Factor Analysis and Validation of the General Health Questionnaire in Women: A General Practice Survey

P. HOBBS, C. B. BALLINGER and A. H. W. SMITH

Summary: As part of a survey of 1517 women aged 20-60 years, a factor analysis and validation study of the General Health Questionnaire (GHQ) was carried out. Although three clinically relevant factors could be isolated, their sensitivity was less than that of the total GHQ score, which was found to be a good measure of current psychiatric disturbance in this community sample.

The General Health Questionnaire (GHQ) has been found to be a useful instrument for screening psychiatric morbidity in the community (Goldberg *et al*, 1976). The primary function of the GHQ is to identify current psychiatric disturbance, which may then be clarified by a standardized interview, such as the Clinical Interview Schedule (Goldberg *et al*, 1970). The GHQ has been used in studies of psychiatric morbidity in general practice (Goldberg and Blackwell, 1970), general medical wards (Maguire *et al*, 1974), new referrals to gynaecological out-patient clinics (Ballinger, 1977; Worsley *et al*, 1977) and in a general population sample (Ballinger, 1975).

Validation studies have been carried out (Goldberg and Blackwell, 1970; Tarnopolsky *et al*, 1979) and attempts made to isolate factors and produce subscales from the standard 60-item or from the 30-item GHQ, to improve its sensitivity and specificity (Goldberg *et al*, 1976; Worsley *et al*, 1978; Goldberg and Hillier, 1979). In the original description, Goldberg (1972) pointed out that the misclassification rate with the GHQ increased when the true prevalence of psychiatric disturbance was particularly high or low, and that the GHQ had its greatest value in the middle range of prevalence. Recently, Benjamin *et al* (1982) found a sensitivity rate of only 54 per cent for the GHQ in a sample of 92 women aged 40–49 years; they suggested that using a 15-item factor from the GHQ with a Lickert scoring method improved its sensitivity. In the present study, which is part of a larger survey of the effects of social, domestic and reproductive stresses on psychiatric symptoms in women, a factor analysis was carried out on 1488 completed GHQs from a sample of 1517 female respondents (29 were poorly completed). A validation study was also carried out on the GHQs completed by 382 respondents, who were also interviewed.

Method

All females between the ages of 20–60 on the list of one group of general practitioners were asked to complete a GHQ. Subjects were allocated consecutive index numbers as the GHQs were posted to them, and all respondents with even index numbers who scored 12 or more were then asked if they were willing to be interviewed. These subjects were designated Group A. A further group was selected for interview by taking the nearest index number subject of matching age, marital status and social class with a score of 11 or less. These women were designated Group B.

All women in Groups A and B were interviewed, using the Clinical Interview Schedule (CIS) described by Goldberg *et al* (1970). A total CIS score was calculated for each subject by adding together the ratings for each individual symptom and double the rating for each manifest abnormality. Any subject scoring 13 or more was counted as a 'case'. An overall severity rating was also made at the end of the interview, and any subject scoring two, three or four was counted as a case. These cut-off points are all as described in Goldberg's original (1972) instructions. A diagnosis was formulated at the end of the interview according to the International Classification of Disease (ICD), eighth edition, 1968.

The data from the GHQ and the standard psychiatric interviews were processed, using the Statistical Package for the Social Sciences (SPSS) (Nie *et al*, 1975) version SPSSH release 8.01 running on a DEC-10 computer at the University of Dundee from October, 1981 onwards. A factor analysis of the 60

GHQ items for 1488 subjects was carried out, using the principal factor with iteration method and vari-max rotation.

Results

Table I shows marital status, social class, mean age and mean GHQ scores for the whole population sample and for the two interview groups. Mean CIS scores are shown for the two interview groups.

Factor analysis of GHQ items

An initial factor analysis of the 60 GHQ items for 1488 subjects isolated 11 factors, before the eigenvalue fell below 1.0; these factors accounted for a total of 60.9 per cent of the variance. Whilst good separation was achieved using them, the clinical usefulness of these factors was limited as many accounted for a very small percentage of the variance. The finding of separate factors for insomnia (factor iv), appearance (factor vii), head symptoms (factor viii) and heat changes (factor x) was of little value, particularly as both factors v and vi reflected depressive symptom-atology, and factors i, ii and iii a more general debility.

A second analysis was run, limiting the number of factors to three. The results are shown in Table II, and suggest a debility factor, a depression factor and a somatic factor, although the separation is far from good. Each GHQ item was then allocated to the factor for which it obtained the highest rotational value, and the subjects were classified into factor groups as follows:

Debility (Group 1): A score of 9 or more on the 32 items for factor i.

Depression (Group 2): A score of 5 or more on the 17 items for factor ii.

Somatic (Group 3): A score of 3 or more on the 11 items for factor iii.

TABLE I
General sample characteristics

Variable	Group A		Group B		Whole sample	
Marital status:	n	(%)	n	(%)	n	(%)
Married	144	(75.8)	142	(74.4)	1110	(73.5)
Widowed	8	(4.2)	10	(5.2)	66	(4.4)
Divorced	11	(5.8)	9	(4.7)	59	(3.9)
Separated	9	(4.7)	11	(5.8)	59	(3.9)
Single	18	(9.5)	19	(9.9)	217	(14.3)
Total	190	(100.0)	191	(100.0)	1511	(100.0)
Social class:	n	(%)	n	(%)	n	(%)
I	10	(5.7)	10	(5.5)	95	(6.9)
II	31	(17.6)	31	(17.0)	248	(18.1)
III	78	(44.3)	82	(45.1)	657	(48.0)
IV	36	(20.5)	41	(22.5)	206	(15.1)
V	8	(4.5)	7	(3.9)	75	(5.5)
Not applicable	13	(7.4)	11	(6.0)	88	(6.4)
Total	176	(100.0)	182	(100.0)	1369	(100.0)
Age:						
Number	191		191		1514	
Mean	38.2		37.9		38.8	
Standard deviation	11.8		11.9		12.5	
GHQ score:						
Number	189		191		1488	
Mean	24.3		2.4		9.6	
CIS score:						
Number	191		191		382	
Mean	12.7		3.9		8.3	

Missing cases: Marital status 6; Social Class 148; Age 3; GHQ score 29; CIS score 0.
Matched variables: Groups A and B were matched for marital status, social class and age.
GHQ and CIS score distributions are strongly skewed to low scores. Parametric measures and tests are inappropriate.

TABLE II

Factor analysis of the GHQ with 3 factors

Factor I: 'Debility', 32 items, 19 with a rotational value (RV) ≥ 0.5
: Eigenvalue 21.18, Variance 82.3%

RV	Item	RV	Item
0.64	42. Enjoying things?	0.52	41. Life a struggle?
0.61	23. No interest?	0.50	44. Feeling bad tempered?
0.60	28. Doing well?	0.50	29. More often late?
0.59	37. Making a start?		————————————
0.58	30. Feeling satisfied?	0.49	36. Able to make decisions?
0.58	16. Enough energy?	0.48	55. Feeling nervous?
0.58	3. Run down?	0.48	35. Feeling useful?
0.57	47. Everything on top of you?	0.47	43. Taking things hard?
0.56	1. Feeling well?	0.46	21. Keeping busy?
0.56	22. Taking longer?	0.44	4. Feeling ill?
0.55	2. Needing a tonic?	0.43	45. Feeling panicky?
0.55	15. Feeling alert?	0.42	31. Able to show affection?
0.55	7. Able to concentrate?	0.42	24. Careful about appearance?
0.54	38. Dreading things?	0.40	33. Able to chat?
0.53	39. Things a strain?	0.39	25. Taking care over clothing?
0.53	49. Feeling unhappy?	0.38	26. Getting out?
		0.35	13. Feeling tired?

Factor II: 'Depression', 17 items, 12 with a rotational value (RV) ≥ 0.5
: Eigenvalue 2.87, Variance 11.1%

RV	Item	RV	Item
0.75	56. Life not worth living?	0.52	46. Facing up to things?
0.74	59. Wish to die?	0.51	54. Feeling happy?
0.71	42. Feeling hopeless?	0.51	50. Feeling confident?
0.70	57. Thinking about death?		————————————
0.70	60. Thinking of suicide?	0.42	27. Managing?
0.61	51. Feeling worthless?	0.42	32. Getting out?
0.59	53. Feeling hopeful?	0.41	48. People looking at you?
0.54	58. Nerves too bad?	0.36	34. Feeling afraid?
0.54	40. Can't overcome things?	0.29	19. Bad dreams?

Factor III: 'Somatic', 11 items, 5 with a rotational value (RV) ≥ 0.5
: Eigenvalue 1.70, Variance 6.6%

RV	Item	RV	Item
0.67	20. Disturbed nights?	0.41	10. Sweating?
0.67	18. Difficulty staying asleep?	0.39	5. Head pains?
0.66	11. Waking early?	0.38	9. Hot and cold?
0.58	12. Waking unrefreshed?	0.37	6. Tight head?
0.57	17. Difficulty getting off to sleep?	0.27	8. Fear of collapse?
————————————			
0.43	14. Worry and insomnia?		

In an attempt to improve the sensitivity of these factors, a second classification, based only on those GHQ items with a rotational value of 0.5 or more, was also employed:

Debility (Group 4): A score of 5 or more on the 19 items for factor i.

Depression (Group 5): A score of 4 or more on the 12 items for factor ii.

Somatic (Group 6): A score of 2 or more on the 5 items for factor iii.

It can be seen that to qualify for a group, more than 25 per cent of the GHQ items for that group must be positive. This threshold is arbitary, and lowering it would increase specificity at the expense of sensitivity, while raising it would have the opposite effect. The correlations between the scores for these six factor groups with the total GHQ score, the total CIS score, and the overall severity score are shown in Table III.

Validation study

Three measures of psychiatric morbidity, obtained from the standardized psychiatric interviews, were

used in the validation study. These were: the total CIS score, the overall severity rating and the clinical diagnosis. Three separate criteria were therefore used to define a 'case':

1. Case = Psychiatric diagnosis.
   Non-case = No psychiatric diagnosis.
2. Case = Severity score of 2 to 4.
   Non-case = Severity of 0 to 1.
3. Case = CIS score of 13 or more.
   Non-case = CIS score of 12 or less.

On the 60-item GHQ, a probable case is identified by a score of 12 or more. The effectiveness of the total GHQ score and its factor grouping in the identification of cases were calculated in terms of sensitivity and specificity. Sensitivity is defined as the proportion of defined cases which were also predicted, and specificity as the proportion of defined non-cases which were also predicted. Ideally, both measures should approach 100 per cent, although the sensitivity value is generally considered more important. The sensitivity

and specificity value and proportion of misclassified cases for the full-scale GHQ, in relation to the three measures of psychiatric morbidity obtained from the interview, are shown in Table IV. These figures relate to the equal number of high and low GHQ scorers interviewed, and should be compared with those corrected for the whole sample (Table VII and below). Table V shows the sensitivity and specificity values for the six groups identified in the factor analysis study, in relation to the three case definitions from the psychiatric interviews. In Table VI, the sensitivity values for the overall GHQ score as well as the six groups obtained on factor analysis are shown in relation to individual diagnoses obtained at the interview.

Corrected prevalence rates

In the total sample of 1488, 452 women scored 12 or more on the 60-item GHQ. This gives an estimated prevalence of 30.4 per cent. In Table VII a corrected prevalence is shown, along with the modified speci-

TABLE III

Spearman Rank Correlation Coefficients

Variable								
1. Total GHQ score	–							
2. Debility score (Group 1)	.96							
3. Debility score (Group 4)	.95	.98						
4. Depression score (Group 2)	.80	.77	.75					
5. Depression score (Group 5)	.78	.77	.75	.94				
6. Somatic score (Group 3)	.81	.67	.66	.53	.53			
7. Somatic score (Group 6)	.72	.61	.61	.49	.48	.86		
8. Total CIS score	.72	.71	.70	.67	.69	.57	.51	
9. Severity score	.73	.72	.71	.69	.69	.59	.52	.86
	1.	2.	3.	4.	5.	6.	7.	8.

All the correlation coefficients are highly significant: $P < 0.001$.

TABLE IV

Sensitivity and specificity of the GHQ (see also Table VII)

GHQ score	Case definition					
	Diagnosis		Severity score		CIS score	
	Non-case	Case	Non-case	Case	Non-case	Case
Non-case	158	32	179	12	186	5
Case	16	161	55	136	81	110
Chi-squared =	213.4		166.9		134.6	
Sensitivity =	84.4%		91.9%		95.7%	
Specificity =	90.8%		76.5%		69.7%	
Misclassified =	13.1%		17.5%		22.5%	

All the chi-squared values are highly significant: $P < 0.0000$.

TABLE V
Case estimation using GHQ factors

	Case definition					
	Sensitivity %			Specificity %		
GHQ factor	Diagnosis	Severity	CIS score	Diagnosis	Severity	CIS score
Total GHQ score:						
11/12 threshold	83.4	91.9	95.7	90.8	76.5	69.7
Debility factor:						
Group 1	75.7	83.1	89.6	92.5	79.5	74.5
Group 4	80.3	87.2	88.7	88.5	75.2	68.2
Depression factor:						
Group 2	34.2	43.9	52.2	98.9	95.3	94.0
Group 5	35.2	45.3	53.0	98.3	94.5	92.9
Somatic factor:						
Group 3	62.7	71.0	73.9	89.1	80.3	75.3
Group 6	54.4	61.5	67.0	89.7	82.9	79.8

TABLE VI
Diagnosis estimation using GHQ factors

	Diagnosis			
	Specificity %		Sensitivity %	
	None	Anxiety neurosis	Depressive neurosis	Depressive psychosis
Total GHQ score:				
Non-cases	90.8	—	—	—
Factor Groups 1, 2, 3:				
Non-cases	79.3	—	—	—
Factor Groups 4, 5, 6:				
Non-cases	83.9	—	—	—
Total GHQ score:				
Cases	—	81.7	86.6	93.3
Debility factor:				
Group 1	—	74.7	76.3	80.0
Group 4	—	84.5	77.3	86.7
Depression factor:				
Group 2	—	18.3	44.3	53.3
Group 5	—	19.7	46.4	60.0
Somatic factor:				
Group 3	—	62.0	65.0	85.7
Group 6	—	49.3	56.7	85.7
Total number (100%)	174	71	97	15

A variety of other diagnostic groups accounted for a further 25 cases.

TABLE VII

The GHQ in prevalence estimations

Whole sample by GHQ score		Interviewed sample by CIS score			Projection for whole sample by CIS score		
		Non-case	Case	Total	Non-case	Case	Total
Non-case	1036	186	5	191	1009	27	1036
Case	452	81	110	191	192	260	452
Total	1488	267	115	382	1201	287	1488
Sensitivity =		95.7%			90.6%		
Specificity =		69.7%			84.0%		
Estimated Prevalence =		—			30.4%		
True Prevalence =		—			19.3%		

The GHQ score as an estimator of 'true cases' projected for the whole sample

Case definition:	Diagnosis	Severity score	CIS score
Sensitivity (%)	70.3	83.2	90.6
Specificity (%)	95.5	88.2	84.0
Estimated Prevalence (%)	30.4	30.4	30.4
True Prevalence (%)	39.3	26.0	19.3

ficity and sensitivity values for the 60-item GHQ, using the three correction factors obtained from the three types of case definition at interview.

Discussion

The use of the GHQ in community surveys of psychiatric morbidity where the proportion of high scores is low has been questioned recently (Benjamin *et al*, 1982). In the original description of the development of the GHQ, Goldberg (1972) noted that it was most effective in case identification in the middle range of prevalence and that in order to calculate the true prevalence of psychiatric disorder from the proportion of high scores on the GHQ at high and low prevalence levels. a correction factor would be need to be applied. This factor could be calculated according to the sensitivity and specificity of the questionnaire for that population, which could be assessed by interviewing a sample of the population to find the proportion of false positives and false negatives identified by the GHQ.

Tarnopolsky *et al* (1979) carried out a validation exercise on the 30-item GHQ in a community survey and, from standardized interviews carried out on a proportion of those subjects obtaining high scores and low scores on the GHQ, predicted that its sensitivity would fall as low as 54 per cent where the proportion of high scores was only 22 per cent. Benjamin *et al* (1982), using the 60-item scale in a population of 92 women between the ages of 40 and 49 years, calculated that the sensitivity of the GHQ was only 54.5 per cent, when 25 per cent of the subjects scored 12 or more.

They concluded that the GHQ is unsuitable as a screening instrument for mental illness in the community, but noted that factor analysis yielded a 15-item GHQ factor which, when used with a more complex scoring system, reduced the level of misclassification; this was contrary to the findings of Goldberg and Hillier (1979).

In the present study, a much larger number of women with a greater age range completed the GHQ, and a sample of high scoring and low scoring subjects were interviewed, as suggested by Tarnopolsky *et al* (1979). As predicted, the sensitivity and specificity of the GHQ were consistently high, in comparison with all methods of defining a case at interview, where the proportion of high scores was 50 per cent (Table IV). When these values were corrected for the proportion of GHQ high scores in the total sample (30.4 per cent), the lowest value calculated for sensitivity was 70.3 per cent and for specificity 84 per cent (Table VII).

In this study, sensitivity was lowest where a 'case' was defined by the application of a diagnostic label (Table VII). This method of case definition would tend to identify individuals with chronic illnesses who were in remission and who would, therefore, only score one on the overall severity rating. This is consistent with Goldberg's original comments (1972) that the GHQ detected current emotional disturbance, and would miss patients with chronic illness who were well at the time of completing the questionnaire. In deciding on the use of self-rating questionnaires in the community, it is clearly important to decide if it is current disturbance which is considered relevant, or whether the

aim is to identify all types of psychiatric disorder—active or in remission—in which case a different type of screening instrument would be indicated.

In the factor analysis of the GHQ, an attempt was made to identify factors which might be more sensitive than the overall GHQ score in the identification of cases or factors which might relate closely to specific diagnosis in the subjects interviewed. The first method of factor analysis yielded eleven factors, only one of which contributed to a major part of the variance, and many of the factors seemed clinically irrelevant. Other factor analytic studies have obtained between five and eleven significant factors, but few of these account for significant amounts of variance (Goldberg et al, 1976; Worsley et al, 1978; Goldberg and Hillier, 1979; Benjamin et al, 1982). These minor factors are likely to be highly sample- and technique-dependent. The second analysis was restricted to three factors, and the first (Groups 1 and 4) contained many general items suggesting 'failure to cope'. This factor was labelled 'debility', and its existence supports Goldberg's view that the largest group of patients identified by the GHQ include those with mild depression and anxiety states who would have been labelled 'neurasthenic' in the past. This factor correlated highly with the total GHQ score (Table III), but was no better than the total GHQ score in overall case identification (Table V). However, it shows a slightly higher sensitivity than the total GHQ score in relation to anxiety neurosis (Table VI).

The second factor (Groups 2 and 4), contained items relating mainly to depression, and the third factor (Groups 3 and 6) items relating mainly to insomnia and somatic complaints. Neither of these factors was more sensitive than the total GHQ score in relation to overall case identification (Table V) or to the identification of a specific diagnostic category (Table VI).

This study would indicate that the GHQ is a valid instrument for detecting the presence of current psychiatric disturbance in a general practice list population of women. However, it further emphasizes the need to validate the GHQ for each new sample. Factor analysis does not appear useful in enhancing the value of the GHQ, and neither the total GHQ score nor the score on any of the factors isolated in this study are particularly useful in identifying specific psychiatric illnesses. However, the GHQ was not designed for this purpose, for which there are other screening instruments which are more appropriate.

References

BALLINGER, C. B. (1975) Psychiatric morbidity and the menopause; screening of a general population sample. British Medical Journal, iii, 344–6.

—— (1977) Psychiatric morbidity and the menopause: Survey of a gynaecological out-patient clinic. British Journal of Psychiatry, 131, 83–9.

BENJAMIN, S., DECALMER, P. & HARAN, D. (1982) Community screening for mental illness: A validity study of the General Health Questionnaire. British Journal of Psychiatry, 140, 174–80.

GOLDBERG, D. P. (1972) The Detection of Psychiatric Illness by Questionnaire: A Technique for the Identification as Assessment of Non-Psychotic Psychiatric Illness. London: Oxford University Press.

—— & BLACKWELL, B. (1970) Psychiatric illness in general practice: A detailed study using a new method of case identification. British Medical Journal, ii, 439–43.

—— COOPER, B., EASTWOOD, M. R., KEDWARD, H. B. & SHEPHERD, M. (1970) A standardised psychiatric interview for use in community surveys. British Journal of Preventive and Social Medicine, 24, 18–23.

—— KAY, C. & THOMPSON, L. (1976) Psychiatric morbidity in general practice and the community. Psychological Medicine, 6, 565–9.

—— RICKELS, K., DOWNING, R. & HESBACHER, P. (1976) A comparison of two psychiatric screening tests. British Journal of Psychiatry, 129, 61–7.

—— & HILLIER, V. F. (1979) A scaled version of the General Health Questionnaire. Psychological Medicine, 2, 139–45.

MACGUIRE, G. P., JULIER, D. L., HAWTON, K. E. & BANCROFT, J. H. J. (1974) Psychiatric morbidity and referral on two general medical wards. British Medical Journal, i, 268–70.

NIE, N. H., HULL, C. H., JENKINS, J. G., STEINBRENNER, K. & BENT, D. H. (1975) SPSS: Statistical Package for the Social Sciences (2nd edition). New York: McGraw-Hill.

TARNOPOLSKY, A., HAND, D. J., MCLEAN, E. K., ROBERTS, H. & WIGGINS, R. D. (1979) Validity and uses of a screening questionnaire (GHQ) in the community. British Journal of Psychiatry, 134, 508–15.

WORSLEY, A., WALTERS, W. A. W. & WOOD, E. C. (1977) Screening for psychological disturbance amongst gynaecology patients. *Australian and New Zealand Journal of Obstetrics and Gynaecology*, **17**, 214–19.

—— —— —— (1978) Responses of Australian patients with gynaecological disorders to the General Health Questionnaire: A factor analytic study. *Psychological Medicine*, **8**, 131–8.

*Paul R. Hobbs, M.B., Ch.B., M.R.C.Psych., *Senior Registrar, Dundee Psychiatric Service, Royal Dundee Liff Hospital, Dundee DD2 5NF, Scotland*

C. Barbara Ballinger, B.Sc., M.B., Ch.B., M.R.C.P., M.R.C.Psych., *Senior Lecturer, Department of Psychiatry, University of Dundee, Ninewells Hospital and Medical School, Dundee DD1 9SY, Scotland*

Anne H. W. Smith, M.B., Ch.B., M.R.C.Psych., *Consultant Psychiatrist, Dundee Psychiatric Service, Royal Dundee Liff Hospital*

* Reprints.