



Published in final edited form as:

*J Affect Disord.* 2006 December ; 96(3): 259–269.

## Validity of the Assessment of Bipolar Spectrum Disorders in the WHO CIDI 3.0

Ronald C. Kessler, Ph.D., Hagop S. Akiskal, M.D., Jules Angst, M.D., Margaret Guyer, PhD, Robert M.A. Hirschfeld, M.D., Kathleen R. Merikangas, Ph.D., and Paul E. Stang, Ph.D.

*From the Department of Health Care Policy, Harvard Medical School (Kessler, Jin); the International Mood Center, University of California San Diego and VA Psychiatry Service, San Diego, California (Akiskal); the Zurich University Psychiatric Hospital, Zurich (Angst); the Massachusetts Mental Health Center (Guyer); the Department of Psychiatry and Behavioral Sciences, University of Texas Medical Branch, Galveston, Galveston, Texas (Hirschfeld); the Intramural Research Program, National Institute of Mental Health (Merikangas); the Department of Health, West Chester State University (Stang); and Galt Associates (Stang).*

### Abstract

**Objective**—Although growing interest exists in the bipolar spectrum, fully structured diagnostic interviews might not accurately assess bipolar spectrum disorders. A validity study was carried out for diagnoses of threshold and sub-threshold bipolar disorders (BPD) based on the WHO Composite International Diagnostic Interview (CIDI) in the National Comorbidity Survey Replication (NCS-R). CIDI BPD screening scales were also evaluated.

**Method**—The NCS-R is a nationally representative US household population survey (n = 9282) using CIDI to assess DSM-IV disorders. CIDI diagnoses were evaluated in blinded clinical reappraisal interviews using the non-patient version of the Structured Clinical Interview for DSM-IV (SCID).

**Results**—Excellent CIDI-SCID concordance was found for lifetime BP-I (AUC = .99,  $\kappa$  = .88, PPV = .79, NPV = 1.0), either BP-II or sub-threshold BPD (AUC = .96,  $\kappa$  = .88, PPV = .85, NPV = .99), and overall bipolar spectrum disorders (i.e., BP-I/II or sub-threshold BPD; AUC = .99,  $\kappa$  = .94, PPV = .88, NPV = 1.0). Concordance was lower for BP-II (AUC = .83,  $\kappa$  = .50, PPV = .41, NPV = .99) and sub-threshold BPD (AUC = .73,  $\kappa$  = .51, PPV = .58, NPV = .99). The CIDI was unbiased compared to the SCID, yielding a lifetime bipolar spectrum disorders prevalence estimate of 4.4%. Brief CIDI-based screening scales detected 67–96% of true cases with positive predictive value of 31–52%.

**Limitation**—CIDI prevalence estimates are still probably conservative, though, but might be improved with future CIDI revisions based on new methodological studies with a clinical assessment more sensitive than the SCID to sub-threshold BPD.

**Conclusions**—Bipolar spectrum disorders are much more prevalent than previously realized. The CIDI is capable of generating conservative diagnoses of both threshold and sub-threshold BPD. Short CIDI-based scales are useful screens for BPD.

---

Address comments to RC Kessler, Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA 02115. Voice: 617-432-3587; Fax: 617-432-3588; Email: kessler@hcp.med.harvard.edu..

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## Keywords

Bipolar Disorders; Bipolar Spectrum; Mania; Hypomania; Composite International Diagnostic Interview (CIDI); Validity; National Comorbidity Survey Replication (NCS-R)

Although the estimated lifetime prevalence of bipolar disorders (BPD) in international adult population surveys using structured diagnostic interviews and standardized diagnostic criteria is only approximately 0.8% for BP-I and 1.1% for BP-II, (Bauer and Pfennig, 2005; Pini et al., 2005; Waraich et al., 2004; Angst, 2004; Tohen and Angst, 2002; Wittchen et al., 2003; Weissman et al., 1996) recent clinical and epidemiological studies suggest that bipolar spectrum disorders might affect up to 6% of the general population (Angst, 1998; Angst et al., 2003; Akiskal et al., 2000; Akiskal and Benazzi, 2005; Judd and Akiskal, 2003; Benazzi and Akiskal, 2001). This estimate is uncertain, though, as bipolar spectrum disorders, which includes not only BP-I and BP-II but also cases with episodes of hypomania of lesser severity or briefer duration than specified in the DSM and ICD criteria, have not been the focus of sustained attention in large-scale community epidemiological studies.

An impediment to resolving this uncertainty is lack of information on the accuracy of fully structured diagnostic interviews to assess sub-threshold BPD. The current report presents results of a clinical reappraisal study to address this issue by evaluating the validity of Version 3.0 of the WHO Composite International Diagnostic Interview, (Kessler and Ustun, 2004) the most widely used fully structured diagnostic interview in psychiatric epidemiology, in assessing both threshold and sub-threshold BPD. Validity is assessed in comparison to blindly administered clinical re-interviews using the non-patient version of the Structured Clinical Interview for DSM-IV (SCID) (First et al., 2002) as the validity standard. Data are also presented on the accuracy of CIDI-based screening scales for BPD.

The clinical reappraisal study was carried out in conjunction with the National Comorbidity Survey Replication (NCS-R) (Kessler and Merikangas, 2004), a nationally representative survey of mental disorders among English-speaking household residents ages 18 and older in the continental US. A previous report of the main NCS-R clinical reappraisal study documented good CIDI-SCID concordance for lifetime diagnoses of most anxiety disorders, substance use disorders, and major depressive disorder, with  $\kappa$  for classes of disorder in the range .48–.54, positive predictive value (PPV; the percent of CIDI cases confirmed by the SCID) in the range .74–.99, and negative predictive value (NPV; the percent of CIDI non-cases confirmed by the SCID) in the range .80–.89 (Kessler et al., 2005b). BPD was not included in the main clinical reappraisal study because of its low prevalence. However, a separate clinical reappraisal study was subsequently carried out explicitly to evaluate BPD. The results of that study are reported here.

## METHODS

### The NCS-R survey design

The NCS-R was administered face-to-face to a sample of 9282 adult respondents between February 2001 and April 2003. The sample was based on a multi-stage clustered area probability design described in more detail elsewhere (Kessler et al., 2004b). Informed consent was obtained verbally prior to data collection. Consent was verbal rather than written to maintain consistency with the baseline NCS (Kessler et al., 1994). The response rate was 70.9%. Respondents were given a \$50 incentive for participation. A probability sub-sample of hard-to-recruit pre-designated respondents was selected for a brief telephone non-respondent survey. Non-respondent survey participants were given a \$100 incentive. The Human Subjects Committees of Harvard Medical School and the University of Michigan both approved these

recruitment and consent procedures. The results of the non-respondent survey were used to create a non-response adjustment weight that was added to more conventional within-household probability of selection and post-stratification weights to create a composite NCS-R weight. A more detailed discussion of NCS-R sampling and weighting is presented elsewhere (Kessler et al., 2004b).

### **CIDI assessment of bipolar disorders**

The World Health Organization's Composite International Diagnostic Interview (CIDI) Version 3.0 (Kessler and Ustun, 2004) is a fully structured lay-administered diagnostic interview. DSM-IV criteria were used to define mania, hypomania, and major depressive episode (MDE). The requirement that symptoms do not meet criteria for a Mixed Episode (Criterion C for mania and Criterion B for MDE) was not operationalized in making these diagnoses. Respondents were classified as having lifetime BP-I if they ever had a manic episode and as having lifetime BP-II if they never had a manic episode, ever had a hypomanic episode, and ever had an episode of MDE. Respondents were classified as having sub-threshold BPD if they met any of the following three sets of criteria: (i) they had a history of recurrent sub-threshold hypomania (at least two Criterion B symptoms, such as grandiosity or decreased need for sleep, along with all other criteria for hypomania) in the presence of MDE; (ii) they had a history of recurrent hypomania in the absence of recurrent MDE; or (iii) they had a history of recurrent sub-threshold hypomania in the absence of inter-current MDE. The reduction in number of required symptoms for a determination of sub-threshold hypomania was confined to two Criterion B symptoms (compared to the DSM-IV requirement of three or four if the mood is only irritable) in order to retain the core features of hypomania in the sub-threshold definition. Recurrent hypomania and sub-threshold hypomania absent inter-current MDE were included in the definition because hypomania in the absence of MDE is part of the DSM-IV definition of BPD NOS. All diagnoses excluded cases with plausible organic causes. For purposes of this paper, we define the bipolar spectrum as a lifetime history of BP-I, BP-II or sub-threshold BPD.

### **The BPD clinical reappraisal sample**

Clinical reappraisal interviews were administered to a probability sub-sample of 40 NCS-R respondents: 10 with CIDI/DSM-IV BP-I, 10 with CIDI/DSM-IV BP-II, 10 with CIDI/DSM-IV sub-threshold BPD, and 10 with no bipolar spectrum disorders who endorsed a CIDI diagnostic stem question for mania-hypomania. The clinical reappraisal sample dataset was weighted to adjust for the fact that CIDI cases were over-sampled, generating a weighted distribution in the clinical reappraisal sample with the same CIDI prevalence estimates for the three disorder classes as in the full NCS-R sample.

It is noteworthy that the clinical reappraisal sample included no respondents who denied the CIDI mania-hypomania diagnostic stem questions, as the prevalence of clinician-diagnosed BPD would have been so low in that sub-sample that a prohibitively large number of clinical reappraisal interviews would have been required to obtain a confidence interval narrow enough to be useful even in the absence of any positive case. To illustrate the problem, assume plausibly that the lifetime prevalence of BP-I in the entire sample was 1.0% and that PPV of the CIDI BPD assessment was well above .5 (Kessler et al., 1997), implying that the prevalence of clinician-assessed BP-I in the sub-sample of survey respondents who failed to endorse a CIDI BPD diagnostic stem question was well below 0.5%. This, in turn, would mean that the expected number of cases of clinician-diagnosed BP-I in clinical reappraisal interviews of respondents in this sub-sample would be zero unless clinical interviews were carried out with at least 200 such respondents, a number far greater than the number we were able to interview in the clinical reappraisal study.

A finding of zero prevalence in a smaller clinical reappraisal sample would be of little value because the confidence interval of this estimate would be consistent with prevalence as high as in the entire sample. This statement can be illustrated concretely by noting that the upper end of the 95% confidence interval of a simple random sample with an observed prevalence of zero is approximately  $3/n$ , where  $n$  is the number of respondents in the sample (Hanley and Lippman-Hand, 1983). Using this formula, we can see that a sample of at least 300 respondents would have been needed in the sub-sample who failed to endorse a CIDI BPD stem question to obtain an upper bound of the confidence interval of 1.0% even if all these respondents were classified as non-cases in the clinical reappraisal interviews. An even larger sample would have been needed to test the much more plausible hypothesis that true prevalence is no more than a small fraction of one percent in this sub-sample. Based on financial constraints on carrying out SCID interviews with this large a number of CIDI stem-question negatives, we concentrated our clinical reappraisal interviews on respondents who endorsed a CIDI BPD diagnostic stem question and assumed conservatively that prevalence would have been zero among other respondents in calculating concordance of CIDI diagnoses with clinical diagnoses.

### **Assessment of bipolar disorders in the clinical reappraisal sample**

BPD was assessed in the clinical reappraisal sample using the lifetime non-patient version of the Structured Clinical Interview for DSM-IV (SCID) (First et al., 2002) by two clinical interviewers with experience treating bipolar disorder. One interviewer was a PhD clinical psychologist with 10 years clinical experience. The other was an MSW with 15 years clinical experience. An expanded version of the standard SCID training program (Gibbon et al., 1981) was used for interviewer training. This program began with completion of the SCID training videotapes and manuals and was then followed by practice interviewing and supervisor (MG) feedback based on audiotapes of interviewer sections. Quality control monitoring throughout the production field period included supervisor review of all hard copy completed SCID interviews and weekly supervisor-interviewer review of completed cases. Clinical experts (HA, RMH) were used to resolve uncertainties in ratings.

The SCID interviews were administered over the telephone by interviewers who were blinded to the CIDI diagnoses. Telephone administration is now widely accepted in clinical reappraisal studies based on evidence of comparable validity to in-person administration (Kendler et al., 1992; Rohde et al., 1997; Sobin et al., 1993). A great advantage of telephone administration is that a centralized and closely supervised clinical interview staff can carry out the interviews throughout the country. A disadvantage is that the roughly 5% of people in the household population of the US without telephones cannot be included in clinical calibration studies when interviews are done by telephone.

### **Assessment of aggregate concordance**

After weighting the clinical reappraisal sample data to be representative of the main sample, we investigated whether CIDI prevalence estimates are comparable to SCID prevalence estimates using McNemar tests to evaluate the statistical significance of differences in the proportions of respondents who were false positives versus false negatives. As with all our significance tests, McNemar tests were carried out using .05-level two-sided evaluations with design-based estimation methods that adjusted for the effects of weighting and clustering and over-sampling of CIDI cases (Kish and Frankel, 1974; Wolter, 1985).

### **Assessment of individual-level concordance**

Individual-level CIDI-SCID diagnostic concordance was next evaluated using two different descriptive measures, Cohen's  $\kappa$  (Cohen, 1960) and the area under the receiver operating characteristic curve (AUC) (Hanley and McNeil, 1982). Although  $\kappa$  is the most widely used measure of concordance in validity studies of psychiatric disorders, it has been criticized

because it is dependent on prevalence and consequently is often low in situations where there appears to be high agreement between low-prevalence measures (Byrt et al., 1993; Cook, 1998; Kraemer et al., 2003). An important implication of this fact is that  $\kappa$  varies across populations that differ in prevalence even when the populations do not differ in sensitivity (SN; the percent of true cases correctly classified by the CIDI) or specificity (SP; the percent of true non-cases correctly classified). As sensitivity and specificity are considered to be fundamental parameters, this means that the comparison of  $\kappa$  across different populations cannot be used to evaluate the cross-population performance of a test.

Critics of  $\kappa$  prefer to assess concordance with measures that are a function of SN and SP. The odds-ratio (OR) meets this requirement, as OR is equal to  $[SN \times SP]/[(1-SN) \times (1-SP)]$  (Agresti, 1996). However, the upper end of the OR is unbounded, making it difficult to use the OR to evaluate the extent to which CIDI diagnoses are consistent with clinical diagnoses. Yules Q has been proposed as an alternative measure to resolve this problem (Spitznagel and Helzer, 1985), as Q is a bounded transformation of OR [ $Q = (OR - 1)/(OR + 1)$ ] that ranges between -1 and +1. Q can be interpreted as the difference in the probabilities of a randomly selected clinical case and a randomly selected clinical non-case that differ in their classification on the CIDI being correctly versus incorrectly classified by the CIDI. The difficulty with Q is that “tied pairs” (i.e., clinical cases and non-cases that have the same CIDI classification) are excluded, which means that Q does not tell us about actual prediction accuracy.

The AUC is a measure that resolves this problem, as AUC can be interpreted as the probability that a randomly selected clinical case will score higher on the CIDI than a randomly selected non-case. Although developed to study the association between a continuous predictor and a dichotomous outcome, the AUC can be used in the special case where the predictor is a dichotomy, in which case AUC equals  $(SN + SP)/2$ . As a result of this useful interpretation, we focus on AUC in our evaluation of CIDI-SCID diagnostic concordance. We also report SN and SP, the key components of AUC in the dichotomous case, as well as PPV, NPV and  $\kappa$ .

### Expanded assessment of concordance using CIDI symptom-level data

We estimated a stepwise logistic regression equation in which SCID diagnoses were treated as dichotomous outcomes and CIDI symptom variables were the predictors in order to determine whether CIDI symptom-level data could significantly improve the prediction of SCID diagnoses compared to prediction from CIDI diagnoses. As discussed in more detail elsewhere (Kessler et al., 2004a), significant improvement of this sort could be used to generate predicted probabilities of SCID diagnoses for each survey respondent who was not in the clinical reappraisal sample. Diagnostic imputations based on these predicted probabilities could then be used to make estimates of the prevalence and correlates of clinical diagnoses in the full sample so as to incorporate the analysis of validity into substantive investigations. For example, it would be possible in this way to carry out parallel analyses of the extent to which the correlates of predicted SCID diagnoses differ from the correlates of CIDI diagnoses.

A second goal in carrying out stepwise regression analysis was to determine whether a short subset of CIDI symptom questions could be selected to serve as a useful screening scale for BPD. Other useful disorder-specific screening scales have been developed from the CIDI (Sheehan et al., 1998; Kessler et al., 2005a). Although BPD screening scales already exist, (Soldani et al., 2005; Hirschfeld et al., 2000) they lack the psychometric properties one would want in a useful screen. For example, a large-scale community survey found that the widely used Mood Disorders Questionnaire (MDQ) detected only 28% of respondents independently classified by the SCID as having bipolar I or II disorders (Hirschfeld et al., 2003). The failure to detect 72% of SCID cases (i.e., an SN of .28) is a serious limitation, as useful screening scales capture the majority of true cases without including so many false positives that second-stage evaluation is not cost-effective. We are aware of no existing BPD screening scale that



has been shown to have such properties in a community survey, although evaluations of the recently developed Hypomania Checklist (HCL-32) (Angst et al., 2005) in clinical samples are very promising (Carta et al., 2006) and are now being extended to community samples in a number of countries.

Given the strong CIDI-SCID diagnostic concordance documented below, the stepwise logistic regression analysis to develop CIDI screening scales was carried out in the full NCS-R sample. The CIDI/DSM-IV diagnoses of BP-I/II and bipolar spectrum disorders were the outcomes. In order to understand the variables used as predictors, it is important to know that the CIDI, like many other psychiatric diagnostic interviews, uses a stem-branch structure to assess disorders. In the case of manic-hypomanic episode, two stem questions are used to operationalize elements of DSM-IV Criterion A (the existence of a distinct period of abnormally and persistently elevated, expansive, or irritable mood, lasting at least one week or any duration if hospitalization is necessary). The first is a rather complex question that asks respondents about euphoria (“Some people have periods lasting several days or longer when they feel much more excited and full of energy than usual. Their minds go too fast. They talk a lot. They are very restless or unable to sit still and they sometimes do things that are unusual for them, such as driving too fast or spending too much money. Have you ever had a period liked this lasting several days or longer?”) The second question asks about irritability (“Have you ever had a period lasting several days or longer when most of the time you were so irritable or grouchy that you either started arguments, shouted at people, or hit people?”). Respondents who say no to both these stem questions are coded as not having a history of either mania or hypomania. The stepwise regression analysis to develop a CIDI screen for BPD consequently excluded respondents who failed to endorse one or more of these screening questions.

In this subsample, the CIDI asks a follow-up Criterion B screening question: “People who have episodes like this often have changes in their thinking and behavior at the same time, like being more talkative, needing very little sleep, being very restless, going on buying sprees, and behaving in ways they would normally think are inappropriate. Did you ever have any of these changes during your episodes of being (excited and full of energy/very irritable or grouchy)” Respondents who say no to this question are coded as not having a history of either mania or hypomania. Those who say yes, in comparison, are asked to think of an episode when they had a large number of these problems and to answer 15 yes-no questions that operationalize the seven DSM-IV Criterion B symptoms of manic-hypomanic episode. Respondents who endorse between zero and two of these 15 questions are coded as not having a history of either mania or hypomania, while those who endorse three or more questions are administered additional questions about episode duration (to operationalize the Criterion A requirement of a seven-day duration for mania and four-day duration for hypomania), severity of role impairment (Criterion D for manic episode and Criteria C-E for hypomania episode), and possible organic causes (Criterion E for manic episode and Criterion F for hypomania episode). For precise question wording, see [www.hcp.med.harvard.edu/wmhcdi](http://www.hcp.med.harvard.edu/wmhcdi).

Based on this CIDI skip logic, the stepwise logistic regression analysis to develop CIDI-based screening scales was carried out in the sub-sample of respondents who endorsed the Criterion B screening question. The analysis focused on the 15 Criterion B symptom questions. Our goal was to determine whether a subset of these questions could be selected that screened for BPD with good accuracy. Separate analyses were carried out in the subsample of respondents who endorsed the euphoria stem question and in the larger subsample of those who endorsed either the euphoria or the irritability stem question in predicting both BP-I/II and bipolar spectrum disorders.

## RESULTS

### Aggregate concordance

The lifetime prevalence estimates (standard error in parentheses) of DSM-IV BP-I, BP-II, and sub-threshold BPD in the weighted SCID clinical reappraisal sample are 1.0% (0.6), 1.7% (0.7), and 1.4% (0.6), respectively, for a total prevalence estimate in the SCID of 4.0% compared to 4.4% in the CIDI. (Table 1) As noted above, these prevalence estimates are conservative, as they are based on the assumption that all main survey respondents who failed to endorse a CIDI BPD stem question would have been classified as non-cases if they had been administered clinical reappraisal interviews. McNemar tests are not significant either for individual diagnoses of BP-I, BP-II, or sub-threshold BPD ( $\chi_1^2 = 0.1-0.3$ ,  $p = .56-.75$ ) or for a summary measure of any bipolar spectrum disorder ( $\chi_1^2 = 0.5$ ,  $p = .49$ ). These results document that the CIDI assessment of DSM-IV BPD prevalence is unbiased in comparison to the SCID.

### Individual-level concordance

Individual-level CIDI-SCID concordance was found to be excellent for any bipolar spectrum disorder, with AUC of .998 and  $\kappa$  of .94. (Table 2) All SCID cases are positive on the CIDI (SN), while 99.5% of SCID non-cases are negative on the CIDI (SP). The proportion of CIDI cases confirmed by the SCID is 88.4% (PPV), while the proportion of CIDI non-cases confirmed as not being cases by the SCID is 100% (NPV). Individual-level concordance is also quite high for BP-I (AUC = .999,  $\kappa = .88$ ), but lower for BP-II (AUC = .834,  $\kappa = .50$ ) and considerably lower for sub-threshold BPD (AUC = .726,  $\kappa = .51$ ) due to comparatively low values of SN (.679 for BP-II, .457 for sub-threshold BPD) in conjunction with high values of SP (.990-.994).

The comparatively low values of SN for BP-II and sub-threshold BPD are due to CIDI-SCID inconsistencies in distinguishing between BP-II and sub-threshold BPD rather than to differences in distinguishing either BP-II or sub-threshold BPD from non-cases. This can be seen clearly by noting that excellent CIDI-SCID concordance exists for a composite diagnosis of either BP-II or sub-threshold BPD (AUC = .961,  $\kappa = .88$ ). SN for this composite diagnosis (.926) is dramatically higher than for either BP-II or sub-threshold BPD alone (.457-.679).

### Concordance using CIDI symptom-level data

The results reported in Table 2 apply to dichotomous CIDI coding schemes; that is, schemes in which each individual is classified either as a case or non-case. Unlike the situation in clinical practice, though, there is no need for dichotomous coding in epidemiological surveys. Classification accuracy can sometimes be improved by assigning predicted probabilities of being a case to individual respondents based on their symptom profiles rather than forcing each respondent to be classified dichotomously as a case or non-case. In order to investigate the implication of this approach for improving CIDI-SCID concordance in the classification of BPD, stepwise logistic regression of CIDI BPD symptom questions was carried out to predict SCID diagnoses of BP-I/II in the clinical reappraisal sample. (Parallel analyses were not carried out for BP-I or for bipolar spectrum disorders because AUC of the dichotomous classification is so high that meaningful improvement with a more refined assessment would be impossible.) The stepwise equation that included all significant (.05-level of significance) symptom-level predictors had an AUC of .985, which is higher than the AUC in Table 2 of .928 for the dichotomous CIDI classification.

## CIDI screening scales

Given the strong CIDI-SCID concordance found in the clinical reappraisal sample, forward stepwise logistic regression using a .05-level entry criterion to develop CIDI screening scales was carried out in the full NCS-R sample using CIDI diagnoses as the outcomes and CIDI symptom questions as predictors. We focused on the subsample of NCS-R respondents who endorsed the Criterion B screening question, using responses to the 15 Criterion B symptom questions as predictors. A subset of nine questions was found to capture the significant associations between the full set of 15 and the CIDI diagnoses of BP-I/II and bipolar spectrum disorder. The same set of nine questions (Table 3) emerged as important in these equations both among respondents who endorsed the CIDI euphoria diagnostic stem question and among the larger subset of respondents who endorsed either the euphoria or irritability stem question in predicting both BP-I/II and bipolar spectrum disorders.

A simple 0–9 count of the number of questions endorsed was cross-classified with CIDI diagnoses to examine dose-response relationships. Counts were collapsed using standard procedures for creating strata to construct stratum-specific likelihood-ratios (Peirce and Cornell, 1993). These strata were then dichotomized so as to create proportions of the population with positive screens 2–3 times the observed proportions of NCS-R respondents with the disorders. The goal in doing this was to determine whether dichotomous versions of these screening scales would detect the majority of respondents classified as cases by the full CIDI while increasing the number of false positives only modestly. In doing this, we were mindful of the fact that a screen can easily detect the majority of cases by using such a low threshold that a large proportion of the population screens in the positive range of the scale. This defeats the purpose of having a screening scale, though, as the critical requirement is to detect cases while keeping the number of false positives low. We consequently sought cut-points that would detect the majority of true cases while having a low proportion of false positives. We defined “low” for this purpose as a predicted prevalence no more than 2–3 times as high as the CIDI prevalence.

The most important statistics for evaluating the screening scales are SN and PPV. The former (SN) tells us the proportion of true cases (i.e., cases of DSM-IV BPD defined by the full CIDI) that can be detected by setting the threshold for screened positives at the place we did, while the latter (PPV) tells us the proportion of screened positives that are true cases. Evaluation of SN and PPV shows that the CIDI screening scales meet the desired requirements of detecting a high proportion of true cases (high SN) while minimizing the number of false positives (high PPV). Depending on whether only one (euphoria) or two (euphoria and irritability) screening questions are used to define the sub-sample that is administered further questions, whether the outcome under consideration is BP-I/II or bipolar spectrum disorders, and whether a broad or narrow threshold is selected, a CIDI screening scale consisting of 11–12 questions can detect between 67.2% and 96.0% of true cases, with a proportion of true cases among the screened positives in the range 31.5–52.0%. (Table 4) Similarly strong associations between the screening scales and full diagnoses were found in replications across a number of practically useful sub-samples, such as the sub-sample of respondents who were high users of primary care services in the year before interview and the sub-sample of respondents with low incomes. (Results not reported, but available on request.)

Stratum-specific coding rules were also developed for the screening scale to assign predicted probabilities of being a true case (PPV) across the range of the 0–9 scale in the total sample and important sub-samples. (Results not reported, but posted at [www.hcp.med.harvard.edu/ncs/bpdscreen](http://www.hcp.med.harvard.edu/ncs/bpdscreen)). Concordance of these dimensional classifications with the full CIDI was good (AUC = .744–.852) As one might expect, PPV for a given screening scale score increased when we focused on sub-samples with high prevalence, such as heavy users of primary care or users of specialty mental health services. These PPV values could be



used to generate estimates of prevalence and correlates of BPD in epidemiological surveys that included the screening scale but not the full CIDI. These prevalence estimates based on imputations from a dimensional classification are likely to be more accurate than those based on a dichotomous classification.

## LIMITATIONS

The results reported here are limited by the fact that the clinical reappraisal sample was small and included no NCS-R respondents who denied the CIDI diagnostic stem questions for mania-hypomania. The issue of omitted NCS-R respondents who denied the CIDI BPD stem questions is of special importance in that this design feature lets us to assume that all SCID cases of BPD were captured in the sub-sample of NCS-R respondents who endorsed one of the two CIDI diagnostic stem questions. There are several reasons to think that this assumption is incorrect. The most obvious of these is that survey respondents are not perfectly consistent in their reports even in response to identical questions, implying that at least some NCS-R respondents who would have been classified as BPD cases in the SCID denied the CIDI BPD stem questions.

Beyond this obvious problem, the CIDI euphoria stem question is a complex multi-component question that might confuse respondents with BPD who experienced some, but not all, of the symptoms described in the question, leading to some number of false negative responses. In addition, the CIDI stem question for euphoria does not emphasize what some experts (Akiskal and Benazzi, 2005) consider the core BPD feature of over-activity, while both the euphoria and irritability stem questions are phrased in the kind of negative way (e.g., thoughts going “too” fast) that might lead respondents with sub-threshold BPD, who often experience their symptoms as positive, to respond negatively. Both stem questions also require a minimum duration of “several days”, which is longer than the data-based definitions in recent studies of bipolar spectrum disorders (Benazzi and Akiskal, In Press). Another limitation is that the SCID is likely to miss some proportion of true BPD cases (Akiskal and Benazzi, 2005). An additional limitation is that the screening scales were evaluated in the same dataset in which they were developed, probably leading to an over-estimation of their concordance with full diagnoses.

Taken together, the above limitations mean that some number of people with true bipolar spectrum disorders were omitted from the analysis because of CIDI stem question false negatives, that some proportion of true cases in the sample might have been misclassified as non-cases because of SCID insensitivity, and that concordance of the CIDI with SCID diagnoses of the remaining cases was likely over-estimated due to absence of cross-validation. Based on these limitations, the SCID and CIDI prevalence estimates of DSM-IV BPD (4.0–4.4%) should be interpreted as conservative and the estimates of CIDI-SCID concordance should be interpreted as anticonservative.

## CONCLUSIONS

Within the context of these limitations, the results reported here suggest that the prevalence of DSM-IV bipolar spectrum disorder is at least 4.0% and, given the limitations noted above, probably higher. The CIDI 3.0 assessment of DSM-IV BPD has good concordance with independent SCID diagnoses both at the aggregate level (i.e., in terms of yielding unbiased estimates of prevalence) and at the individual level (i.e., in terms of classifying individual cases). The results also show that a fairly short (11–12) sub-set of CIDI questions can be used to create very useful screening scales for BP-I/II as well as screening scales for bipolar spectrum disorders, although the validity of this screen might be improved by modifying the CIDI BPD diagnostic stem questions in the ways described in the previous paragraph.

With regard to CIDI-SCID concordance, the results show that concordance is considerably higher for the classification of BP-I than BP-II, but that a highly accurate classification can be

made for a composite diagnosis of either BP-II or sub-threshold BPD. The CIDI does considerably less well, in comparison, distinguishing between SCID cases of BP-II and cases of sub-threshold BPD. This weakness appears to apply, though, only to the classification scheme that requires each respondent to be assigned dichotomously to a single diagnostic category. When the classification scheme is refined to assign each respondent a predicted probability of each diagnostic category, the CIDI provides a much more accurate distinction between BP-I/II and non-cases, where sub-threshold cases are included in the category of non-cases. Cross-validation in an independent dataset would be especially useful in evaluating this last conclusion, though, as it is based on a comparison between two small sub-samples of cases.

The results regarding the accuracy of the CIDI screening scales, in comparison, are based on the full NCS-R sample of 9282 respondents. The absence of cross-validation remains an issue that can only be addressed in an independent replication. However, the large size of the NCS-R sample made it possible to replicate the results regarding screening scale accuracy in theoretically important sub-samples. The finding that screening accuracy remains consistently high across all these sub-samples provides strong indirect support for the value of BPD screening scales based on the CIDI. It is noteworthy that these scales detected between 67% and 96% of true cases. This compares very favorably to the 28% of true cases detected by the widely-used MDQ screening scale for BPD (Hirschfeld et al., 2003). As noted earlier in the paper, though, another very promising screening instrument, the Hypomania Checklist (HCL-32) (Angst et al., 2005), is currently being tested in a number of countries in community surveys and might prove to be more useful than the CIDI in this regard.

It is important to recognize that the PPV of the CIDI-based screening scales, as that of any screening scale, is likely to vary across populations as a function of prevalence. This means that the estimates of PPV found here cannot be assumed to hold in all settings. For example, PPV might be higher in general medical samples and considerably higher in specialty mental health outpatient samples. This means that independent validation studies should ideally be carried out whenever these (or other) screening scales are being used to make prevalence estimates in new populations. In the absence of independent validation studies, estimates of PPV have been generated for a number of important sub-populations in the NCS-R (e.g., primary care users weighted by number of visits in the past year; low-income residents of urban areas, etc.) and are posted on the NCS web site ([www.hcp.med.harvard.edu/ncs/bpdscreen](http://www.hcp.med.harvard.edu/ncs/bpdscreen)).

#### Acknowledgements

The National Comorbidity Survey Replication (NCS-R) is supported by the National Institute of Mental Health (NIMH; U01-MH60220) with supplemental support from the National Institute of Drug Abuse, the Substance Abuse and Mental Health Services Administration, the Robert Wood Johnson Foundation (Grant # 044780), and the John W. Alden Trust. Additional support for preparation of this paper was provided by BristolMyersSquibb. Collaborating NCS-R investigators include Ronald C. Kessler (Principal Investigator, Harvard Medical School), Kathleen Merikangas (Co-Principal Investigator, NIMH), James Anthony (Michigan State University), William Eaton (The Johns Hopkins University), Meyer Glantz (NIDA), Doreen Koretz (Harvard University), Jane McLeod (Indiana University), Mark Olfson (Columbia University College of Physicians and Surgeons), Harold Pincus (University of Pittsburgh), Greg Simon (Group Health Cooperative), T Bedirhan Ustun (World Health Organization), Michael Von Korff (Group Health Cooperative), Philip Wang (Harvard Medical School), Kenneth Wells (UCLA), Elaine Wethington (Cornell University), and Hans-Ulrich Wittchen (Max Planck Institute of Psychiatry). The views and opinions expressed in this report are those of the authors and should not be construed to represent the views of any of the sponsoring organizations, agencies, or US Government. A complete list of NCS publications and the full text of all NCS-R instruments can be found at <http://www.hcp.med.harvard.edu/ncs>. Send correspondence to [NCS@hcp.med.harvard.edu](mailto:NCS@hcp.med.harvard.edu). The NCS-R is carried out in conjunction with the World Health Organization World Mental Health (WMH) Survey Initiative. We thank the staff of the WMH Data Collection and Data Analysis Coordination Centers for assistance with instrumentation, fieldwork, and consultation on data analysis. These activities were supported by the John D. and Catherine T. MacArthur Foundation, the Pfizer Foundation, the US Public Health Service (1R13MH066849, R01-MH069864, and R01 DA016558), Eli Lilly and Company, GlaxoSmithKline, Ortho-McNeil Pharmaceutical, Inc. and the Pan American Health Organization. A complete list of WMH publications and instruments can be found at (<http://www.hcp.med.harvard.edu/wmhcdi>).

## References

- Akiskal HS, Benazzi F. Optimizing the detection of bipolar II disorder in outpatient private practice: toward a systematization of clinical diagnostic wisdom. *J Clin Psychiatry* 2005;66:914–921. [PubMed: 16013908]
- Akiskal HS, Bourgeois ML, Angst J, Post R, Moller H, Hirschfeld R. Reevaluating the prevalence of and diagnostic composition within the broad clinical spectrum of bipolar disorders. *J Affect Disord* 2000;59 (Suppl 1):S5–S30. [PubMed: 11121824]
- Angst J. The emerging epidemiology of hypomania and bipolar II disorder. *J Affect Disord* 1998;50:143–151. [PubMed: 9858074]
- Angst J. Bipolar disorder--a seriously underestimated health burden. *Eur Arch Psychiatry Clin Neurosci* 2004;254:59–60. [PubMed: 15146333]
- Angst J, Adolfsson R, Benazzi F, Gamma A, Hantouche E, Meyer TD, Skeppar P, Vieta E, Scott J. The HCL-32: towards a self-assessment tool for hypomanic symptoms in outpatients. *J Affect Disord* 2005;88:217–233. [PubMed: 16125784]
- Angst J, Gamma A, Benazzi F, Ajdacic V, Eich D, Rossler W. Toward a redefinition of subthreshold bipolarity: epidemiology and proposed criteria for bipolar-II, minor bipolar disorders and hypomania. *J Affect Disord* 2003;73:133–146. [PubMed: 12507746]
- Bauer M, Pfennig A. Epidemiology of bipolar disorders. *Epilepsia* 46 Suppl 2005;4:8–13.
- Benazzi F, Akiskal H. The duration of hypomania in bipolar-II disorder in private practice: methodology and validation. *J Affect Disord*. In Press
- Benazzi F, Akiskal HS. Delineating bipolar II mixed states in the Ravenna-San Diego collaborative study: the relative prevalence and diagnostic significance of hypomanic features during major depressive episodes. *J Affect Disord* 2001;67:115–122. [PubMed: 11869758]
- Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol* 1993;46:423–429. [PubMed: 8501467]
- Carta MG, Hardoy MC, Cadeddu M, Murru A, Campus A, Morosini PL, Gamma A, Angst J. The accuracy of the Italian version of the Hypomania Checklist (HCL-32) for the screening of bipolar disorders and comparison with the Mood Disorder Questionnaire (MDQ) in a clinical sample. *Clin Pract Epidemiol Ment Health* 2006;2:2. [PubMed: 16524481]
- Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 1960;20:37–46.
- Cook, RJ. Kappa and its dependence on marginal rates. In: Armitage, P.; Colton, T., editors. *The Encyclopedia of Biostatistics*. Wiley; New York, NY: 1998. p. 2166-2168.
- First, MB.; Spitzer, RL.; Gibbon, M.; Williams, JBW. Biometrics Research. New York State Psychiatric Institute; New York, NY: 2002. Structured Clinical Interview for DSM-IV Axis I Disorders, Research Version, Non-patient Edition (SCID-I/NP).
- Gibbon M, McDonald-Scott P, Endicott J. Mastering the art of research interviewing. A model training procedure for diagnostic evaluation. *Arch Gen Psychiatry* 1981;38:1259–1262. [PubMed: 7305606]
- Hanley JA, Lippman-Hand A. If nothing goes wrong, is everything all right? Interpreting zero numerators. *JAMA* 1983;249:1743–1745. [PubMed: 6827763]
- Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29–36. [PubMed: 7063747]
- Hirschfeld RM, Holzer C, Calabrese JR, Weissman M, Reed M, Davies M, Frye MA, Keck P, McElroy S, Lewis L, Tierce J, Wagner KD, Hazard E. Validity of the mood disorder questionnaire: a general population study. *Am J Psychiatry* 2003;160:178–180. [PubMed: 12505821]
- Hirschfeld RM, Williams JB, Spitzer RL, Calabrese JR, Flynn L, Keck PE Jr, Lewis L, McElroy SL, Post RM, Rappaport DJ, Russell JM, Sachs GS, Zajecka J. Development and validation of a screening instrument for bipolar spectrum disorder: the Mood Disorder Questionnaire. *Am J Psychiatry* 2000;157:1873–1875. [PubMed: 11058490]
- Judd LL, Akiskal HS. The prevalence and disability of bipolar spectrum disorders in the US population: re-analysis of the ECA database taking into account subthreshold cases. *J Affect Disord* 2003;73:123–131. [PubMed: 12507745]

- Kessler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ. A population-based twin study of major depression in women. The impact of varying definitions of illness. *Arch Gen Psychiatry* 1992;49:257–266. [PubMed: 1558459]
- Kessler RC, Abelson J, Demler O, Escobar JI, Gibbon M, Guyer ME, Howes MJ, Jin R, Vega WA, Walters EE, Wang P, Zaslavsky A, Zheng H. Clinical calibration of DSM-IV diagnoses in the World Mental Health (WMH) version of the World Health Organization (WHO) Composite International Diagnostic Interview (WMHCIDI). *Int J Methods Psychiatr Res* 2004a;13:122–139. [PubMed: 15297907]
- Kessler RC, Adler L, Ames M, Demler O, Faraone S, Hiripi E, Howes MJ, Jin R, Secnik K, Spencer T, Ustun TB, Walters EE. The World Health Organization Adult ADHD Self-Report Scale (ASRS): a short screening scale for use in the general population. *Psychol Med* 2005a;35:245–256. [PubMed: 15841682]
- Kessler RC, Berglund P, Chiu WT, Demler O, Heeringa S, Hiripi E, Jin R, Pennell BE, Walters EE, Zaslavsky A, Zheng H. The US National Comorbidity Survey Replication (NCS-R): design and field procedures. *Int J Methods Psychiatr Res* 2004b;13:69–92. [PubMed: 15297905]
- Kessler RC, Berglund P, Demler O, Jin R, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* 2005b;62:593–602. [PubMed: 15939837]
- Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen HU, Kendler KS. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. *Arch Gen Psychiatry* 1994;51:8–19. [PubMed: 8279933]
- Kessler RC, Merikangas KR. The National Comorbidity Survey Replication (NCS-R): background and aims. *Int J Methods Psychiatr Res* 2004;13:60–68. [PubMed: 15297904]
- Kessler RC, Rubinow DR, Holmes C, Abelson JM, Zhao S. The epidemiology of DSM-III-R bipolar I disorder in a general population survey. *Psychol Med* 1997;27:1079–1089. [PubMed: 9300513]
- Kessler RC, Ustun TB. The World Mental Health (WMH) survey initiative version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *Int J Methods Psychiatr Res* 2004;13:93–121. [PubMed: 15297906]
- Kish L, Frankel MR. Inferences from complex samples. *J Roy Stat Soc* 1974;36:1–37.
- Kraemer HC, Morgan GA, Leech NL, Gliner JA, Vaske JJ, Harmon RJ. Measures of clinical significance. *J Am Acad Child Adolesc Psychiatry* 2003;42:1524–1529. [PubMed: 14627890]
- Peirce JC, Cornell RG. Integrating stratum-specific likelihood ratios with the analysis of ROC curves. *Med Decis Making* 1993;13:141–151. [PubMed: 8483399]
- Pini S, de Queiroz V, Pagnin D, Pezawas L, Angst J, Cassano GB, Wittchen HU. Prevalence and burden of bipolar disorders in European countries. *Eur Neuropsychopharmacol* 2005;15:425–434. [PubMed: 15935623]
- Rohde P, Lewinsohn PM, Seeley JR. Comparability of telephone and face-to-face interviews in assessing axis I and II disorders. *Am J Psychiatry* 1997;154:1593–1598. [PubMed: 9356570]
- Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R, Dunbar GC. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 1998;59 (Suppl 20):22–33. [PubMed: 9881538]quiz 34–57.
- Sobin C, Weissman MM, Goldstein RB, Adams P, Wickramaratne PJ, Warner V, Lisch JD. Diagnostic interviewing for family studies: comparing telephone and face-to-face methods for the diagnosis of lifetime psychiatric disorders. *Psychiatr Genet* 1993;3:227–234.
- Soldani F, Sullivan PF, Pedersen NL. Mania in the Swedish Twin Registry: criterion validity and prevalence. *Aust N Z J Psychiatry* 2005;39:235–243. [PubMed: 15777359]
- Spitznagel EL, Helzer JE. A proposed solution to the base rate problem in the kappa statistic. *Arch Gen Psychiatry* 1985;42:725–728. [PubMed: 4015315]
- Tohen, M.; Angst, J. Epidemiology of bipolar disorder. In: Tsuang, M.; Tohen, M., editors. *Textbook in Psychiatric Epidemiology*. Wiley; New York, NY: 2002. p. 427-444.
- Waraich P, Goldner EM, Somers JM, Hsu L. Prevalence and incidence studies of mood disorders: a systematic review of the literature. *Can J Psychiatry* 2004;49:124–138. [PubMed: 15065747]

- Weissman MM, Bland RC, Canino GJ, Faravelli C, Greenwald S, Hwu HG, Joyce PR, Karam EG, Lee CK, Lellouch J, Lepine JP, Newman SC, Rubio-Stipec M, Wells JE, Wickramaratne PJ, Wittchen H, Yeh EK. Cross-national epidemiology of major depression and bipolar disorder. *JAMA* 1996;276:293–299. [PubMed: 8656541]
- Wittchen HU, Mhlig S, Pezawas L. Natural course and burden of bipolar disorders. *Int J Neuropsychopharmacol* 2003;6:145–154. [PubMed: 12890308]
- Wolter, K. *Introduction to Variance Estimation*. Springer-Verlag; New York, NY: 1985.



**Table 1**

Significance of difference between SCID and CIDI lifetime prevalence estimates of DSM-IV bipolar disorders in the weighted clinical reappraisal sample (n = 40)

	SCID prevalence		McNemar test <sup>1</sup>
	%	(se)	$\chi^2$
I. Individual diagnoses			
BP-I	1.0	(0.6)	0.2
BP-II	1.7	(0.7)	0.3
Subthreshold BPD	1.4	(0.6)	0.1
II. Combinations of diagnoses			
BP-I/II	2.6	(1.0)	0.6
BP-II or subthreshold BPD	3.1	(1.0)	0.1
Any bipolar spectrum disorders	4.0	(1.2)	0.5

<sup>1</sup>None of the McNemar tests is significant at the .05 level.

Individual-level concordance of CIDI with SCID diagnoses of lifetime DSM-IV bipolar disorders in the weighted clinical reappraisal sample (n = 40)

Table 2

	AUC	$\kappa$	(se)	SN	(se)	SP	(se)	PPV	(se)	NPV	(se)
I. Individual diagnoses											
BP-I	.999	.88	(.26)	1.0	--	.998	(.001)	.789	(.174)	1.0	--
BP-II	.834	.50	(.38)	.679	(.189)	.990	(.005)	.408	(.183)	.997	(.002)
Subthreshold BPD	.726	.51	(.34)	.457	(.143)	.994	(.004)	.585	(.190)	.990	(.005)
II. Combinations of diagnoses											
BP-I/II	.928	.69	(.25)	.868	(.105)	.989	(.005)	.583	(.145)	.998	(.002)
BP-II or subthreshold BPD	.961	.88	(.14)	.926	(.047)	.995	(.004)	.847	(.110)	.998	(.001)
Any bipolar spectrum disorders	.998	.94	(.09)	1.0	--	.995	(.004)	.884	(.088)	1.0	--

**Table 3**

## Questions used in the CIDI-based BPD screening scales

**I.** Stem questions

1. Some people have periods lasting several days or longer when they feel much more excited and full of energy than usual. Their minds go too fast. They talk a lot. They are very restless or unable to sit still and they sometimes do things that are unusual for them, such as driving too fast or spending too much money. Have you ever had a period liked this lasting several days or longer?<sup>1</sup>
2. Have you ever had a period lasting several days or longer when most of the time you were so irritable or grouchy that you either started arguments, shouted at people, or hit people?

**II.** Criterion B screening question

1. People who have episodes like this often have changes in their thinking and behavior at the same time, like being more talkative, needing very little sleep, being very restless, going on buying sprees, and behaving in ways they would normally think are inappropriate. Did you ever have any of these changes during your episodes of being (excited and full of energy/very irritable or grouchy)?

**III.** Criterion B symptom questions

Think of an episode when you had the largest number of changes like these at the same time. During that episode, which of the following changes did you experience?

1. Were you so irritable that you either started arguments, shouted at people, or hit people?<sup>2</sup>
2. Did you become so restless or fidgety that you paced up and down or couldn't stand still?
3. Did you do anything else that wasn't usual for you – like talking about things you would normally keep private, or acting in ways that you'd usually find embarrassing?
4. Did you try to do things that were impossible to do, like taking on large amounts of work?
5. Did you constantly keep changing your plans or activities?
6. Did you find it hard to keep your mind on what you were doing?
7. Did your thoughts seem to jump from one thing to another or race through your head so fast you couldn't keep track of them?
8. Did you sleep far less than usual and still not get tired or sleepy?
9. Did you spend so much more money than usual that it caused you to have financial trouble?

<sup>1</sup>If this question is endorsed, the irritability stem question is skipped and the respondent goes directly to the Criterion B screening question.

<sup>2</sup>This question is asked only if the euphoria stem question is endorsed.

**Table 4**

Individual-level concordance of CIDI screening scales with SCID/DSM-IV diagnoses of lifetime DSM-IV the total NCS-R sample (n = 9282)

Outcome	AUC	$\kappa$	(se)	SN
I. Among respondents who endorsed the CIDI euphoria stem question				
A. BP-I/II				
Narrow <sup>I</sup>	.826	.50	(.03)	.672
Broad <sup>I</sup>	.926	.48	(.02)	.888
B. Bipolar spectrum disorders				
Narrow <sup>I</sup>	.847	.54	(.02)	.734
Broad <sup>I</sup>	.881	.53	(.02)	.815
II. Among respondents who endorsed the CIDI euphoria or irritability stem question				
A. BP-I/II				
Narrow <sup>I</sup>	.848	.59	(.02)	.728
Broad <sup>I</sup>	.950	.56	(.02)	.962
B. Bipolar spectrum disorders				
Narrow <sup>I</sup>	.846	.49	(.03)	.717
Broad <sup>I</sup>	.948	.46	(.02)	.940

<sup>I</sup>The narrow cut-point was selected in each case to approximate as closely as possible a situation in which the number of screened positives was two times the number of true cases. The broad cut-point was selected to approximate a situation in which the number of screened positives was three times the number of true cases.