

Validity across translations of short survey psychiatric diagnostic instruments: CIDI-SF and CIS-R versus SCID-I/NP in four European countries

Ondine Pez · Fabien Gilbert · Adina Bitfoi · Mauro Giovanni Carta ·
Vesna Jordanova · Carmen Garcia-Mahia · Raimundo Mateos-Alvarez ·
Martin Prince · Bogdana Tudorache · Catherine Blatier · Viviane Kovess-Masfety

Received: 12 January 2009 / Accepted: 12 October 2009 / Published online: 3 November 2009
© Springer-Verlag 2009

Abstract

Objectives The aims of the study are: first, to compare two short diagnostic instruments, CIDI-SF and CIS-R, with respect to the structured clinical interview for non-patient (SCID-I/NP) for anxiety and depressive disorders; and second, to evaluate the influence of four languages, Italian, Romanian, Spanish and French, on the concordance tests. **Methods** A total of 120 participants from Italy and Romania, 119 from Spain and 141 from France ($N = 500$) were recruited randomly in a local primary care research centre (GPs or medical centres). The instruments were administered during a unique session: the lay instruments by students in psychology and the SCID by experienced

psychiatrists. Kappa, sensitivity/specificity/negative (NPV) and positive predictive values (PPV), ROC curve (AUC) and the Youden Index (Y) were calculated.

Results Results were better for the CIDI-SF than the CIS-R for anxious disorders, depressive disorders and any of them. The results were identical to that obtained by the CIDI 3.0 for the three categories and constant across the languages except for depressive disorders in Romania and France.

Conclusions CIDI-SF is a cost-effective instrument and could be easily integrated into health surveys; its performance values are better across languages than values proposed by the CIS-R and could be increased by inclusion of few additional information.

O. Pez · V. Kovess-Masfety (✉)
University of Paris Descartes, EA 4069, 12 rue de l'école de
médecine, 75006 Paris, France
e-mail: viviane.kovess-masfety@ext.parisdescartes.fr

O. Pez
e-mail: pez_ondine@hotmail.com; scmheproject@gmail.com

F. Gilbert
MGEN Foundation for Public Health, 3 square Max-Hymans,
75748 Paris Cedex 15, France
e-mail: fgilbert@mgen.fr

A. Bitfoi · B. Tudorache
The Romanian League for Mental Health, SOS. Mihai Bravu No.
90-96, BL.D17, SC.4, ET.8, AP149 Sector 2,
021331 Bucharest 39, Romania
e-mail: adina_bitfoi@yahoo.com

B. Tudorache
e-mail: lrsm@mnt.ro

M. G. Carta
Division of Psychiatry, Department of Public Health, University
of Cagliari, Viale merello 22, 09123 Cagliari, Italy
e-mail: mgcarta@tiscali.it

C. Garcia-Mahia · R. Mateos-Alvarez
Faculty of Medicine Santiago de Compostela, Santiago de
Compostela, 15782 Santiago de Compostela, Spain
e-mail: mrmateos@usc.es

M. Prince
Centre for Public Mental Health, Institute of Psychiatry,
King's College of London, De Crespigny Park,
London SE5 8AF, UK
e-mail: m.prince@iop.kcl.ac.uk

V. Jordanova
Section of Epidemiology, Institute of Psychiatry,
King's College of London, London, UK
e-mail: jordanova2000@yahoo.com

C. Blatier
Faculty of Psychology, University of Grenoble,
Pierre-Mendès-France, U.F.R. Sciences de l'Homme et de la
Société, B.P. 47 X, Grenoble, France
e-mail: Catherine.Blatier@upmf-grenoble.fr

Keywords Cross-cultural validation · Short diagnostic disorders instruments: CIDI-SF, CIS-R · SCID-I/NP · Anxious and depressive disorders · Local primary care patients

Introduction

Differences in prevalence estimates for depression and anxiety between countries have been frequently reported [38] and generally attributed to methodological differences: different instruments, diagnostic criteria, level of impairment and sampling methodologies; however, even in surveys using common instruments, criteria and design, such as ESEMED [16] or WMH [19], differences have to be interpreted with caution since little is known on the stability of the instruments across translations and cultures. A recent study [19] has established concordance between CIDI [39] and SCID [17] by comparing English, French, Italian and Spanish data but without exploring possible cross-cultural and inter-language differences.

In addition, the formulation of mental health questions could fit better some cultures than others. A closely based DSM-IV criteria approach, as that used in the CIDI [39], may be an efficient approach in some cultures, while, on the contrary, an approach based on primary care physician types of questions, such as those used by the CIS-R [28] in the UK, may be more convenient in others.

The objective of our study was to compare the psychometric properties of two short diagnostic instruments, the CIDI-SF [23] and the CIS-R [28], with respect to a diagnostic reference, the Structured Clinical Interview for Non-Patient (SCID-I/NP [17]) across four countries: Italy, Romania, Spain and France taking explicitly into account results across languages on a large re-interview sample (500 cases).

The short form version of the CIDI has been selected in accordance with the EU monitoring group recommendations (ECHI [24]). This questionnaire that could be quickly completed was derived from the CIDI and designed in the USA [23]. It has been widely used since for epidemiological studies in Canada [4, 31–33] and Europe [20, 25–27] and compared with the CIDI [23, 33]. The short form version of the CIDI has been demonstrated to be generally reliable, although both the CIDI and the CIDI-SF tend to overestimate the prevalence of major depressive disorders [33]. CIDI-SF psychometrics properties have been studied in Finland with respect to the SCAN: results were fair, but the specificity of the CIDI-SF was higher and agreement better for a more global affective disorder category than for major depressive episode [1].

The CIS-R [28, 29] has been widely used in the UK as part of the National Survey of Psychiatric Morbidity [10, 11, 14], as well as in other countries such as Brazil [6, 12],

Chile [3] and Taiwan [30], and has been used to compare the prevalence of mental health disorders between different ethnic communities in Great Britain [5, 8, 13, 36].

The respective validity of the CIS-R and the full version of the CIDI has been already assessed in the UK [9, 20] with respect to the schedule for clinical assessment in neuropsychiatry (SCAN) [37]. This last study [21] using concordance and the receiver operating characteristic (ROC) analysis found that the CIDI was a highly valid assessment of common mental disorders, whilst the CIS-R was moderately valid contrary to what was found in the Brugha study [9]. These differences have been attributed to different duration of periods between interviews, since in the protocol used by Jordanova et al. both interviews were conducted during the same session.

Methods

To compare the psychometric properties of the CIDI-SF vis-à-vis the CIS-R to a semi-structured clinical interview, the SCID-I/NP, we conducted face-to-face interviews in four countries (Italy, Romania, Spain and France).

The CIDI-SF and the CIS-R, two structured instruments with open and closed questions on the most common DSM-IV mental health problems, were administered by trained lay interviewers (graduate students in psychiatry or psychology).

The SCID-I/NP is a semi-structured instrument which permits the evaluation of the most common mental health problems of the DSM-IV Axis 1. The SCID-I/NP was administered by experienced psychiatrists.

For practical reasons, only depressive and anxiety disorders were considered. Seven DSM-IV diagnostic categories [2] were selected in the CIDI-SF 12-month prevalence version (Table 1): major depression, generalised anxiety, specific phobia, social phobia, agoraphobia, panic attack and obsessive–compulsive disorders. Since CIS-R provides diagnostic categories according to the ICD-10 classification [40], we used syntaxes implemented in previous surveys in the UK [20] to generate equivalent DSM-IV diagnostic categories (Table 1). The CIS-R yields information on 1-month prevalence rates. For the SCID-I/NP, diagnoses were identified for three types or groups of disorder (Table 1). These are ‘any mood disorder’, including major depressive disorder, mood disorder due to a general medical condition and dysthymia, ‘any anxiety disorder’, including any phobia (specific, social, agoraphobia), general anxiety disorder, panic disorder, obsessive–compulsive disorder, anxiety disorder due to general medical condition and ‘any diagnosis’, which refers to at least one of the above diagnoses being present.

Table 1 Diagnostic categories

	CIDI-SF 12 months	CIS-R 1 month	SCID-I/NP 1 month
Any mood disorder	MDE	MDE	MDE MDE due to a GMC Dysthymia
Any anxiety disorder	GAD Specific phobia Social phobia Agoraphobia Panic attack OCD	GAD Specific phobia Social phobia Agoraphobia Panic attack OCD	GAD ^a Specific phobia Social phobia Agoraphobia Panic attack ^a OCD ^a

GAD general anxiety disorder, *GMC* general medical condition, *MDE* major depressive disorder, *OCD* obsessive–compulsive disorder

^a Including anxiety disorders due to a general medical condition

Translations

The instruments were specifically translated and then back translated for validation using a standard methodology [34]. Translations were performed by bilingual specialists with experience in the mental health domain. For the CIDI-SF, we extracted items from validated translations of the CIDI 3.0 full version (French and Spanish) when this was available, which were subsequently validated by the reference bilingual translator for the study.

Study population

The study included 500 volunteers aged 18 years or more attending primary care clinics in France, Italy, Spain and Romania.

Table 2 Interview methods

Countries	Number	Sample	Interviewers
Italy	120	Adult population consulting an individual GP	Two tandems of an experienced physician and a resident in psychiatry, était un interne dans ce cas on dit resident in psychiatry
Romania	120	Adult population attending a primary care centre	Two groups of two psychiatrists working in a university clinic with experience in structured and semi-structured interviews
Spain	119	Adult population attending a primary care centre	One senior psychiatrist, with experience in structured and semi-structured interviews Three PhD (two in psychiatry and one in psychology) students
France	141	Adult population consulting an individual GP or a specialist at the MGEN centre ^a	One psychiatrist with experience in structured and semi-structured interviews Four psychology MSc students and one PhD psychologist
Total	500		

^a Public health mutual insurance centre

Interviews

The interviewees administering the questionnaires are presented in Table 2.

Lay interviewers met the potential respondents, presented the project, asked agreements and began the interviews with the survey instruments. The psychiatrist was not present during that phase and inversely the interviewer did not attend the psychiatrist interview. Depending on the availability of the psychiatrist, the survey interviews were administered before or after the psychiatrists' interviews with the SCIDI-I/NP; both were done on the same day.

All the respondents who wanted to participate answered the entire survey instruments. About 60 interviews were done by each interviewer for Italy and for Romania, about 35 for France and about 40 for Spain.

Statistical analysis

The socio-demographic characteristics of the sample and the diagnoses were compared between countries using the χ^2 test. To measure the risk of having a disorder, we calculated prevalence rates, which is the proportion of individuals with a disorder in a sample. We then investigated whether the CIDI-SF, the CIS-R or the SCID-I/NP prevalence estimates differed between countries.

To quantify the association between each survey instrument (CIS-R or CIDI-SF) and the gold standard (The SCID-I/NP), we used several statistical measures of the performance of a binary classification test (presence/absence of a disorder). Sensitivity (Se; proportion of true positives correctly identified by the CIDI-SF or the CIS-R) and specificity (Sp; the proportion of true negatives identified by the CIDI-SF or the CIS-R) were the intrinsic

values measured. Positive (PPV; the proportion of CIDI-SF or CIS-R cases confirmed by SCID-I/NP) and negative predictive values (NPV; the proportion of CIDI-SF or CIS-R non-cases confirmed by SCID-I/NP) were determined for the two test instruments.

Individual-level CIDI-SF/CIS-R versus SCID-I/NP diagnostic concordance was evaluated using the area under the receiver operator characteristic curve (AUC [18]). The AUC is calculated as a dichotomous outcome between non-cases and cases, such as $(Se + Sp)/2$. This measure can be interpreted as the probability that a clinical case, which has been randomly selected, will score higher on the CIDI-SF and the CIS-R than a non-case randomly selected. The AUC values are between 0 and 1: 1 equals a perfect performance and 0 the absence of performance. Above 0.9, the concordance level is excellent; between 0.7 and 0.9 is considered as indicative and useful; between 0.5 and 0.7, concordance is fair.

We used also the Youden index ($Y = Se + Sp - 1$) which provides a single numerical estimation of the overall diagnostic effectiveness and to summarise the accuracy of the test instrument. A value of 1 means that the test is perfect.

The level of concordance between the SCID-I/NP and the test instruments (CIDI-SF and CIS-R) was evaluated by the kappa concordance test. The kappa measures the agreement between two different judgements or instruments. In psychiatry, kappa of more than 0.40 is considered as a good agreement between both instruments. The test of equality of ROC areas (χ^2) was calculated to measure differences between countries and between test instruments compared to the SCID-I/NP [15]. This specific χ^2 is calculated in a non-parametric approach by comparing the areas under two correlated receiver operating curves and depending on the number of observations.

We analysed the results in a global way and cross-cultural one.

Results

Socio-demographic features of the study sample

With respect to gender (female 67.9%), age (mean: 50 ± 18.0 years) and marital status (26.21% 'never married'; 55.04% 'married'; 18.75% 'widowed or divorced'), the four participating countries are comparable. For employment status, differences are observed in the proportion of retired subjects (43.6% French, 50% Romanian, but only 22% Italian and 21% Spanish). With respect to educational level, France differs from other countries in having a majority of participants (74.8%) with higher education. This is a consequence of the fact that the

primary care study centre in France is an Health Management Organisation (HMO) for teachers and related professions.

Estimated prevalence rates of the SCID-I/NP versus the CIDI-SF diagnosis instruments

Concerning the SCID-I/NP estimated prevalence rates, no significant difference between countries is found for the mood disorders. The estimated prevalence range between 4.3% for France to 12.5% for Italy. However, divergences are found for anxiety and overall disorders. Italy presents higher estimated prevalence: for any anxiety disorders (from 6.38% for France to 25.83% for Italy) and any diagnoses (9.9% for France and 30.8% for Italy).

For any anxiety and for any disorders, prevalence estimated by the CIDI-SF is higher than the SCID-I/NP, but follow the same divergences between countries. In Table 3, with the CIDI-SF, the estimated prevalence of any mood disorder is significantly higher ($\chi^2(3) = 9.78$, $p < 0.01$) in France and Spain (14.2 and 21%, respectively) than in Italy and Romania (7.5 and 11.7%, respectively). Discrepancy is also observed for any anxiety disorder ($\chi^2(3) = 12.55$, $p < 0.01$) and for 'any diagnosis' ($\chi^2(3) = 14.23$, $p < 0.01$).

Measure of the performance for the CIDI-SF

For any countries, performance tests are good to excellent, for mood, anxiety and any disorders. The Youden indexes, respectively, 0.46, 0.44 and 0.48, highlight the idea that sensitivity (from 0.56 to 0.76) and specificity (from 0.72 to 0.90) balance each other, and whatever the type of disorders, the CIDI-SF gives important information on the concordance level. The AUC values between 0.72 and 0.73 confirm that the performance is indicative. The kappa values, respectively, 0.36, 0.33 and 0.37, are quite good whatever the type of disorders. The PPV is quite fair (0.15–0.67). On the other hand, the NPV is very high (0.84–1).

Looking at the sensitivities and specificities values, we note divergences between disorder types. For the mood disorders, the sensitivity is about 0.56 and the specificity 0.90. In other words, the CIDI-SF capacity to detect the true positives is very fair in comparison with its capacity to detect true negatives. For the anxiety disorders, if a better detection of true positives is found, there is a minor loss of detection for the true negatives. For any diagnosis, sensibility and specificity are quite similar, respectively, 0.76 and 0.72, and point out that globally 3/4 of the individuals are well classified, regardless of whether they are true positives or true negatives. For any diagnosis, the CIDI-SF is as sensitive as it is specific. On the contrary, for the

Table 3 Consistency of current DSM-IV CIDI-SF 12-month prevalence and SCID-I 1-month prevalence diagnoses in the cross-cultural sample

	SCID	CIDI-SF	Se	Sp	PPV	NPV	Youden index	AUC	Kappa	
			(95% CI)	(95% CI)	(95% CI)	(95% CI)		(95%CI)	(95% CI)	
All mood disorders										
All countries	8.2%	13.6%	0.56	0.90	0.34	0.96	0.46	0.73	0.36	
			(0.49–0.72)	(0.87–0.93)	(0.23–0.46)	(0.94–0.98)	0.46	0.73	(0.65–0.81)	
Italy	12.5%	11.7%	0.60	0.95	0.64	0.94	0.55	0.78	0.57	
			(0.32–0.84)	(0.89–0.98)	(0.35–0.87)	(0.88–0.98)	0.55	0.78	(0.65–0.91)	
Romania	10.0%	7.5%	0.25	0.94	0.33	0.92	0.19	0.60	0.22	
			(0.05–0.57)	(0.88–0.98)	(0.08–0.7)	(0.85–0.96)	0.19	0.60	(0.47–0.73)	
Spain	6.7%	21.0%	1.00	0.85	0.32	1.00	0.85	0.92	0.43	
			(0.63–1)	(0.77–0.91)	(0.15–0.53)	(0.96–1)	0.85	0.92	(0.89–0.96)	
France	4.3%	14.2%	0.50	0.87	0.15	0.98	0.37	0.69	0.18	
			(0.12–0.88)	(0.81–0.93)	(0.03–0.38)	(0.93–1)	0.37	0.69	(0.47–0.91)	
<i>p</i> value ^a	0.081	0.021						0.000		
All anxiety disorders										
All countries	17.4%	31.8%	0.68	0.76	0.37	0.92	0.44	0.72	0.33	
			(0.57–0.77)	(0.71–0.8)	(0.3–0.45)	(0.88–0.95)	0.44	0.72	(0.66–0.77)	
Italy	25.8%	27.5%	0.68	0.87	0.64	0.89	0.54	0.77	0.53	
			(0.49–0.83)	(0.78–0.93)	(0.45–0.8)	(0.8–0.94)	0.54	0.77	(0.68–0.86)	
Romania	23.3%	22.5%	0.50	0.86	0.52	0.85	0.36	0.68	0.36	
			(0.31–0.69)	(0.77–0.92)	(0.52–0.32)	(0.76–0.92)	0.36	0.68	(0.58–0.78)	
Spain	16.0%	33.6%	0.79	0.75	0.38	0.95	0.54	0.77	0.37	
			(0.54–0.94)	(0.65–0.83)	(0.23–0.54)	(0.88–0.99)	0.54	0.77	(0.67–0.87)	
France	6.4%	41.8%	1.00	0.62	0.15	1.00	0.62	0.81	0.17	
			(0.66–1)	(0.53–0.7)	(0.07–0.27)	(0.96–1)	0.62	0.81	(0.77–0.85)	
<i>p</i> value ^a	0.001	0.006						0.119		
All diagnoses										
All countries	20.8%	38.0%	0.76	0.72	0.42	0.92	0.48	0.73	0.37	
			(0.67–0.84)	(0.67–0.76)	(0.35–0.49)	(0.88–0.95)	0.48	0.73	(0.69–0.79)	
Italy	30.8%	32.5%	0.70	0.84	0.67	0.84	0.55	0.77	0.54	
			(0.53–0.84)	(0.75–0.91)	(0.5–0.81)	(0.77–0.93)	0.55	0.77	(0.69–0.86)	
Romania	25.0%	27.5%	0.63	0.84	0.58	0.87	0.48	0.74	0.46	
			(0.44–0.80)	(0.75–0.91)	(0.39–0.75)	(0.79–0.94)	0.48	0.74	(0.64–0.83)	
Spain	19.3%	42.0%	0.87	0.69	0.40	0.96	0.56	0.78	0.39	
			(0.66–0.97)	(0.59–0.78)	(0.26–0.55)	(0.88–0.99)	0.56	0.78	(0.69–0.86)	
France	9.9%	48.2%	1.00	0.57	0.21	1.00	0.57	0.79	0.21	
			(0.77–1)	(0.48–0.66)	(0.12–0.32)	(0.95–1)	0.57	0.79	(0.74–0.83)	
<i>p</i> value ^a	0.001	0.003						0.839		

^a Chi² *p* value between countries

mood and anxiety disorders, the CIDI-SF is more specific than sensitive.

Comparisons of the different values of the discriminative indexes of countries give more precise results and several observations could be made. For the mood disorders, good intrinsic values were found for Italy and Spain, but not for France and Romania. In Italy, for instance, the results are better balanced: the specificity and NPV are high, whilst the sensitivity and PPV are correct. These results indicate a very good capacity to detect both true positives and true negatives. Moreover, the agreement between the two instruments is significant (0.57). The Spanish results are very good for the sensitivities, specificities and NPV, whilst the PPV is poor (0.32). For Spain, the probability that an individual is detected as being a case by both the SCID-I/NP and the CIDI-SF is fair. This is a direct consequence of the very few observations (6.7%) estimated by the SCID-I/NP. France and the Romania report results illustrating an important capacity to detect true negatives, but an important incapacity to detect true positives (sensitivity and PPV, fair). These discrepancies are also found in the test of equality of the AUC, where the performance value variability between countries was found ($\chi^2(3) = 29.53, p < 0.01$).

For the anxiety disorders, the Italian, Spanish and Romanian results were quite good. The Italian results could particularly detect true positives well. The important number of CIDI-SF estimated cases has a weak influence on the PPV value. However, no performance (AUC values) differences were found between countries ($\chi^2(3) = 5.84, NS$). In the French sample, because the number of estimated cases by the CIDI-SF was overestimated, the instrument had excellent sensitivity, NPV and AUC for the anxiety disorders, but poor PPV and kappa values. In brief, the CIDI-SF better detects the true negatives than true positives.

For any diagnoses, because this section groups together the mood and anxiety disorders, and because a big proportion of cases are from the anxiety disorders category, the same results are found. In brief, good results are observed for the Italian sample, correct ones for the Romanian ones, quite fair ones for the Spanish ones and very poor ones for the French sample for the PPV and kappa values. No performance discrepancies (AUC values) were found between countries ($\chi^2(3) = 0.84, NS$).

In conclusion, results comparing the discriminative performance and the intrinsic values of the CIDI-SF are quite good if only sensitivities, specificities, AUC and Youden indexes are considered. However, because of an overestimation of probable cases by the CIDI-SF, the agreement values are very poor. This observation, which is visible for the whole sample (all countries pooled together), is the direct effect of results in the French and the Spanish samples, whilst for the two other countries, results are

usually significant. Concerning the AUC equality, differences between countries are only found for the mood disorders.

Estimated prevalence rates of the CIS-R diagnosis instruments versus SCID-I/NP

On comparing the SCID-I/NP and the CIDI-SF, the CIS-R gives fewer estimated prevalence rates. For all countries, CIS-R estimates mood disorders of about 4.82%, any anxiety disorder of about 12.6% and overall disorders of about 14.2%. No significant differences are found for the estimated prevalence rates between countries. It is important to note that for France, the CIS-R gives higher prevalence rates than the SCID-I/NP does for any anxious disorders and for any diagnosis. For Spain, higher prevalence is found for any mood disorders.

Measure of the performance for the CIS-R

The classification tests are better for the mood disorders and for any diagnosis than for any anxiety disorders (Table 4). It seems that specificity (0.88–1) and the NPV (0.77–0.97) are very good. However, sensitivities (0.22–0.44) and PPV (0.11–0.67) are quite fair (except for PPV for Italy and any mood disorders). In fact, the detection of true negatives and the assurance that individuals who have a positive result on the test instrument are true positives is very poor. The PPVs are correct (0.54 for all diagnoses) and are the direct consequence of the fewer CIS-R observations compared to those estimated by the SCID-I/NP. As a consequence, the probability of detecting a case with both the CIS-R and SCID-I/NP is more relevant. However, agreements between both instruments are low to moderate for any anxiety disorders (from 0.07 to 0.37).

In the cross-cultural analyses, results for any mood disorders are really close except for the PPV. These values vary depending on the difference in the number of cases between the SCID-I/NP and the CIS-R. More the differences are important between the number of true positives detected by the CIS-R and SCID-I/NP, better are the PPVs.

When the number of cases is fewer with the lay questionnaire than with the gold standard (SCID-I/NP), the probability of detecting a case by both is all the more greater, such that the number of cases detected by CIS-R is smaller and the number of cases detected by SCID-I/NP is considerable. In Italy for instance, the number of cases detected with the CIS-R is four times less important than the number detected by the SCID-I/NP. As a result, PPV is equal to 1, which means that the totality of individuals detected by the CIS-R are also detected by the SCID-I/NP.

For any anxiety disorders, results are comparable to the mood disorders for the Italian and Spanish samples. There

Table 4 Consistency of current DSM-IV CIS-R and SCID-I 1 month prevalence diagnoses in the cross-cultural sample

	SCID	CIS-R	Se	Sp	(95% CI)	PPV	(95% CI)	NPV	(95% CI)	Youden index	AUC	(95% CI)	Kappa	(95% CI)
All mood disorders														
All countries	8.2%	4.8%	0.32	0.98	(0.18–0.48)	0.54	(0.33–0.74)	0.94	(0.92–0.96)	0.29	0.65	(0.57–0.72)	0.36	(0.21–0.52)
Italy	12.5%	3.3%	0.27	1.00	(0.07–0.55)	1.00	(0.4–1)	0.91	(0.84–0.95)	0.27	0.63	(0.52–0.95)	0.39	(0.12–0.66)
Romania	10.0%	6.7%	0.33	0.96	(0.1–0.65)	0.50	(0.16–0.84)	0.93	(0.86–0.97)	0.29	0.65	(0.51–0.79)	0.35	(0.06–0.63)
Spain	6.7%	7.6%	0.38	0.95	(0.09–0.76)	0.33	(0.08–0.70)	0.95	(0.90–0.99)	0.32	0.66	(0.48–0.84)	0.30	(–0.0–0.61)
France	4.3%	2.1%	0.33	0.99	(0.04–0.78)	0.67	(0.09–0.99)	0.97	(0.93–0.99)	0.32	0.66	(0.46–0.87)	0.43	(0.02–0.83)
<i>p</i> value ^a	0.081	0.129									0.992			
All anxiety disorders														
All countries	17.4%	12.6%	0.32	0.92	(0.22–0.43)	0.44	(0.32–0.58)	0.86	(0.83–0.9)	0.24	0.62	(0.57–0.67)	0.27	(0.16–0.38)
Italy	25.8%	14.2%	0.35	0.93	(0.19–0.55)	0.65	(0.38–0.86)	0.81	(0.72–0.88)	0.29	0.64	(0.55–0.73)	0.34	(0.14–0.53)
Romania	23.3%	13.3%	0.29	0.91	(0.13–0.49)	0.50	(0.25–0.75)	0.81	(0.71–0.88)	0.20	0.60	(0.51–0.70)	0.23	(0.03–0.43)
Spain	16.0%	10.1%	0.37	0.95	(0.16–0.62)	0.58	(0.28–0.85)	0.89	(0.81–0.94)	0.32	0.66	(0.55–0.77)	0.37	(0.14–0.61)
France	6.4%	12.8%	0.22	0.88	(0.03–0.6)	0.11	(0.01–0.35)	0.94	(0.89–0.98)	0.10	0.55	(0.4–0.7)	0.07	(–0.12 to 0.26)
<i>p</i> value ^a	0.001	0.798									0.617			
All diagnoses														
All countries	20.8%	14.2%	0.37	0.92	(0.27–0.47)	0.54	(0.41–0.66)	0.85	(0.81–0.88)	0.28	0.64	(0.59–0.69)	0.32	(0.22–0.42)
Italy	30.8%	15.0%	0.38	0.95	(0.23–0.55)	0.78	(0.52–0.94)	0.77	(0.68–0.85)	0.33	0.67	(0.58–0.75)	0.39	(0.21–0.56)
Romania	25.0%	15.0%	0.33	0.91	(0.17–0.53)	0.56	(0.31–0.79)	0.80	(0.71–0.88)	0.24	0.62	(0.53–0.71)	0.28	(0.09–0.48)
Spain	19.3%	14.3%	0.44	0.93	(0.23–0.66)	0.59	(0.33–0.82)	0.87	(0.79–0.93)	0.36	0.68	(0.57–0.79)	0.40	(0.19–0.62)
France	9.9%	12.8%	0.29	0.89	(0.08–0.58)	0.22	(0.06–0.48)	0.92	(0.86–0.96)	0.18	0.59	(0.46–0.71)	0.16	(–0.06–0.37)
<i>p</i> value ^a	0.001	0.947									0.637			

^a Chi² *p* value between countries

is a minor difference in the Romanian sample: the inter-instrument agreement is fair (from 0.23 to 0.35) with a moderate PPV (from 0.50 to 0.56). In the French sample, the kappa (0.07–0.16) and PPV (0.11 for all mood disorders and 0.22 for any disorders) are very poor. However, for the French sample, kappa and PPV are better for the mood disorder than the other one. The detection is very uncommon: the CIS-R detects two times more individuals with mood disorders than the SCID-I/NP does, and two times less individuals with anxiety disorders.

For any diagnosis, the Italian and Spanish are acceptable, fairly good in the case of Romanian and mediocre for the French sample. In this last sample, the PPV and the kappa are very fair. In brief, it seems that the CIS-R is more specific than sensitive. The cases observed by the CIS-R are fairest than detected by the SCID-I/NP, which involves good PPV and good probability to detect true cases. Concerning the inter-instrument agreement, values are from fairly good to poor. However, the CIS-R appears stable between countries when the equality of AUC is compared (test of equality of ROC areas; NS).

Comparison of performance detection between both lay questionnaires compared to the SCID-I/NP

Globally, the CIDI-SF gives significantly better AUC values than the CIS-R does for the anxiety disorders (cf. Table 5; $\chi^2(3) = 7.91$; $p < 0.01$) and for any diagnoses

Table 5 Test of equality of the ROC areas comparison between lay questionnaires and by country

	CIDI-SF	CIS-R	Test of equality of ROC areas
All mood disorders			
All countries	0.73	0.65	0.084
Italy	0.78	0.63	0.025*
Romania	0.60	0.65	0.658
Spain	0.92	0.66	0.005**
France	0.69	0.66	0.776
All anxiety disorders			
All countries	0.72	0.62	0.005**
Italy	0.77	0.64	0.069
Romania	0.68	0.60	0.162
Spain	0.77	0.66	0.075
France	0.81	0.55	0.001**
All diagnoses			
All countries	0.74	0.64	0.002**
Italy	0.77	0.67	0.084
Romania	0.74	0.62	0.027*
Spain	0.78	0.68	0.093
France	0.79	0.59	0.003**

* $p < 0.05$; ** $p < 0.01$

($\chi^2(3) = 9.75$; $p < 0.01$). For the mood disorders, there is no AUC difference between the CIS-R and the CIDI-SF.

In a cross-cultural analysis, the CIDI-SF has significantly better AUC values except for the Romanian one in the mood disorders categories. However, even when the CIS-R is likely to have better AUC values, there is no significant difference in terms of comparison of ROC areas ($\chi^2(1) = 0.20$; NS) For the mood disorders, the CIDI-SF's AUC is significantly better for Spain ($\chi^2(1) = 5.01$; $p < 0.05$) and for Italy ($\chi^2(1) = 7.99$; $p < 0.01$), whilst for France and Romania no statistic differences are found.

For the anxiety disorders, a strong difference is found for the French AUC only (CIDI-SF: 0.81 and CIS-R: 0.55; $\chi^2(1) = 11.56$; $p < 0.01$). Italy and Spain are close to being also significant.

For the overall diagnosis category, CIDI-SF AUC for France and Romania are significantly different and better than the CIS-R ones (respectively, $\chi^2(1) = 8.97$; $p < 0.01$ and $\chi^2(1) = 4.84$; $p < 0.05$).

In brief, CIDI-SF globally gives better AUC for the anxiety disorders and any diagnoses. Some cross-cultural differences have been found and confirm this tendency. The test of equality of ROC areas takes also into consideration the number of observations, which can also explain why in some situations the differences are significant, but not in another situation (i.e. for all diagnoses when countries are pooled together compared to Italy or Spain).

Discussion

Compared to the SCID-I/NP, results concerning the CIDI-SF and the CIS-R are quite different. On one hand, the CIDI-SF is likely to be sensitive and specific, but on the other, the PPV is very fair. The CIDI-SF gives doubtful detection of false positives. The risk in using the CIDI-SF is that it detects more cases than expected. On the other hand, with the CIS-R which appears to be more specific than sensitive, the risk is of underestimation of the number of cases. The fact that the CIS-R has medium PPV indicates also a medium reliability in detecting true positives.

In addition, divergences between countries have been found concerning the performance of both instruments in detecting cases. The Italian and the Spanish results are globally good, whereas the results from Romania are good with the CIDI-SF and medium with the CIS-R. The results are globally poor for the French survey and for both instruments.

In consequence, if we require a good instrument for detecting cases, even if the instrument overestimates the number of cases and classifies non-cases as cases, the CIDI-SF might be preferred. On the contrary, if the main objective is to detect true negatives and have a better

confidence concerning the true positives, but with the certainty that some cases will be missed, the CIS-R may be chosen. One important factor that allows choosing between both lay questionnaires is the fact that the CIDI-SF gives globally better AUC values.

The results obtained by the WMH authors on the comparison between the CIDI 3.0 and SCID are very similar to those obtained when AUC is considered: for mood disorder, 0.73 versus 0.75; for anxiety disorders, 0.72 versus 0.73; and for any diagnoses, 0.73 versus 0.76. In addition, those results are very similar across countries and languages except for mood disorders in one of the countries.

The results using kappa are lower than in the WMH study; however, kappa is very sensitive to prevalence and when a country (France) is removed, the results become much closer.

It is striking that the abridged version gives such a closer result than the full CIDI version at least for AUC. It is also noteworthy that results are stable across countries, but it could be argued that all the countries included spoke 'latine' languages, which are much more similar than English, and a fortiori very different languages, such as Chinese or Yoruba.

The CIDI-SF seems to have the same tendency than the full CIDI, the sensitivity of which has also been considered to be inadequate [1] but to a higher extent. This may be explained by the absence of organic exclusion or bereavement items in this questionnaire, and questions that evaluate the clinical significance of the reported symptoms [33]. In addition, difficulty in assessing 'clinically significant distress or impairment' was one of the reasons why the authors of the CIDI-SF did not take these notions into account when developing CIDI-SF syntax. These authors argued that a person who would present dysphoria or anhedonia 'for most of the day, nearly everyday' for 2 weeks or more would be considered to show clinically significant distress [35].

An alternative explanation for the low sensitivity of the CIDI-SF may relate to differences in the time periods considered by the different instruments. Both the SCID-I/NP and the CIS-R provide data on 1-month prevalence rates (the SCID-I/NP provides lifetime prevalence rates as well), whereas the CIDI-SF provides 12-month prevalence rates. However, this is unlikely to be the principal explanation for the discrepancy, since it is the CIDI-SF, rather than the CIS-R, that shows a better concordance with the SCID-I/NP. Moreover, since anxiety and depressive disorders tend to be chronic, the time base of the prevalence estimate is unlikely to be a significant source of bias.

One possibility is that the thresholds or cut points used in the CIDI-SF to determine the detection may vary depending on the countries or are incorrect for such population. For instance, for the CIDI-SF major depression

disorder, the threshold is superior or equal to 3, which corresponds to a probability of 0.55 to have the same trouble with the full CIDI [22].

In addition, this primary care population was selected because it was expected to be sufficiently representative of the general population, and to contain sufficient cases of mental health disorders to generate prevalence rates that could be compared in a meaningful way between countries.

All subjects are successively exposed to all three instruments. Although the questionnaires are proposed in a random order, so as not to influence responses and to reduce fatigue or boredom associated with responding to multiple questions, we cannot assess whether replies to any of the questionnaires are influenced by fatigue. Fatigue bias may thus contribute to inconsistency between questionnaires. For example, tired participants may have wanted to interrupt the interview, and thus endeavoured to minimise symptoms and to respond in a negative way. Differences between diagnostic instruments may thus be explained by the fact that during the second phase of the interview, respondents may be less likely to report symptoms [7] or may simply deny diagnostic questions [22].

The attitude of the respondent to the instrument can also be a source of bias. We noted that subjects' attitudes to the CIS-R are incompatible with our initial hypothesis that the CIS-R would be better accepted by the sample population because of its friendly question structure compared to the CIDI-SF, which is often perceived as a clinical interview that could negatively influence responses. In fact, both interviewers and interviewees considered the CIS-R to be less accessible than the CIDI-SF. In particular, participants found that the CIS-R was too long and too fastidious, which incited some participants to respond in a negative way and to stop the interview prematurely. We could thus hypothesise that participants' attitudes to the instrument used may influence replies to diagnostic questions. Interestingly, these remarks are similarly reported in all countries.

Another one possible explanation is the effect of interviewing methods on the respondents' responses. The experienced clinicians who use the semi-structured instrument have the possibility to go deep into the symptoms and to take time to reformulate questions. With the structured interviewing method, the lay interviewers and the respondents are guided by the questionnaire and items choices, which restrict the possibility of obtaining precise information.

In addition, the interviewers and the way the instruments have been proposed can also be a source of bias. As we mentioned previously, the interviewers have a tendency to consider the CIS-R to be much longer and fastidious than the CIDI-SF. This feeling may impact on the way the CIS-R had been administered. The fact that the CIS-R obtains fewer observations can be explained by the difficulties of

some interviewers in tackling it. One of the problems with the CIS-R is the difficulty to follow the different steps and to calculate for each diagnosis section the number of points, before jumping to another diagnosis section. On the contrary, for the CIDI-SF, no calculation is needed between the diagnosis sections and the interviewers have to follow simple skips after each question. The CIDI-SF is clearly easier to administrate and can be easily self-administered.

Conclusion

In brief, lay questionnaires, such as the CIDI-SF and the CIS-R, are used to reduce cost by eliminating the need for an experienced clinician and to reduce the time duration of the administration. However, because of the need of acceptable and feasible length and low cost investment, limitation and loss of validity can be found.

From our results, we have quite acceptable validity and performance values for both test instruments compared to the gold standard instrument, though it is not perfect. It is clear that the considerable differences in the interviewing methods used for the SCID-I/NP and both lay questionnaires can explain a part of the surprising differences between prevalence rates observed. In addition, we are quite suspicious about the validity of the SCID-I/NP in large-scale general population surveys. Test and re-test reliability studies on the SCID are far from perfect and describe strong variability in the kappa values in the general population [17].

Though our results suggest a lack of accuracy of case identification for the lay questionnaires, our choice of an instrument is the CIDI-SF. In our study, the performance of the CIDI-SF as a lay diagnostic instrument is globally better for the anxiety disorders and, by consequence, for the overall diagnoses than the CIS-R, compared to the SCID-I/NP, though performance disparities are found between countries. The CIDI-SF results are surprisingly quite similar to the CIDI 3.0 and better than CIS-R, compared to the SCID-I/NP. In addition, the CIDI-SF is well accepted by the interviewers and the interviewees. Consequently, the CIDI-SF seems to be a good choice for determining globally the prevalence of the most frequent diagnoses, mainly depressive and anxiety disorders in the general population when time is restricted, for example in a phone survey. However, researchers have to take into consideration the CIDI-SF's tendency to overestimate the false positives. This overestimation can be a disadvantage to some public researches. Nonetheless, it is cautious to choose an instrument that overidentifies false positives than an instrument that overidentifies false negatives. It is also important to highlight that the respondents considered to be

false positive by the lay questionnaires feel symptoms, which are highly associated with the presence of the corresponding disorder. The possible limitations of this survey are the use of GP's population that is more likely to have borderline disorders, the type of interviewing methods used and, finally, the cutoff points used by the CIDI-SF, which could explain the overestimation.

Furthermore, better estimating the cutoff point's level of sensibility, adding some impairment measures and adding bereavement and organic restriction items may increase the performance of the CIDI-SF questionnaire for the selected diagnoses.

Thus, we would like to argue that our results and prevalence rates gathered in a large-scale survey with short instruments like the CIDI-SF and the CIS-R must and can be interpreted in a cautious manner. Information has to be given on the instrument limits and methodology biases. However, the short instruments are good and useful tools to provide preliminary pictures of the general population's mental health. Given the importance of the burden of mental health and its high co-morbidity with many physical disorders, especially chronic illnesses, such short instruments could easily fit into any health-related survey.

Conflict of interest statement The paper presents results of the MINDFUL project conducted with the EA 4069 Paris Descartes University and the MGEN Foundation for Public Health, which received funding from the European Union, within the framework of the Public Health Program Grant No. 2003119.

References

1. Aalto-Setälä T, Haarasilta L, Marttunen M, Tuulio-Henriksson A, Poikolainen K, Aro H et al (2002) Major depressive episode among young adults: CIDI-SF versus SCAN consensus diagnoses. *Psychol Med* 32:1309–1314
2. American Psychiatric Association (2002) Diagnostic and statistical manual of mental disorders (DSM-IV-TR), 4th edn. American Psychiatric Association, Washington, DC
3. Araya R, Rojas G, Fritsch R, Acuna J, Lewis G (2001) Common mental disorders in Santiago, Chile: prevalence and socio-demographic correlates. *Br J Psychiatry* 178:228–233
4. Beaudet MP (1997) Dépression. *Health Rep* 7:171–185
5. Bhui K, Bhugra D, Goldberg D (2000) Cross-cultural validity of the Amritsar Depression Inventory and the General Health Questionnaire amongst English and Punjabi primary care attenders. *Soc Psychiatry Psychiatr Epidemiol* 35(6):248–254
6. Botega NJ, Pereira WA, Bio MR, Garcia C Jr, Zomignani MA (1995) Psychiatric morbidity among medical in-patients: a standardized assessment (GHQ-12 and CIS-R) using 'lay' interviewers in a Brazilian hospital. *Soc Psychiatry Psychiatr Epidemiol* 30(3):127–131
7. Bromet EJ, Dunn LO, Connell MM, Dew MA, Schulberg HC (1986) Long-term reliability of diagnosing lifetime major depression in a community sample. *Arch Gen Psychiatry* 43:435–440
8. Brugha TS, Jenkins R, Bebbington P, Meltzer H, Lewis G, Farrell M (2004) Risk factors and the prevalence of neurosis and

- psychosis in ethnic groups in Great Britain. *Soc Psychiatry Psychiatr Epidemiol* 39:939–946
9. Brugha TS, Meltzer H, Jenkins R, Bebbington PE, Taub NA (2005) Comparison of the CIS-R and CIDI lay diagnostic interviews for anxiety and depressive disorders. *Psychol Med* 35(7):1089–1091
 10. Brugha TS, Weich S, Singleton N, Lewis G, Bebbington PE, Jenkins R, Meltzer H (2005) Primary group size, social support, gender and future mental health status in a prospective study of people living in private households throughout Great Britain. *Psychol Med* 35:708–714
 11. Burton JW, Singleton N, Melzer D (2004) The mental health of early retirees: national interview survey in Britain. *Soc Psychiatry Psychiatr Epidemiol* 40:99–105
 12. Cho H, Costa E, Menezes P, Chalder T, Bhugra D, Wessely S (2007) Cross-cultural validation of the Chalder Fatigue Questionnaire in Brazilian primary care. *J Psychosomatic Res* 62(3):301–304
 13. Commander MJ, Sashi Dharan SP, Odell SM, Surtees PG (1997) Access to mental health care in an inner city health district. I: pathways into and within-specialist psychiatric services. *Br J Psychiatry* 170:312–316
 14. Cooke D, Newman S, Sacker A, Devellis B, Bebbington P, Meltzer H (2007) The impact of physical illnesses on non-psychotic psychiatric morbidity: data from the household survey of psychiatric morbidity in Great Britain. *Br J Health Psychol* 12(Pt 3):463–471
 15. DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating curves: a nonparametric approach. *Biometrics* 44:837–845
 16. ESEMeD/MHEDEA 2000 Investigators (2004) Prevalence of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta Psychiatr Scand* 109(Suppl 420):21–27
 17. First MB, Spitzer RL, Gibbon M, Williams JBW (2002) Structured clinical interview for DSM-IV Axis I Disorders, Research Version, Non-Patient Edition (SCID-I/NP). Biometrics Research, New York State Psychiatric Institute, New York. http://www.scid4.org/psychometric/scidI_validity.html
 18. Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143:29–36
 19. Haro JM, Arbabzadeh-Bouchez S, Brugha TS, De Girolamo G, Guyer ME, Jin R, Lepine JP, Mazzi F, Reneses B, Vilagut G, Sampson NA, Kessler RC (2006) Concordance of the composite international diagnostic interview Version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO World Mental Health Surveys. *Int J Methods Psychiatr Res* 15(4):167–180
 20. Isometsa E, Aro S, Aro H (1997) Depression in Finland: a computer-assisted telephone interview study. *Acta Psychiatr Scand* 96(2):122–128
 21. Jordanova V, Wickramasinghe C, Gerada C, Prince M (2004) Validation of two survey diagnosis interviews among primary care attendees: a comparison of CIS-R and CIDI with SCAN ICD-10 diagnosis categories. *Psychol Med* 34:1013–1023
 22. Kessler RC, Wittchen HU, Abelson J, Zhao S (2000) Methodological issues in assessing psychiatric disorder with self-reports. In: Stone AA, Turrkan JS, Bachrach CA, Jobe JB, Kurtzman HS, Cain VS (eds) *The science of self-report: implications for research and practice*. Lawrence Erlbaum Associates, Mahwah, pp 229–255
 23. Kessler RC, Andrews G, Mroczek D, Üstün B, Wittchen HU (1998) The World Health Organization composite international diagnostic interview short form (CIDI-SF). *Int J Methods Psychiatr Res* 7:171–185
 24. Korkeila J, Lehtinen V, Bijl R, Dalgrad OS, Kovess V, Morgan A, Salize HJ (2003) Establishing a set of mental health indicators for Europe. *Scand J Public Health* 31:451–459
 25. Kovess V, The State of Mental Health in The European Union Collaborators. (2004). *The State of Mental Health in the European Union*. European Communities. http://ec.europa.eu/health/ph_projects/2001/monitoring/fp_monitoring_2001_frep_06_en.pdf. Accessed 26 July 2007
 26. Kovess V, Fournier L, Lesage A, Amiel-Lebigre F, Caria A (2001) Two validation studies of the CIDI: a simplified version of the CIDI. *Psychiatric Netw* 4:10–24
 27. Lehtinen V, Sohlman B, Kovess-Masféty V (2005) Level of positive mental health in the European Union: results from the Eurobarometer 2002 survey. *Clin Pract Epidemiol Ment Health* 1:9
 28. Lewis G, Pelosi AJ (1990) *Manual of the revised clinical interview schedule (CIS-R)*. Institute of Psychiatry, London
 29. Lewis G, Pelosi AJ, Araya R, Dunn G (1992) Measuring psychiatric disorder in the community: a standardized instrument for use by lay interviewers. *Psychol Med* 22:465–486
 30. Liu SI, Mann A, Cheng A, Tjung JJ, Hwang LC (2004) Identification of common mental disorders by general medical doctors in Taiwan. *Gen Hosp Psychiatr* 26(4):282–288
 31. National population health survey- public use microdata files-household component. (1998). Statistics Canada, Ottawa. Cat no. 82M0009XCB
 32. Patten SB, Brandon-Christie J, Devji J, Sedmak B (2000) Performance of the composite international diagnostic interview short form for major depression in a community sample. *Chronic Dis Can* 21:68–72
 33. Patten SB, Wang JL, Beck CA, Maxwell CJ (2005) Measurement issues related to the evaluation and monitoring of major depression prevalence in Canada. *Chronic Dis Can* 26(4):100–106
 34. Perneger TV, Leplège A, Etter JF (1999) Cross-cultural adaptation of a psychometric instrument: two methods compared. *J Clin Epidemiol* 52(11):1037–1046
 35. Walters EE, Kessler RC, Nelson CB, Mroczek D (2002) Scoring the World Health Organization's composite international diagnostic interview short form (CIDI-SF). <http://www.who.int/msa/cidi/cidifscorint.pdf>
 36. Weich S, Nazroo J, Sproston K, McManus S, Blanchard M, Erens B, Karlsen S, King M, Lloyd K, Stanfeld S, Tyrer P (2004) Common disorders and ethnicity in England: the EMPIRIC study. *Psychol Med* 34(8):1543–1551
 37. Wing JK, Babor T, Brugha TE, Burke J, Cooper JE, Giel R, Jablenski A, Regier D, Sartorius N (1990) SCAN: schedules for clinical assessment in neuropsychiatry. *Arch Gen Psychiatry* 47:589–593
 38. Wittchen HU, Jacobi F (2005) Size and burden of mental disorders in Europe: a critical review and appraisal of 27 studies. *Eur Neuropsychopharmacol* 15:357–376
 39. World Health Organisation (1990) *The composite international diagnostic interview (version 1.0)*. World Health Organisation, Geneva
 40. World Health Organization (1992) *The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines*. World Health Organisation, Geneva