Performance of the GHQ-12 in relation to current and lifetime CIDI psychiatric diagnoses

GHQ-12 IN RELATION TO CIDI DIAGNOSES

Jorge J Caraveo-Anduaga*
Nora A Martínez*
Gabriela Saldívar*
Jorge L López J*
Ma. Teresa Saltijeral*

Summary

Background. The validity of the General Health Questionnaire (12 item version) has been extensively studied recently as a case detector for current mental illness in general medical settings. Consideration not only of the total score but also of the item contents of the responses, that include negative and positive aspects of mental health, and the study of the factor structure of the instrument in diverse cultures is focused.

Methods. Validity and thorough analysis of the performance of the GHQ-12 was done using the Composite International Diagnostic Interview as external clinical criterion, permitting caseness to be defined in accordance with two diagnostic systems, both for current and lifetime disorders. Performance of each item was illustrated in ROC curves and measured in terms of odds ratios; finally, factor-structures were obtained.

Results. Validity was similar to recent reports. Performance of each item showed that the first four items were only significantly associated with current disorders. Factor analysis with varimax rotation identified 3 factors explaining 59% of the common variance for current disorders, while for lifetime disorders, 2 factors explained 45% of the variance.

Conclusions. Results suggest a continuous measure within the structure of the instrument with three components: a common way of expressing distress; a possible vulnerability core, and a third component that seems to be more related with either previous or current psychiatric disorders, having practical implications for further research and epidemiological surveillance.

Key words: Mental disorders, epidemiology, primary care, screening, GHQ, CIDI.

Resumen

Antecedentes. La validez del Cuestionario General de Salud (versión de 12 reactivos), CGS-12, como instrumento útil para detectar los probables casos en la práctica médica general, ha sido recientemente objeto de un estudio muy extenso en diferentes países. Sin embargo, la puntuación general es sólo un aspecto de las cualidades del instrumento y se ha sugerido que debieran estudiarse las respuestas específicas a los reactivos, en virtud de que en ellos se consideran aspectos tanto positivos como negativos en relación con la salud mental, así como explorar la estructura factorial del instrumento en diferentes culturas.

Método. En este trabajo se estudió la validez del CGS-12 utilizándolo como criterio externo de la Entrevista Internacional Diagnóstica Compuesta (CIDI), permitiendo que la definición de caso se hiciera con base en dos sistemas de clasificación, y la información se analizara no sólo en relación con la psicopatología presente en el momento de la entrevista, sino también en relación con la existencia de trastornos psiquiátricos alguna vez en la vida. Este abordaje constituye una aproximación diferente y original respecto de los otros estudios. El funcionamiento de cada reactivo se ilustró por medio de gráficas ROC, y su asociación se midió en términos del *odds ratio*; finalmente, se obtuvieron las estructuras factoriales.

Resultados. Las coeficientes de validez para el instrumento en general fueron similares a los obtenidos en otros estudios. El análisis del funcionamiento de cada reactivo mostró que los primeros cuatro sólo se asociaron significativamente con la presencia de psicopatología al momento del estudio, mientras que los ocho reactivos restantes se asociaron con la presencia de trastornos en la vida. El análisis factorial con rotación varimax identificó tres factores que explican el 59% de la varianza para los casos actuales, mientras que para los trastornos en la vida, dos factores explicaron el 45% de la varianza.

Conclusiones. Los resultados obtenidos sugieren la existencia de un gradiente de severidad en el contenido del instrumento que se manifiesta por medio de tres componentes: una forma común de expresión de malestar; un posible factor de vulnerabilidad; y un tercer grupo de síntomas que parece estar más relacionado tanto con la presencia de psicopatología actual como con trastornos presentados alguna vez en la vida. Los hallazgos sugieren implicaciones prácticas tanto para la vigilancia epidemiológica de la salud mental como para su aplicación en otros estudios epidemiológicos.

Palabras clave: Trastornos mentales, epidemiología, atención primaria, tamizaje, CGS, CIDI.

* Division of Epidemiological and Social Research. Instituto Mexicano de Psiquiatría. Calz. México-Xochimilco # 101, San Lorenzo Huipulco, 14370, México, D.F.

Introduction

The General Health Questionnaire (10) is one of the most widely used screening self-administered instruments for detecting psychiatric disorders and for assessing mental distress. Validity studies have been performed in different countries and for both clinical and general populations (13). Threshold for case definition not only varies with the setting (34) but also seems to be related with the clinical instrument and diagnostic criteria which are used to evaluate the concurrent validity.

More accurate structured diagnostic instruments for epidemiological and clinical research, such as the Diagnostic Interview Schedule (DIS) (30), the Composite International Diagnostic Interview (CIDI) (31) and the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) (40) have been developed and used recently (20), incorporating computer algorithms for different diagnostic criteria such as the Diagnostic and Statistical Manual of the American Psychiatric Association, 3rd edition revised (DSM-III-R) and the International Classification of Diseases, 10th edition (ICD-10). Unlike the Clinical Interview Schedule (CIS) (9) or the Present State Examination (PSE) (38) the CIDI allows the study of current and lifetime prevalence for several psychiatric disorders, thus favoring another approach to study the performance of screening instruments such as the GHQ.

As epidemiological studies have become more comprehensive incorporating many different variables, there is a need for short instruments that can accurately measure the degree of mental distress in the population (19). A recent World Health Organization study in 15 centers around the world communicated some of the psychometric properties of the GHQ-12 to detect cases of current mental illness in general medical settings, showing the value of this brief version of the instrument (14). Consideration not only of the total score obtained in the GHQ but also of the item contents and the study of the factor structure of the instrument in diverse cultures has been emphasized, since these studies may show differences in the way symptoms cluster in different groups of patients and in the general population (39,36,22).

Dimensionally, the GHQ-12 includes questions about positive and negative aspects of mental health. Positive mental health items ask about normal functioning and contribute to the score when answered "less than usual" while the negative mental health items contribute to the score when answered "more than usual". In this sense, these latter represent symptoms, while the first are related to behavioral and perceived social functioning. Lewis (23) signaled the need to investigate the concepts of positive and negative mental health for the interpretation of GHQ factor studies as well as their relevance and potential utility in community and clinical settings. Furthermore, Lewis and Araya (24) found a higher mean on the positive scale in a sample of British primary care attendees, while the opposite occurred in a Chilean sample, thus suggesting an ascertainment bias based on culture, although further research is needed.

The objectives of this paper are: 1) to study the performance of the GHQ-12 as a screening instrument in a primary care clinic and compare it with a highly structured diagnostic interview, the CIDI; 2) to obtain the odds ratio for each GHQ-12 item in accordance with two diagnostic classification systems: ICD-10 and DSM-III-R; 3) to identify groups of items which better indicate the presence of a disorder, and, 4) based on these analyses, to confirm or reject the bias assessment hypothesis about the questionnaire.

Method

The study was performed at a Health Center located in the southwest of Mexico City, using a two-stage design. The target population were adults between 18 and 64 years of age, living in the geographical area served by the primary care facility (PHC), and who attended it during an 8 week period (March-May, 1992). In the first stage, one out of every two attendants was included in the sample, and for each, the GHQ-12 was administered by research assistants. In the second stage, blind to previous results, a group of clinicians administered the Composite International Diagnostic Interview (CIDI 1.0) to all high score's (≥ 3) on the GHQ-12, and to every tenth low score. All sections from the CIDI were included except tobacco and most of the questions about somatization. Psychiatric patients in current treatment were excluded.

The systematic sampling method was chosen based on the average size of the population registered at the PHC, and taking into account the availability of research personnel, both for the first and second phase of the study. Also, for the second stage, we had to consider the space and time consumption to perform the clinical interviews. Time was estimated between 45 minutes and one hour and a half.

Instruments: The GHQ-12 is the shortest version of the GHQ (10). Validity studies of the GHQ-60 have been performed on different population samples in Mexico (4,28,27,29,32). Even though the questionnaire is self-administered, it has been reported that in subjects with low schooling it is best to use trained interviewers (5), as we did in this study.

The CIDI is a highly structured diagnostic interview designed for application to different cultures, by lay or clinical interviewers trained in the use of the instrument (31). Five clinicians, two psychiatrists and three clinical psychologists, received a one week training program following the same procedures as at Washington University, where one of us, J.C, was trained. High inter-ratter reliability, mean ICC = 0.87, was accomplished at the end of this phase and supervision was close throughout the field work.

The CIDI used was the 1.0 core version. A Spanish translation prepared in Chile was obtained from Dr. Vicente et. al.* and carefully reviewed comparing it with the authorized English version. Only some words were changed in order to adapt it to the common use of Spanish in Mexico. The diagnostic computer program

Table 1
Sex and age distribution of the study population

Attended Population	N	%	Mean Age	s.d.*
Males	123	12.5	35.2	13.1
Females	856	87.4	33.1	12.7
Total	979	100	33.3	12.8
Sample	N	%	Mean Age	s.d*
Males	64	13.1	32.4	12.3
Females	424	86.8	32.9	12.4
Total	488	100	32.9	12.3

*Standard deviation.

Table 2
Sociodemographic characteristics[1]

	N	%
Sex		
Male	58	12.6
Female	403	87.4
Age		
18-25	174	37.7
26-35	125	27.1
36-45	73	15.7
46-55	61	13.3
56-64	28	6.0
Marital Status		
Single	130	28.2
Married/Common-law	276	59.9
Separated/Divorced/Widowed	55	11.9
Education		
None	47	10.2
Primary	234	50.7
Jr. High School	103	22.3
High School	63	13.4
University	14	3.0
Occupation		
Professional/Technician	6	1.3
Employee	28	6.1
Qualified/Not qualified Workers	162	35.2
Under employee	14	3.0
Housewife	215	46.6
Student	18	3.9
Other	18	3.9
*Monthly Income**		
One minimal salary or less	304	65.9
Two minimal salaries	34	7.4
More than two minimal salaries	29	6.3
Not Applicable	90	19.5
Unknown	4	0.9

*Monthly minimal income in these data equivalent to 143 U.S. dollars
[1]First-Stage

for the CIDI was provided by WHO authorities. The following diagnoses were included: depression, dysthymia, mania, agoraphobia, social phobia, simple phobia, panic disorders, generalized anxiety, obsessive-compulsive disorder, hypochondria's, organic brain syndrome, substance abuse and dependence.

A 10 weight factor for low score's was employed in the analyses to correct their under-sampling. Validity coefficients were calculated for different cut-off points on the GHQ-12, determine the optimal threshold for case definition, maximizing sensitivity and specificity of the instrument as a whole. Then, odds ratios using the Epi 5 program (7) were obtained of each item as related to ICD-10 and DSM-III-R diagnostic criteria, both for current and lifetime cases. Finally, the structure of the instrument was studied using factor analysis with varimax rotation.

Results

During the 8 week field period, 979 adult subjects were attended at the PHC, from which 488 were selected for the study. Table 1 shows the attending population and the selected sample. There were no significant statistical differences between the attending population and the sample with regard to age (t = 1.03, df 976, p = ns) and gender (t = .34, df 976, p = ns), nor to age by gender (t = 2.47, df 123, p = ns; t = .18, df 851, p = ns).

Further, 27 subjects were excluded since 12 were psychiatric patients in current treatment, and the rest were not located or attended the facility for other reasons. The 27 excluded subjects did not differ from the final sample in terms of age (t = 1.55, df 485, p = ns) and sex (t = 1.44, df 485, p = ns).

Sociodemographic characteristics of 461 patients, none of which refused to participate in the first stage of the study are shown in Table 2.

There was a clear female predominance, representing 87% of the attendants; mean age for both males and females was 32 years of age (s.d. 12.4), 60% were married or living in common-law, and most females were housewives; 51% of the sample had only elementary school level and 66% reported a monthly income equivalent to 143 US dollars or less.

Twenty-six low-score's on the GHQ refused the clinical interview, and were replaced with the next person on the list. Refusals and replacements were compared by sex, age, marital status, education, occupation and monthly income; none were significantly different, assuring that bias in the selection was not introduced by this maneuver.

Also, 16 high-score's on the GHQ refused the second phase. Comparison by sociodemographic characteristics and by GHQ-12 score did not reveal statistically significant differences between refusals and subjects interviewed.

Finally, a total of 435 subjects was included in the analyses, the weighted number of low-score's on the GHQ-12 was 270, and 165 high-score's interviewed. The most frequent diagnoses found were depression (including depressive episodes and dysthymia) and anxiety disorders. A report on diagnosis prevalence will be presented elsewhere.

The prevalence of each GHQ-12 item showed that items 1,2,3 and 10 were most frequently reported, with just a few differences between females and males; the

3

Table 3
GHQ-12 Item prevalence by sex[1]

GHQ-12	Males (n = 58) %	Females (n = 377) %	total (n = 435) %
1. Lost sleep (–)	27.6	45.9*	43.4
2. Under strain (–)	56.9*	38.2	40.7
3. Poor concentration (+)	37.9**	24.4	26.2
4. Not useful (+)	6.9	12.5	11.7
5. Not facing problems (+)	5.2	6.9	6.7
6. Not capable of decisions (+)	5.2	5.3	5.3
7. Burdened by difficulties (–)	10.3	8.8	9.0
8. Unhappy (+)	17.2	13.5	14.0
9. Not enjoying actives (+)	10.3	8.5	8.7
10. Sad and depressed (–)	44.8	45.4	45.3
11. Losing confidence (–)	8.6	13.3	12.6
12. Worthless (–)	10.3	17	16.1

Mental Health items: (–) Negative and (+) Positive
[1] Weighted Sample
*Tp ≥ .05
**Tp ≥ .001

former reported more sleep disturbances and the latter more tension and difficulties in concentration (Table 3).

Validity of the GHQ-12: The 2/3 cut-off point was found to be the best suited for our study population as expressed in the balance between the coefficient values obtained for sensitivity and specificity (Table 4).

Below this cut-off point, specificity decreased considerably without an improvement in sensitive regarding 'caseness' defined as in DSM-III-R, ICD-10 and total clinical diagnoses. Clinical diagnoses included 16 women with sexual dysfunction's, not diagnosed through the CIDI diagnostic program, as only some items from the somatization section were used during the interview. Interestingly, for lifetime prevalence, specificity slightly varied while the positive predictive value (PPV) incre-

ased. Performance of the instrument increasing the cut-off point (Figure 1) clearly showed a considerable drop in sensitivity, reaching values below 49% for lifetime disorders.

Items performance: The performance of each GHQ-12 item is illustrated in ROC curves providing a description of disease delectability for current and lifetime disorders according to the diagnostic systems. In current disorders (Figure 2) the ROC curves show how the first three GHQ-12 items along with item 10 are responsible for the highest screening performance in terms of sensitivity, while the rest are more specific. Among the latter, items 11 and 12 are in the upper range while items 5 to 7 are the most specific. The same pattern is observed for lifetime disorders (Figure 3),

Table 4
GHQ-12 Validity coefficients

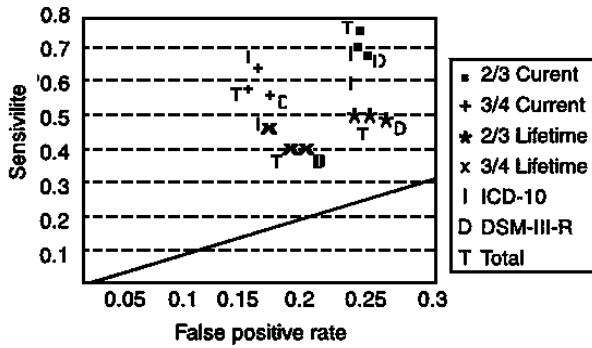| Validity coefficients	ICD-10		DSM-III-R		Total	
	C*	LT**	C*	LT**	C*	LT**
Sensitivity	69	55	67	48	73	49
Specificity	75	74	73	71	75	73
False Positive Rate	25	26	27	29	25	27
False Negative Rate	31	45	33	52	27	51
P.P.V.	54	59	48	57	50	64
N.P.V.	85	70	85	63	89	59
Prevalence	30	40	28	45	26	49
Misclassification rate	27	34	29	39	26	39

* Current
** Life Time

Figure 1. Performance of the GHQ-12 for current and lifetime disorders using two cut-off points.

although with a slight decrease in sensitivity of the more specific items, while the position in the ROC curve of the more sensitive items denote an increase in the false positive rate.

Additionally, strength of the association expressed in terms of the odds ratio (Tables 5 & 6), shows that the first four items from the GHQ-12, the first two in the negative branch and the other two in the positive branch of mental health, seem to discriminate patients with
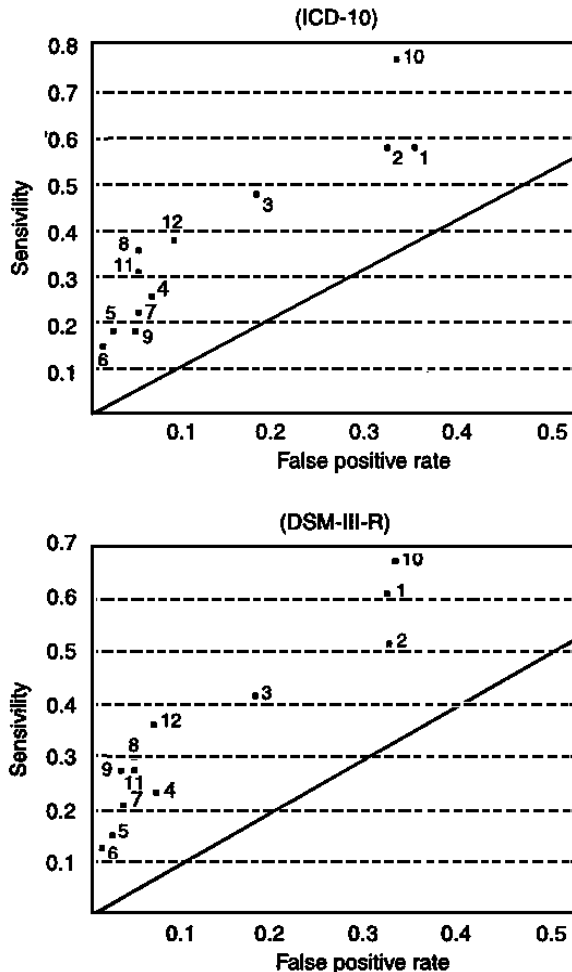

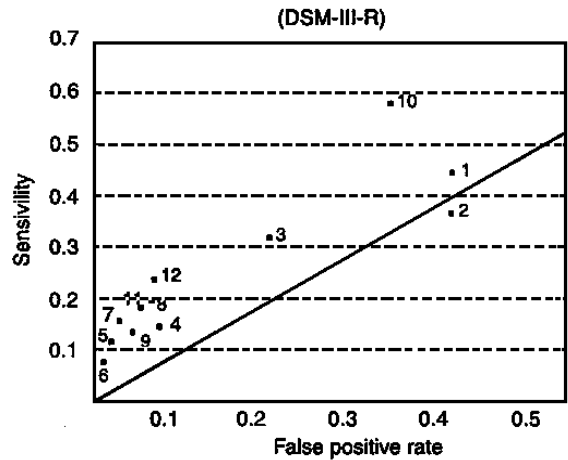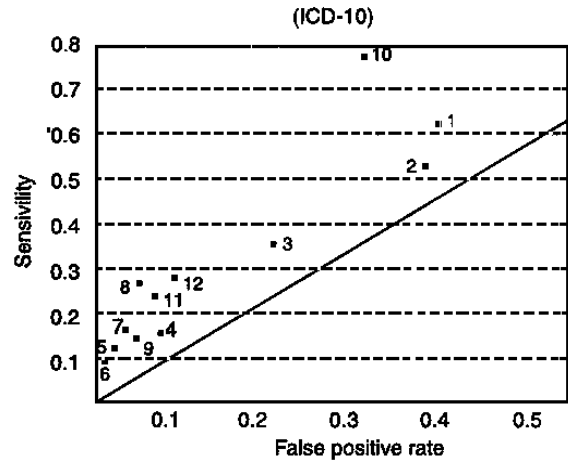**Figure 2.** Performance of the GHQ-12 in current disorders (ICD-10).

Figure 3. Performance of the GHQ-12 in lifetime disorders (ICD-10).

current disorders from those with lifetime prevalence, as in the latter the odds ratios dropped losing or showing only marginal statistical significance (Table 6).

The difference is more evident with the DSM-III-R diagnostic criteria than with the ICD-10. Items 5 and 6, both in the positive branch of mental health, showed the widest variability, although with high odds ratios. Item 7, from the negative branch, showed slight differences in diagnostic classifications, and in both current and lifetime disorders. It is worth to note that odds ratios for items 5 to 7 were among the highest for lifetime disorders, and clearly in accord with the DSM-III-R (Table 6). Item 8, from the positive branch, and item 10 from the negative, both equivalent in contents, were found with higher odds ratios in the ICD-10 than in the DSM-III-R criteria. This could be explained because ICD-10 takes into account mild depressive episodes. Finally, items 11 and 12, both in the negative branch, showed stronger association with ICD-10 diagnoses.

Factor analysis: The questionnaire can be scored in two basic manners: Likert's and Goldberg's. When using Likert's technique each item is treated as a continuous variable (1,2,3,4) and the factor structure obtained reflects mental distress, while using Goldberg's score

Table 5
Odds ratio in current disorders

Items GHQ	ICD-10	DSM-III-R	TOTAL
Lost sleep	2.39 (1.51 - 3.79)	2.94 (1.86 - 4.65)*	3.24 (2.07 - 5.10)*
Under strain	2.81 (1.77 - 4.47)	2.04 (1.30 - 3.20)	2.31 (1.49 - 3.59)
Poor concentration	4.00 (2.45 - 6.53)	2.96 (1.83 - 4.78)	3.65 (2.26 - 5.88)
Not useful	4.28 (2.25 - 8.16)	3.52 (1.86 - 6.67)	3.74 (1.98 - 7.11)
Not facing problems	6.31 (2.67 - 15.14)	6.80 (2.83 - 16.72)	7.24 (2.94 - 18.40)
Not capable of decisions	9.33 (3.34 - 27.30)	10.94 (3.70 - 34.62)	9.33 (3.34 - 27.30)
Burdened by difficulties	5.52 (2.65 - 11.60)	5.66 (2.69 - 12.00)	5.52 (2.65 - 11.60)
Unhappy	8.60 (4.58 - 16.27)	4.97 (2.73 - 9.10)	8.60 (4.58 - 16.27)
Not enjoying activities	4.62 (2.22 - 9.69)	3.72 (1.80 - 7.73)	4.21 (2.02 - 8.85)
Sad and depressed	6.45 (3.83 - 10.92)	3.75 (2.34 - 6.03)	4.52 (2.83 - 7.24)
Losing confidence	7.46 (3.90 - 14.36)	5.53 (2.94 - 10.48)	5.85 (3.08 - 11.18)
Worthless	6.21 (3.49 - 11.10)	5.96 (3.35 - 10.64)	5.57 (3.14 - 9.95)

* 95% Confidence Interval

the responses become dichotomous (0,0,1,1) and results will show the "unusual" manifestations, or in other words, the morbid expressions on which our interest is focused. Although the former score is recommend for factor analysis, the latter can be employed as long as data variables are not skewed (6). To be certain about this assumption, the mean and median of each item, and of the instrument as a whole, were obtained using Likert scoring method, showing that both measures were of the order of 2 below. The overall mean was 2.395 (s.d. 2.4) and the median was 1.0. Also, Pearson's product-moment correlation between pairs of items using dichotomic scores were reviewed showing a range between 0.20 and 0.52, thus assuring a normal distribution, so that factor analysis using Goldberg scoring method could be performed reliably.

Factor analysis using varimax rotation showed 3 factors explaining 59% of the variance for subjects with current disorders diagnosed by either diagnostic classification. The criterion to include the items in the factors was a loading of .50 or higher. In current disorders, the items grouped in the first factor expressed feelings of sadness and despair, while the items in the second factor conspicuously denoted inhibition expressed as 'not being able' in the contents of the items, and the third factor exclusively signaled a sleep disturbance (Table 7).

For lifetime disorders, 2 factors were obtained explaining 45% of the variance. Conspicuously, in the first factor appeared items 5 to 7, which were observed in the second factor for current disorders, and shown to be more specific in the ROC curves, while the more sensitive items appeared here in the second factor.

Table 6
Odds ration lifetime disorders

Items GHQ	ICD-10	DSM-III-R	TOTAL
Lost sleep	1.63 (1.09 - 2.45	1.15 (0.77 - 1.72)*	1.61 (1.08 - 2.40)
Under strain	1.29 (0.86 - 1.95)	0.86 (0.57 - 1.29)	0.88 (0.59 - 1.31)
Poor concentration	2.11 (1.84 - 3.33)	1.55 (0.99 - 2.44)	1.67 (1.06 - 2.63)
Not useful	2.08 (1.11 - 3.91)	1.46 (0.78 - 2.73)	1.54 (0.82 - 2.90)
Not facing problems	4.16 (1.70 - 10.51)	4.28 (1.69 - 11.28)	4.27 (1.61 - 11.98)
Not capable of decisions	4.42 (1.60 - 12.82)	4.83 (1.65 - 15.16)	4.23 (1.64 - 18.52)
Burdened by difficulties	4.70 (1.93 - 9.22)	4.32 (1.96 - 9.77)	3.80 (1.68 - 8.87)
Unhappy	5.10 (2.67 - 9.81)	2.50 (1.38 - 4.56)	3.72 (1.93 - 7.24)
Not enjoying activities	3.07 (1.45 - 6.55)	2.59 (1.23 - 5.53)	2.73 (1.26 - 6.03)
Sad and depressed	4.84 (3.14 - 7.46)	2.71 (1.80 - 4.09)	4.30 (2.82 - 6.58)
Losing confidence	3.83 (2.01 - 7.37)	2.66 (1.42 - 5.02)	2.56 (1.35 - 4.92)
Worthless	2.92 (1.67 - 5.13)	2.61 (1.49 - 4.60)	2.22 (1.26 - 3.93)

* 95% Confidence Interval

Table 7
Table 7
Factor-structure of the GHQ-12 in current and lifetime
ICD-10 or DSM-III-R DIAGNOSES

Current		Lifetime	
Factor 1		*Factor 1*	
General Dysphoria	35%	*Inhibition*	33.2%
10. Sad and depressed	.79	6. Not capable of decisions	.76
3. Poor concentration	.76	5. Not facing problems	.72
2. Under strain	.73	7. Burdened by difficulties	.67
12. Worthlessness	.62	12. Worthlessness	.56
11. Losing confidence	.58	11. Losing confidence	.52
9. Not enjoying activities	.53	8. Unhappy	.52
Factor 2		*Factor 2*	
Social Functioning	13.4%	*Distress*	11.2%
6. Not capable of decisions	.84	2. Under strain	.71
5. Not facing problems	.74	3. Poor concentration	.69
7. Burdened by difficulties	.77	1. Lost sleep	.65
		10. Sad and depressed	.57
Factor 3			
Sleep Disturbance	9.9%		
1. Lost sleep	.93		

* 95% Confidence Interval

Finally, for comparative purposes, a factor analysis regardless of caseness was performed using Likert's score, showing that also 3 factors were obtained explaining 55.3% of the variance, and with a Cronbach's alpha of 0.82. In the first factor, with the exception of only one, all positive items were gathered, while the second factor included the first three items, those with higher sensitivity, and finally in the third factor, items 7,11 and 12 were included, all of them exploring negative aspects of mental health and with higher specificity with respect to caseness.

Table 8
Factor structure of the GHQ-12 regardless of caseness, using likert scoring

Factor 1	
Social functioning	*35.6%*
6. *Not capable of decisions*	82
8. Unhappy	72
5. Not facing problems	63
4. Not useful	57
9. Not enjoying activities	54
Factor 2	
Distress	*12.4%*
1. Lost sleep	75
2. Under strain	69
3. Poor concentration	66
Factor 3	
Despair	*9.0 %*
11. Losing confidence	83
12. Worthlessness	67
7. Burdened by difficulties	64

Discussion

The sample obtained for this study is representative of the kind of population attended at the PHC facilities in Mexico City, as shown in a previous report (29), where women were also more frequent among the attendants and in the same proportion as in the present report. This has been observed in other studies from Latin America and other countries (18,3,33), although gender differences are not so conspicuous. In Mexico, this finding is explained because other institutions, such as the Mexican Institute of Social Security, provide medical services for workers and their families, while PHC facilities like the one where this study was performed, are for the population with no other affiliation.

The validity of the GHQ-12 in primary care settings in Latin America was previously studied in Brazil by Mari and Williams (26) using the CIS to confirm caseness. The authors reported sensitivity of 85%, specificity of 79%, PPV 82%, and overall misclassification rate of 18%, all of which are superior to the indexes obtained in the present study. However, our results are very similar and consistent with those obtained by previous studies in Mexico using the 60,30 and 28 item versions of the GHQ (28,29,15), where sensitivity and specificity

both ranged between 73% to 79%, and with a report on the validity of the GHQ-12 in a PHC in Chile (1). Mari and William's (25) commented, in a study where the 30 item GHQ version was used obtaining lower validity indexes, sensitivity 75%, specificity 62% and overall misclassification rate 32%, that the best version of the instrument seems to depend on practical considerations, where sociocultural differences as well as the external criteria for validity should be considered. It must be remembered that prevalence has an effect on the predictive power of the test (2). Results by Mari and Williams (26) were obtained in a population with 53% prevalence. In the present study prevalence was 26%; the PPV and the NPV, both of which indicate the utility of the test, would have been in the order of 77% and 71%, respectively, with a 53% prevalence. The PPV increases notably and the NPV shows a significant drop. On the other hand, using the indexes reported in Brazil but with a 26% prevalence, the PPV would be 59% and NPV 94%. Finally, compared to validity coefficients recently reported in 15 centres where the mean prevalence was 24% (14), our results are similar to those at the lower end.

External criteria also affect the prevalence and the validity coefficients. The criteria to define a "case" is less stringent with the CIS than with more highly structured diagnostic instruments such as the CIDI, where the algorithms of two diagnostic classifications are incorporated. Notably, validity coefficients of the GHQ-12 with respect to both classifications showed only very slight differences. Furthermore, the validity coefficients as well as the estimated prevalence for all current disorders in accord to the DSM-III-R were almost identical to results obtained by Gureje & Obikoya (17) using the CIDI in a population of primary care patients in Nigeria.

Factor analysis identified 3 factors explaining 59% of the total variance in patients with current disorders, which is quite similar to results obtained by Goldberg et al. (11) using the complete version of the GHQ where 6 factors explained 53.5% of the variance, and later, using the 28-item version where 4 factors explained 59% (12). Also, results from the present study are slightly superior to a previous report in Mexico, where Medina-Mora et. al. (28) using the best 30 items elicited 6 factors explaining 46.6% of the variance, and with 28 items, 4 factors explained 52.4%. In the same report, these authors pointed out that their results indicate the possibility of obtaining shorter and balanced versions of the GHQ without losing efficiency. Our results confirm this impression.

We should like to emphasize two original features of the present study; the first is related with the CIDI as external criterion and its structure which permits the identification of lifetime prevalence and not only current cases. The second feature is the performance of each item expressed in terms of the odds ratio, both for current and lifetime disorders, that as far as we are aware, has not and been previously studied. The odds ratio is used in cross-sectional studies to measure the strength of the relationship between a suspected risk factor (int this case detection factors) and a mental disorder. The odds for a group with a suspected risk factor is

defined as the likelihood of having the disorder divided by the likelihood of not having it. An odds ratio is the ratio of the odds for the risk group divided by the odds of the comparison group. If these two odds values are equal, the odds ratio will be 1.0, the null value, which additionally should not be contained within the interval confidence limits in order to consider the odds ratio significant (21). At this point, the GHQ was originally only considered useful as a screening instrument for current cases. Results from this study indicate that the GHQ-12 is also useful for detecting lifetime disorders, without changing the usual score, specially through the report of the last 8 items from the GHQ-12 version, while the first four items showed significant association only with current disorders (Table 5).

Based on these results some questions arose: first, what is the role of these items with respect to current and lifetime caseness? William's et al. (37) proposed a system of caseness based on three major types of axes unlikely to be independent: symptomatology, personality and social functioning. The first and third are openly assessed by the questionnaire while personality disorders are not aimed to be detected by the GHQ. Nevertheless, as these latter disorders may or may not be accompanied by intermittent or persistent dysthymic symptoms, and in this way manifested in terms of an individual's behavior and social functioning, some manifestations could appear as latent traits. From our study, factor analysis detected a subgroup, items 5 to 7, that remained together and specific for lifetime and current disorders. For lifetime prevalence, the factor in which the items were included explained 33.2% of the variance, while for current disorders it explained 13.4%. In other words, these results suggest that responses to this subgroup of items seem to indicate a more persistent state. If so, could it be that the positive response to these items are the consequence of having a psychiatric disorder? or, do they represent psychological vulnerability, or both? On this issue, what evidence is available from previous studies, and what is new from this report?

Vázquez-Barquero et al. (36) compared the factor structures of the different versions of the GHQ. Six general factors emerged from the complete 60 item version: "social functioning/optimism, general dysphoria, general illness, negativism, sleep disturbance and depressive thoughts". Items from the 12 item version appeared distributed among the first two and in the fourth factor. Further, only two factors were derived from the disembodied GHQ-12: "general dysphoria" where items from negativism were incorporated, and "social functioning". The relationship between the factor scores and PSE assessments to define caseness indicated that "sleep disturbance" was a good discriminator as well as "general dysphoria", while "social functioning" showed a moderate level of discrimination. In our study, three factors were obtained for those with current disorders, two of them with similar structure as the results from the above mentioned study, and the presence of sleep disturbance as another relevant semiological issue only explored by one item in the GHQ-12, and also, a highly sensitive indicator of current cases as shown in our study by means of the odds ratio. As the PSE assessment is based on information from the previous month, the re-

sults are equivalent to current cases elicited through the CIDI as shown in this report, and where disabilities in social functioning also appeared as a second factor behind general dysphoria.

Duncan-Jones et al. (8), illustrating the utility of latent trait models, analyzed the GHQ-12 functioning and raised the issue of dimensionally, as they found a low correlation between observed scores for subsets formed from the positive and from the negative items. Also, by simulation they found that the latent correlation between the subsets was much higher but less than unity, concluding that the two parts of the test measure different facets of psychological distress. Dimensionally, Vázquez-Barquero et al. (36) using Likert scoring, gathered all positive items on the social functioning/optimism factor found in their study, while all negative items appeared in the general dysphoria factor. Regardless of caseness, and using this same type of scoring, we obtained very similar results. On the other hand, using dichotomic scoring, both negative and positive items appeared intermingled in the factors, although negative items predominated in the first factor obtained in current cases. Goodchild & Duncan-Jones (16) showed that positive and negative items are good indicators of presents state although these latter are stronger, results that have been reported consistently in other studies and also confirmed in the present report.

Jacobsen et al. (19) presented combinations of the best four and six items from the GHQ-20 as indicators of current mental distress selected by multiple regression analysis and by competent physicians. Among the best items "feeling unhappy and depressed" and "losing confidence", both found in the present report in the structure of the first factor in current cases, were elicited by statistical analyses while the rest of the items are not included in the GHQ-12. Physicians came up with various combinations including practically all items from the GHQ-12, and correlation coefficients between the scores based on the 20 items and those selected by the five clinicians were very high in all examined subgroups. Interestingly, two of the psychiatrists selected mainly negative mental health items, while the other psychiatrist and the two general practitioners selected an equal number of positive and negative aspects. Furthermore, the combinations obtained by regression analysis were predominantly, but not exclusively, negative, thus suggesting that in evaluating mental distress, indicators arise from a combination of both negative and positive subscales as shown in our results.

The fact that different combinations of GHQ items worked well in both populations from northern Norway, was explained indicating that different kinds of common mental problems generally have similar impact on the quality of life and every-day activities (19). Our results go a little further in a different geographical and cultural context by showing the performance of each item and eliciting subgroups related to current and lifetime prevalence of psychiatric disorders.

Following Lewis' (23) encouragement to further explore the relevance in clinical settings of positive and negative dimensions in interpreting the factor studies of the GHQ, we have found that, dimensionally, the highly specific subgroup of items elicited in our study

are predominantly positive and, inversely, the subgroup of highly sensitive items are predominantly negative. In terms of the odds ratios, the first four GHQ-12 items were significantly associated with the presence of a current disorder but not with lifetime cases, thus leaving the rest as more specific indicators of poor mental health. For lifetime prevalence, the first factor included the three disability items, 5 to 7, found for current cases separately, as well as "feeling worthless" and "losing confidence", which remained in this same factor, and including the item "not feeling reasonably happy". For the second factor, the four most frequent and sensitive items in our sample, as shown in the ROC curves, were gathered. Also, the factor structure using Likert scoring grouped items 1 to 3 into a separate factor. Moreover, further analysis of our data, not included in this paper, showed that these first three GHQ-12 items conformed a consistent factor present among those with and without psychiatric diagnosis in the studied population, thus suggesting a common way of expressing distress, relatively independent of the presence of psychopathology as measured in this study.

On the other hand, the subgroup of highly specific items mainly exploring positive aspects of mental health, items 5 to 7, remained together in the structure of the factors elicited in this report, representing in their contents an important aspect of diminished positive mental health: inhibition. Interestingly, the factor structure using Likert scoring identified items 7,11 and 12 in the third factor, explaining 9% of the variance, regardless of caseness. These items showed the lowest mean and median values of the whole sample and covariance of the two latter items has been presented by Duncan-Jones et al. (8) showing the utility of latent trait models in psychiatric epidemiology.

Together, these findings suggest that, within the GHQ-12 structure lies a continuous measure of the health-illness process and evidence from this report has identified: first, a component indicating a common way of expressing distress (item 1 to 3 and item 10); a second component indicating a possible vulnerability core (items 5 to 7), and a third component that seems to be more related with either previous or current psychiatric disorders (items 7,11 and 12). Practical implications could be that subjects responding positively to these two latter groups of items should be monitored for surveillance, or considered as being in risk for developing psychiatric disorders. Also epidemiological studies not specially focused on psychiatric disorders but interested in evaluating potential mental health risks, could use these more specific groups of items for surveillance. Also, in psychosocial and social medicine research, the components found in this study could be explored further in association with other variables. Actually, this task is currently under way on a representative sample of the adult general population in Mexico City, and preliminary analyses (Caraveo et al., in preparation) have found the same factor structures as reported in this paper.

Finally, the last point for discussion is whether the GHQ-12 is a culturally biased measure in Latin America. Lewis & Araya (24) presented data showing an ascertainment bias of the GHQ-12 explained almost entirely by

the difference in the negative scale of the instrument, while the difference between the Chilean and British samples in the positive scale was not statistically significant. The threshold determined in that study was conspicuously high (4/5) as compared to the British sample (1/2) or to the one found in the present report (2/3). As a whole, the mean positive score in the British sample was higher than the negative score (7.5 vs 4.9). In this regard, from our stand point (Tables 5 & 6), the positive and negative mental health items showed that in either diagnostic classification the mean odds ratio of the positive items were consistently higher than the mean of the negative items for either current or lifetime disorders; and even for current disorders, the mean odds ratio of the positive items, three and four, was higher than the mean of the negative items, first and second, that as a whole, were identified as sensitive indicators for current cases. All together, results presented in this paper do not support the affirmative conclusion of an ascertainment bias put forward by Lewis & Araya, and it is worth to note that the empirical data come from Mexico, which is also a Spanish Latin American country, and not as in the case of Brazil, where Portuguese is the official language.

Nevertheless, the discrepancies were intriguing specially because the population included in this study is quite comparable to the Chilean primary care sample, and that current use of the Spanish language is quite similar in both countries; also, because a previous study in Mexico using the GHQ-30 (28) found that the best threshold for that version was the same as the one determined by Goldberg (10) in the United Kingdom, and which is almost the same case in the present report. So, we further examined the Spanish version of the GHQ-12 used in Chile (35) and found that the only, but highly important, difference between our Spanish versions was the time-frame for asking the questions. All Mexican versions of the GHQ include in the general instructions that items are referred to the previous month, as used in this study, or to a two-week period as required when the CIS was employed as ultimate criterion measure of psychiatric disorder. Furthermore, once the time-frame has been clearly established, expressions like recently or lately do not appear beside the questions as in the Chilean version. Time perception is certainly not only an issue influenced by cultural bac-kground, but also differently conceptualized by people. Last few weeks, recently and lately, are unespecific time-related terms which are subject to different interpretations, and next to questions asking about distress this could easily lead to a positive bias response, even among members with equal education and not medically ill.

The above-mentioned wording difference seems to be the most plausible explanation for the discrepancy between our results and those obtained in Chile, emphasizing the importance of carefully considering time-frame in the structure of questionnaires for cross-cultural research. The issue does not represent an objection to the more general point addressed by Lewis & Araya (24) about further exploring the performance of the GHQ-12 and the patterns of scores in different cultural and socioeconomic settings. Nevertheless, based on our experience, we suggest it would be worth to study the factor structure of the instrument comparing extents groups with and without lifetime and current psychopathology, and for this the CIDI is the best instrument to be used as external criterion.

In conclusion, analyses of the performance of the GHQ-12 as related to the CIDI show evidence of three components within the structure of the instrument with practical implications for sociomedical research as well as for detecting caseness and epidemiological surveillance. Also, this paper presents evidence that the instruments is not culturally biased, but instead, it is a consistent measure very useful for cross-cultural research.

REFERENCES

1. ARAYA R, WYNN R, LEWIS G: Comparison of two self administered psychiatric questionnaires (GHQ-12 and SRQ-20) IN PRIMARY CARE IN CHILE. *Soc Psychiat and Psychiatric Epidem,* 27:168-173, 1992.
2. BALDESSARINI RJ, FINKLESTEIN S, ARANA GW: The predictive power of diagnostic tests and the effect of prevalence of illness. *Arch Gen Psychiatry,* 40:569-573, 1983.
3. BARRET JE, BARRETT JA, OXMAN TE, GERBER PD: The prevalence of psychiatric disorders in a primary care practice. *Arch Gen Psychiatry,* 45:1100-1106, 1988.
4. CAMPILLO SC, MEDINA-MORA ME, MARTINEZ LP, CARAVEO AJ: Cuestionario (GHQ) de detección de posibles casos psiquiátricos en una comunidad. *Cuadernos Científicos CEMESAM,* 11:43-44, 1979.
5. CASTRO RR, MEDINA-MORA ME: Poder discriminativo de un cuestionario que detecta padecimientos emocionales entre sujetos que requieren y no requieren atención especializada, con bajo nivel de escolaridad. *Revista de Enseñanza e Investigación en Psicología,* 8:229-235, 1982.
6. COMREY AL, LEE HB: *A First Course in Factor Analysis.* Second edition. Lawrence Erlbaum Associates, Publishers, Nueva Jersey, 1992.
7. DEAN AG, DEAN JA, BURTON AH, DICKER RC: Epi Info, Version 5: a word processing, database, and statistics program for epidemiology on microcomputers. USD, Incorporated, Stone Mountain, Georgia, 1990

8. DUNCAN-JONES P, GRAYSON DA, MORAN PA: The utility of latent trait models in psychiatric epidemiology. *Psychol Medicine,* 16:391-405, 1986.
9. GOLDBERG DP, COOPER B, EASTWOOD MR, KEDWARD HB, SHEPHERD M: A standardised psychiatric interview for use in community surveys. *Brit J Prev Soc Medicine,* 24:18-23, 1970.
10. GOLDBERG DP: *The Detectionof Psychiatric Illness by Questionnaire.* Institute of Psychiatry. Oxford University Press, London, 1972.
11. GOLDBERG DP, KAY C, THOMPSON L: Psychiatric morbidity in general practice and the community. *Psychol Medicine* 6:565-569, 1976.
12. GOLDBERG DP, HILLER VF: A scaled version of the General Health Questionnaire. *Psychol Medicine,* 9:139-145, 1979.
13. GOLDBERG DP, WILLIAMS P: *User's Guide to the General Health Questionnaire.* NFER-NELSON, Windsor, 1988.
14. GOLDBERG DP, GATER R, SARTORIUS N, ÜSTTÜN TB, PICCINELLI M, GUREJE O, RUTTER C: The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychol Medicine,* 27:191-197, 1997.
15. GOMEZ EM, MORALES CF, ARETIA PA, GUTIERREZ E: Detección de Alteraciones emocionales en pacientes obstétricas y ginecológicas. *Ginecología y Obstetricia de México,* 58:112-116, 1990.
16. GOODCHILD ME, DUNCAN-JONES P: Chronicity and the General Health Questionnaire. *Brit J Psychiatry,* 146. 55-61, 1985.
17. GUREJE O, OBIKOYA B: The GHQ-12 as a screening tool in a primary care setting. *Soc Psychiat Psychiatric Epidem,* 25:276-280, 1990.
18. HARDING TW, DE ARANGO MV, BALTAZAR J, CLIMENT CE, IBRAHIM HHA, LADRIDO-IGNACIO L, SRINIVASA-MURTHY R, WIG NN: Mental disorders in primary health care: a study of their frequency and diagnosis in four developing countries. *Psychol Medicine,* 10:231-241, 1980.
19. JACOBSEN BK, HASVOLD T, HOYER G, HANSEN V: The General Health Questionnaire: How many items are really necessary in population surveys?. *Psychol Medicine,* 25:957-961, 1995.
20. JANCA A, ÜSTÜN TB, SARTORIUS N: New versions of the World Health Organization instruments for the assessment of mental disorders. *Acta Psychiat Scandinavica,* 90:73-83, 1994.
21. KLEINBAUM DG, KUPPER LL, MORGENSTEN H: *Epidemiologic Research. Principles and Quantitative Methods.* Lifetime Learning Publications. Belmont, 1982.
22. KRAUSE I-B, ROSSER RM, KHIANI ML, LOTAY NS: Psychiatric morbidity among Punjabi medical patients in England measured by General Health Questionnaire. *Psychol Medicine,* 20:231-241, 1990.
23. LEWIS G, ARAYA RI: Is the General Health Questionnaire (12 item) a culturally biased measure of psychiatric disorder? *Soc Psychiat and Psychiatric Epidem,* 30:20-25, 1995.
24. LEWIS G: Dimensions of neurosis. *Psychol Medicine,* 22:1011-1018, 1992.
25. MARI JJ, WILLIAMS P: Minor psychiatric disorder in primary care in Brazil: a pilot study. *Psychol Medicine,* 14:223-227, 1984.
26. MARI JJ, WILLIAMS P: A comparison of the validity of two Psychiantric screening questionnaires (GHQ-12 and SRQ-20) in Brazil, using Relative Operating Characteristic (ROC) analysis. *Psychol Medicine,* 15:651-659, 1985.
27. MARTINEZ LP, MEDINA-MORA ME, PADILLA GP, CARAVEO AJ, CAMPILLO SC: Prevalencia de trastornos emocionales en personas de edad avanzada. *Salud Pública de México,* 25:612-619, 1983.
28. MEDINA-MORA ME, PADILLA GP, CAMPILLO SC, MAS CC, EZBAN M, CARAVEO AJ, CORONA J: The factor structure of the GHQ: a scaled version for a hospital's general practice service in México. *Psychol Medicine,* 13:355-361, 1983.
29. PADILLA GP, MAS CC, EZBAN M, MEDINA-MORA ME, PELAEZ O: Frecuencia de trastornos mentales en pacientes que acuden a la consulta general de un centro de salud. *Salud Mental* 7(3):72-78, 1984.
30. ROBINS LN, HELZER JE, CROUGHAN J, RATCLIFF KS: National Institute of Mental Health Diagnostic Interview Schedule: Its history, characteristics and validity. *Arch Gen Psychiatry,* 38:381-389, 1981.
31. ROBINS LN, WING JK, WITTCHEN HU, HELZER JE, BABOR TF, BURKE J, FARMER A, JABLENSKI A, PICKENS R, REGIER DA, SARTORIUS N, TOWLE LH: The Composite International Diagnostic Interview. An epidemiological instrument suitable for use in conjuntion with different diagnostic systems and in different cultures. *Arch Gen Psychiatry,* 45:1069-1077, 1988.
32. ROMERO MM, MEDINA-MORA ME: Validez de una versión del Cuestionario General de Salud, para detectar psicopatología en estudiantes universitarios. *Salud Mental,* 10(3):90-97, 1987.
33. SARACENO B, TERZIAN E, MONTERO-BARQUERO F, TOGNONI G: Salud mental en el nivel de atención primaria: estudio multicéntrico en seis países centroamericanos. *Psiquiatría Pública,* 5:73-94, 1993.
34. TARNOPOLSKY A, HAND DJ, MCLEAN EK, ROBERTS H, WIGGINS RD: Validity and uses of a screening questionnaire (GHQ) in the community. *Brit J Psychiatry,* 134:508-515, 1979.
35. TORRES P, ALVAREZ V: Validez de un cuestionario para detectar desordenes emocionales (GHQ-12) en estudiantes universitarios: comunicación preliminar. *Revista Psiquiatría,* 4:255-264, 1987.
36. VAZQUEZ-BARQUERO JL, WILLIAMS P, DIEZ-MANRIQUE JF, LEQUERICA J, ARENAL A: The factor structure of the GHQ-60 in a community sample. *Psychol Medicine,* 18. 211-218, 1988.
37. WILLIAMS P, TARNOPOLSKY A, HAND D: Case definition and case identification in psychiatric epidemiology: review and assessment. *Psychological Medicine,* 10:101-114, 1980.
38. WING JK, COOPER JE, SARTORIUS N: *The Measurement and Classification of Psychiatric Symptoms.* Cambridge University Press, Cambridge, 1974.
39. WORSLEY A. WALTERS WAW, WOOD EC: Responses of Australian patients with gynecological disorders to the General Health Questionnaire: a factor analytic study. *Psychol Medicine,* 8:131-138, 1978.
40. WORLD HEALT ORGANIZATION: *Schedules for Clinical Assessment in Neuropsychiatry* (SCAN). World Healt Organization, Ginebra, 1992.