Composite International Diagnostic Interview screening scales for DSM-IV anxiety and mood disorders


Psychological Medicine / FirstView Article / October 2012, pp 1 - 13
DOI: 10.1017/S0033291712002334, Published online:

Link to this article: http://journals.cambridge.org/abstract_S0033291712002334

How to cite this article:

Request Permissions : Click here
Composite International Diagnostic Interview screening scales for DSM-IV anxiety and mood disorders


¹ Department of Health Care Policy, Harvard Medical School, Boston, MA, USA; ² Department of Psychiatry, University Hospitals Case Medical Center, Case Western Reserve University, Cleveland, OH, USA; ³ Clinical Services, EPI-Q, Inc., Oakbrook Terrace, IL, USA; ⁴ School of Medicine, University of Washington, Seattle, WA, USA; ⁵ Lindner Center of HOPE and Department of Psychiatry, University of Cincinnati College of Medicine, Cincinnati, OH, USA; ⁶ Depression Clinical and Research Program and the Bipolar Clinic and Research Program, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; ⁷ Columbia University School of Social Work, New York, NY, USA; ⁸ Department of Psychiatry, University of California San Diego, San Diego, CA, USA; ⁹ Departments of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia Veterans Affairs Medical Center, and the University of Pittsburgh Medical Center, Philadelphia and Pittsburgh, PA, USA; ¹⁰ Institute of Clinical Psychology and Psychotherapy, Technische Universität Dresden, Dresden, Germany

Background. Lack of coordination between screening studies for common mental disorders in primary care and community epidemiological samples impedes progress in clinical epidemiology. Short screening scales based on the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI), the diagnostic interview used in community epidemiological surveys throughout the world, were developed to address this problem.

Method. Expert reviews and cognitive interviews generated CIDI screening scale (CIDI-SC) item pools for 30-day DSM-IV-TR major depressive episode (MDE), generalized anxiety disorder (GAD), panic disorder (PD) and bipolar disorder (BPD). These items were administered to 3058 unselected patients in 29 US primary care offices. Blinded SCID clinical reinterviews were administered to 206 of these patients, oversampling screened positives.

Results. Stepwise regression selected optimal screening items to predict clinical diagnoses. Excellent concordance (area under the receiver operating characteristic curve (AUC)) was found between continuous CIDI-SC and DSM-IV/SCID diagnoses of 30-day MDE (0.93), GAD (0.88), PD (0.90) and BPD (0.97), with only 9–38 questions needed to administer all scales. CIDI-SC versus SCID prevalence differences are insignificant at the optimal CIDI-SC diagnostic thresholds ($\chi^2 = 0.0–2.9$, $p = 0.09–0.94$). Individual-level diagnostic concordance at these thresholds is substantial (AUC 0.81–0.86, sensitivity 68.0–80.2%, specificity 90.1–98.8%). Likelihood ratio positive (LR+) exceeds 10 and LR− is 0.1 or less at informative thresholds for all diagnoses.

Conclusions. CIDI-SC operating characteristics are equivalent (MDE, GAD) or superior (PD, BPD) to those of the best alternative screening scales. CIDI-SC results can be compared directly to general population CIDI survey results or used to target and streamline second-stage CIDIs.

Received 21 June 2012; Accepted 5 September 2012

Key words: Bipolar disorder, Composite International Diagnostic Interview (CIDI), generalized anxiety disorder, major depression, panic disorder, screening scales, validity.

Introduction

Although research on the community epidemiology of mental disorders (i.e. general population incidence, prevalence, risk factors, consequences) is an active area of investigation (Susser et al. 2006; Kessler & Üstün, 2008b; Tsuang et al. 2011), research on clinical epidemiology (i.e. prevalence, severity, long-term course in treatment samples) is underdeveloped, especially in primary care settings. Indeed, the most important clinical epidemiological study of mental disorders in primary care remains the World Health Organization (WHO) Collaborative Study on Psychological Problems in General Health Care (Üstün & Sartorius, 1995), a study carried out nearly two decades ago that led to few extensions (e.g. Wittchen et al. 2002; Barkow et al. 2003; Kisely & Simon, 2005;
Kisely et al. 2006). More sustained long-term clinical epidemiological studies exist in specialty treatment samples (Katz et al. 1979; Bruce et al. 2005), but those studies are now outdated because of changes in the composition of patient populations since the studies were initiated (Kessler et al. 2005).

We know from primary care screening studies that untreated mental disorders are common in primary care (Lowe et al. 2008; Gili et al. 2011). However, screening studies tell us little about the natural history of these disorders, as screening studies typically focus on current prevalence or treatment response. Yet information is needed on episode recurrence and onset of secondary disorders to understand the public health significance and long-term cost-effectiveness of primary care screening, outreach and treatment quality improvement (Barrett et al. 2005; Konnopka et al. 2009). This integration of primary care with public health is now an area of considerable policy interest (Committee on Integration to Improve Population Health, 2012).

One way to build a critical mass of such data would be to blend longitudinal clinical epidemiological studies with community epidemiological surveys. For example, several new community epidemiological surveys in the WHO World Mental Health (WMH) Survey Initiative (Kessler & Üstün, 2008a) are using a dual-frame sampling approach with parallel samples of (i) patients in primary care (both with and without detected and undetected mental disorders) and (ii) other household residents in the same communities. This design facilitates comparisons of illness prevalence course among treated and untreated cases by collecting successive snapshots of current prevalence of disorders over multiple points in time.

Screening scales will be used in the primary care segment of these surveys as the first stage in a two-stage approach to oversample patients with current mental disorders for second-stage interviews. Screening scale responses are being ‘preloaded’ in the computerized scripts of the second-stage interviews to control question skip logic (e.g. skipping sections based on negative screening responses; expanding questions based on positive screening responses). The screening scales used for this purposes are based on the WHO Composite International Diagnostic Interview (CIDI; Kessler & Üstün, 2004), the diagnostic interview used in the WMH surveys and most other psychiatric epidemiological surveys throughout the world. Psychometric analyses of these disorder-specific CIDI screening scales (CIDI-SC) have been reported previously for adult attention deficit/hyperactivity disorder (ADHD; Kessler et al. 2007), insomnia, (Kessler et al. 2010a) and overall serious mental illness (SMI; Kessler et al. 2010b). The current report presents comparable results for the CIDI-SC scales of major depressive episode (MDE), bipolar disorder (BD), generalized anxiety disorder (GAD) and panic disorder (PD). Although the results presented are for cross-sectional rather than longitudinal analyses, they are relevant for the longitudinal design described above because the latter is made up of a series of cross-sectional snapshots.

Method

Screening scale development

The CIDI

The CIDI is a fully structured research diagnostic interview developed for use by trained lay interviewers to generate diagnoses of lifetime and recent DSM-IV-TR/ICD-10 disorders (Robins et al. 1988). Clinical reappraisal studies document generally good concordance of CIDI diagnoses with blinded clinical diagnoses (Wittchen, 1994; Kessler et al. 1998). The CIDI uses extensive skip logic to reduce interview length. This skip logic is also used in the CIDI-SC based on the assumption that tablet computers will be used to administer, score and print out summary screening scale results.

Expanding the CIDI item pool

All CIDI symptom questions operationalize DSM/ICD criteria using simple descriptive language (Robins et al. 1988). However, validation studies find some CIDI questions less concordant than others with independent clinical assessments (Wittchen et al. 1995; Kessler et al. 2006; Green et al. 2011). We consequently expanded the CIDI item pool in developing the CIDI-SC scales by reviewing a wide range of other diagnostic instruments to generate alternative symptom questions. The expanded question set was reviewed iteratively, with diagnostic experts using their judgment to pinpoint alternative questions they considered potentially useful and to help revise and prioritize indicators. Previous methodological research has shown that such iterative expert review is often the most useful form of pretesting (Converse & Presser, 1986; Presser & Blair, 1994; Groves et al. 2009).

Pilot testing

Once preliminary symptom questions were generated, a convenience sample of 15 psychiatric out-patients with each diagnosis was administered the disorder-specific symptom questions. Cognitive debriefing interviews (Willis, 2005) assessed problems in conceptual understanding and question wording. These interviews were conducted by professional cognitive
interviewers using the ‘think aloud’ method (Presser et al. 2004) to elicit initial respondent reactions and collect alternative terminologies for confusing phrases. The results were presented to the diagnostic experts for review and final question revision.

The clinical reappraisal study

The sample

The revised questions were then administered to 3058 patients sampled from 29 primary care offices selected to include practices in both urban and rural areas in all four US Census Regions (Northeast, South, Midwest and West). No other stratification criteria were used in selecting practices. Practices were recruited through the Primary Care Network (www.primarycarenet.org). The original sample design called for a quota sample of 100 completed interviews in each of 30 offices with an unselected consecutive sample of patients. This sample size was selected to allow for the second-stage assessment of at least 30 screened positives for even the least common disorder (BPD) assuming plausible prevalence estimates and second-stage response rates. However, because one selected office dropped out after office recruitment ended, other offices in the same sample stratum were asked to continue data collection for 2 days beyond the time they met their quota, yielding a sample slightly larger than the originally targeted 3000 and based on 29, not 30, offices.

Respondent recruitment began by giving a ‘study fact brochure’ to patients as they checked in that explained the study as a test of a new screening questionnaire for common anxiety and mood disorders. The brochure explained that the study needed people aged ≥ 18 years both with and without the disorders to complete a 15-min laptop computer questionnaire in the waiting room; that participants would be remunerated US$25; that some participants would be asked to participate in a telephone follow-up interview that could take up to 1 h to complete; and that telephone respondents would receive an additional US$50. The brochure emphasized that responses were confidential and decisions about participation would not affect health-care treatment or benefits. Patients who informed the office receptionist that they were interested in the study then provided written informed consent and received a laptop computer to complete the questionnaire in the waiting room. Telephone numbers provided in the questionnaire were used to contact respondents and administer clinical reappraisal interviews within 3 days of the visit. The Human Subjects Committee of the New England Institutional Review Board (www.neirb.com) approved these recruitment, consent and field procedures.

The clinical reappraisal interview

Each CIDI-SC respondent was classified as ‘very likely’, ‘likely’, ‘possible’ or ‘no’ on each screening scale. A probability subsample of 30 respondents classified ‘very likely’ and 20 classified ‘likely-possible’ was selected for each scale with replacement and administered the clinical reappraisal interview. The sampling fraction varied across disorders due to prevalence differences to make the sample well-distributed across practices. Fifty patients who screened ‘no’ on all screening scales were also interviewed. The total clinical reappraisal sample of 206 is less than 50 × 5 = 250 because some respondents were independently selected for multiple disorders.

The clinical interview was an abridged Research Version, Non-Patient Edition of the Structured Clinical Interview for DSM-IV (SCID-I; First et al. 2002) focused on the four syndromes under study: 30-day MDE and GAD and lifetime and 30-day PD and mania/hypomania. Experienced SCID interviewers administered the interviews under the supervision of a study collaborator (P.E.K.) blinded to CIDI-SC responses. 30-day PD was defined as lifetime PD with 30-day panic attacks and/or persistent concern about additional attacks, worry about implications/consequences of attacks, or significant change in behavior due to attacks. 30-day BPD was defined as lifetime mania/hypomania with either 30-day MDE or 30-day mania/hypomania. SCID diagnoses were made without diagnostic hierarchy rules but with organic exclusions. Organic exclusions were not made in the screening scales. Each SCID disorder was classified as severe or non-severe to determine whether CIDI-SC could differentially detect more severe cases. BP-I versus BP-II defined BPD severity whereas the distinction between severe and non-severe cases of the other disorders was based on SCID interviewer assessments of whether there were (i) many symptoms more than needed for diagnosis, (ii) several symptoms that were particularly severe and/or (iii) marked impairment in social or occupational functioning associated with the disorder.

Analysis methods

The clinical reappraisal sample was weighted to adjust for oversampling of patients screened as ‘very likely’, ‘likely’ or ‘possible’. Iterative stepwise logistic regression was then used (0.05-level entry criterion) to predict SCID diagnoses from CIDI-SC symptoms to determine the minimum CIDI-SC question set needed to approximate SCID diagnoses. An unweighted
summary CIDI-SC score was created for each diagnosis from this minimum symptom set and receiver operating characteristic curve (ROC) analysis (Margolis et al. 2002) was used to estimate the area under the ROC curve (AUC) for each scale. The AUC is the probability of correctly identifying SCID cases from CIDI-SC scores in paired comparisons of randomly selected pairs of SCID cases and non-cases, where CIDI-SC tied scores are assigned a 50% chance of correct classification (Kraemer, 1992). The AUC has a predicted value of 0.5 when the screening scale is completely unrelated to the true score and 1.0 when perfectly related. CIDI-SC scores were not weighted to avoid overfitting in the absence of a large enough sample for cross-validation.

Each CIDI-SC score was then collapsed so that SCID prevalence estimates increased monotonically across screening scale strata but did not differ significantly within strata using the logic of stratum-specific likelihood ratio (LR) analysis (Pepe, 2003). McNemar χ² tests then tested the significance of differences between CIDI-SC and SCID prevalence estimates. Significance tests were based on Taylor series design-based standard errors to adjust for data weighting (Wolter, 1985).

Individual-level concordance was evaluated using the AUC and Cohen’s κ (Cohen, 1960). Although κ is the traditional measure used in psychiatric research, it is not emphasized here because it is varies across populations that differ in prevalence even when sensitivity (SN; the percentage of true cases correctly classified) and specificity (SP; the percentage of true non-cases correctly classified) are constant (Cook, 1998). The AUC, in comparison, is a function of SN and SP, which are considered the fundamental parameters of agreement. The AUC equals (SN+SP)/2 when the screen is dichotomous. AUC scores between 0.5 and 1.0 are often interpreted in parallel with κ as slight (AUC = 0.5–0.6, κ = 0.0–0.2), fair (AUC = 0.6–0.7, κ = 0.2–0.4), moderate (AUC = 0.7–0.8, κ = 0.4–0.6), substantial (AUC = 0.8–0.9, κ = 0.6–0.8) and almost perfect (AUC > 0.9, κ > 0.8) (Landis & Koch, 1977). We also report total classification accuracy (TCA), the proportion of all respondents whose CIDI-SC and SCID classifications are consistent.

In addition, we report disaggregated measures of operating characteristics, including SN and SP, positive predictive value (PPV; proportion of screened positives confirmed by the SCID), negative predictive value (NPV; proportion of screened negatives confirmed as non-cases by the SCID), LR positive [(LR + ) SN/(1 - SP)] and LR negative [LR–; (1 - SN)/SP]]. LR + and LR– assess the relative proportions of screened positives versus screened negatives confirmed as cases (LR + ) or non-cases (LR–). LR + values ≥ 5 and LR– values ≤ 0.2 are generally considered useful, whereas LR+ values ≥ 10 and LR– values ≤ 0.1 are considered sufficient to rule in/out diagnoses (Haynes et al. 2006).

**Comparison with other widely used screening scales**

To compare CIDI-SC operating characteristics with other screening scales, a 1990–2012 Medline search of screening scale validity studies was carried out using search terms ‘screening’, ‘validity’, ‘sensitivity’, ‘specificity’, ‘case finding’ and ‘AUC’ crossed with ‘depression’, ‘bipolar disorder’, ‘manic depression’, ‘generalized anxiety disorder’ and ‘panic disorder’. We focused on studies where screening scales were compared to blinded clinical reappraisal interviews in samples of patients, community residents or internet users. Only key studies were considered.

**Results**

**Stepwise logistic regression**

Separate stepwise logistic regression analyses were used to predict each SCID disorder from the corresponding CIDI screening items.

**MDE**

Three CIDI questions were entered stepwise to predict 30-day dysphoria (sad–depressed, down–discouraged) and anhedonia (little–no interest in day-to-day activities) in the total sample. Among respondents with dysphoria and/or anhedonia, five additional questions were entered to screen for other DSM-IV Criterion A symptoms of MDE or the Criterion C requirement of clinically significant distress or impairment. The AUC for the continuous CIDI-SC scale with these eight questions was 0.93.

**GAD**

Two CIDI questions were entered to screen for 30-day DSM-IV GAD Criterion A (excessive anxiety–worry about multiple events–activities) in the total sample. Among respondents with dysphoria and/or anhedonia, five additional questions were entered to screen for other DSM-IV Criterion A symptoms of MDE or the Criterion C requirement of clinically significant distress or impairment. The AUC for the continuous CIDI-SC scale with these seven questions was 0.88.

**PD**

Two CIDI questions were entered to screen for having lifetime attacks of intense fear or discomfort that came on very suddenly in the total sample. Among
respondents with such attacks, seven follow-up questions were entered about psycho-physiological symptoms. Among patients with such symptoms, an additional question asked about symptoms reaching a peak within 10 min and two question asked about the Criterion A1 DSM-IV PD requirement that attacks be recurrent and unexpected. Four questions asked about Criterion A2 that attacks be followed by a month of persistent concern about another attack, worry about implications or significant change in behavior. Final questions then asked about 30-day prevalence. The AUC for the continuous CIDI-SC scale with these 15 symptom questions crossed with reports of 30-day recency was 0.90.

BPD

Two CIDI questions were entered to screen for lifetime DSM-IV mania-hypomania Criterion A (distinct periods of abnormally persistently elevated, expansive or irritable mood) in the total sample. Among respondents who endorsed at least one such question, four additional questions were entered to screen for Criterion B (more talkative than usual, racing thoughts, psychomotor agitation, excessive involvement in activities having high potential for painful consequences) and two for Criterion D (mania)/E (hypomania) involving presence–absence of marked impairment or hospitalization. A final question then asked about 30-day prevalence. The AUC for the continuous CIDI-SC scale with these 15 symptom questions crossed with reports of 30-day recency was 0.90.

### Concordance of DSM-IV CIDI-SC and SCID diagnoses

CIDI-SC versus SCID prevalence estimate differences are insignificant for all disorders at optimal (for estimating prevalence) CIDI-SC thresholds ($\chi^2 = 0.0–2.9$, $p = 0.09–0.98$) (Table 1). Aggregate diagnostic concordance at these thresholds is substantial for all disorders ($AUC = 0.81–0.86$), with proportions of SCID cases detected (SN) of 68.0–80.2%. The proportions of SCID non-cases classified correctly (SP) are 90.1–98.8%. Lower SN than SP is expected for thresholds designed to estimate prevalence without bias when only a minority of patients has a disorder, in which case LR+ is more informative than SN. LR+ is > 10 for three of the four CIDI-SC at the optimal thresholds, indicating that screened positives are much more likely than screened negatives to be confirmed as SCID cases. LR+ is 8.1 for MDE, an informative but not definitive value. The proportions of screened positives at the optimal thresholds confirmed as SCID cases (PPV) are in the range 48.2% (GAD) to 73.7% (BPD) (Table 2).

The screen for MDE, the only one where LR+ is <10, can be made more conservative by raising the threshold (LR+ = 24.5, PPV = 85.9%), but at the cost of reducing SN from 80.2% to 46.5%. All four CIDI-SC can be made less conservative by lowering their thresholds, increasing SN to between 94.8% (BPD) and 100% (GAD and PD), but at the cost of increasing the estimated prevalence and decreasing LR+ and PPV. The only disorder where this conservative change is efficient is BPD, with an estimated prevalence

<table>
<thead>
<tr>
<th>CIDI-SC</th>
<th>SCID</th>
<th>McNemar*</th>
<th>AUC</th>
<th>$\kappa$</th>
<th>TCA</th>
<th>SN (s.e.)</th>
<th>SP (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDE</td>
<td>23.8 (3.5)</td>
<td>19.7 (3.3)</td>
<td>2.9</td>
<td>0.85</td>
<td>0.65</td>
<td>88.2</td>
<td>80.2</td>
</tr>
<tr>
<td>GAD</td>
<td>10.8 (1.7)</td>
<td>7.7 (1.7)</td>
<td>2.5</td>
<td>0.81</td>
<td>0.52</td>
<td>91.9</td>
<td>68.0</td>
</tr>
<tr>
<td>PD</td>
<td>13.7 (2.6)</td>
<td>10.7 (2.5)</td>
<td>2.4</td>
<td>0.85</td>
<td>0.62</td>
<td>91.9</td>
<td>76.4</td>
</tr>
<tr>
<td>BPD</td>
<td>4.4 (1.0)</td>
<td>4.4 (1.2)</td>
<td>0.0</td>
<td>0.86</td>
<td>0.73</td>
<td>97.7</td>
<td>74.0</td>
</tr>
</tbody>
</table>

*Prevalence estimates based on the CIDI-SC do not differ significantly from those based on the SCID for any of the diagnoses ($p = 0.09–0.98$).
increasing from 4.4% to 7.0%, LR+ and PPV both remaining high (31.6, 59.1%) and SN increasing from 74.0% to 94.8%.

Although SP is above 90% for all disorders, this is not a definite rule-out when only a small minority of respondents has the disorder, in which case LR× is more informative. LR× is 0.2 for only two diagnoses at the optimal threshold (MDE and PD) whereas LR× is never below 0.2, meaning that none of the diagnoses can be ruled out confidently with CIDI-SC scores below the optimal diagnostic threshold. However, thresholds can be lowered to produce LR× values less than 0.1 for all disorders, although at the cost of reducing SN.

For MDE, 54.1% of patients can be ruled out [i.e. at a threshold where 45.9% (100%–45.9% = 54.1%) of patients screen positive] with LR− 0.1 (NPV = 99.7%). For GAD, 55.7% of patients can be ruled out with LR− 0.0 (NPV = 100%). For PD, 86.3% of patients can be ruled out with LR− 0.2 and 37.9% with LR− 0.0. None of the PD screened negatives at the lower thresholds and 2.8% at the next lowest threshold had a SCID diagnosis. For BPD, 93% of patients can be ruled out with LR− 0.1 (NPV = 99.8%).

Severe and non-severe cases

SN is higher for severe than non-severe cases of all four diagnoses at the optimal threshold (Table 3). SN is 85.4–92.9% for severe MDE, PD and BPD versus 68.9–69.6% for non-severe cases and 70.8% versus 59.6% for severe and non-severe GAD. However, none of the severe versus non-severe SN differences are statistically significant because of the small numbers of cases ($\chi^2 = 0.2–2.4$, $p = 0.12–0.68$).

Comparisons with other screening scales

MDE

The nine-item Patient Health Questionnaire (PHQ-9; Spitzer et al. 1999) is the most widely used major depression screening scale. Reviews of many PHQ-9 primary care validity studies (Gilbody et al. 2007; Wittkampf et al. 2007; Kroenke et al. 2010; Manea et al. 2012) show a central tendency of the AUC to be 0.85–0.88, which does not differ meaningfully from the CIDI-SC MDE AUC of 0.85. (See online Appendix Tables 1–4 for detailed results.) CIDI-SC SN and SP (0.80, 0.90) are also in the middle of the PHQ-9 ranges (0.77–0.88, 0.88–0.94). The AUC of other MDE screening scales is generally lower (0.72–0.84) and LR+ uninformative (Zigmond & Snaith, 1983; Broadhead et al. 1995; Farvolden et al. 2003; Hunter et al. 2005; Donker et al. 2009; Gaynes et al. 2010; Houston et al. 2011). One exception was found in a community survey of the Center for Epidemiologic Studies Depression scale (CES-D; Radloff, 1977), with an AUC of 0.94 (Beekman et al. 1997), but other CES-D validity studies found considerably lower AUC, at 0.76–0.82 (Schulberg et al. 1985; Klinkman et al. 1997; Thomas et al. 2001).

Table 2. CIDI screening scale (CIDI-SC) classification of DSM-IV/SCID cases and non-cases at different thresholds on the CIDI-SC (n = 206)*

<table>
<thead>
<tr>
<th></th>
<th>p (s.e.)</th>
<th>SN (s.e.)</th>
<th>LR+ (s.e.)</th>
<th>PPV (s.e.)</th>
<th>SP (s.e.)</th>
<th>LR− (s.e.)</th>
<th>NPV (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. MDE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conservative</td>
<td>10.7 (1.9)</td>
<td>46.5 (8.3)</td>
<td>24.5</td>
<td>85.9 (4.6)</td>
<td>98.1 (0.6)</td>
<td>0.5</td>
<td>88.2 (3.2)</td>
</tr>
<tr>
<td>Optimal</td>
<td>23.8 (3.5)</td>
<td>80.2 (7.6)</td>
<td>8.1</td>
<td>66.5 (7.0)</td>
<td>90.1 (2.5)</td>
<td>0.2</td>
<td>94.9 (2.2)</td>
</tr>
<tr>
<td>Anti-conservative</td>
<td>45.9 (4.7)</td>
<td>99.2 (0.8)</td>
<td>3.0</td>
<td>42.6 (6.1)</td>
<td>67.2 (5.1)</td>
<td>0.0</td>
<td>99.7 (0.3)</td>
</tr>
<tr>
<td>II. GAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td>10.8 (1.7)</td>
<td>68.0 (10.0)</td>
<td>11.1</td>
<td>48.2 (6.7)</td>
<td>93.9 (1.2)</td>
<td>0.3</td>
<td>97.2 (1.4)</td>
</tr>
<tr>
<td>Anti-conservative</td>
<td>44.3 (4.9)</td>
<td>100 (–)</td>
<td>2.5</td>
<td>17.3 (3.7)</td>
<td>60.3 (5.2)</td>
<td>0.0</td>
<td>100 (–)</td>
</tr>
<tr>
<td>III. PD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td>13.7 (2.6)</td>
<td>76.4 (10.3)</td>
<td>12.3</td>
<td>59.5 (8.6)</td>
<td>93.8 (1.4)</td>
<td>0.2</td>
<td>97.1 (1.4)</td>
</tr>
<tr>
<td>Anti-conservative</td>
<td>62.1 (4.9)</td>
<td>100 (–)</td>
<td>1.7</td>
<td>17.2 (4.0)</td>
<td>42.5 (5.2)</td>
<td>0.0</td>
<td>100 (–)</td>
</tr>
<tr>
<td>IV. BPD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td>4.4 (1.0)</td>
<td>74.0 (13.8)</td>
<td>61.7</td>
<td>73.7 (9.0)</td>
<td>98.8 (0.5)</td>
<td>0.3</td>
<td>98.8 (0.8)</td>
</tr>
<tr>
<td>Anti-conservative</td>
<td>7.0 (1.8)</td>
<td>94.8 (3.4)</td>
<td>31.6</td>
<td>59.1 (13.4)</td>
<td>97.0 (1.4)</td>
<td>0.1</td>
<td>99.8 (0.2)</td>
</tr>
</tbody>
</table>

CIDI, Composite International Diagnostic Interview; SCID, Structured Clinical Interview for DSM-IV; p, proportion of patients who screened positive on the CIDI-SC at the designated threshold; SN, sensitivity of the CIDI-SC at the designated threshold; LR+, likelihood ratio positive of the CIDI-SC at the designated threshold; PPV, positive predictive value of the CIDI-SC at the designated threshold; SP, specificity of the CIDI-SC at the designated threshold; LR−, likelihood ratio negative of the CIDI-SC at the designated threshold; NPV, negative predictive value of the CIDI-SC at the designated threshold; MDE, major depressive episode; GAD, generalized anxiety disorder; PD, panic disorder; BPD, bipolar disorder; s.e., standard error.
GAD

The CIDI-GAD AUC (0.81) is in the middle of the range for the GAD screening scales reviewed (0.74–0.85) (Broadhead et al. 1995; Farvolden et al. 2003; Kroenke et al. 2007; Donker et al. 2009, 2011; Houston et al. 2011). However, CIDI-SC SN and SP (0.68, 0.94) are closest to those of one specialty treatment screening scale, the Web-Based Depression and Anxiety Test (WB-DAT; Farvolden et al. 2003). Other screening scales have higher SN (0.83–0.93) but much lower SP (0.45–0.82). CIDI-SC and WB-DAT consequently have much higher LR+ (11.1, 10.5) than other scales (1.7–4.9), indicating higher confirmation of screened positives. This can be illustrated using Bayes’ theorem to calculate post-test probability of SCID GAD for screened positives (Altman & Bland, 1994), which shows that for a true GAD prevalence of 5–15%, confirmation of screened positives would be only 21–46% for screening scales but much higher for WB-DAT (36–65%) and CIDI-SC (37–66%). Caution is needed in interpreting the WB-DAT results, however, as they were obtained in a specialty treatment setting.

PD

The CIDI-SC PD AUC (0.85) is at the upper end of the PD screening scales reviewed (0.69–0.88) (Broadhead et al. 1995; Stein et al. 1999; Farvolden et al. 2003; Lowe et al. 2003; Hunter et al. 2005; Bunevicius et al. 2007; Kroenke et al. 2007; Donker et al. 2009). CIDI-SC has among the highest LR+ (12.3) along with WB-DAT (12.5) and one of two GAD-7 (19.0) validity studies (Lowe et al. 2003). The high LR+ in that GAD-7 study, however, is offset by a much lower LR+ (3.9) in a second much larger GAD-7 study (Kroenke et al. 2007). The scales with high LR+ are much more distinct for their high SP (0.94–0.96) than high SN. If we assume that the true PD prevalence is in the range 5–15% in primary care and SN–SP estimates are accurate, confirmation of screened positives would be 35–65% for CIDI-SC and WB-DAT, 17–77% for the GAD-7, and no higher than 20–45% for other screening scales.

BPD

Although the Mood Disorder Questionnaire (MDQ; Hirschfeld et al. 2000) is by far the most widely used BPD screening scale, the vast majority of MDQ studies focus on patients in treatment for depression and investigate whether those with BPD can be distinguished from non-bipolar depressives (Hirschfeld et al. 2000, 2005; Miller et al. 2004, 2011; Weber Rouget et al. 2005; Parker et al. 2008, 2012; Twiss et al. 2008; Zimmerman et al. 2009). This focus reflects the fact that the MDQ was developed to address the under-detection of BPD among depressed patients (Hirschfeld & Vornik, 2004). We are aware of only two MDQ validity studies that evaluated ability to distinguish patients with BPD from all other patients (including those without depression) in settings other than a specialty clinic (Hirschfeld et al. 2003; Dodd et al. 2009). These studies were both carried out in community samples. The MDQ AUC was fairly low in both studies (AUC=0.62) compared to much higher AUCs (0.86–0.96) for the CIDI-SC BPD scale at its two informative thresholds.

Only one other BPD screening scale, the Mood Swings Questionnaire (MSQ; Parker et al. 2006), had an AUC as high as the CIDI-SC, but this was in a study in a mental health specialty clinic among patients presenting for treatment of depression. Two subsequent studies in that same clinic produced lower MSQ AUC estimates (0.73–0.81; Parker et al. 2008, 2012). Other BPD screening scales reviewed had lower AUC (0.66–0.81; Hunter et al. 2005; Gaynes et al. 2010).

Table 3. CIDI screening scale (CIDI-SC) sensitivity (SN) and likelihood ratio positive (LR+) for detecting severe and non-severe DSM-IV/SCID cases (n=206)

<table>
<thead>
<tr>
<th></th>
<th>Severe</th>
<th>Non-severe</th>
<th>Total</th>
<th>Severe v. non-severe^a</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SN (S.E.)</td>
<td>LR+</td>
<td>SN (S.E.)</td>
<td>LR+</td>
</tr>
<tr>
<td>MDE</td>
<td>92.9 (4.1)</td>
<td>9.4</td>
<td>69.6 (12.6)</td>
<td>7.0</td>
</tr>
<tr>
<td>GAD</td>
<td>70.8 (15.2)</td>
<td>11.6</td>
<td>59.6 (14.7)</td>
<td>9.8</td>
</tr>
<tr>
<td>PD</td>
<td>90.8 (7.5)</td>
<td>14.6</td>
<td>68.9 (14.7)</td>
<td>11.1</td>
</tr>
<tr>
<td>BPD</td>
<td>85.4 (10.0)</td>
<td>71.2</td>
<td>69.0 (18.6)</td>
<td>57.5</td>
</tr>
</tbody>
</table>

CIDI, Composite International Diagnostic Interview; SCID, Structured Clinical Interview for DSM-IV; SN, sensitivity of the CIDI-SC at the designated threshold; LR+, likelihood ratio positive of the CIDI-SC at the designated threshold; MDE, major depressive episode; GAD, generalized anxiety disorder; PD, panic disorder; BPD, bipolar disorder; S.E., standard error.

^a Although SN is consistently higher for severe than non-severe cases, none of these differences is statistically significant (p=0.12-0.68).
The advantage of CIDI-SC over these other scales can be traced to high CIDI-SC SN at its anti-conservative threshold (0.95). Although, as noted earlier, high SN is often accompanied by low LN+, this is not the case with CIDI-SC BPD, where LR+ is 31.6 at the anti-conservative threshold and 62–85% of screened positives would be confirmed as SCID cases if the true BPD prevalence was in the range 5–15%.

Discussion

CIDI-SC operating characteristics are equivalent to the best alternative screening scales for MDE and GAD and superior to other screening scales for PD and BPD. CIDI-SC results can be compared directly to general population epidemiological CIDI surveys because CIDI-SC items all come from the CIDI. Such nested screening scales can be useful in targeting and streamlining CIDI follow-up interviews by ‘pre-loading’ CIDI-SC responses into the CIDI computerized interview program to guide interview question skip logic. Such an integrated computerized CIDI interviewing system is currently in development and includes options for self-administering CIDI-SC on tablet computers in primary care waiting rooms, web-based CIDI-SC self-administration to track treatment response, and interviewer-based CIDI interview administration using pre-loaded CIDI-SC responses.

The fact that AUCs of continuous CIDI-SC scales in ROC analyses (0.88–0.97) are considerably higher than AUCs of dichotomized CIDI-SC scales at their unbiased thresholds (0.81–0.86) means that meaningful variation in SCID prevalence exists throughout the CIDI-SC scale ranges. One implication, as shown in the comparative analyses of LR+ and LR− at multiple thresholds, is that different thresholds can be useful for screening in than screening out cases. Importantly, the CIDI-SC has excellent LR+ and LR− at multiple informative thresholds. Furthermore, continuous CIDI-SC scores can be converted into predicted probabilities of clinical diagnoses in epidemiological studies to yield more accurate estimates of prevalence than by dichotomizing scores and classifying each respondent as either a definite case or a non-case. This predicted probability approach is discussed in more detail elsewhere (Kessler et al. 2010b).

Despite these positive findings, several study limitations are noteworthy. First, CIDI-SC SN is lower for GAD than other diagnoses. Disaggregation shows that this is because CIDI-SC have difficulty operationalizing the DSM-IV requirement that worries be excessive. CIDI-SC questions for this requirement have a higher threshold than the SCID. A similar result was found in an earlier study of the full CIDI (Wittchen et al. 1995). Concerns exist about clinician ability to determine when worries are excessive (Ruscio et al. 2005), leading to the suggestion that more concrete guidance be given in DSM-5 about defining excessiveness (Andrews et al. 2010). Although such guidance does not appear in currently proposed DSM-5 criteria (www.dsm5.org/ProposedRevisions), new behavioral requirements (proposed DSM-5 Criterion D) of marked avoidance, time–effort, procrastination or seeking reassurance might help to establish a threshold for excessiveness that could be the basis for improving revised CIDI-SC GAD SN.

Second, although we want to use CIDI-SC results to create a cross-walk between general population CIDI epidemiological surveys and primary care CIDI-SC screening studies, no guarantee exists that CIDI-SC operating characteristics will be similar in community epidemiological surveys and primary care samples. It is consequently important to include CIDI-SC in future CIDI community surveys and validate their operating characteristics relative to diagnoses based on the full CIDI and SCID. Such methodological studies are currently underway in new CIDI surveys in the WHO WMH Survey Initiative (Kessler & Üstün, 2008a).

Third, our clinical reappraisal sample was relatively small because of funding limitations, precluding cross-validation, subgroup analysis, or analysis of information values across the range of continuous CIDI-SC scores to evaluate sensitivity to change. These limitations make it especially important to replicate the current study in independent primary care samples, to investigate the stability of the encouraging results reported here and to carry out analyses of the clinical sensitivity of variation in continuous CIDI-SC scores to assess the severity of anxiety and depression. Larger replication studies could also help to establish an empirical foundation for determining whether even shorter versions of CIDI-SC might be developed based on computerized adaptive testing (Gibbons et al. 2011).

Supplementary material

For supplementary material accompanying this paper visit: www.hcp.med.harvard.edu/wmhcidi/resources.php.

Acknowledgments

Support for this study was provided by AstraZeneca. The sponsor had no role in the design and conduct of the study; collection, management, analysis and interpretation of the data; and preparation, review or approval of the manuscript.
Declaration of Interest

R. C. Kessler has consulted for AstraZeneca, Analysis Group, Bristol–Myers Squibb, Cerner-Galt Associates, Eli Lilly and Company, GlaxoSmithKline Inc., HealthCore Inc., Health Dialog, Integrated Benefits Institute, John Snow Inc., Kaiser Permanente, Matria Inc., Mensante, Merck & Co. Inc., Ortho-McNeil Janssen Scientific Affairs, Pfizer Inc., Primary Care Network, Research Triangle Institute, Sanofi-Aventis Group, Shire US Inc., SRA International Inc., Takeda Global Research & Development, Transcept Pharmaceuticals Inc., and Wyeth-Ayerst; has served on advisory boards for Appliance Computing II, Eli Lilly and Company, Mindsite, Ortho-McNeil Janssen Scientific Affairs, Plus One Health Management and Wyeth-Ayerst; and has had research support for his epidemiological studies from Analysis Group Inc., Bristol–Myers Squibb, Eli Lilly and Company, EPI-Q, GlaxoSmithKline, Johnson & Johnson Pharmaceuticals, Ortho-McNeil Janssen Scientific Affairs., Pfizer Inc., Sanofi-Aventis Groupe, and Shire US Inc. J. R. Calabrese has received research grant support from Abbott, AstraZeneca, Bristol–Myers Squibb, Cephalon, Eli Lilly and Company, GlaxoSmithKline, Janssen, Repligen, Sunovion/DSPA, Takeda and Wyeth; has consulted to or served on advisory boards of AstraZeneca, Bristol–Myers Squibb, Cephalon, Sunovion/DSPA, Forest, GlaxoSmithKline, Janssen, Johnson and Johnson, Lundbeck, Neurosearch, OrthoMcNeil, Otsuka, Pfizer, Repligen, Schering-Plough, Servier, Solvay, Supernus, Synosia, and Wyeth; and has provided CME lectures supported by Abbott, AstraZeneca, Bristol–Myers Squibb, GlaxoSmithKline, Janssen, Johnson and Johnson, Lundbeck, Merck, Sanofi Aventis, Schering-Plough, Pfizer, Solvay, and Wyeth. P. A. Farley, M. A. Jewell and A. C. Shillington are employees of EPI-Q, the organization that implemented the primary care screening and recruitment for the SCID clinical reappraisal interviews in this study. In addition, they carry out contract research for AstraZeneca, Cephalon, Merck & Co. Inc., Sanofi-Aventis, GlaxoSmithKline, Genentech, Biogen, Roche, Transcept Pharmaceuticals Inc., Lundbeck, Shire US Inc., Takeda, Novartis, Pfizer, Abbott, and Adolor. Their compensation related to these activities is limited to their salary. They also own stock in EPI-Q. P. E. Keck is employed by the University of Cincinnati College of Medicine and University of Cincinnati Physicians, the organization that carried out the clinical reappraisal interviews in this study. He is presently or has been in the past year a principal or co-investigator on research studies sponsored by Alkermes, AstraZeneca, Cephalon, GlaxoSmithKline, Eli Lilly and Company, Marriott Foundation, National Institute of Mental Health (NIMH), Orexigen, Pfizer Inc., and Shire. He has been reimbursed for consulting to, in the past 2 years: 2011: Pamlab, 2012: Bristol–Myers Squibb. Patents: Dr Keck is a co-inventor on United States Patent No. 6,387,956: Shapira NA, Goldsmith TD, Keck PE Jr. (University of Cincinnati) Methods of treating obsessive-compulsive spectrum disorder comprises the step of administering an effective amount of tramadol to an individual. Filed 25 March 1999; approved 14 May 2002. Dr Keck has received no financial gain from this patent. A. A. Nierenberg has consulted for the American Psychiatric Association, Appliance Computing Inc. (Mindsite), Basileia, Brain Cells Inc., BRANDeis University, Bristol–Myers Squibb, Dey Pharmaceuticals, Dainippon Sumitomo, Eli Lilly and Company, EpiQ, L.P./Mylan Inc., Novartis, PGx Health, Shire, Schering-Plough, Takeda Pharmaceuticals, and Targacept; consulted through the MGH Clinical Trials Network and Institute (CTNI) for AstraZeneca, Brain Cells Inc., Dianippon Sumitomo/Sepracor, Johnson and Johnson, Labopharm, Merck, Methylation Science, Novartis, PGx Health, Shire, Schering-Plough, Targacept and Takeda/Lundbeck Pharmaceuticals; received grant/research support from NIMH, PamLabs, Pfizer Pharmaceuticals, and Shire; received honoraria from Belvoir Publishing, University of Texas Southwestern Dallas, Hillside Hospital, American Drug Utilization Review, American Society for Clinical Psychopharmacology, Baystate Medical Center, Columbia University, CRICO, Dartmouth Medical School, IMEDEX, Israel Society for Biological Psychiatry, Johns Hopkins University, MJ Consulting, New York State, Medscape, MBL Publishing, National Association of Continuing Education, Physicians Postgraduate Press, SUNY Buffalo, University of Wisconsin, University of Pisa, University of Michigan, University of Miami, APSARD, ISBD, SciMed, Slack Publishing and Wolters Kluwer Publishing; was currently or formerly on the advisory boards of Appliance Computing Inc., Brain Cells Inc., Eli Lilly and Company, Johnson and Johnson, Takeda/Lundbeck, Targacept, and InfoMedic; owns stock options in Appliance Computing Inc. and Brain Cells Inc.; has copyrights to the Clinical Positive Affect Scale and the MGH Structured Clinical Interview for the Montgomery Asberg Depression Scale exclusively licensed to the MGH Clinical Trials Network and Institute (CTNI); and has a patent extension application for the combination of bupropion, bupropion, and melatonin for the treatment of depression. M. B. Stein has a financial interest/arrangement or affiliation with one or more organizations that could be perceived as a real or apparent conflict of interest in
the context of the subject of this presentation. He receives or has in the past 3 years received Research Support from: Hoffmann-La Roche; is currently or in the past 3 years has been a Consultant for Care Management Technologies; and receives payment for editorial work from Depression and Anxiety (Publisher: Wiley-Blackwell). M. E. Thase has served as an advisory/consultant for Alkermes, AstraZeneca, Bristol–Myers Squibb Company, Eli Lilly and Company, Dey Pharma, L.P., Forest Laboratories, Gerson Lehman Group, GlaxoSmithKline (ended 2008), Guidepoint Global, H. Lundbeck A/S, MedAvante Inc., Merck and Co. Inc. (formerly Schering Plough and Organon), Neurontics Inc., Novartis (ended 2008), Otsuka, Ortho-McNeil Pharmaceuticals (Johnson & Johnson), Pamlab, L.L.C., Pfizer (formerly Wyeth Ayerst Pharmaceuticals), PGX Inc., Shire US Inc., Sunovion Pharmaceuticals Inc., Supernus Pharmaceuticals, Takeda, Transcept Pharmaceuticals; has received grant support from Agency for Healthcare Research and Quality, Eli Lilly and Company, Forest Pharmaceuticals, GlaxoSmithKline (ended July 2010), National Institute of Mental Health, Otsuka Pharmaceuticals, Sepracor Inc. (ended January 2009); is or has been on the Speakers bureaus for AstraZeneca (ended June 2010), Bristol–Myers Squibb Company, Dey Pharmaceutical, Eli Lilly and Company (ended June 2009), Merck and Co. Inc., Pfizer (formerly Wyeth Ayerst Pharmaceuticals); holds equity in MedAvante Inc.; receives royalties from American Psychiatric Foundation, Guilford Publications, Herald House and W.W. Norton & Company Inc.; and has a spouse employed by BMS and Pfizer/Wyeth).

References


