an Indian hospital. *British Journal of Psychiatry*, **154**, 504-509.

Dhadphale, M. and Ellison, R. H. (1983). The frequency of mental disorders in the outpatients of two Nyanza hospitals. *The Central African Journal of Medicine*, **29** (2), 29-32.

Dhadphale, M., Ellison, R. H. and Griffin, L. (1982). Frequency of mental disorders among outpatients at a rural district hospital in Kenya. *The Central African Journal of Medicine*, **28** (4), 85-89.

Dhadphale, M., Ellison, R. H. and Griffin, L. (1983). The frequency of psychiatric disorders among patients attending semi-urban and rural general out-patient clinics in Kenya. *British Journal of Psychiatry*, **142**, 379-383.

Diop, B., Collingnon, R., Gueye, M. and Harding, T. W. (1982). Diagnosis and symptoms of mental disorder in a rural area of Senegal. *African Journal of Medicine and Science*, **11**, 95-103.

Duncan-Jones, P., Grayson, D. A. and Moran, P. A. P. (1986). The utility of latent trait models in psychiatric epidemiology. *Psychological Medicine*, **16**, 391-405.

El-Rufaie, O.E.F. and Absood, G.H. (1994). Validity Study of the Self-Reporting Questionnaire (SRQ-20) in Primary Health Care in the United Arab Emirates. *International Journal of Methods in Psychiatric Research*, **4**, 45-53.

Egdell, H. G. (1983). Mental health care in the developing world. *Tropical Doctor*, **13**, 149-152.

Erdreich, L. S. and Lee, E. T. (1981). Use of relative operating characteristic analysis in epidemiology. *American Journal of Epidemiology*, **114** (5), 649-662.

Foulds, G.A. and Hope, K. (1968). *Manual of the Symptom Sign Inventory (SSI).* University of London Press: London.

Freeman, M., Seris, N., Matabula, E. and Price, M. (1991). *An evaluation of mental health services in the South Eastern Transvaal.* Johannesburg: Centre for Health Policy, University of the Witwatersrand, **10**, 28-35.

Goldberg, D.P., Cooper, B., Eastwood, M.R., Kedward, M.B. and Shepherd, M. (1970) *British Journal of Preventive and Social Medicine.* **24**, 18.

Goldberg, D. P. (1972). *The Detection of Psychiatric Illness by Questionnaire.* Oxford University Press, London.

Goldberg, D. and Huxley, P. (1992). *Common Mental Disorders, a Bio-Social Model.* Routledge, London.

Goldberg, D. and Williams, P. (1988). *A User's Guide to the General Health Questionnaire.* NFER-NELSON:Windsor, Great Britain.

Hall, A. and Williams, H. (1987). Hidden psychiatric morbidity - Part I: a study of prevalence in an out-patient population at Bindura Provincial Hospital. *Central African Journal of Medicine*, **33**, 239-243.

Harding, T. W. (1976). Validating a method of psychiatric case identification in Jamaica. *Bulletin of the World Health Organization*, **54**, 225-231.

Harding, T. W., De Arango, M. V., Baltazar, J., Climent, C. E., Ibrahim, H. H. A., Ladrido-Ignacio, L., Srinivasa Murthy, R. and Wig, N. N. (1980). Mental disorders in primary health care: a study of their frequency and diagnosis in four developing countries. *Psychological Medicine*, **10**, 231-241.

Harding, T. W., Climent, C. E., Diop, Mb., Giel, R., Ibrahim, H. H. A., Srinivasa Murthy, R., Suleiman, M. A. and Wig, N. N. (1983). The WHO collaborative study on strategies for extending mental health care: II: the development of new research methods. *American Journal of Psychiatry*, **140**, 1474-80.

Hoeper, E., et al (1984). The usefulness of screening for mental illness. *The Lancet* i, 33-35.

Iacoponi, E. and Mari, J. J. (1989). Reliability and factor structure of the Portuguese version of Self-Reporting Questionnaire. *The International Journal of Social Psychiatry*, **35** (3), 213-222.

De Jong, J. T. V. M. (1987). *A Descent into African Psychiatry.* Royal Tropical Institute, the Netherlands.

De Jong, J. T. V. M., De Lein, G. A. J. and Ten Horn, S. G. H. H. M. (1986). A baseline study on mental disorders in Guinea-Bissau. *British Journal of Psychiatry*, **148**, 27-32.

Kaplan, H. I. and Sadock, B. J. (Eds.) (1989). *Comprehensive Textbook of Psychiatry*. Volume 1, fifth edition, Williams and Wilkins, Baltimore.

Kigamwa, A. (1991). Psychiatric morbidity and referral rate among medical in-patients at Kenyatta National Hospital. *East-African Medical Journal*, **68** (5), 383-388.

Kortmann, F. (1987). Problems in communication in transcultural psychiatry. The Self-Reporting Questionnaire in Ethiopia. *Acta Psychiatrica Scandinavica*, **75**, 563-570.

Kortmann, F. (1990). Psychiatric case finding in Ethiopia: Shortcomings of the Self-Reporting Questionnaire. *Culture, Medicine and Psychiatry*, **14**, 381-391.

Kortmann, F. and Ten Horn, S. (1988). Comprehension and motivation in responses to a psychiatric screening instrument. Validity of the SRQ in Ethiopia. *British Journal of Psychiatry*, **153**, 95-101.

Krishnaswamy, S. and Kyaw, O. H. N. (1985). A pilot study on the use of the World Health Organisation Self-Reporting Questionnaires (SRQ) on a Malaysian population. *Singapore Medical Journal*, **26** (6), 431-434.

Ladrido-Ignacio, L., Climent, C. E., De Arango, M. V. and Baltazar, J. (1983). Research screening instruments as tools in training health workers for mental health care. *Tropical and Geographical Medicine*, **35**, 1-7.

Lemeshow, S., Hosmer Jr., D. W., Klar, J. and Lwanga, S. K. (1990). *Adequacy of Sample Size in Health Studies*. Published on behalf of WHO by John Wiley and Sons, Chichester.

Lewis, G., Pelosi, A.J., Araya, R.I. and Dunn, G. (1992). Measuring psychiatric disorder in the community: The development of a standradised assessment for lay interviewers. *Psychological Medicine*. **22**, 465-486.

Lima, B. R., Chavez, H., Samaniego, N. and Pai, S. (1992). Psychiatric disorders among emotionally distressed disaster victims attending primary mental health clinics in Ecuador. *Bulletin of PAHO*, **26** (1), 60-66.

Mari, J. J. and Williams, P. (1985). A comparison of the validity of two psychiatric screening questionnaires (GHQ-12 and SRQ-20) in Brazil, using Relative Operating Characteristic (ROC) analysis. *Psychological Medicine*, **15**, 651-659.

Mari, J. J. and Williams, P. (1986). A validity study of a psychiatric screening questionnaire (SRQ-20) in primary care in the city of Sao Paulo. *British Journal of Psychiatry*, **148**, 23-26.

Mari, J. J., Iacoponi, E., Williams, P., Simões, O. and Silva, J., B., T. (1987). Detection of psychiatric morbidity in the primary medical care setting in Brazil. *Revista de Saúde Pública*, **21** (6), 501-507.

Orley, J. and Wing, J. K. (1979). Psychiatric disorders in two African villages. *Archives of General Psychiatry*, **36**, 513-520.

Penayo, U., Kullgren, G. and Caldera, T. (1990). Mental disorders among primary health care patients in Nicaragua. *Acta Psychiatrica Scandinavica*, **82** (1), 82-85.

Rahim, S. I. A. and Cederblad, M. (1989). Epidemiology of mental disorders in young adults of a newly urbanised area in Khartoum, Sudan. *British Journal of Psychiatry*, **155**, 44-47.

Reichenheim, M. E. and Harpham, T. (1991). Maternal mental health in a squatter settlement in Rio de Janeiro. *British Journal of Psychiatry*, **159**, 683-690.

Rumble, S., Parry, C. D. H., Zwarenstein, M. and Swartz, L. (1993). *Prevalence of psychiatric morbidity in the adult population of Mamre.* Paper presented at a conference of the Epidemiological Association of Southern Africa (August 1993), Durban, South Africa.

Russell, D., Peplau, L.A. and Cutrona, C.E. (1980). The revised UCLA loneliness scale: concurrent and discriminant validity evidence. *Journal of personality and Social Psychology*, **39**, 472-480.

Salleh, M. R. (1990). Psychiatric morbidity in schizophrenic relatives- Use of Self-Reporting Questionnaires (SRQ). *Singapore Medical Journal*, **31** (5), 457-462.

Sanson-Fischer, R. W. and Martin, C. J. (1981). Standardized interviews in psychiatry: Issues of reliability. *British Journal of Psychiatry*, **139**, 138-143.

Sen, B. (1987). An analysis of the nature of depressive phenomena in primary health care utilising multivariate statistical techniques. *Acta Psychiatrica Scandinavica*, **76** (1), 28-32.

Sen, B. and Williams, P. (1987). The extent and nature of depressive phenomena in primary health care; a study in Calcutta, India. *British Journal of Psychiatry*, **151**, 486-493.

Sen, B., Wilkinson, G. and Mari, J. J. (1987). Psychiatric morbidity in primary health care; a two-stage screening procedure in developing countries: choice of instruments and cost-effectiveness. *British Journal of Psychiatry*, **151**, 33-38.

Stewart, M., Tudiver, F., Bass, M. J., Dunn, E. V. and Norton, P. G. (Eds.) (1992). *Tools for Primary Care Research, research methods for primary care.* Volume 2, Sage publications: California.

Theorell, T.G.T. (1982). Advances in cardiology: psychological problems before and after myocardial infarction. *Review of Research on Life Events and Cardiovascular Illness*, **29**, 140-147.

Thom, R. G. M., Zwi, R. M. and Reinach, S. G. (1993). The prevalence of psychiatric disorders at a primary care clinic in Soweto, Johannesburg. *South African Medical Journal*, **83**, 653-656.

Tafari, S., Aboud, F. E. and Larson, C. P. (1991). Determinants of mental illness in a rural Ethiopian population. *Social Science and Medicine*, **32** (2), 197-201.

Tulsky, D. S. (1990). An introduction to test theory. *Oncology*, **4** (5), 43-48.

Upadhyaya, A., Creed, F. and Upadhyaya, M. (1990). Psychiatric morbidity among mothers attending well-baby clinic: a cross-cultural comparison. *Acta Psychiatrica Scandinavica*, **81** (2), 148-151.

Wing, J. K., Cooper, J. E. and Sartorius, N. (1974). *The Measurement and Classification of Psychiatric Symptoms*. Cambridge University Press: Cambridge.

Williams, P., Wilkinson, G. and Rawnsley, K. (Eds.) (1989). *The Scope of Epidemiological Psychiatry*. London, Routledge.

World Health Organization (1975). WHO expert committee on mental health: organization of mental health services in developing countries. *WHO Technical Report Series 564*, Geneva, WHO.

World Health Organization (1977). *International Classification of Diseases, Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death*. Geneva, WHO.

World Health Organization (1984). Mental health care in developing countries: a critical appraisal of research findings. *WHO Technical Report Series 698*, Geneva, WHO.

World Health Organization (1992). *ICD-10, The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*. Geneva, WHO.

Zwi, R. and Thom, R. (1992). The prevalence of psychiatric disorder in a primary care clinic in Soweto. *Urbanization and Health Newsletter*, **13**, 29-32.