Brit. J. Psychiat. (1976), **129**, 61–7

A Comparison of Two Psychiatric Screening Tests

By D. P. GOLDBERG, K. RICKELS, R. DOWNING and P. HESBACHER

Summary. A comparison is made between the General Health Questionnaire (GHQ) and the Symptom Checklist (SCL) as psychiatric screening tests in community-based research projects. Both are shown to correlate equally well with independent clinical assessment, and the differences between them mainly reside in the form of their response scales. The GHQ works best as a screening test, since it has fewer false positives associated with its use, but it may miss those with long-standing disorders. The SCL tends not to miss long-standing disorders and furnishes diagnostic sub-scales if these are required. Both tests function better with men than with women and with whites than with blacks, but neither is affected by social class or age of the respondent. The study revealed high correlations between the symptoms of anxiety and depression, and indicated some possible differences between the symptom clusters seen in whites and in blacks.

Studies of psychiatric illness in general practice in England (Shepherd *et al*, 1966) and the United States (Locke and Gardner, 1969) have shown considerable variation in rates for the diagnosis of psychiatric illness between individual practitioners. Many investigators have used some form of symptom check list or questionnaire in an effort to reduce this variation. Very few of these questionnaires have acceptable validity, and no study has so far compared the relative efficiency of two such instruments.

'The General Health Questionnaire'(GHQ) in its British form (Goldberg, 1972) is a self-administered questionnaire consisting of 60 questions concerned with psychological distress or altered behaviour. For each item the respondent is asked to compare his recent state with his usual state, and an item is only counted as being present if it is being experienced 'more than usual'. The present study used a 30-item version with 3 items rephrased for North American use.

The Lipman-Rickels Scale is a 35-item version of the Hopkins symptom distress checklist (SCL) (Derogatis *et al*, 1974) and consists of physical and psychological symptoms, each scored on a frequency scale running from 'not at all' to 'extremely'.

Both scales cover symptom areas such as psychophysiological symptoms and minor affective disorder, but the GHQ also has items dealing with role satisfaction and outwardly observable behaviour. The main differences between them lie in their different response scales:

GHQ scales:	Not at all	Same as usual	Rather more than usual	Much more than usual
GHQ scoring:	0	0	1	1
SCL scales:	Not at all	A little	Quite a bit	Extremely
SCL scoring:	0	1	2	3

The 30-item GHQ therefore generates a score between 0 and 30, while the 35-item SCL generates a score between 0 and 105.

The aim of the present study is to compare the two questionnaires using an independent clinical assessment by a psychiatrist, and to investigate the effects of demographic variables on the screening characteristics of the questionnaires. Techniques of principal component analyses will be used to compare patterns of symptomatology in black and white respondents.

https://doi.org/10.1192/bjp.129.1.61 Published online by Cambridge University Press

Method

The survey was carried out in the practices of five general practitioners who had previously co-operated in research projects with the Psychopharmacology Research Unit. The practices were chosen in such a way that respondents would represent a wide range of social class for both blacks and whites.

In each practice consecutive attenders were asked to fill in the two questionnaires while they waited to see their doctor. A notice in the waiting room explained to patients that their doctor was co-operating with the University of Pennsylvania in carrying out a health survey. Demographic details were recorded on a separate card by the receptionist, and the general practitioner recorded his clinical impressions on this card when he saw the patient in his usual way. At the end of his interview with them, the general practitioner asked suitable patients whether they minded seeing another doctor, and if they agreed they were introduced to the research psychiatrist (D.G.) who was sitting in another office. The two questionnaires and the general practitioner's survey form were kept by the receptionist, and were at no stage seen by the research psychiatrist. The latter then administered the Clinical Interview Schedule of the General Practice Research Unit (Goldberg *et al*, 1970), which has been found to have an inter-rater reliability of +0·92 when used by experienced psychiatrists. The interview starts with an unstructured section which is used both to establish rapport with the patient and to form an impression of his present complaint and previous health, and five-point ratings are then made about symptoms the patient has experienced the previous week in 10 symptom areas. After the patient has left, 12 further ratings are made concerning abnormalities which have been manifested during the interview, and the 22 individual ratings can then be summed to produce an 'objective severity score'. In addition, it is usual for the psychiatrist who has carried out the interview to make a single subjective rating of the severity of psychiatric disturbance on a five-point scale ranging from 0—no disturbance and 1—subclinical disturbance to 2—mild, 3—moderate and 4—severe psychiatric disturbance.

In each practice the research psychiatrist interviewed patients selected by the general practitioner until 50 interviews had been completed. Since all the doctors concerned were in private practice they had to be allowed to use their own discretion about whom to send in for an interview. They were told not to look at the questionnaire, and to send patients in for interview whether or not they considered them to be 'cases'. As things turned out, the doctors usually sent their next patient in for an interview whenever the psychiatrist told the receptionist that he was free, but they also often asked patients to wait and see the psychiatrist when they considered them to be 'cases'. In one practice in the 'ghetto' area many patients were illiterate, and in these cases a research assistant read the questions out to the patients.

When the questionnaires and general practitioner assessment cards were collated and scored it was found that there were missing data for 6 patients; the following results are therefore based on 244 patients.

Results

In the present study, the main criterion assessment of psychiatric status is the subjective five-point rating made on the basis of the standardized psychiatric interview. When the results were plotted out (Fig 1), each screening test was found to correlate well with this clinical assessment, and indeed the screening tests

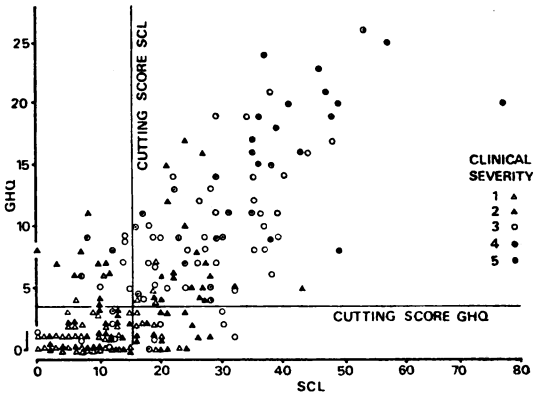


Fig 1.

https://doi.org/10.1192/bjp.129.1.61 Published online by Cambridge University Press

correlated fairly well with one another ($r = +0{\cdot}78$).

From the point of view of screening, patients can be divided into correctly identified cases and normals (true positives and true negatives) and normals with high scores (false positives) and cases with low scores (false negatives).

Symptom checklist (SCL)		
High scores (>16)	45	71
Low scores (<15)	114	14
	Normals	Cases

General Health Questionnaire (GHQ)		
High scores (>4)	36	73
Low scores (<3)	123	12
	Normals	Cases

These figures can be used to calculate the screening characteristics of the two questionnaires shown in Table I.

When these figures are examined, it can be seen that the GHQ emerges as a better discriminator than the SCL, mainly because there are fewer false positives (normals with high scores) associated with it, therefore giving the GHQ a higher specificity (true negatives $\times 100$/normals). The sensitivities (true positives $\times 100$/cases) of the two questionnaires are fairly similar. Two further screening characteristics—'hits positive' rate and overall misclassification rate—are dependent upon prevalence, but they have also been calculated for this particular population. The hits positive rate is the probability that an individual with a high score will be found to be a case at clinical interview. Although the GHQ is superior to the SCL in this respect it is of interest that the 60-item GHQ in a comparable British setting would have a hits positive rate of 81 per cent.

TABLE I

A comparison of the screening characteristics of the two questionnaires

	Symptom Checklist	General Health Questionnaire
Sensitivity	84%	86%
Specificity	72%	77%
Hits positive rate (at 35% prevalence) ..	61%	67%
Overall misclassification rate (at 35% prevalence)	24%	20%
Product moment correlation between questionnaire score and clinical assessment ..	+0·70	+0·70

The different response scales of the questionnaires have resulted in different individuals being misclassified, and it was thought of interest to examine the clinical interview schedules of the patients in an attempt to categorize the types of case misclassified by each questionnaire.

False positives

False positives on both questionnaires (n = 23). Seventy per cent of these patients fell just below the threshold required for a clinically significant disturbance, so that they had been assessed as 'subclinically disturbed' by the research psychiatrist using his overall subjective scale of disturbance. Two other measures were available: the objective severity score derived by summing the 22 ratings made during the interview, and the general practitioner's rating. Nine patients (39 per cent) would have been counted as cases by the former criterion, and twelve (52 per cent) by the latter. A further five patients received ratings as being 'suspicious, defensive' during the interview, and this may be related to a reluctance to discuss symptomatology with a strange doctor, even though symptoms had been complained of on both the questionnaires. There were also five older patients (mean age 67) with multiple physical complaints who at interview were reticent about psychological problems.

False positives on Symptom Checklist only (n = 21). This group consisted mainly of patients coming up for routine visits, and seemed to score highly on the SCL by virtue of somatic symptoms; or by endorsing 'a little' to many of the

https://doi.org/10.1192/bjp.129.1.61 Published online by Cambridge University Press

questions and thereby cumulating a high score. Two thirds of these patients were thought to have a subclinical disturbance on the standard interview, but the two other measures of morbidity showed this group to be less disturbed than the previous group: only three patients would have been counted as cases by the objective severity score, and six by the general practitioner. Six patients were rated at the extreme 'normal' end of the scale by both research psychiatrist and GP.

False positives on General Health Questionnaire only (n = 12). Eleven of these patients were rated as having a subclinical disturbance, although, as in the previous group, the other two measures of morbidity rated them as 'well': only two were cases on the objective overall score, and only one was thought disturbed by the GP. Most of the patients appeared to have minor, transient disturbances.

False negatives

False negatives on both questionnaires (n = 6). Three of these patients received morbid ratings for 'suspicious, defensive' behaviour at interview. One patient was reluctant to come in at all and asked if her name could be erased. Later in the interview she began to cry, but when asked about this said she had a head cold. Another patient, asked about her long pauses before replying, said 'I'll consider each question before I answer it', but eventually went on to describe in-patient psychiatric treatment. The third patient had mild anxiety symptoms concerned with her husband. The interview was unusual in that the husband insisted on coming in with the patient and frequently interrupted her. The interviewer's notes read: 'She is deaf, but not as much as she makes out. There is doubt about the reliability of this information, since her husband is checking everything she says.' Her general practitioner rated her as having a moderately severe psychiatric illness, but she scored zero on the SCL and 1 on the GHQ. The remaining three patients had had their symptoms for longer than a year: more than 15 years in two cases. The GHQ tends to miss patients with very long-standing disorders because of the nature of its response scale.

False negatives on Symptom Checklist only (n = 8). These patients formed a heterogeneous group, but there was little doubt about the correctness of the psychiatric assessment: none of these patients was borderline, and indeed their objective overall severity scores were all well above the threshold as well. In several of these cases it seemed that the GHQ was detecting the patients because of the items dealing with outwardly observable behaviour. The general practitioners were aware of half of these cases, while the others were 'hidden illnesses'.

False negatives on GHQ only (n = 6). These were all mild, long-standing illnesses, lasting from 'several years' to 25 years. Only one was rated as psychiatrically ill by the general practitioner. The failure of the GHQ to detect these patients was undoubtedly due to the length of the history causing the patient to endorse 'same as usual', and therefore not being scored as a case.

TABLE II

The effect of demographic variables on the mis-classification rates of the two questionnaires

	n	Symptom Checklist 'SCL'	General Health Questionnaire 'GHQ'
Sex			
Males ..	80	19%	19%
Females ..	167	28%	21%
*Race**			
Whites ..	148	22%	18%
Blacks ..	97	28%	25%
Marital status			
Single ..	48	24%	25%
Married+ widowed ..	177	24%	18%
Divorced+ separated ..	22	35%	27%

* 2 patients were Oriental.

The effect of demographic variables on screening

The best way of comparing two screening instruments on a given population is to compare the overall misclassification rates. This rate is the sum of the incorrectly classified patients

https://doi.org/10.1192/bjp.129.1.61 Published online by Cambridge University Press

divided by the total number of patients who were assessed at the case-finding stage of the screening process. Some variables—such as age and social class (whether measured by Hollingshead and Redlich's Index or by the Registrar General's Classification)—have no effect on the misclassification rates. Table II sets out the three variables which were found to have some effect. It can be seen that while both questionnaires do better with males than with females the GHQ owes its superiority to the SCL to the different rates for females. Both questionnaires function better with whites than with blacks, possibly because they were constructed for white populations, or because, as we shall see, the structure of symptomatology is different for the two races.

Correlations between ratings

Some of the more interesting correlations between variables are shown on Table III. Sex and age have been omitted, since they did not significantly correlate with any of the other variables. It can be seen that each of the questionnaires correlates with the morbid anxiety or depression rated at interview; and indeed these morbid affects correlate +0·63 with one another. The two overall severity scores are highly correlated with one another, but this is hardly surprising since the interviewer does not make his subjective overall severity score until his 22 ratings are complete, so the two scores are in no sense independent measures. Hollingshead and Redlich's social index (high index—low social class) and race (white = 1, black = 2) are shown to be correlated in the present sample, but otherwise there is a low, positive correlation of approximately +0·20 between these variables and most of the others.

Principal components analysis with the GHQ

It was thought of interest to carry out multivariate analyses with the GHQ, since considerable published data already exist concerning the SCL (Derogatis *et al*, 1970). Analyses were carried out for each race separately for the 1,310 respondents who had completed the questionnaire as the first stage of screening. In each race four factors accounted for a little over 50 per cent of the total variance, and varimax rotations were found to be more meaningful than the original, unrotated structure.

Table IV shows that in each race the first factor is concerned with depression and anxiety, and accounts for about 21 per cent of the variance. The items which define the factor for each race and even the item loads themselves are strikingly similar. It should perhaps be emphasized that both rotated and unrotated solutions have been examined up to the 7-factor solution for each race, and that there is no solution which produces anxiety items on one dimension and depressive items on another. (The original principal components analysis carried out on English respondents (Locke & Gardner, 1969) also showed anxiety and depression on the same dimension.) It will be recalled that the earlier section of the paper also showed a correlation between anxiety and depression as rated at psychiatric interview.

TABLE III

Correlation between some of the variables measured in the study

	GHQ score	Subjective overall severity	Objective overall severity	Anxiety at interview	Depression at interview	Social index	Race (w = 1) (b = 2)
SCL score	+0·78	+0·70	+0·75	+0·53	+0·62	+0·24	+0·19
GHQ score		+0·70	+0·71	+0·55	+0·67	+0·23	+0·18
Subjective overal severity			+0·89	+0·73	+0·75	+0·22	+0·21
Objective overall severity				+0·77	+0·80	+0·17	+0·19
Anxiety					+0·63	+0·14	+0·15
Depression						+0·21	+0·24
Social index							+0·53

https://doi.org/10.1192/bjp.129.1.61 Published online by Cambridge University Press

TABLE IV

The factor structure of symptoms for Blacks and Whites (varimax rotations of the four factor solutions)

	1	2	3	4	Cumulative variance
Whites					
Variance ..	20·6%	11·7%	9·8%	9·4%	51·5%
Item loads ..	Life entirely hopeless 0·73	Restless nights 0·72	Busy and occupied 0·69	No warmth or affection 0·64	
	Feels worthless 0·73	Losing sleep, worry 0·70	Making decisions 0·55	Can't get on with others 0·64	
	Lost confidence 0·71	Not alert 0·61	Managing as well as most people 0·54	Not feeling happy 0·58	
	Life not worth living 0·72	No energy 0·57	Playing a useful part in things 0·54	Can't enjoy activities 0·55	
	Depressed, unhappy 0·66	Can't concentrate 0·53			
	Scared, panicky 0·65				
	Nerves too bad 0·61				
	Nervous and hung up 0·58				
	Felt under strain 0·46				
Provisional name ..	'Depression and anxiety'	'Insomnia and anergia'	'Social functioning'	'Anhedonia'	
Blacks					
Variance ..	21·3%	12·4%	11·0%	8·5%	53·2%
Item loads ..	Life entirely hopeless 0·77	No warmth or affection 0·65	No energy 0·63	Losing sleep, worry 0·76	
	Feels worthless 0·76	Can't face problems 0·58	Not alert 0·63	Restless, disturbed nights 0·74	
	Lost confidence 0·74	Not hopeful about future 0·54	Not enjoying activities 0·60		
	Life not worth living 0·72	Not busy and occupied 0·53	Feels useless 0·55		
	Depressed, unhappy 0·62	Not feeling happy 0·52			
	Scared, panicky 0·70				
	Nerves too bad 0·63				
	Nervous and hung up 0·60				
	Felt under strain 0·49				
Provisional name ..	'Depression and anxiety'	'Anhedonia'	'Anergia'	'Insomnia'	

https://doi.org/10.1192/bjp.129.1.61 Published online by Cambridge University Press

It will be seen that the remaining dimensions are fairly similar, except that in whites there is a factor which appears to relate to defective social functioning, and this does not appear in blacks.

Discussion

Choice of a particular self-administered psychiatric screening test will depend on the needs of a given research project: both questionnaires furnish scores which correlate equally well with independent clinical assessment, and the patients misclassified by both constitute an inherent limitation of the brief, self-administered pencil and paper test. An experienced clinician will always make observations which cannot be built into the test, and there will always be patients with borderline disorders about whom doubt will arise.

The GHQ has a slightly better overall performance as a screening test, mainly because it does not give so many false positives among female patients as the SCL, and possibly because it strays outside the conventional symptom areas and includes items of outwardly observable behaviour. Its main shortcoming compared with the SCL concerns its tendency to miss some patients with long-standing disorders.

The SCL tends not to miss long-standing disorders, but it tends to misclassify 'middle-users' of its response scale who play safe by endorsing 'a little' to many of the items and thus manage to accumulate a score in the morbid range. The SCL has the advantage of providing separate sub-scales for anxiety, depression, psychophysiological symptoms and so on, which are not available with the GHQ.

Just as the GHQ works better with the British populations on which it was originally calibrated, so both questionnaires function better as screening tests with whites than with blacks. Neither questionnaire appears affected by age or social class.

In these data, anxiety and depression emerge as highly correlated affects and have a strikingly similar factor structure for the two races. In contrast, a dimension of defective social functioning appeared in the analysis for whites but not for blacks, whose remaining symptom clusters were concerned with anhedonia, anergia and insomnia.

Acknowledgements

This study was supported in part by USPHS Research Grants MHO-8957 and MHO-8958.

We wish to thank Dr D. Clark, Dr W. Rial, Dr H. Rosenfield, Dr A. Segal and Dr B. Zamostien who allowed this work to be carried out in their practices in Philadelphia.

References

Derogatis, L. R., Lipman, R. S., Covi, L., Rickels, K. & Uhlenhuth, E. H. (1970) Dimensions of outpatient neurotic pathology: comparison of a clinical versus an empirical assessment. *J. consult. & clin. Psychol.*, **34**, 164–71.

—— —— Rickels, K., Uhlenhuth, E. H. & Covi, L. (1974) The Hopkins Symptom Checklist (HSCL): A Measure of Primary Symptom Dimensions. In *Psychological Measurements in Psychopharmacology. Mod. Probl. Pharmacopsychiat.*, **7**, 79–110 (ed P. Pichot). Basel: Karger.

Goldberg, D. P. (1972) *The Detection of Psychiatric Illness by Questionnaire*. Oxford University Press, No. 21.

—— Cooper, B., Eastwood, M. R., Kedward, H. B. & Shepherd, M. (1970) A standardized psychiatric interview suitable for use in community surveys. *Brit. J. prev. soc. Med.*, **24**, 18.

Locke, B. Z. & Gardner, E. A. (1969) Psychiatric disorders among the patients of general practitioners and internists. *Public Health Reports*, **84**, no. 2, 167–73.

Shepherd, M., Cooper, B., Brown, A. C. & Kalton, G. (1966) *Psychiatric Illness in General Practice*. London.

David Goldberg, D.M., F.R.C.P., F.R.C.Psych., *Professor of Psychiatry, University of Manchester, Withington Hospital, Manchester, M20 8LR*

Karl Rickels, M.D., *Professor of Psychiatry and Pharmacology, University of Pennsylvania; and Director of the Psychopharmacology Research Unit and Chief of Psychiatry, Philadelphia General Hospital*

Robert Downing, Ph.D., *Assistant Professor of Psychology in Psychiatry, University of Pennsylvania*

Peter Hesbacher, Ph.D., *Assistant Professor of Sociology in Psychiatry, University of Pennsylvania*

Address of the last three authors: 203 Piersol Building, University Hospital, Philadelphia, Pa. 19104

(*Received 29 September; revised 19 November 1975*)

https://doi.org/10.1192/bjp.129.1.61 Published online by Cambridge University Press