VU Research Portal

Low-Intensity Screening and Treatment for Common Mental Disorders

Donker, T.

2010

document version
Publisher's PDF, also known as Version of record

citation for published version (APA)
Donker, T. (2010). *Low-Intensity Screening and Treatment for Common Mental Disorders*. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam].

Chapter 4

Brief self-rated screening for depression on the internet

Abstract

The internet offers promising possibilities for the quick screening of depression for treatment and research purposes. This paper aims to validate three self-rated measures to screen for depression on the internet: SID (single-item depression scale), CES-D (Center for Epidemiological Studies Depression scale) and K10 (Kessler psychological distress scale). Of the 502 subjects aged 18-80 who rated the SID, CES-D and K10 measures on the internet, 157 (31%) subjects were also interviewed by telephone using the WHO Composite International Diagnostic Interview (CIDI) for DSM-IV-disorders.

Cronbach's alpha for both web self-rated measures CES-D and K10 was 0.90. The SID correlated 0.68 (P<.001) with the CES-D and with the K10. The CES-D correlated 0.84 with the K10 (P<.001). Subjects with a DSM-IV diagnosis for any depressive disorder had significantly higher means (P<.001) on the three self-rated measures for depressive symptoms than subjects without a diagnosis of any depressive disorder. Using any depressive disorder as the gold standard, the area under the curve (AUC) of the SID was 0.71 (95% CI: 0.63-0.79), which was significantly lower than the AUC of the CES-D (AUC: 0.84; 95% CI: 0.77-0.90, P=0.003) and of the K10 (AUC: 0.81; 95% CI: 0.73-0.88, P=0.0024). The AUCs for the K10 and CES-D did not differ significantly from each other. One limitation is that the CIDI-interviews were not recorded, so inter-rater reliability could not be calculated.The CES-D and K10 are reliable, valid tools for care providers to quickly screen depressive patients on the internet and for researchers to collect data.

Keywords: Depression; Screening instruments; Internet use; Validation

1. Introduction

Estimations of 12-month prevalence rates of depressive disorders vary between 6.6% (Andrews et al., 2001) to 11% (Kessler et al., 1994) in the general population and 12% in a primary care population (Sartorius et al., 1996). Depressive disorders are among the top four leading causes of disease burden worldwide (Lopez and Murray, 1998). They cause serious disability (van Schaik et al., 2007), reduced quality of life (Cuijpers et al., 2004), and incur huge economic costs (Cuijpers et al. 2007b). Their identification and treatment are therefore important. Screening for such disorders, using reliable and valid self-report questionnaires, needs to be improved and made more user-friendly, both to researchers and health care providers.

Detection would improve with simpler, shorter, more reliable and valid depression self-rating (U.S.Preventive Services Task Force, 2002). This kind of improved self-rating would be conducted more readily by subjects (Cuijpers et al., 2009) and save time for health care providers who have many competing demands on their time.

Screening conducted via the internet offers easy and quick access to large numbers of users at low cost (Austin et al., 2006; Buchanan, 2003). Collecting data on the internet saves researchers time and organization, minimizes data-entry errors by allowing automatic transcription into a computerized database (Coles et al., 2007), and can reduce missing values by making responses to all items obligatory before submission (Austin et al., 2006). Moreover, people sometimes disclose more sensitive information in computer-based compared to face-to-face interviews (Buchanan 2002; Davis 1999; Joinson 1999) as 'the computer has no eyebrows' (Marks et al., 2007). However, there are several disadvantages of web-based screening as well. For example, the anonymous nature of the Internet allows people to participate frivolously or with malicious intent, which can affect the data quality. Furthermore, regarding ethical principles, it is more difficult to assess the subjects' identities or their reactions to the research experience online in case surveys might upset the test-taker (Kraut et al., 2004).

Though several studies have found equivalent psychometric properties in web-based versus paper-pencil questionnaires (Andersson et al., 2003; Carlbring et al., 2007a; Houston et al., 2001; Spek et al., 2008), other studies did not (reviewed by Buchanan (2002), which means that the psychometric equivalence cannot be assumed (Buchanan 2002; 2003). One factor which might affect the reliability and validity of self-ratings on the internet is variability in presentation of the test across different computers due to technical discrepancies between different hardware and software configurations (Austin, et al., 2006; Buchanan and Smith, 1999). Furthermore, the heterogeneity (e.g. age, education, socio-economic status) of internet users is increasing which may introduce unknown confounding variables, possibly adding to 'noise' in the data and reducing the proportion of variance in responses accounted for by differences in whatever (e.g. depression) one is trying to measure (Buchanan and Smith, 1999). Variations in the amount of control over the testing environment (e.g. at home versus the lab, and in rater mood or fatigue) might influence the validity of web questionnaires (Buchanan and Smith, 1999; Davis, 1999), as they do too for paper-pencil administration. And, as mentioned earlier, social-desirability effects might be less pronounced for web-based than paper-pencil administration (Joinson, 1999). It is therefore necessary to check the

validity of rating each measure on the internet before adopting it (Buchanan, 2002; Buchanan and Smith, 1999).

This study aims to validate internet-based screening of depression by three self-rated measures – the Center for Epidemiological Studies Depression Scale (CES-D; Radloff, 1977) and the Kessler psychological distress scale (K10; Kessler and Mrozeck, 1994), and the Single-Item Depression (SID; McKenzie and Marks, 1999) scale. Selection was based on the psychometric properties of the paper-pencil versions, their understandability, and their availability without charge. Diagnosis in a standard diagnostic interview was used as the 'gold standard'.

2. Methods

2.1 Participants and procedure

Data for this study were collected as part of a larger investigation of a brief, web-based screener (WSQ) for common mental disorders (detailed in Donker et al., submitted). In short, participants were recruited from the general population by using Internet banners (Google, Dutch internet-sites on mental health issues). We targeted adults aged 18 or older who were anxious, depressed or thought of themselves as drinking too much alcohol – the kind of people for whom the WSQ is intended. We expected them to have a high rate of common mental disorders. In order to test the false and true negative rate of the WSQ, we also recruited 20 undergraduate psychology students who were not required to have symptoms, using banners at the Vrije Universiteit (VU) students-web page for participation in VU studies. The study included participants who completed the self-ratings and gave informed consent; it excluded people who reported suicide plans (detected by a self-rated question on the WSQ). It was approved by the Medical Ethics Committee at the VU Medical Centre in Amsterdam.

The study consisted of two parts:

1. A total of 502 participants responded on the internet to demographic questions, the SID, CES-D and K10 -details below. Participants were allowed to return to any of the measures for review or changes.
2. *Of the 502 who completed the internet measures, a total of 157 participants also completed a telephone DSM-IV-diagnostic interview with a Composite International Diagnostic Interview (CIDI)-trained interviewer (CIDI lifetime, WHO version 2.1(World Health Organization, 1997).*

A total of 687 people applied for the study, of whom 185 (27%) were excluded on the following grounds: high suicide risk (n=5), no written informed consent (n=22) or refusal to participate (n=158). This left 502 participants. Of these, 118 refused to participate in the diagnostic telephone interview. A further 227 could not be contacted leaving 157 participants. A CIDI-trained interviewer phoned 157 participants within a mean of 13 days to conduct a CIDI assessment of the presence of a current (last 6 months) DSM-IV diagnosis (American Psychiatric Association, 1994) of major depressive disorder and dysthymia. The CIDI interviews were not recorded.

2.2 Measures

Participants rated the online measures in a fixed order. They were allowed to return to any of the measures for review or changes.

2.2.1 SID

The Dutch version of the single-item Depression Scale (SID; McKenzie and Marks, 1999) has one self-rated item ("How much are you troubled by feeling miserable or depressed"). It inquires about current symptoms and can be scored on a range of 0 to 8 (0=hardly at all, 2 = slightly, 4 = definitely, 6 = markedly, 8= very severely disturbing/disabling). When compared with the 21 item Revised Beck Depression Inventory-21 (BDI; Beck et al., 1979), the English version of the SID had acceptable reliability and validity ($r = 0.71-0.78$; cut-off score 4: sensitivity 0.73, specificity: 0.87; cut-off score 6: sensitivity: 0.71, specificity: 0.90).

2.2.2 CES-D

The Dutch version of the Center for Epidemiological Studies Depression scale (CES-D; Radloff, 1977) has 20 self-rated items inquiring about symptoms over the past week; each item is scored 0-3; total score range is 0-60. Among different populations, the paper-pencil CES-D has acceptable psychometric properties with a cut-off score of 16 (sensitivity: 0.84-1.00, specificity: 0.69-0.90) (Beekman et al., 1997; Parikh et al., 1988; Wada et al., 2007; Whooley et al., 1997). The Dutch Internet CES-D is also reliable and valid with a cut-off score of 22 (Cronbach's alpha: 0.93; AUC: 0.90; sensitivity: 0.90; specificity: 0.74) in an adolescent population (Cuijpers et al., 2008).

2.2.3 K10

The Kessler psychological distress scale (K10; Kessler and Mroczek, 1994) has 10 self-rated items inquiring about symptoms over the past month. Each item scored 1-5; total score range 10-50 (Andrews and Slade, 2001) and has good psychometric properties in screening for depressive disorders (Cairney et al., 2007; Furukawa et al., 2003; Kessler et al., 2003). The K10 was translated into Dutch using the World Health Organisation (WHO) translation protocol applied for the Epidemiology of Mental Disorders (ESEMeD) study. The Dutch paper-pencil version of the K10 is also reliable (Cronbach's alpha: 0.93), and valid, with a cut-off score of 20 (AUC: 0.87; sensitivity: 0.90; specificity: 0.74) (Donker et al., in press).

2.2.4 Diagnoses

We used the Lifetime version 2.1 of the CIDI (World Health Organization, 1997) in its Dutch version (Janca et al., 1994; Smeets and Dingemans, 1993) as a 'gold standard' to assess the presence of DSM-IV disorders in the last 6 months (Major Depression, Dysthymia). The CIDI is reliable and valid (Andrews and Peters, 1998; Wittchen, 1994). It was administered by telephone, by trained CIDI interviewers who were qualified psychologists or master-level psychology students. There are several studies providing qualified justification for the use of telephone interviews to

collect data of mental health disorders (Evans et al., 2004; Rohde et al., 1997; Willemse et al., 2004).

2.3 Statistical analyses

We tested the reliability of the three screening measures (SID, CES-D, K10) by their internal consistency (Cronbach's α), and calculated their associations by Pearson correlation coefficients.

Construct validity was assessed by testing whether scores on the SID, CES-D and K10 differed significantly between subjects with and without DSM-IV depressive disorders, using t-tests on the means and standard deviations of each self-rated measure separately.

Predictive validity was calculated by analyzing whether the Area Under the Curve (AUC) of sensitivity versus (1 – specificity) of the SID differed significantly from the AUC of the CES-D and the K10. The AUC of sensitivity versus (1 –) specificity measures a scale's accuracy; it equals the probability that a randomly chosen case will score higher than a randomly chosen non-case (Cairney et al. 2007). AUCs of 0.5 to 0.7 are said to reflect low accuracy, 0.7 to 0.9 moderate accuracy, and 0.9 to 1.0 high accuracy (Fischer et al., 2003). The AUC curve is used to generate the AUC summary statistic.

We calculated sensitivity (the probability that someone who has the disorder screens positive) and specificity (the probability that someone without the disorder screens negative). In addition, we determined the best cut-off score by deriving the set which generates the highest sum of sensitivity and specificity coefficients (Leentjens, et al., 2000).

To establish whether scores on the SID, CES-D and K10 differed significantly between subjects with and without a CIDI-interview, we conducted t-tests on the mean and standard deviation of each screening instrument separately.

Our analyses used diagnoses subjects reached within the last 6 months in the categories 'any depressive disorder' (major depressive disorder, dysthymia and minor depression) and 'depressive plus anxiety' (any depressive disorder plus at least one co-morbid anxiety disorder).

All analyses used SPSS version 15.0 for Windows, except that equivalence of the AUCs was tested with STATA SE9.

3. Results

3.1 Demographics

The total sample (N=502) had a mean age of 43 (SD 13, range 18-80); and 285 (57%) of the subjects were female; the majority was Dutch (n= 474, 94%) and 217 (43%) subjects received medium education (Intermediate Vocational Training [community college], school of higher general secondary education or pre-university education). Of the 157 subjects who had a CIDI interview, the mean age was 43 (SD 15, range 18-80); 89 (57%) were female; the majority was Dutch (n=146, 94%) and 73 (47%) subjects received medium education. Of the 157 subjects who were CIDI interviewed, 52 (33%) subjects met DSM-IV criteria for any depressive disorder (within the past six months). Of them, 44 (28%) patients met DSM-IV criteria for depressive disorder but did have a co-morbid anxiety disorder.

Of the subjects with any depressive disorder, 46 (29%) had a major depressive disorder, 9 subjects had a dysthymia (6%) and 8 subjects met the DSM-IV criteria for a minor depression (5%).

3.2 Reliability (Cronbach's alpha) of the SID, CES-D and K10

In the total sample (N=502), the reliability (Cronbach's alpha) for the CES-D was 0.92, and for the K10 it was 0.90. As the SID consists of only one item, Cronbach's alpha could not be calculated.

3.3 Correlations between the SID, CES-D and K10

Correlations among the three measures ranged from 0.68 of the SID with the CES-D and with the K10 (both P<0.001) to 0.84 of the CES-D with the K10 (P<.001).

3.4 Comparison of means and standard deviations between CES-D, K10, SID and CIDI DSM-IV diagnosis of depression and anxiety

As shown in Table 1, subjects with a CIDI DSM-IV diagnosis for any depressive disorder (n=52) had significantly higher means on all three self-rated measures for depressive symptoms than did subjects without any depressive disorders (SID: $t[149.29]=5.60$, P<0.001; CES-D: $t[155]=8.04$, P<0.001; K10 $t[155]=7.11$, P<0.001). Similarly, subjects with a depressive disorder plus at least one anxiety disorder (n=44) also had significantly higher means on the three self-rated measures for depressive symptoms than did subjects without any depressive disorder (SID: $t[138.35]=5.86$, P<0.001; CES-D: $t[155]=6.98$, P<.001; K10: $t[155]=6.67$, P<0.001).

Table 1. Means and standard deviations of the internet-based self-rated SID, CES-D and K10, in subjects with a current DSM-IV depressive disorder (6-month prevalence, n=157)

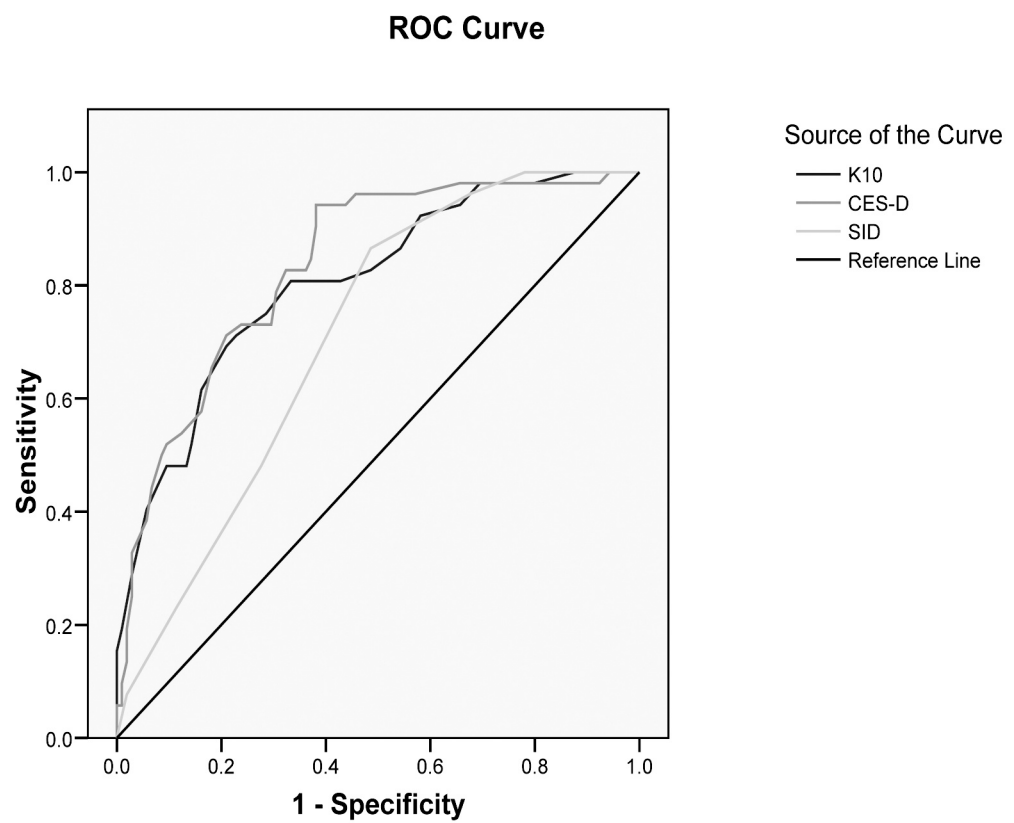|	N	%	SID			CES-D			K10		
			M	SD		M	SD		M	SD	
<u>No depr or anx disorder</u>	55	35	3.5	2.0		17.8	11.0		23.0	7.3	
Any depressive disorder	52	33.1	5.6	1.2	*	34.0	9.1	*	33.5	6.9	*
Depr + anx disorder[1]	44	28	5.7	1.1	*	34.1	9.5	*	33.9	6.2	*

[1] subjects who had a depressive disorder plus at least one co-morbid anxiety disorder
*significant at $p<0.001$

3.5 AUC of the SID, CES-D and K10

Using any CIDI depressive disorder as a gold standard, the AUC of the SID was 0.71 (95% CI: 0.63-0.79), which was significantly lower than the AUC of the CES-D (AUC: 0.84; 95% CI: 0.77-0.90, χ^2 (1) =8.89, P=0.003) and K10 (AUC: 0.81; 95% CI: 0.73-0.88, χ^2 (1) =5.09, P=0.0024) (Figure 1, Table 2). The areas under the curve for the K10 and CES-D did not differ significantly from each other (χ^2 (1) =1, P=.32).

Figure 1. AUC-curves of the SID, CES-D and K10 for any depressive disorder

ROC Curve

ROC = Receiver Operator Characteristic
* Diagonal line represents reference line that is of no diagnostic value

Table 2. Area Under the Curve (AUC) for the internet-based self-rated SID, CES-D and K10 in subjects with a current DSM-IV depressive disorder (6-month prevalence, n=157)

	SID AUC (95% CI)	CES-D AUC (95% CI)	K10 AUC (95% CI)
Any depressive disorder*	0.71 (0.63-0.79)	0.84 (0.77-0.90)	0.81 (0.73-0.88)
Depressive + anxiety disorder[1]*	0.72 (0.64-0.80)	0.82 (0.74-0.89)	0.81 (0.74-0.89)

[1] subjects who had a depressive disorder plus at least one co-morbid anxiety disorder
*p<.05 AUC for the SID was smaller than the AUC for the CES-D and K10

3.6 Sensitivity and specificity and optimal cut-off points for the SID, CES-D and K10

Table 3 shows sensitivities and specificities for the optimal cut-off points for each of the three web-based self-rated measures for any depressive disorder. Using a cut-off score where the sum of sensitivity and specificity was maximum, the best cut-off score for the SID was 5 (sensitivity: 0.87; specificity: 0.51) and for the CES-D it was 22 (sensitivity: 0.94; specificity: 0.62). Several cut-off points yielded the highest sensitivity + specificity coefficients for the K10 (cut-off score 29, 31, 32; sensitivity: 0.69-0.81; specificity: 0.67-0.79).

Table 3. Sensitivity (sens) and specificity (spec) for different cut-off scores of internet-based self-rated SID, CES-D and K10 for subjects with any depressive disorder (6-month prevalence, n=157)

	SID			CES-D			K10		
	Cut-off	sens	spec	Cut-off	sens	spec	Cut-off	sens	spec
Any depressive disorder	1	1.00	0.07	15	0.98	0.31	18	0.98	0.20
	2	1.00	0.11	16	0.98	0.34	19	0.98	0.24
	3	1.00	0.22	17	0.96	0.43	20	0.98	0.25
	4	0.96	0.32	18	0.96	0.49	21	0.98	0.31
	5	0.87	0.51 ¹	19	0.96	0.54	22	0.94	0.34
	6	0.48	0.72	20	0.94	0.56	23	0.92	0.42
	7	0.23	0.89	21	0.94	0.59	24	0.87	0.46
	8	0.08	0.98	22	0.94	0.62 ¹	25	0.85	0.49
	9	0.00	1.00	23	0.90	0.62	26	0.83	0.51
				24	0.85	0.63	27	0.81	0.57
				25	0.83	0.64	28	0.81	0.60
				26	0.83	0.68	29	0.81	0.67 ¹
				27	0.79	0.70	30	0.75	0.71
				28	0.73	0.71	31	0.71	0.77 ¹
				29	0.73	0.72	32	0.69	0.79 ¹
				30	0.73	0.76	33	0.62	0.84
				31	0.72	0.79	34	0.52	0.86
				32	0.65	0.82	35	0.48	0.87

¹ cut-off score for which the sum of sensitivity and specificity is maximum

3.7 Differences between CIDI Interviewed and non-interviewed sub-samples

Demographic variables and scores on the SID and CES-D did not differ significantly between subjects who had a CIDI diagnostic interview and those who did not. However, those who had a CIDI interview scored significantly lower on the K10 (CIDI interviewed: M=25 [SD 8.2] vs. CIDI not interviewed: M=27 [SD 8.3], p=.04).

4. Discussion

Findings from our study suggest that both the web-based CES-D and web-based K10 yield reliable (Cronbach's α 0.90-0.92) and valid (AUC 0.81-0.84) self-ratings for depressive disorders which are similar to the paper-pencil ratings (Beekman et al., 1997; Donker et al., in press). The SID was moderately accurate (AUC 0.71) and a cut-off score of 5 gave high sensitivity (0.87) but lower specificity (0.51) compared to findings from previous research (McKenzie and Marks, 1999). Reducing the number of cases of depression that are missed increases the sensitivity of self-rated measures. The brevity and high sensitivity of the SID makes it suitable for routine use by hard-pressed professionals in primary care or mental health settings. Short questionnaires are also completed more readily by subjects (Cuijpers et al., 2009). The SID is a useful first step. If subjects screen positive on the SID (e.g. cut-off score ≥3: sensitivity 1.00, specificity 0.22), then a second step

can be a further screening with the CES-D or K10. Collecting scores on internet-rated measures can save researchers much time and effort.

We found that the optimal cut-off score (highest sum of sensitivity plus specificity coefficients) in our study for the web-based CES-D and web-based K10 (scores of 22 and 29 respectively) were much higher than for the paper-pencil versions (16 and 20 respectively [Beekman et al., 1997; Donker et al., in press]), but for the web-based CES-D accorded with previous research (cut-off score of 22; Cuijpers et al., 2008). Generally, sensitivities of the SID, CES-D and K10 were higher (0.81 to 0.94), and specificities were lower (0.51 to 0.67) with their internet versions than their paper-pencil versions obtained from previous research (sensitivity 0.73 to 1.00; specificity: 0.69 to 0.90 [Beekman et al., 1997, Donker et al., in press, McKenzie and Marks, 1999]). For detecting depressive disorders in people who were likely to have co-morbid anxiety disorders, the CES-D was no better than the K10, though both were better than the SID (P<0.001).

The high internal consistencies of the K10 and the CES-D on the internet resembled those of their paper-pencil versions; each measure in both versions seems to tap a single construct. However, the higher cut-off scores on their web-based compared to the paper-pencil versions may reflect participants' greater anonymity when rating sensitive issues on the net at home, which might increase honesty and self-disclosure. This might also explain our study's lower specificity compared to previous research (Cuijpers, et al., 2007a; Donker et al., in press; McKenzie and Marks, 1999), as the gold-standard CIDI diagnostic interviews conducted by humans on the phone; less anonymity could yield less self disclosure, more 'false' positives, and less specificity. Moreover, the trained interviewers questioned subjects thoroughly about symptoms, which too could account for our low specificity. Another explanation for the lower specificity of the three internet-rated measures might be that we have used 6-month prevalence rates as gold standard, whereas the SID assesses current symptoms, the CES-D inquires about symptoms of the past week and the K10 assesses symptoms of the past month, rather than symptoms during the previous 6 months. Therefore, specificity might be higher when the SID, CES-D and K10 are validated against concurrent DSM-IV diagnoses.

Participants were allowed to return to any of the measures for review or changes. Previous research suggests that by allowing backtracking (and adjust answers before submitting) in computerized questionnaires generates less social desirability (Richman, et al., 1999). Furthermore, there is growing evidence for the equivalence of paper and computerized forms of easily administered questionnaires that allow backtracking (e.g., Kobak et al., 1990, Rosenfeld et al., 1992).

One limitation of this study is that the CIDI interviews were not recorded, so inter-rater reliability could not be calculated. Another is that subjects always completed the SID, CES-D and K10 in a fixed order, and all CIDI interviews were administered after the internet-questionnaires, so order effects could not be ruled out. A third limitation may be that internet ratings were not done in a controlled setting, precluding control of the test situation regarding noise, distractions, mood, fatigue or intoxication which might increase rating variability (Buchanan and Smith, 1999) and reduce the generalisability of the findings. Similar lack of control, however, might apply to internet and paper-pencil ratings done in other 'real life' situations. Also, as the self-report measures were rated only by internet, there was

no direct statistical comparison with paper-pencil ratings. However, in a previous study of Ogles et al. (1998), a correlation of 0.96 was found for paper-pencil and computerized CES-D scores. This suggests that data from paper-pencil administration and computerized administration are identical for the CES-D. As far as we know, there are no direct comparisons between paper-pencil and computerized versions of the K10 and SID available. Also, there is no consensus on optimal levels of sensitivity and specificity, as these depend on the test's aim, costs and benefits (Smits et al., 2007). The sensitivity and specificity of a test are fixed features but the false-positive and false-negative rates depends on the prevalence of the disorder and are therefore depending on the setting in which the test is used. The benefits of a test are clear: detecting true positive and true negatives cases. The costs associated with false-positive and false-negative outcomes are more difficult to establish and are beyond the scope of this paper. It involves patient burden not averted by treatment in case of a false-negative and patient burden due to a false diagnosis and treatment in case of a false- positive. Another limitation is that, as described earlier, 6-month prevalence rates of DSM-IV diagnoses were used, whereas the SID assesses current symptoms, the CES-D inquires symptoms of the past week and the K10 inquires about symptoms over the past month. Ideally, the three self-rated measures should be validated against concurrent DSM-IV diagnoses. Another limitation is that subjects who had a CIDI interview scored significantly lower on the K10 than those who did not a CIDI interview, so the K10 results might be less generalisable to other populations. However, the differences in mean scores on the K10 between those who had a CIDI interview and those who did not were small (2 points), so the influence on sensitivity and specificity might be negligible.

Furthermore, a prevalence rate of 33.1% of subjects with a current CIDI DSM-IV depressive disorder was found in this study. Estimations of 12-month prevalence rates vary considerably between 6.6% (Andrews et al., 2001) to 11% (Kessler et al., 1994) in the general population and 12% in a primary care population (Sartorius et al., 1996). Furthermore, we found a prevalence rate of 28% with a depressed and a co-morbid anxiety disorder in our population, which is considerably higher compared to previous research in which a prevalence rate of 5% for co-morbid depressive and anxiety disorders was found among primary care participants (Sartorius et al. 1996). However, the prevalence rate of depressive and co-morbid anxiety disorders does not affect sensitivity and specificity. Positive and Negative Predictive Values are, however, influenced prevalence. When prevalence is high, which might be the case in self-selected samples such as those in this study, "true" negatives" will have a greater impact, and when prevalence is low, "true positives" have a higher impact on the NPV and PPV. When prevalence is low, a positive diagnosis from the CES-D, K10 and SID should be regarded with caution. Finally, making subjects rate all items may yield complete data at the cost of some of those ratings perhaps being unreliable when subjects do not agree with any of the options presented.

It would be worth assessing the value of the self-rated measures in primary care populations and specialized mental health care settings, and evaluating whether adding one or two questions (e.g. "do you want help for your problems") to the SID would increase its specificity.

In sum, though internet ratings yielded different results (in this case higher cut-off scores and sensitivity scores with lower specificity scores) compared to paper-

pencil ratings of the same measures, the CES-D and K10 can be reliable and valid tools to quickly screen for depression on the internet.

Contributors

All authors contributed to the design of this study. TD drafted the manuscript. All authors contributed to the further writing of the manuscript. All authors read and approved the final manuscript.