British Journal of Psychiatry (1989), **154**, 378–382

PROQSY: a Computerised Technique for Psychiatric Case Identification in General Practice

GREG WILKINSON and A. C. MARKUS

Two-stage screening for minor psychiatric morbidity was carried out with a consecutive sample of 256 patients aged 15–69 consulting an experienced GP in a semirural setting over a seven-week period. Patients completed a short GHQ, and the GP made a standardised assessment of the reason for consultation and the degree of psychological disorder. A 40% random subsample were evaluated using PROQSY, a newly developed computerised psychiatric assessment based on the CIS. PROQSY is acceptable to patients in general practice, and yields an estimate of the prevalence of minor psychiatric morbidity (33%), and validity indices for the 12-item, 28-item, and 30-item GHQs, comparable with those found in similar studies, when the CIS has been administered by an interviewer. PROQSY may be recommended as a research tool, providing a criterion measure of minor psychiatric morbidity in general practice, where it may reduce time and manpower, and eliminate interviewer bias.

Two-stage screening for minor psychiatric morbidity in general practice using the General Health Questionnaire (GHQ) (Goldberg, 1972) as a first-stage screening instrument and the Clinical Interview Schedule (CIS) (Goldberg *et al*, 1970*a*) as a second-stage criterion measure has become established (Goldberg & Blackwell, 1970*b*; Goldberg, 1978; Skuse & Williams, 1978; Skuse & Williams, 1984; Wright & Perini, 1987; Goldberg & Williams, 1988).

This report is based on the above technique, and documents the use of PROQSY, a newly developed computerised version of the CIS (Lewis *et al*, 1988), as a criterion measure of minor psychiatric morbidity in a semirural general-practice setting. Relative Operating Characteristic (ROC) analysis was used to compare the validity of the screening instruments and the GP's assessments. In addition, patients' views on the use of PROQSY were elicited using a brief standardised interview.

Method

Setting

The survey took place at a modern, purpose-built health centre in Thame, an Oxfordshire market town, where the population has risen over the past 20 years from 5000 to 10 000. The environment is semirural. Over 10 000 patients from Thame and its surrounding villages are registered with the practice, which has been in existence for over 30 years. There are six GP partners, and one trainee GP. A.C.M. is one of the senior partners. The practice operates a 10-minute appointments system, and patients may consult with the doctor of their choice, subject to availability.

Procedure

First stage. Over a six-week period, all consecutive patients registered with the practice and consulting the GP were asked to complete a short form of the GHQ (Goldberg, 1972), which comprised the items contained in the 28-item and 30-item versions of the GHQ (thus including the 12-item GHQ), while awaiting their appointment with the doctor.

After each consultation, the GP, who was blind to GHQ scores, made a standardised assessment of the patient's reason for consultation and the degree of psychological problem present.

Second stage. A 40% random subsample of patients were selected for assessment using PROQSY. PROQSY scores were calculated as for the CIS, by summing individual symptom scores. A cut-off score of 9/10 using PROQSY was chosen in advance as the optimum for case/non-case discrimination, on the basis of validity studies comparing PROQSY with the CIS (Lewis *et al*, 1988).

PROQSY was administered in a quiet, open area of the health centre, beside a window but out of sight of other consulters and staff. An IBM-compatible microcomputer with a standard keyboard and monochrome screen was used. Patients were seated comfortably and G.W. waited with them until they had completed the instructions section and the first question ("How old are you?"). Thereafter GW returned occasionally to ask if there were any difficulties. The nature of the survey and the computerised assessment was discussed with patients after the procedure was completed.

Validity of PROQSY

Lewis *et al* (1988) have shown that in both primary and secondary health care settings there is good agreement between PROQSY assessments and those of the CIS administered by psychiatrists, both in assessing overall severity and in defining 'cases' of psychiatric disorder.

https://doi.org/10.1192/bjp.154.3.378 Published online by Cambridge University Press

In addition, when individual symptoms elicited by PROQSY and the CIS were compared, the levels of agreement found were similar to those found from interobserver studies of standardised interviews.

ROC analysis

The application of ROC analysis (Swets, 1964; Egan, 1975) to studies of this kind has been found to be feasible and straightforward, and superior to the conventional method of presentation of validity data (Mari & Williams, 1985).

The most useful summary index of the discriminating ability of an instrument is the area under the ROC curve (Swets, 1979), which represents the probability that a random pair of normal and abnormal subjects will be correctly rated as regards their disease state (Hanley & McNeil, 1982). This index varies from 0.5 for a screening test whose ability to discriminate between 'cases' and 'non-cases' is no better than chance, to 1.0 for a test with perfect discrimination. A method of comparing the areas under ROC curves derived from the same sample has been described by Hanley & McNeil (1983). A computer program for ROC analysis has been developed by Dorfman & Alf (1969) to estimate indices of accuracy, and has been modified by Metz *et al* (1984) to estimate parameters for the fitted ROC curve.

Patients' views on the use of PROQSY

To ascertain patients' views on the use of PROQSY, 50 consecutive patients were questioned immediately after completing the PROQSY assessment. The interview was structured and responses were classified on four-point Likert scales. In addition, a record was made of respondents' spontaneous comments about their experience of using PROQSY. A further 33 consecutive patients were asked whether, in retrospect, it was easier to disclose answers to questions about emotional symptoms to their doctor or to PROQSY.

Results

Characteristics of the sample

A total of 256 consulters (152 males and 104 females) were eligible to take part in the survey. One male and one female did not complete the first-stage questionnaires: thus the response rate in the first stage was 99.2%. The mean ages of the male and female participants were 39.1 years (s.d. 15.6; range 15–69) and 36.3 years (s.d. 13.7; range 15–66). Of the 100 first-stage participants (54 male and 46 female) randomly selected for assessment using PROQSY, four men and two women did not take part in the second stage: thus the response rate in the second stage was 94%. The males and females completed PROQSY in a mean of 15.4 min (s.d. 4.1; range 7–25) and 18 min (s.d. 6.3; range 12–35). No person who began the assessment stopped before it was completed.

General practitioner's assessments

The GP completed assessments on all participants. Reason for consultation was classified by the GP as follows:

Entirely physical problem	158	(62%)
Primary physical problem; associated psychological problem	27	(11%)
Primary psychological problem; associated physical problem	10	(4%)
Unrelated physical and psychological problems	51	(20%)
Entirely psychological problem	5	(2%)
No problem, miscellaneous, unclassifiable	3	(1%)

TABLE I
GP's assessment of degree of psychological problem

Psychological problem	*Male*	*Female*	*Total*	*12-item GHQ*		*28-item GHQ*		*30-item GHQ*		*PROQSY*	
				0–1	*2+*	*0–3*	*4+*	*0–2*	*3+*	*0–9*	*10+*
Nil	91	57	148	120	28	109	39	116	32	33	7
Mild – not influencing management at present	26	14	40	30	10	28	12	28	12	9	4
Mild – influencing management at present	12	16	28	12	16	11	17	10	18	3	5
Moderate – not influencing management at present	10	5	15	9	6	5	10	7	8	4	3
Moderate – influencing management at present	11	11	22	4	18	4	18	4	18	1	3
Severe – not influencing management at present	0	0	0	0	0	0	0	0	0	0	0
Severe – influencing management at present	1	0	1	0	1	1	0	0	1	0	1
Totals	151	103	254	175	79	158	96	165	89	50	23

https://doi.org/10.1192/bjp.154.3.378 Published online by Cambridge University Press

TABLE II
Cross-tabulation of reason for consultation with degree of psychological problem

Reason for consultation	*Degree of psychological problem*				
	Nil		*Mild or greater*		*Total*
Entirely physical	146	91 M, 55 F	2	0 M, 2 F	148
Physical and psychological	15	7 M, 8 F	91	53 M, 38 F	106
Totals	161		93		254

$\chi^2 = 186.4$ (total); 120.0 (males); 63.8 (females); all d.f. = 1; all $P < 0.001$.

TABLE III
Areas under ROC curves

	Total		*Male*		*Female*	
	Mean	*s.d.*	*Mean*	*s.d.*	*Mean*	*s.d.*
12-item GHQ	0.84	(0.07)	0.77	(0.11)	0.91	(0.07)
28-item GHQ	0.89	(0.04)	0.92	(0.05)	0.88	(0.07)
30-item GHQ	0.89	(0.05)	0.86	(0.07)	0.90	(0.07)
GP assessment	0.74	(0.07)	0.70	(0.11)	0.79	(0.09)

Although conditions classified as entirely psychological problems were uncommon, some degree of psychological problem coexisted with a physical problem in 37% of the sample. There was no association between sex and the reason for consultation.

Table I shows the GP's classification of the degree of psychological problem present. The GP rated 42% (106/254) of the sample as having at least a mild degree of psychological problem. There was no association between sex and degree of psychological problem.

Table II shows reason for consultation (categorised as either entirely physical, or physical and psychological) cross-tabulated with degree of psychological problem (categorised as nil, or mild and greater), and that there were highly significant positive associations between these variables overall and for both sexes. Thus, patients whose consultations were categorised as for physical and/or psychological problems tended to be categorised as having mild or greater degree of psychological problem.

Validity of the GHQ and GP assessments using ROC analysis

ROC analysis was restricted to the 40% random subsample selected for the second stage of the survey. The areas under the ROC curves are shown in Table III and Fig. 1. Curves were obtained for GHQ and GP's assessments completed on the day of the consultation and within seven days of consultation. These sets of curves were very similar, so only data from those assessments completed on the day of consultation ($n = 73$) are shown here.

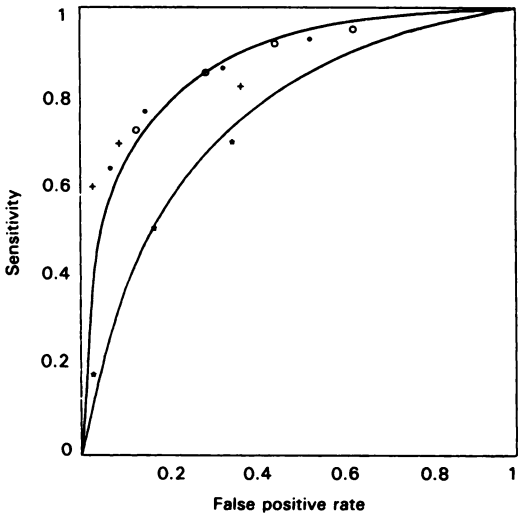


FIG. 1 A comparison between ROC curves for the 28-item (o) and 30 (item (°) GHQs and the GP's assessments (*). Co-ordinates for the 12-item (+) GHQ are also shown.

The GHQ uniformly has higher area values than the GP's assessments, but the areas under the ROC curves were not statistically significantly different from each other – using a method of comparing the areas under ROC curves derived from the same cases (Hanley & McNeil, 1983).

Optimum cut-off scores from ROC curves

The optimum cut-off points for the 12-item, 28-item, and 30-item GHQs and the GP's assessments were 1/2, 3/4, 2/3, and 0/1, respectively, with false positive and false negative misclassifications being regarded as of equal weight.

Estimated validity coefficients of the GHQ and GP assessments

Using these cut-off scores, validity coefficients were calculated for the sample (Diamond & Lillienfeld, 1962) (Table IV). As can be seen, validity coefficients from the GHQ data were similar but generally superior to those for the GP's assessments.

Estimated prevalence of psychiatric morbidity

Table V shows the expected distribution of psychiatric morbidity in the sample calculated according to the method suggested by Skuse & Williams (1984) using figures obtained from the 30-item GHQ and with the cut-off scores for the GHQ and PROQSY quoted above. The estimated prevalence of psychiatric morbidity was thus 33% (82.91/254) (95% confidence interval 27–39%). Using the same method, it was 32% in males and 34% in females).

https://doi.org/10.1192/bjp.154.3.378 Published online by Cambridge University Press

TABLE IV
Validity coefficients of the 12-item, 28-item, and 30-item GHQ and GP assessments

	12-item GHQ			*28-item GHQ*			*30-item GHQ*			*GP*		
	Total	*Male*	*Female*	*Total*	*Male*	*Female*	*Total*	*Male*	*Female*	*Total*	*Male*	*Female*
Sensitivity	75	73	80	86	88	88	78	75	83	71	71	71
Specificity	91	90	91	81	86	78	86	83	89	73	75	69
Positive predictive value	80	78	81	65	42	58	73	67	81	56	56	54
Negative predictive value	88	87	91	93	90	95	89	88	91	84	85	82
Overall misclassification	15	16	12	17	40	20	17	19	12	28	27	30
Kappa	0.67	0.63	0.72	0.61	0.27	0.56	0.63	0.56	0.73	0.41	0.42	0.37

TABLE V
Estimated prevalence of psychiatric morbidity

	Subsample assessed using PROQSY (n = 73)			*Total sample (n = 254)*		
	Number	*PROQSY + ve*		*Number*	*Estimated psychiatric cases*	
		Number	*Proportion (p)*		*Proportion (p)*	*Number*
GP -ve; GHQ -ve	30	2	0.07	116	0.07	7.73
GP -ve; GHQ +ve	10	5	0.5	32	0.5	16
GP +ve; GHQ −ve	19	4	0.21	49	0.21	10.32
GP +ve; GHQ +ve	14	12	0.86	57	0.86	48.86
Totals	73	23	—	254	—	82.91

Patients' opinions about the use of PROQSY

The 25 males and 25 females interviewed agreed that the on-screen instructions for the use of PROQSY were easy to understand (41 "very easy" and 9 "fairly easy") and that PROQSY was easy to complete (28 "very easy" and 22 "fairly easy"). All thought that they had given an accurate description of themselves using PROQSY (14 "very accurate" and 36 "fairly accurate"): several respondents suggested that there was a need for a greater number of response categories in PROQSY and that they would have liked to qualify some of their responses; others stressed that their circumstances were unusual at the time of the PROQSY assessment, and this affected the accuracy of their descriptions. All thought that PROQSY was acceptable for routine use in the health centre (36 "very acceptable" and 14 "fairly acceptable"): several commented that it was a welcome development. There was no statistical association between the respondents' assessments of the accuracy or acceptability of PROQSY and sex or their PROQSY scores.

A further 33 consecutive patients (20 males and 13 females) were interviewed (by GW) immediately after completing the PROQSY assessment, again using a standardised interview. They were asked whether, in retrospect, it was easier to disclose their answers to their doctor or to PROQSY. First, under the condition that they had a satisfactory relationship with their doctor: 46% indicated their doctor; 36% were indifferent between the two; and 18% indicated PROQSY. Secondly, under the condition that they did not have a satisfactory relationship with their doctor: 18% indicated their doctor; 9% were indifferent between the two; and 73% indicated PROQSY.

Discussion

This two-stage screening survey documents the use of PROQSY as a criterion measure of minor psychiatric morbidity in general practice. It also appears to be the first validity study of the GHQ from a semirural British general practice setting. The results observed show similar validity indices for the 12-item, 28-item, and 30-item GHQs, and their general superiority to standardised GP assessments.

The GP's assessment of reason for consultation was broadly similar to those reported in previous British studies (Goldberg & Blackwell, 1970*b*; Skuse & Williams, 1984; Wright & Perini, 1987). This was also true as regards his assessment of degree of psychological problem, and it is notable that the 42% rated as having a mild or greater degree of psychological problem corresponds exactly to the percentage assessed as having subclinical or greater degree of psychiatric disorder in the study conducted by Skuse & Williams (1984).

The validity indices calculated for the GP were substantial, though generally inferior to those of the different versions of the GHQ. In comparison to the figures for 'identification index' and 'accuracy' quoted by Marks *et al* (1979) for 91 Manchester GPs

https://doi.org/10.1192/bjp.154.3.378 Published online by Cambridge University Press

using the 60-item GHQ, the GP here had a higher sensitivity and an equivalent 'accuracy'. Taken together, these findings suggest that this GP has an ability to detect minor psychiatric morbidity which is comparable to that reported for other GPs, so supporting the generalisability of this work.

PROQSY was acceptable for use in this setting. In all, 197 computerised assessments were carried out during office hours over a seven-week period. None of the patients who refused to take part in the study indicated that this was on account of PROQSY. Indeed, patients had few difficulties with PROQSY. However, three participants required some help keying in their responses: one man with severe depression who said that he would be unable to complete the interview unaided; a male with a spastic paralysis; and an obsessional woman who was taking an excessive time deciding about which response to give and then to key it in.

ROC analysis has been used already in a few studies employing the GHQ (Mari & Williams, 1985; Bridges & Goldberg, 1986; Bellantuono *et al*, 1987; Surtees, 1987). The areas under the ROC curves obtained here closely resemble the values obtained in these other conventional studies. Differences between the three versions of the GHQ in area under the ROC curve or in other validity indices were marginal. Overall, our results confirm the observation that all applications of ROC analysis to the GHQ have given values for the area under the curve of greater than 0.80 (Goldberg & Williams, 1988).

The GP identified 42% of the sample as having at least a mild psychological problem; the different versions of the GHQ identified 31–38% as being probable psychiatric cases; and the estimated prevalence of psychiatric disorder based on PROQSY was 33%. This may be compared to Skuse & Williams' (1984) finding of an estimated prevalence of 34%, using conventional methods in a study conducted with a principal general practitioner in an outer-London practice. As has been found in conventional studies, the GP and GHQ tended to overidentify psychiatric disorder as measured by this technique.

Acknowledgements

This work was carried out under the direction of Professor Michael Shepherd at the General Practice Research Unit, Institute of Psychiatry, London, with the support of the Department of Health and Social Security. We thank the staff of Thame Health Centre and the participants for their help.

References

BELLANTUONO, C., FIORI, R., ZANOTELLI, R., *et al* (1987) Psychiatric screening in general practice in Italy: a validity study of the GHQ. *Social Psychiatry*, **22**, 113–117.

BRIDGES, K. W. & GOLDBERG, D. P. (1986) The validation of the GHQ–28 and the use of the MMSE in neurological in-patients. *British Journal of Psychiatry*, **148**, 548–553.

DIAMOND, E. L. & LILLIENFELD, A. M. (1962) Effects of errors in classification and diagnosis in various types of epidemiological studies. *American Journal of Public Health*, **52**, 1137–1144.

DORFMAN, D. D. & ALF, E. (1969) Maximum-likelihood estimation of parameters of signal detection theory and determination of confidence intervals: rating method data. *Journal of Mathematical Psychology*, **6**, 487–496.

EGAN, J. P. (1975) *Signal Detection Theory and ROC Analysis*. New York: Academic Press.

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

—— (1978) *The Manual of the General Health Questionnaire*. Windsor: NFER Publishing Company.

——, COOPER, B., EASTWOOD, M. R., *et al* (1970*a*) A standardised psychiatric interview for use in community surveys. *British Journal of Preventive and Social Medicine*, **24**, 18–23.

—— & BLACKWELL, B. (1970*b*) Psychiatric illness in general practice: a detailed study using a new method of case identification. *British Medical Journal, ii*, 439–443.

—— & WILLIAMS, P. (1988) *A User's Guide to the General Health Questionnaire*. Windsor: NFER Publishing Company.

HANLEY, J. A. & MCNEIL, B. J. (1982) The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. *Radiology*, **143**, 29–36.

—— & —— (1983) A method of comparing the areas under Receiver Operating Characteristic curves derived from the same cases. *Radiology*, **148**, 839–843.

LEWIS, G., PELOSI, A. J., GLOVER, E., *et al* (1988) The development of a computerised assessment for minor psychiatric disorder. *Psychological Medicine* **18**, 737–745.

MARI, J. de J. & 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.

MARKS, J. N., GOLDBERG, D. P. & HILLIER, V. F. (1979) Determinants of the ability of general practitioners to detect psychiatric illness. *Psychological Medicine*, **9**, 337–353.

METZ, C. E., WANG, P. L. & KRONMAN, H. B. (1984) *ROCFIT*. Chicago: Department of Radiology and the Franklin McLean Memorial Hospital Research Institute, University of Chicago.

SKUSE, D. & WILLIAMS, P. (1984) Screening for psychiatric disorder in general practice. *Psychological Medicine*, **14**, 365–377.

SURTEES, P. G. (1987) Psychiatric disorder in the community and the General Health Questionnaire. *British Journal of Psychiatry*, **150**, 828–835.

SWETS, J. A. (1964) *Signal Detection and Recognition by Human Observers*. New York: John Wiley & Sons.

—— (1979) ROC analysis applied to the evaluation of medical imaging techniques. *Investigative Radiology*, **14**, 365–377.

WRIGHT, A. F. & PERINI, A. F. (1987) Hidden psychiatric illness: use of the General Health Questionnaire in general practice. *Journal of the Royal College of General Practitioners*, **37**, 164–167.

*Greg Wilkinson, MRCP, MRCPsych, *Senior Lecturer, Institute of Psychiatry*; A. C. Markus, MA, BM, BCh, FRCGP, MRCP, *GP, Thame Health Centre and Department of Community Medicine and General Practice, University of Oxford*

*Correspondence: *De Crespigny Park, Denmark Hill, London SE5 8AF*

https://doi.org/10.1192/bjp.154.3.378 Published online by Cambridge University Press