

ORIGINAL PAPER

I. Sandanger · T. Moum · G. Ingebrigtsen
O.S. Dalgard · T. Sørensen · D. Bruusgaard

Concordance between symptom screening and diagnostic procedure: the Hopkins Symptom Checklist-25 and the Composite International Diagnostic Interview I

Accepted: 11 November 1997

Abstract The definition of case is a core issue in psychiatric epidemiology. Psychiatric symptom screening scales have been extensively used in population studies for many decades. Structured diagnostic interviews have become available during recent years to give exact diagnoses through carefully undertaken procedures. The aim of this article was to assess how well the Hopkins Symptom Checklist-25 (HSCL-25) predicted cases by the Composite International Diagnostic Interview (CIDI), and find the optimal cut-offs on the HSCL-25 for each diagnosis and gender. Characteristics of concordant and discordant cases were explored. In a Norwegian two-stage survey mental health problems were measured by the HSCL-25 and the CIDI. Only 46% of the present CIDI diagnoses were predicted by the HSCL-25. Comorbidity between CIDI diagnoses was found more than four times as often in the concordant cases (cases agreed upon by both instruments) than in the discordant CIDI cases. Concordant cases had more depression and panic/generalized anxiety disorders. Neither the anxiety nor the depression subscales improved the prediction of anxiety or depression. The receiver operating characteristic

(ROC) curves confirmed that the HSCL-25 gave best information about depression. Except for phobia it predicted best for men. Optimal HSCL-25 cut-off was 1.67 for men and 1.75 for women. Of the discordant HSCL-25 cases, one-third reported no symptoms in the CIDI, one-third reported symptoms in the CIDI anxiety module, and the rest had symptoms spread across the modules. With the exception of depression, the HSCL-25 was insufficient to select individuals for further investigation of diagnosis. The two instruments to a large extent identified different cases. Either the HSCL-25 is a very imperfect indicator of the chosen CIDI diagnoses, or the dimensions of mental illness measured by each of the instruments are different and clearly only partly overlapping.

Introduction

The agreement between psychiatric symptom screening and medical diagnosis on an individual basis, for clinical or research purposes, has seldom been reported to be better than 50% (Dohrenwend 1994). The HSCL-25 (Derogatis et al. 1974) has been used to identify psychiatric illness in primary care (Hesbacher et al. 1980; Hansson et al. 1994), in family planning services (Winokur et al. 1984), among refugees (Mollica et al. 1987; Felsman et al. 1990; McKelvey et al. 1993a), and among migrants (McKelvey et al. 1993b; Mouanoutoua and Brown 1995). There is good agreement between scores on the HSCL-25 and physicians' ratings of emotional distress (Rickels et al. 1976; Hough et al. 1990; Nettelbladt et al. 1993). The HSCL-25 has been extensively used in Norway to identify mental health problems and their relation to psychosocial risk factors. Using the HSCL-25 ≥ 1.75 as a criterion it has been shown that significant differences exist between communities and groups situated inland compared with those on the coast (Dalgard et al. 1995), and between fishermen and fishermen's wives, compared to other

I. Sandanger (✉) · D. Bruusgaard
Institute of General Practice and Community Medicine,
Department of Social Insurance Medicine,
University of Oslo, P.O. Box 1130 Blindern,
N-0317 Oslo, Norway

T. Moum
Department of Behavioural Sciences in Medicine,
University of Oslo, Oslo, Norway

G. Ingebrigtsen
Psychiatric Centre, Lofoten Hospital, Lofoten Islands,
Nordland, Norway

O.S. Dalgard
National Institute of Public Health,
Oslo, Norway

T. Sørensen
Psychiatric Institute, University of Oslo,
Oslo, Norway

working groups in the same municipality (Sørensen 1987; Sørensen et al. 1994).

The CIDI (Robins et al. 1988, 1994) was developed from the Diagnostic Interview Schedule (DIS) through the co-operation of nations through the World Health Organization, and is presently in use all over the world (Cottler 1995). The CIDI has shown good feasibility in the general populations, high interrater reliability, and has been subjected to tests of reliability and some validity tests with satisfactory results (Janca et al. 1992; Wittchen 1994).

“The keystone of epidemiology is the requirement to define cases of disorder” (Bebbington 1997). A diagnostic procedure like the CIDI reveals substantially more information about the condition than the HSCL-25 rating, but we do not know how these different ways of assessing illness relate to each other. They may both tap some underlying, joint phenomenon, in which case the HSCL-25 ought to predict the CIDI diagnosis, or two different, but partly overlapping dimensions of mental health problems. In the latter case, divergence between the two would be expected. The two instruments, the HSCL-25 and the CIDI, have never been compared before in an epidemiological survey.

This paper will address the following research questions: How well does a simple screening procedure such as the HSCL-25 predict psychiatric morbidity in the population, as assessed by a carefully undertaken and comprehensive method such as the CIDI? Which CIDI diagnoses are most likely to be identified by the HSCL-25, and Can the agreement between HSCL-25 \geq 1.75 cases and CIDI cases be improved by relying on the specific anxiety and depression subscales in the HSCL-25? What are the characteristics of the discordant cases, and what are the optimal HSCL-25 cut-offs for the diagnoses for women and men respectively?

Methods

Sample, response rate and design

In a Norwegian population survey in 1989–1991 we enquired about mental health problems, psychosocial risk factors, help seeking and other health parameters. We approached a random sample of 3656 individuals (age 18 years or more) from a borough in Oslo and the islands of Lofoten in northern Norway, drawn by the Central Bureau of Statistics. Of these, 929 persons were eliminated because they were deceased, had moved, or were impossible to locate after at least 12 attempts. Personal contact was obtained with 2727 individuals, of whom 2015 were personally interviewed, giving a response rate of 74% at stage 1. The gender and age composition of the sample was very similar to the population from which they were drawn. Stage 1 consisted of an in-person interview, which included a self-administered version of the HSCL-25 with the interviewer present. An HSCL-25 score equal to or greater than 1.55 was considered “a possible psychiatric case” (Rickels et al. 1976; Murphy 1981) and used as the selection criterion for stage 2. In addition a random sample of those with lower HSCL-scores was selected. Of the 797 persons in the new sample, 617 responded, a completion rate of 77% (415 with an HSCL-25 score of 1.55 or more, and 202 below). Persons with language problems were excluded from stage 2; 11 persons were excluded in the present ana-

lyses because of uncompleted HSCL-25 ratings at stage 2. Thus, on average 11.5 months after stage 1, 30% of the original interview population received a second HSCL-25 rating and the CIDI. The stage 2 data were adjusted for the sampling procedure to reflect the population sample and used to test concordance between the HSCL-25 and the CIDI.

Measures and variables

The screener, HSCL-25

Twenty-five questions about the presence and intensity of anxiety and depression symptoms over the previous week were scored on a scale from 1 (not bothered) to 4 (extremely bothered). Some of the symptoms might also be interpreted as somatoform. The HSCL-25 score was calculated as the sum score of items divided by number of items answered. A score equal to or larger than 1.75 was defined “a case” (Winokur et al. 1984).

Psychiatric diagnoses, the CIDI

Diagnoses were derived from the ICD-10 Diagnostic Criteria for Research Diagnoses (WHO 1989), assessed with the Composite International Diagnostic Interview (CIDI) version 1.0. The CIDI is built of modules for each diagnostic category, and information about individual symptoms can be used separately. It was carried out by trained lay interviewers, mostly women health professionals. The CIDI focuses first on lifetime ever symptoms and later in the interview limits itself to different time periods during the “previous year”. Symptoms are probed and somatic or alcohol- or drug-related problems ruled out. In this presentation we have used the modules for depressive disorders, anxiety disorders and somatoform disorders, and restricted diagnoses to symptoms present within the previous 2 weeks. The CIDI results were summarised in four diagnostic categories: (1) depression (F31.3–F34.1), (2) panic and anxiety disorders (F41.0–F41.8), (3) phobias (F40.0–F40.2), and (4) somatoform disorders (F44–F45.4). The phobias were separated from the other anxiety disorders because the planned avoidance of phobic situations can prevent phobic attacks and anxiety, in contrast to panic attacks. Exclusion criteria were used according to the diagnostic hierarchy.

Receiver operating characteristic (ROC) curve

The ROC curve is the curve representing all possible combinations of sensitivity and specificity for a test (the screener) and the classification (the diagnosis). The area under the curve represents the amount of information that the screener contains about the diagnosis. The diagonal is the line of no information. Comparing the area under curve (AUC) for the ROC curves of the screener in different groups or for different diagnoses shows whether the screener has more information about one group/diagnosis than another. The point on the curve in the most upper left corner has the highest sensitivity and specificity and represents the optimal cut-off for the screener in finding diagnoses (Hanley and McNeil 1982; Murphy et al. 1987).

Statistical methods

All results were adjusted for the selection procedure from stage 1 to stage 2, to represent a population sample. This was done by adjusting for the sampling variable HSCL-25 less than or equal to/above 1.55 in the first step in all analyses. To test for differential statistical association between dependent and independent variables in the two sampling strata, an interaction term between the sampling factor and other independent variables was added as a last step of the regression analyses. Adjusted ROC curves were made by taking the sampling variable into the equation from which

Table 1 Gender and age distribution of the samples at stage 1 ($n = 2015$) and stage 2 ($n = 606$)

	Men Stage 1 (47%) Stage 2 (41%)	Women Stage 1 (53%) Stage 2 (59%)
18–34 years	36% 27%	40% 36%
35–49 years	30% 33%	28% 29%
50–66 years	19% 21%	17% 20%
67+ years	15% 19%	15% 15%

Table 2 Correspondence between Hopkins Symptom Checklist-25 score of 1.75 or more ($\text{HSCL-25} \geq 1.75$) and Composite International Diagnostic Interview (CIDI) diagnosis in percentages (adjusted to population sample)

	Diagnosis by the CIDI row percent (column percent)	No diagnosis by the CIDI row percent (column percent)	
$\text{HSCL-25} \geq 1.75$	43% (46%)	57% (19%)	100%
$\text{HSCL-25} < 1.75$	16% (54%) (100%)	84% (89%) (100%)	100%

the 800-STATA (Stata Statistical Software, Release 5.0 College Station, Stata Corporation, TX) created the curves and the sensitivity and specificity. Crosstables, logistic regressions, sensitivity and specificity were used for analysing the relation between the HSCL-25 and the CIDI diagnoses. Non-linearity was tested by adding the squared HSCL-25 in a last step of logistic regressions. To explore whether the HSCL-25 subscales for anxiety and depression did a better job than the total scale in predicting diagnoses was done by looking at the improvement of the -2 log likelihood of the model when one of the subscales was added after the total score in two-step logistic regressions. The predictability of CIDI diagnoses from HSCL-25 mean scores was explored through ROC curves, comparing the performance of the HSCL-25 for each diagnosis by looking at both the shape of and the area under the curve (AUC). Comparisons of AUCs were made by z -scores according to Hanley and McNeil (1983) and given as P -values. Significance testing was done by two-tailed t -tests, Chi-square statistics and 95% confidence intervals. The Statistical Package for the Social Sciences (SPSS) version 6.1 and 800-STATA were used for the analyses and the ROC curves. The CIDI data were computed in the CIDI computer programme (Wittchen 1993) and later transferred to the SPSS.

Results

Table 1 shows the gender and age distributions in the stage 1 and stage 2 samples. The distributions were similar in the samples. Table 2 shows that 46% of the CIDI cases were identified by the HSCL-25 (concordant cases), 54% of the CIDI cases were not identified (discordant cases). There were significantly more

HSCL-25 cases and CIDI cases among women than men. Discordant CIDI cases were more frequently found among respondents younger than 40 years of age ($P < 0.05$).

Among single diagnoses, the highest HSCL-25 score was found for depression (1.66), followed by panic and generalized anxiety disorder (1.44), phobias (1.36) and somatoform disorder (1.36). Comorbidity resulted in higher means (Table 3).

Among the respondents who were cases according to both instruments, there were more cases of depression and panic or generalized anxiety disorder than in persons with CIDI diagnoses who were not classified as HSCL-25 cases. Disorders in two or more of the four diagnostic categories were found in 36% of the concordant cases, whereas two or more diagnoses were found in only 8% of the discordant CIDI cases (Table 4). Of the HSCL-25 cases who did not receive a CIDI diagnosis, on average one-third did not report symptoms in any CIDI section, one-third had symptoms in the anxiety section only, and the rest had symptoms spread across the different modules. Depression symptoms alone or combined with symptoms in other sections were found in about 10%. A combination limited to anxiety and depression symptoms did not occur. The most frequent HSCL-25 symptoms among discordant HSCL-25 cases are given in Table 4.

In bivariate analyses the HSCL-25 picked out 74% of the depressions, 50% of the panic/generalized anxiety disorders, 38% of the phobias and 29% of the somatoform disorders. The odds ratio for having a CIDI diagnosis when being an HSCL-25 case, using not being an HSCL-25 case as reference group, was accordingly strongest for depression, followed by panic and generalized anxiety disorders, phobias, and somatoform disorders (Table 5). Having any of the CIDI diagnoses had

Table 3 Mean HSCL-25 for single CIDI diagnoses and comorbidity states (means adjusted for the sampling factor to reflect the population sample)

CIDI diagnoses ^a	n	Mean HSCL-25	(95% CI)
No diagnosis	471	1.25	(1.19–1.30)
Depression only	12	1.66	(1.46–1.87)
Phobias only	46	1.36	(1.25–1.47)
Generalized anxiety or panic disorder only	14	1.44	(1.25–1.64)
Somatoform disorder only	34	1.28	(1.15–1.40)
Depression + phobia	7	2.02	(1.76–2.26)
Depression + panic/gen anx	4	1.46	(1.11–1.81)
Phobia + somatoform	6	1.37	(1.08–1.66)
Phobia + panic/gen anx	7	1.77	(1.50–2.03)
Panic/gen anx + somatoform	2	1.74	(1.23–2.24)
Depression + panic/gen anx + phobia	1	2.10	(1.39–2.80)
Depression + phobia + somatoform	2	2.50	(2.00–2.99)

^a All combinations of diagnoses are included in the table to account for the whole sample

Table 4 Percentage of diagnoses and symptoms in concordant and discordant cases (adjusted to population sample)

	Concordant cases ^a		Discordant CIDI cases ^a		Discordant HSCL-25 cases
CIDI diagnoses	Depressions	33%	Depressions	8%	Non
	Panic/gen anx	28%	Panic/gen anx	15%	
	Phobias	55%	Phobias	48%	
	Somatoform	28%	Somatoform	37%	
	Comorbidity	36%	Comorbidity	8%	
Most frequent HSCL-25 symptoms	Feeling fearful		Tense, keyed up		Nervousness, shakiness
	Nervousness, shakiness		Headaches		Tense, keyed up
	Feeling blue		Low in energy		Low in energy
	Worrying too much		Worrying too much		Feeling blue
	Everything is an effort		Everything is an effort		Worrying too much

^a More than one diagnosis possible per person

Table 5 CIDI diagnoses as dependent variables and HSCL-25 ≥ 1.75 caseness as independent, adjusted to population sample, presented as odds ratios (*I* crude and *II* adjusted for comorbidity, sex and age)

	<i>I</i>		<i>II</i>	
	OR (95% CI)	r^2	Adjusted OR (95% CI)	r^2
Depression	14.04 (4.66–42.24)	0.203	9.82 (3.18–30.32)	0.251
Panic/generalized anxiety	4.37 (1.90–10.03)	0.125	3.66 (1.50–8.95)	0.152
Phobia	3.61 (2.05–6.35)	0.082	2.75 (1.50–5.04)	0.108
Somatoform disorder	2.17 (1.11–4.24)	0.035	2.21 (1.09–4.48)	0.060
Any diagnosis	3.89 (2.50–6.04)	0.139	3.82 (2.44–5.97)	0.126

The analyses were based on 602 persons due to insufficient data in one of the variables in 4 respondents

^a r^2 according to Nagelkerke (Mittlböck and Schemper 1996)

an odds ratio of 3.89. Blockwise adjustments for gender and age had no significant influence on the results. Adjustment for comorbidity between the four diagnoses in multivariate analyses decreased the odds ratios somewhat for all diagnoses except somatoform disorders. Thus the HSCL-25 was significantly better in predicting depression than phobia and somatoform disorder ($P < 0.05$).

By adding the squared HSCL-25 in logistic regressions with each of the CIDI diagnoses as dependent variables, only the model for panic and generalized anxiety disorders was significantly improved (improvement in log likelihood 7.738, $P = 0.0209$). None of these disorders was found at the lowest end of the HSCL-25 scale.

Anxiety and depression subscales

Adding one of the subscales to the total HSCL-25 scale did not improve the statistical model with regard to the prediction of CIDI anxiety or depression, meaning that the anxiety and depression subscales of the HSCL-25 did not predict the CIDI anxiety and depression diagnoses better than the entire HSCL-25 scale (Table 6). However, the anxiety subscale did result in a better prediction of somatoform disorders. Bearing in mind the comorbidity between diagnoses, we repeated the analyses for single diagnoses with similar results.

Receiver operating characteristic (ROC) curves: optimal HSCL-25 cut-off for CIDI diagnoses

The ROC curves confirmed that the HSCL-25 gave most information about depression (AUC = 0.8515), followed by panic and generalized anxiety disorders (AUC = 0.7881), phobias (AUC = 0.6932), and somatoform disorders (AUC = 0.6350) (Table 7). The ROC curve for all diagnoses combined lay between the curves for panic and generalized anxiety disorder and phobia (not shown). The AUCs for both the depression curve and the panic and generalized anxiety curve were significantly larger than the AUC for somatoform disorders ($P = 0.002$, $P = 0.006$).

The ROC curves also showed that the HSCL-25 tended to predict depression and panic and generalized anxiety disorder better for men than for women. Optimal cut-off for depression and panic and generalized anxiety disorder, calculated as the highest product of sensitivity and specificity, was 1.75 for women and 1.67 for men (Fig. 1). Using the lower cut-off for men increases their caseness by HSCL-25 from 10.3% to 12.5%, but this is still significantly lower than the 20.6% for women.

To check the influence of minor problems on the relation between HSCL-25 and CIDI, the intensity scores of the HSCL-25 items were dichotomized into 1 (1+2) meaning “not bothered” and 2 (3+4) meaning “yes, bothered”. The ROC curve derived from this new HSCL-25 (not shown) was practically identical to the original

Table 6 Improvement in likelihood of prediction of diagnosis using HSCL-25 anxiety and depression subscales as compared with HSCL-25 sum score (adjusted for sampling factor to represent the population sample)

	Depression	Panic/gen anx	Phobia	Somatoform
Log likelihood for HSCL-25				
Sum score	166.499	200.228	385.550	307.180
OR	15.54	4.87	6.03	2.16
Improvement in log likelihood when adding HSCL-25 anxiety subscale	0.106 (<i>P</i> = 0.7451)	0.201 (<i>P</i> = 0.6538)	0.001 (<i>P</i> = 0.9754)	14.65 (<i>P</i> = 0.0001)
OR HSCL-25 sum score	11.38	7.37	5.91	0.08
OR HSCL-25 anxiety subscale	1.36	0.66	1.02	23.39
Improvement in log likelihood when adding HSCL-25 depression subscale	0.105 (<i>P</i> = 0.7457)	0.208 (<i>P</i> = 0.6487)	0.002 (<i>P</i> = 0.9666)	14.89 (<i>P</i> = 0.0001)
OR HSCL-25 sum score	24.58	2.62	6.28	227.95
OR HSCL-25 depression subscale	0.63	1.87	0.96	0.01

Using two-step logistic regressions, a significant increase in log likelihood when adding a subscale in step two means that the subscale does significantly better job in predicting the diagnosis than the HSCL-25 sum score in step one. The direction of the effect is seen from the coefficient in the equation. The sign is positive for

the HSCL-25 anxiety subscale and negative for the depression subscale with regard to somatoform disorder, i.e. the anxiety subscale was more strongly related to somatoform disorder and the depression subscale less strongly related than the whole HSCL-25 scale

Table 7 Area under Receiver-operating characteristic curves (*AUC*) for HSCL-25 and diagnoses by the CIDI. Optimal HSCL-25 cut-off for each diagnosis with corresponding sensitivity and specificity, adjusted for sampling factor

	Depressive disorder		Panic and generalized anxiety disorder		Phobia		Somatoform disorder		All diagnoses	
Women										
AUC (SE) ^a	0.8198	(0.0517)	0.7343	(0.0542)	0.6937	(0.0455)	0.5808	(0.0546)	0.6857	(0.0336)
Optimal cut-off	1.75		1.75		1.44		1.96		1.67	
Sensitivity and specificity	0.81	0.70	0.67	0.73	0.65	0.67	0.53	0.67	0.62	0.70
Men										
AUC (SE)	0.9428	(0.0326)	0.8634	(0.039)	0.6732	(0.0666)	0.7223	(0.0635)	0.7523	(0.0439)
Optimal cut-off	1.67		1.67		1.63		1.40		1.63	
Sensitivity and specificity	1.0	0.79	0.80	0.80	0.57	0.79	0.79	0.63	0.67	0.81
All										
AUC (SE)	0.8515	(0.0400)	0.7881	(0.0365)	0.6932	(0.0370)	0.6350	(0.0420)	0.7146	(0.0261)
Optimal cut-off	1.75		1.67		1.63		1.67		1.67	
Sensitivity and specificity	0.81	0.78	0.71	0.73	0.67	0.69	0.52	0.73	0.58	0.78

^aStandard errors are calculated for the unadjusted AUCs. The AUC for the depression curve and the panic and generalized anxiety curve is significantly larger than the AUC for somatoform

disorders (*P* = 0.002, *P* = 0.006). Optimal cut-off is the HSCL-25 score with the maximum product of sensitivity and specificity

ROC curve for all diagnoses, meaning that the explanation for discordant HSCL-25 cases and low prediction of diagnoses did not lie in having several weaker symptoms on the screener, resulting in a high sum score.

Discussion

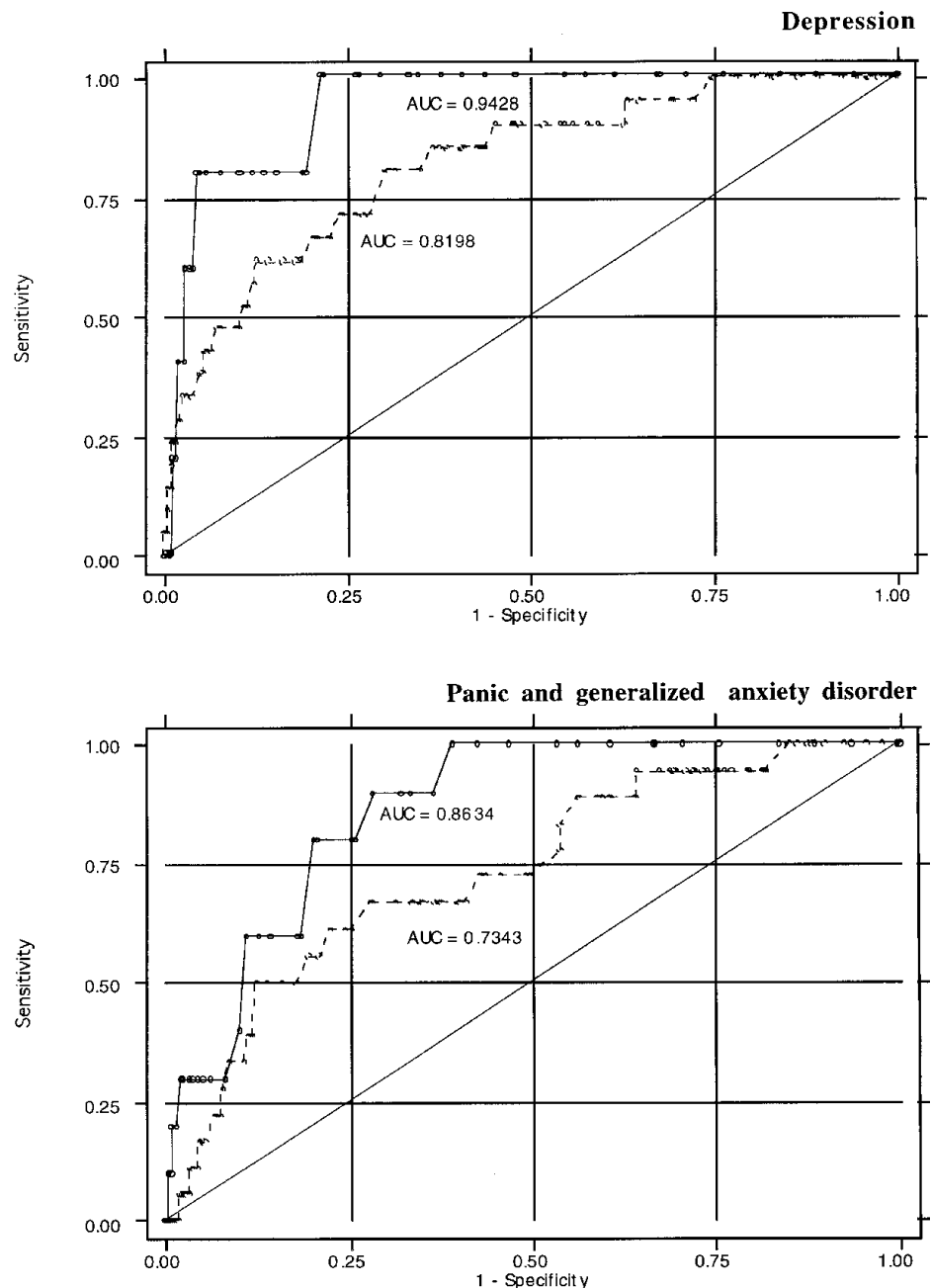
Method

The questions in the HSCL-25 were answered privately by pencil and paper with the interviewer present, whilst the CIDI was a face-to-face interview. A much lower

score on the HSCL-25 in the general population occurred in Norway when the list of symptoms was presented orally by an interviewer, rather than the respondent checking off her/his symptoms on the sheet by her/himself (Moum 1995, Personal communication), probably due to a social desirability bias (Ross and Mirowsky 1984). The CIDI questions might also have been answered more often with “yes” if the respondent had answered alone, rather than to an interviewer. A reliability test between an ordinary CIDI interview and a self-administered CIDI-Auto interview (Andrews 1993; Peters et al. 1993) could test this.

A considerable disagreement in caseness between current-symptom-based approaches like the HSCL-25

Fig. 1 Receiver operating characteristic curves for the prediction of depression, panic and generalized anxiety disorder, phobia and somatoform disorder by the HSCL-25, for each gender, adjusted for sampling factor from stage 1 to stage 2 (*AUC area under curve, stapled lines women, whole lines men*)

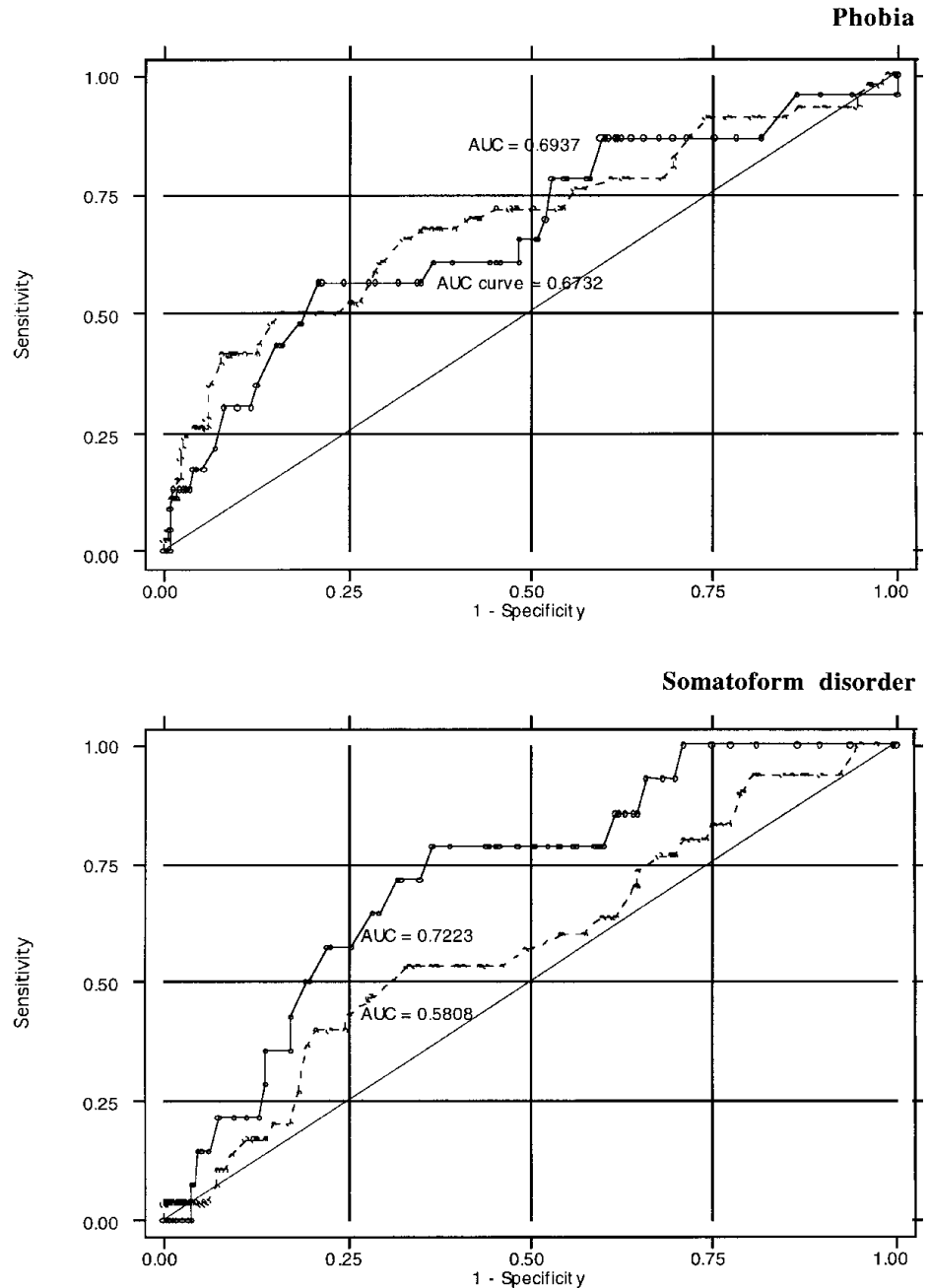


and life time-based methods such as the CIDI has been found (Simon et al. 1995), even when they both refer to the present, and could contribute to low agreement between the two.

The HSCL-25 is a continuous measure of 25 symptoms rated according to their intensity, measuring "how much of it has s/he got", implying that everybody has some. It does not distinguish between essential features and associated symptoms for disorders. The everyday language of the HSCL-25 may also lead to recognition and positive response more easily than the more specific CIDI questions. As such, a high HSCL-25 score can be the result of discomfort from stress and strains. The

diagnoses are "prototypes" of the disorders asking "has s/he got it?" An ill person may fall outside the diagnostic categories. Further, the diagnostic interview exercises a social control that directs the respondents towards a particular view of their illness, which may restrain the response (Waitzkin 1991). The possibility also exists that psychiatric syndromes that occur in the population may consist of symptom constellations other than those experienced in hospitals, which are the basis for definitions of diagnoses (Leighton et al. 1966). This is supported by the fact that high clinical validity for an instrument defined and tested in a patient population has often not been achieved in field studies in general populations

Fig. 1 (Contd.)



(Dohrenwend 1994). Lastly, the validity of a diagnostic instrument such as the CIDI will always “be limited to the validity inherent in the diagnostic criteria on which that instrument is based” (Wittchen 1994). Here we touch on the basic problem of whether or not the present scientific taxonomy is realistic for psychiatry (Margolis 1994). It may be that the limits of psychiatric knowledge leads to classifications and operational definitions that reflect features of varying importance of psychiatric conditions, but which are insufficient to serve as “the truth” about what they are.

The question about the validity of self-appraisal of health arises with both instruments, but more so with

the HSCL-25, because the probing in the CIDI helps rule out unimportant symptoms or symptoms caused by somatic illness, alcohol and drugs. Because health and illness are normatively defined (Twaddle 1974) and the population studied is relatively homogeneous, there is high agreement about what is a health problem. Large Norwegian population studies have shown self-evaluated health to be a good predictor of health care utilisation (Moum 1992; Fylkesnes 1993), and a similar result has also been found in England (Williams et al. 1986).

Prediction of diagnoses

The HSCL-25 ≥ 1.75 correlated most strongly with depression, followed by panic and generalized anxiety disorders, which it was constructed for. Phobias differed from the other anxiety disorders in having a low sensitivity and specificity by the HSCL-25. This could be due to the possibility that persons with phobias can avoid anxiety if they avoid phobic situations.

It seems reasonable that the cases with comorbidity should have the highest HSCL-25 scores. This group included most of the depressions and panic and generalized anxiety disorders. One may of course ask whether comorbidity in fact covers different coexisting entities of disorders, or whether it rather represents different dimensions of an underlying phenomenon which can be treated in one way, and which should have been identified as one disease. Except for panic and generalized anxiety disorders, the diagnoses were spread along the HSCL-25 scale, indicating that the HSCL-25 and the CIDI measured overlapping, but different, dimensions of the mental health problems in question. The HSCL-25 is "the temperature measure" (Frank 1973), indicating that there is some problem, but not defining what it is. We interpret it as a measure of psychological pain or distress. The diagnosis by the CIDI is a psychiatric disorder classified according to the ICD-10 criteria. Obviously distress and diagnosis need not be present at the same time.

Optimal cut-off was lower for men (1.67) than for women (1.75), and sensitivity and specificity was better for men. This raises the question about gender-specific symptomatology (Murphy 1995).

Strangely enough, the constellation of symptoms that theoretically should be the more essential features of depression and anxiety did not do a better job in predicting the diagnoses of those disorders than a mix of symptoms from the whole HSCL-25. A considerable amount of comorbidity exists between depression and other psychiatric disorders, and this has been used to explain why screeners have done well in validation studies on mixed groups of psychiatric patients (Murphy 1990). Comorbidity is, however, not a good explanation of why the subscales did not work better than the full scale in our material, because they did not work better for single diagnoses either. The HSCL-25 anxiety subscale predicted somatoform disorder better and the depression subscale predicted it less well than the whole scale. One reason might be that the anxiety questions are very close to somatic symptoms ("Faintness, dizziness, or weakness", "heart pounding or racing", "trembling", "headaches"), or that these persons are living with a high fear of having cancer or other life-threatening conditions.

The areas under the ROC curves are generally acknowledged as good above 0.80 (Le Gall et al. 1993), which means that depression was well discriminated by the HSCL-25. The area under the curve for panic and generalized anxiety disorders was acceptable, and both

curves were significantly more informative about the diagnosis than the curve for somatoform disorders. For phobias it was marginal, and for somatoform disorders it was too small for acceptable discrimination. These ROC results supported what the HSCL-25 was constructed for, detecting anxiety and depression. Inspecting the ROC curves, one should consider decreasing the cut-off level for an HSCL-25 case to 1.67 for men.

The experiment of constructing a ROC curve based on dichotomized items in the HSCL-25 was in accordance with similar experiments with other instruments, such as comparing a dichotomous and Likert scoring of the General Health Questionnaire (Goldberg 1972) with a dichotomous and a frequency-weighted score of the Health Opinion Survey (Murphy 1990), showing that report of minor problems is not the explanation of high-symptom scores without diagnoses.

Persons with high HSCL-25 scores and no diagnosis may of course suffer from disorders not investigated by us, such as post traumatic stress disorder (Ingebrigtsen et al. 1995), recurrent brief depression (Weiller et al. 1994), "subthreshold depression" (Sherbourne et al. 1994) or adjustment disorder. However, very few of the respondents in this group had recorded symptoms in the CIDI depression section, and still fewer had a combination of anxiety and depression symptoms. "Subthreshold depression" or "mixed anxiety depression disorder" is therefore not a good explanation of HSCL-25 cases without diagnosis in this material. However, the CIDI questions were not designed for them. The most frequent HSCL-25 items in this group were a mix of nervousness, feeling tense, blue, worried and feeling everything is an effort. These are perhaps more likely to be symptoms of distress and fatigue than criteria for a diagnosis. The best guess about discordant HSCL-25 cases may be that they are reflecting the discomfort of particular strains or stressors, either in the process of moving into a diagnosis, or containing the burden of psychosocial strain or somatic illness (Sandanger 1993; Sandanger et al. 1995).

Turning to the group of persons with diagnoses with low HSCL-25 scores, we see that they seldom suffered from depression or had more than one diagnosis. Phobias and somatoform disorders were frequent in this group. With the exception of "headaches", their most frequent HSCL-25 symptoms were the same as in the other cases, but occurred more seldom. It is as if they do not suffer so much from their disorder, or they do not express their suffering in the same language.

The definition of illness or disorder evolves in a cultural setting. It depends not only on symptoms, suffering, loss of functioning, or strain on the patient, their family, health services or society, but also on the growing accessibility of treatment, causing a shift from "life problems" to "illness". With this shift we may well see in the future a growing proportion of newly identified diagnoses in the group which today are classified as only HSCL-25 cases.

Conclusion

The HSCL-25 as a screener for diagnoses of depression, panic/generalized anxiety disorder, phobia and somatoform disorder, as defined by the ICD-10 criteria in the CIDI, was acceptable only for depression, and best for men. The optimal cut-offs, which were 1.67 for men and 1.75 for women, are a reminder of the need for gender-specific validation of psychiatric instruments. Concordant cases showed psychiatric comorbidity more than four times as often as discordant CIDI cases, which means that they showed more psychopathology. The HSCL-25 and the CIDI can be viewed as measures of only partly overlapping dimensions of mental health problems. The meaning and significance of each type of caseness in epidemiological surveys are not certain.

References

- Andrews G (1993) Computerised CIDI (CIDI-Auto). CIDI-SCAN-IPDE Newsletter 3: 262–263
- Bebbington P (1997) What's in a name? Psychiatric epidemiology and social psychiatry at the turn of the century. *Soc Psychiatry Psychiatr Epidemiol* 32: 1–2
- Cottler LB (1995) Published articles related to the DIS & CIDI. DIS Newsletter 12(1)
- Dalgard OS, Brevik JI, Hjemmen A (1995) emphasis. Health profile in the local community, with particular on social risk factors and mental health. In: Dalgard O S, Døhlie E, Ystgaard M (eds) Social network, health and society (in Norwegian). Norwegian University Press, Oslo
- Derogatis LR, Lipman RS, Rickels K, Uhlenluth EH, Covi L (1974) The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. *Behav Sci* 19: 1–15
- Dohrenwend BP (1994) The problem of validity in field studies in psychological disorders. In: Mezzich JE, Jorge MR, Salloum IM (eds) Psychiatric epidemiology. Assessments, concepts and methods. Johns Hopkins University Press, Baltimore London
- Felsman JK, Leong FT, Johnson MC, Felsman IC (1990) Estimates of Psychological distress among Vietnamese refugees: adolescents, unaccompanied minors and young adults. *Soc Sci Med* 31: 1251–1256
- Frank JD (1973) Persuasion and healing: a comparative study of psychotherapy, rev edn. Schocken, New York
- Fylkesnes K (1993) Determinants of health care utilization – visits and referrals. *Scand J Community Med* 21: 40–49
- Goldberg DP (1972) The detection of psychiatric illness by questionnaire: a technique for the identification and assessment of non-psychotic psychiatric illness. Oxford University Press, London
- Goldberg DP, Huxley P (1980) Mental illness in the community. The pathway to psychiatric care. Tavistock, London
- Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143: 29–36
- Hanley JA, McNeil BJ (1983) A method of comparing areas under receiver operating characteristic curves derived from the same cases. *Radiology* 148: 839–843
- Hansson L, Nettelbladt P, Borgquist L, Nordström G (1994) Screening for psychiatric illness in primary care. A cross-sectional study in a Swedish health district. *Soc Psychiatry Psychiatr Epidemiol* 29: 83–87
- Hesbacher PT, Rickels K, Morris RJ, Newman H, Rosenfeld H (1980) Psychiatric illness in family practice. *J Clin Psychiatry* 41: 6–10
- Hough RL, Landsverk JA, Jacobson GF (1990) The use of psychiatric screening scales to detect depression in primary care patients. In: Attkisson CC, Zich JM (eds) Depression in primary care: screening and detection. Routledge, New York
- Ingebrigtsen G, Sandanger I, Sørensen T, Dalgard OS (1995) When disaster strikes. In: Dalgard OS, Døhlie E, Ystgaard M (eds) Social network, health and society (in Norwegian). Norwegian University Press, Oslo
- Janca A, Robins LN, Cottler LB, Early TS (1992) Clinical observation of assessment using the Composite International Diagnostic Interview. *Br J Psychiatry* 160: 815–818
- Le Gall J-R, Lemeshow S, Saulnier F (1993) A new simplified acute physiology score (SAPS II) based on European/North American multicenter study. *J Am Med Assoc* 22/23: 2957–2963
- Leighton AH, Leighton DC, Danley RA (1966) Validity in mental health surveys. *Can Psychiatr Assoc J* 11: 167–178
- Margolis J (1994) Taxonomic puzzles. In: Sadler JZ, Wiggins OP, Schwartz MA (eds) Philosophical perspectives on psychiatric diagnostic classification. The Johns Hopkins University Press, Baltimore
- McKelvey RS, Webb JA, Mao AR (1993a) Premigratory risk factors in Vietnamese Amerasians. *Am J Psychiatry* 150: 470–473
- McKelvey RS, Mao AR, Webb JA (1993b) Premigratory expectations and mental health symptomatology in a group of Vietnamese Amerasian youth. *J Am Acad Child Adolesc Psychiatry* 32: 414–418
- Mittlböck M, Schemper M (1996) Explained variation for logistic regression. *Stat Med* 15: 1987–1997
- Mollica RF, Wyshak G, de Marneffe D, Khuon F, Lavelle J (1987) Indochinese version of the Hopkins Symptom Checklist-25: a screening instrument for psychiatric care of refugees. *Am J Psychiatry* 144: 497–500
- Mouanoutoua VL, Brown LG (1995) Hopkins Symptom Checklist Hmong version: a screening instrument for psychological distress. *J Person Assess* 64: 376–383
- Moum T (1992) Self-assessed health among Norwegian adults. *Soc Sci Med* 35: 935–947
- Murphy JM (1981) Psychiatric instrument development for primary care research: patient self-report questionnaire. National Institute of Mental Health Contract No. 80M014280101D, Division of Biometry and Epidemiology, Rockville
- Murphy JM (1990) Depression screening instruments: history and issues. In: Attkisson CC, Zich JM (eds) Depression in primary care. Routledge, New York London
- Murphy JM (1995) What happens to depressed men? *Harvard Rev Psychiatry* 3: 47–49
- Murphy JM, Berwick DM, Weinstein MC, Borus JF, Budman SH, Klerman G (1987) Performance of screening and diagnostic tests. Application of receiver operating characteristic (ROC) analysis. *Arch Gen Psychiatry* 44: 550–555
- Nettelbladt P, Hansson L, Stefanson CG, Borgquist L, Nordström G (1993) Test characteristics of the Hopkins Symptom Checklist-25 (HSCL-25) in Sweden, using the Present State Examination (PSE-9) as a caseness criterion. *Soc Psychiatry Psychiatr Epidemiol* 28: 130–133
- Peters L, et. al (1993) A procedural validity study of the CIDI-Auto. WHO CIDI Training and References Centre, Sydney
- Rickels K, Garcia CR, Lipman RS, Derogatis LR, Fisher EL (1976) The Hopkins Symptom Checklist: assessing emotional distress in obstetric-gynecological practice. *Primary Care* 3: 751–764
- Robins LN, Wing JK, Wittchen H-U, Helzer JE, Babor TF, Burke J, Farmer A, Jablenski A, Pickens R, Regier DA, Sartorius N, Towle LH (1988) The Composite International Diagnostic Interview. *Arch Gen Psychiatry* 45: 1069–1077
- Robins L, Wing JK, Wittchen H-U, Helzer JE, Babor TF, Burke J, Farmer A, Jablenski A, Pickens R, Regier DA, Sartorius N, Towle L (1994) The Composite International Diagnostic Interview. In: Mezzich JE, Jorge MR, Salloum IM (eds) Psychiatric epidemiology. Assessments concepts and methods. Johns Hopkins University Press, Baltimore London

- Ross CE, Mirowsky J (1984) Socially-desirable response and acquiescence in a cross-cultural survey of mental health. *J Health Soc Behav* 25: 189–197
- Sandanger I (1993) The Composite International Diagnostic Interview for psychiatric diagnoses as an expounding instrument in the Norwegian Community Diagnosis Project: mental health and physical illness. *Int J Methods Psychiatr Res* 3: 137–142
- Sandanger I, Sørensen T, Dalgard OS, Ingebrigtsen G (1995) Childhood strain, present strain and social support. Influence on mental health. In: Dalgard O S, Døhlie E, Ystgaard M (eds) *Social network, health and society* (in Norwegian). Norwegian University Press, Oslo
- Sherbourne CD, Wells KB, Hays RD, Rogers W, Burnam A, Judd LL (1994) Subthreshold depression and depressive disorder: clinical characteristics of general medical and mental health specialty outpatients. *Am J Psychiatry* 151: 1777–1784
- Simon GE, Maier W, Üstün B, Linden M, Boyer P (1995) Research diagnosis of current depressive disorder: a comparison of methods using current symptoms and lifetime history. *J Psychiatr Res* 6: 457–465
- Sørensen T (1987) *Mental health on the northcoast* (in Norwegian) Nordland Fylkeskommune, Fylkeshelsestjenestens kontor, Bodø
- Sørensen T, Ingebrigtsen G, Bøe N (1994) The price of endurance. A community psychiatric perspective on mental health in four coastal municipalities on northern Norway. In: Otterstad O, Jentoft S (eds) (in Norwegian). Ad Notam Gyldendal, Oslo
- Twaddle AC (1974) The concept of health status. *Soc Sci Med* 8: 29–38
- Waitzkin H (1991) *The politics of medical encounters: how patients and doctors deal with social problems*. Yale University Press, New Haven
- Weiller E, Boyer P, Lepine JP, Lecrubier Y (1994) Prevalence of recurrent brief depression in primary care. *Eur Arch Psychiatry Clin Neurosci* 244: 174–181
- Williams P, Tarnopolsky A, Hand D, Sheperd M (1986) Minor psychiatric morbidity and general practice consultation: the West London survey. *Psychol Med Monogr [Suppl 9]*
- Winokur A, Winokur DF, Rickels K, Cox DS (1984) Symptoms of emotional stress in family planning service: stability over a four-week period. *Brit J Psychiatry* 144: 395–399
- Wittchen H-U (1993) Computer scoring of CIDI diagnoses. *Int J Methods Psychiatr Res* 3: 101–107
- Wittchen H-U (1994) Reliability and validity study of the WHO Composite International Diagnostic Interview (CIDI): a critical review. *J Psychiatr Res* 28: 57–84
- World Health Organization (1989) *I.C.D.-10, Chap 5. Mental and behavioral disorders. Diagnostic criteria for research* (April 1989 draft for field trials). WHO division of Mental Health, Geneva