CIDI 3.0 Bipolar screening scales

Ronald C. Kessler, Ph.D.
Harvard Medical School

January 2007

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 development of the CIDI BPD screening scales was provided by BristolMyersSquibb. The NCS-R is carried out in conjunction with the World Health Organization World Mental Health (WMH) Survey Initiative. The staff of the WMH Data Collection and Data Analysis Coordination Centers helped 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.

Screening scales for DSM-IV bipolar disorder (BPD) were developed from version 3.0 of the WHO Compositive International Diagnostic Interview (CIDI 3.0)[1] based on secondary analysis of symptom-level data collected in the US National Comorbidity Survey Replication (NCS-R).[2] The scales have either 11 or 12 questions, depending on whether or not the irritability diagnostic stem question is included. Respondents who fail to endorse either of the first one (euphoria) or two (euphoria or irritability) questions are skipped out of the remainder of the series. Failure to endorse the third question after endorsing one of the first two also leads to a skip out of the remainder of the questions. Respondents who endorse the third question are administered 9 additional symptom questions. The complete set of 11-12 questions takes no more than three minutes to complete. In general population samples, it can be expected that close to 90% of the sample will skip out by the end of the third question, which means that average administration time would be less than one minute. Diagnoses based on the screening scales have excellent concordance with diagnoses based on the full CIDI. (Table 2) CIDI diagnoses, in turn, have excellent concordance with clinical diagnoses based on blinded SCID[3] clinical reappraisal interviews.[4]

(Tables 1 and 2 about here)

Based on the fact that strong CIDI-SCID concordance was found for the DSM-IV BPD in the NCS-R clinical reappraisal sample,[4] forward stepwise logistic regression using a .05-level entry criterion was used to develop the CIDI BPD screening scales in the full NCS-R sample. CIDI diagnoses were used as the outcomes and CIDI symptom questions were used as the predictors in these stepwise analyses. We focused on the subsample of NCS-R respondents who endorsed at least one of the two CIDI BPD Criterion A diagnostic stem questions, which asked about episodes of euphoria or irritability, and the CIDI BPD Criterion B screening question,

which asked for a single yes/no response regarding whether any other associated symptoms occured during the episodes of euphoria or irritability. These three questions are reproduced in Parts I and II of Table 1.

The full CIDI administers 15 additional Criterion B mania-hypomania symptom questions to respondents who endorsed the above questions. Our stepwise analysis investigated whether a smaller subset of these 15 questions might accurately predict complete CIDI diagnoses, noting that the latter require not only endorsement of a sufficient number of Criterion B symptom questions (a minimum of three in the presence of euphoria and four in the presence of irritability without euphoria) but also Criteria C and D symptoms. A subset of nine CIDI Criterion B symptom questions was found to capture the significant associations between the full set of 15 and the CIDI diagnoses of BP-I/II disorder. These nine questions are reproduced in Part III of Table 1. The same set of nine questions emerged as the important ones 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.

Scores based on a simple 0-9 count of the number of questions endorsed were cross-classified with CIDI diagnoses to examine dose-response relationships. Counts were collapsed using standard procedures for creating strata to construct stratum-aspecific likelihood-ratios.[5] 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 which, in the case of BP-I/II, require a history of major depressive episode in addition to a history of hypomania) 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 minimimizing the number of false positives (high PPV). Depending on whether only one (euphoria) or two (euphoria and irritablity) screening questions are used to define the sub-sample that is administered further questions, whether the outcome under considerarion 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 (depending on whether or not the irritability stem question is used) can detect between 67.2% and 96.2% of true DSM-IV cases of BPD, with a proportion of true cases among the screened positives in the range 31.5-52.0%. (Table 2) 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.

(Tables 3-6 about here)

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. (Tables 3-6) Concordance of these dimensional classifications with the full CIDI is good (AUC = .744-.852) As one might expect, PPV for a given screening scale score increases when we focus on sub-samples with high prevalence, such as heavy users of primary care or users of specialty mental health services. These PPV values can be used to assign predicted probabilities of a DSM-IV diagnosis of BPD to respondents who completed the screening scales by focusing on the table that deals with the outcome of interest -- either DSM-IV BP-I/II (Tables 3-4) or BP spectrum disorder (Tables 5-6) -- and on whether the irritability stem question was administered (Tables 4 and 6) or not administered (Tables 3 and 5) and by selecting the reference population that is most relevant to the sample under investigation. Prevalence estimates based on imputations from these dimensional classifications are likely to be more accurate than those based on a dichotomous classification.

With regard to the decision whether or not to use the irritability stem question: it should be noted that even though concordance with a full CIDI diagnosis is higher when the irritability stem question is used rather than not (Table 2), the improvement is small and the inclusion of the complex skip rules required when this is done makes it very difficult to have patients self-administer this version of the screening scale using a paper and pencil form (although this can easily be done with computerized self-administration, as in internet surveys). Users who need a

paper and pencil self-administered screener consequently might want to consider the version that does not use the irritability stem question.

REFERENCES

1. 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.

2. Kessler RC, Merikangas KR. The National Comorbidity Survey Replication (NCS-R): background and aims. *Int J Methods Psychiatr Res.* 2004;13:60-68.

3. First MB, Spitzer RL, Gibbon M, Williams JBW. *Structured Clinical Interview for DSM-IV Axis I Disorders, Research Version, Non-patient Edition (SCID-I/NP).* New York, NY: Biometrics Research, New York State Psychiatric Institute; 2002.

4. Kessler RC, Aksikal HS, Angst J, Guyer M, Hirschfeld RMA, Merikangas KR, Stang PE. Validity of the assessment of bipolar spectrum disorders in the WHO CIDI 3.0. *J Affect Disord.* in press.

5. Peirce JC, Cornell RG. Integrating stratum-specific likelihood ratios with the analysis of ROC curves. *Med Decis Making.* 1993;13:141-151.

Table 1. 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 or 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? [1]
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?[2]
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 thing 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?

---

[1] If this question is endorsed, the irritability stem question is skipped and the respondent goes directly to the Criterion B screening question.
[2] This question is asked only if the euphoria stem question is endorsed.

Table 2. Individual-level concordance of CIDI screening scales with DSM-IV/CIDI diagnoses of lifetime DSM-IV bipolar disorders in the total NCS-R sample (n = 9282)

Outcome	AUC	κ	(se)	SN	(se)	PPV	(se)
I. Among respondents who endorsed the CIDI euphoria stem question							
A. BP-I/II							
Narrow[1]	.826	.50	(.03)	.672	(.031)	.417	(.030)
Broad[1]	.926	.48	(.02)	.888	(.020)	.342	(.027)
B. Bipolar spectrum disorders							
Narrow[1]	.847	.54	(.02)	.734	(.023)	.462	(.023)
Broad[1]	.881	.53	(.02)	.815	(.020)	.421	(.025)
II. Among respondents who endorsed the CIDI euphoria or irritability stem question							
A. BP-I/II							
Narrow[1]	.848	.59	(.02)	.728	(.025)	.520	(.024)
Broad[1]	.950	.56	(.02)	.962	(.010)	.419	(.020)
B. Bipolar spectrum disorders							
Narrow[1]	.846	.49	(.03)	.717	(.031)	.388	(.029)
Broad[1]	.948	.46	(.02)	.940	(.019)	.315	(.026)

[1]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.

Table 3. AUC and stratum-specific PPV of the CIDI-based BPD screening scale for DSM-IV BP-I/II based on endorsement of the CIDI euphoria diagnostic stem question in the total sample and a number of critical sub-samples

	The sample in which PPV is calculated											
	Total		High PCP[1]		WT PCP[1]		SMH[1]		EMPL[1]		LOW INC[1]	
Imputation stratum	**PPV**	**(se)**	**PPV**	**(se)**	**PPV**	**(se)**	**PPV**	**(se)**	**PPV**	**(se)**	**PPV**	**(se)**
Negative Criterion B screen[2]	0.1	(0.0)	0.2	(0.2)	0.3	(0.1)	0.7	(0.3)	0.1	(0.1)	0.2	(0.1)
Number Endorsed/Positive Criterion B screen[2]												
0	0.3	(0.3)	0.0	(0.0)	0.3	(0.3)	0.0	(0.0)	0.4	(0.4)	0.0	(0.0)
1	0.3	(0.3)	0.0	(0.0)	0.3	(0.3)	0.0	(0.0)	0.4	(0.4)	0.0	(0.0)
2	0.3	(0.3)	0.0	(0.0)	0.3	(0.3)	0.0	(0.0)	0.4	(0.4)	0.0	(0.0)
3	3.3	(1.1)	4.9	(4.8)	3.5	(1.7)	13.8	(6.0)	2.4	(0.9)	0.0	(0.0)
4	3.3	(1.1)	4.9	(4.8)	3.5	(1.7)	13.8	(6.0)	2.4	(0.9)	0.0	(0.0)
5	22.0	(3.5)	18.4	(8.6)	22.0	(4.7)	42.0	(8.9)	18.8	(4.3)	17.4	(7.4)
6	31.0	(4.3)	42.5	(12.9)	42.9	(11.9)	48.2	(9.2)	26.6	(4.8)	40.5	(5.9)
7	41.5	(5.8)	52.0	(9.4)	52.6	(8.3)	55.2	(9.5)	39.5	(5.3)	40.5	(5.9)
8	56.3	(6.7)	52.0	(9.4)	52.6	(8.3)	69.6	(11.1)	52.0	(9.0)	40.5	(5.9)
9	67.5	(14.9)	100.0	(0.0)	85.3	(8.8)	88.2	(11.2)	57.7	(18.2)	50.3	(22.2)
AUC	.844		.845		.841		.774		.849		.821	
(n)	(9282)		(578)		(5225)		(821)		(6314)		(695)	

[1]High PCP = respondents who saw a primary care physician at least 12 times in the year before interview; WT PCP = respondents who was a primary care physician at least ones in the year before interview weighted for number of visits; SMH = respondents who received specialty mental health treatment in the year before interview; LOW INC = respondents with family incomes at or below the federal poverty line.

[2]Negative stem = respondents who either failed to endorse the CIDI BPD euphoria diagnostic stem question or endorsed the diagnostic stem question but failed to endorse the Criterion B screening question; Number endorsed/Positive Criterion B screen = number of Criterion B symptom questions endorsed out of a maximum of nine among respondents who endorsed the Criterion B screening question. Equivalence of PPV for some numbers is due to these sub-samples being collapsed because of insignificant differences in PPV.

Table 4. AUC and stratum-specific PPV of the CIDI-based BPD screening scale for DSM-IV BP-I/II based on endorsement of the CIDI euphoria or irritability diagnostic stem questions in the total sample and a number of critical sub-samples

	The sample in which PPV is calculated											
	Total		High PCP[1]		WT PCP[1]		SMH[1]		EMPL[1]		LOW INC[1]	
Imputation stratum	**PPV**	**(se)**	**PPV**	**(se)**	**PPV**	**(se)**	**PPV**	**(se)**	**PPV**	**(se)**	**PPV**	**(se)**
Negative Criterion B screen[2]	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Number Endorsed/Positive Criterion B screen[2]												
0	0.2	(0.2)	0.0	(0.0)	0.2	(0.2)	0.0	(0.0)	0.3	(0.3)	0.0	(0.0)
1	0.2	(0.2)	0.0	(0.0)	0.2	(0.2)	0.0	(0.0)	0.3	(0.3)	0.0	(0.0)
2	0.2	(0.2)	0.0	(0.0)	0.2	(0.2)	0.0	(0.0)	0.3	(0.3)	0.0	(0.0)
3	3.0	(0.9)	3.6	(3.5)	3.0	(1.3)	10.4	(4.7)	2.2	(0.8)	0.0	(0.0)
4	3.0	(0.9)	3.6	(3.5)	3.0	(1.3)	10.4	(4.7)	2.2	(0.8)	0.0	(0.0)
5	19.7	(3.2)	17.0	(8.1)	20.8	(4.3)	39.0	(6.3)	16.0	(3.8)	15.5	(6.7)
6	26.8	(3.5)	33.4	(11.0)	37.2	(10.8)	39.0	(6.3)	24.5	(4.2)	31.1	(8.7)
7	39.5	(5.9)	52.6	(14.9)	50.2	(10.9)	55.2	(9.5)	37.2	(5.4)	36.1	(10.9)
8	56.2	(6.2)	54.9	(14.2)	53.7	(12.0)	71.0	(10.9)	51.7	(8.3)	44.2	(10.1)
9	67.5	(14.9)	100.0	(0.0)	85.3	(8.8)	88.2	(11.2)	57.7	(18.2)	50.3	(22.2)
AUC	.852		.865		.854		.800		.858		.848	
(n)	(9282)		(578)		(5225)		(821)		(6314)		(695)	

[1]High PCP = respondents who saw a primary care physician at least 12 times in the year before interview; WT PCP = respondents who was a primary care physician at least ones in the year before interview weighted for number of visits; SMH = respondents who received specialty mental health treatment in the year before interview; LOW INC = respondents with family incomes at or below the federal poverty line.

[2]Negative stem = respondents who either failed to endorse the CIDI BPD euphoria diagnostic stem question or endorsed the diagnostic stem question but failed to endorse the Criterion B screening question; Number endorsed/Positive Criterion B screen = number of Criterion B symptom questions endorsed out of a maximum of nine among respondents who endorsed the Criterion B screening question. Equivalence of PPV for some numbers is due to these sub-samples being collapsed because of insignificant differences in PPV.

Table 5. AUC and stratum-specific PPV of the CIDI-based BPD screening scale for DSM-IV bipolar spectrum disorder based on endorsement of the CIDI euphoria diagnostic stem question in the total sample and a number of critical sub-samples

	The sample in which PPV is calculated											
	Total		High PCP[1]		WT PCP[1]		SMH[1]		EMPL[1]		LOW INC[1]	
Imputation stratum	**PPV**	**(se)**	**PPV**	**(se)**	**PPV**	**(se)**	**PPV**	**(se)**	**PPV**	**(se)**	**PPV**	**(se)**
Negative Criterion B screen[2]	0.7	(0.1)	2.0	(0.9)	1.2	(0.3)	2.8	(0.8)	0.8	(0.2)	0.7	(0.2)
Number Endorsed/Positive Criterion B screen[2]												
0	3.2	(1.6)	0.0	(0.0)	2.0	(1.4)	0.0	(0.0)	4.5	(2.2)	10.5	(7.7)
1	3.2	(1.6)	0.0	(0.0)	2.0	(1.4)	0.0	(0.0)	4.5	(2.2)	10.5	(7.7)
2	9.3	(3.1)	15.5	(9.8)	11.6	(5.1)	0.0	(0.0)	8.5	(3.4)	18.2	(6.0)
3	23.6	(4.2)	25.1	(13.3)	26.4	(5.3)	33.1	(11.0)	25.3	(5.2)	18.2	(6.0)
4	32.1	(3.9)	25.8	(11.3)	40.3	(11.7)	53.5	(9.9)	33.7	(4.8)	18.2	(6.0)
5	44.2	(3.4)	43.3	(7.4)	47.7	(6.2)	60.1	(5.2)	40.8	(3.1)	47.4	(6.6)
6	44.2	(3.4)	43.3	(7.4)	47.7	(6.2)	60.1	(5.2)	40.8	(3.1)	47.4	(6.6)
7	57.0	(6.1)	55.2	(14.7)	61.2	(10.4)	60.1	(5.2)	56.2	(6.0)	47.4	(6.6)
8	67.2	(6.9)	70.4	(12.5)	66.1	(12.5)	76.5	(9.4)	63.2	(9.2)	57.7	(12.1)
9	67.2	(6.9)	70.4	(12.5)	66.1	(12.5)	76.5	(9.4)	63.2	(9.2)	57.7	(12.1)
AUC	.744		.736		.732		.699		.730		.718	
(n)	(9282)		(578)		(5225)		(821)		(6314)		(695)	

[1]High PCP = respondents who saw a primary care physician at least 12 times in the year before interview; WT PCP = respondents who was a primary care physician at least ones in the year before interview weighted for number of visits; SMH = respondents who received specialty mental health treatment in the year before interview; LOW INC = respondents with family incomes at or below the federal poverty line.

[2]Negative stem = respondents who either failed to endorse either of the two CIDI BPD diagnostic stem questions or endorsed a diagnostic stem question but failed to endorse the Criterion B screening question; Number endorsed/Positive Criterion B screen = number of Criterion B symptom questions endorsed out of a maximum of nine among respondents who endorsed the Criterion B screening question. Equivalence of PPV for some numbers is due to these sub-samples being collapsed because of insignificant differences in PPV.

Table 6. AUC and stratum-specific PPV of the CIDI-based BPD screening scale for DSM-IV bipolar spectrum disorder based on endorsement of either the CIDI euphoria or irritability diagnostic stem questions in the total sample and a number of critical sub-samples

	The sample in which PPV is calculated											
	Total		High PCP[1]		WT PCP[1]		SMH[1]		EMPL[1]		LOW INC[1]	
Imputation stratum	**PPV**	**(se)**	**PPV**	**(se)**	**PPV**	**(se)**	**PPV**	**(se)**	**PPV**	**(se)**	**PPV**	**(se)**
Negative Criterion B screen[2]	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Number Endorsed/Positive Criterion B screen[2]												
0	2.5	(1.3)	0.0	(0.0)	1.7	(1.1)	0.0	(0.0)	3.5	(1.8)	7.5	(5.4)
1	2.5	(1.3)	0.0	(0.0)	1.7	(1.1)	0.0	(0.0)	3.5	(1.8)	7.5	(5.4)
2	7.2	(2.3)	9.1	(6.3)	7.1	(3.7)	0.0	(0.0)	7.1	(2.7)	10.5	(5.1)
3	21.5	(3.7)	29.6	(8.5)	25.7	(5.0)	30.4	(9.9)	23.4	(4.5)	10.5	(5.1)
4	30.6	(3.1)	29.6	(8.5)	37.1	(10.4)	45.4	(8.6)	31.7	(3.9)	21.2	(7.2)
5	46.2	(3.1)	44.0	(6.2)	47.2	(5.6)	62.3	(4.6)	43.8	(3.2)	45.6	(6.1)
6	46.2	(3.1)	44.0	(6.2)	47.2	(5.6)	62.3	(4.6)	43.8	(3.2)	45.6	(6.1)
7	58.4	(5.7)	55.2	(14.7)	61.4	(9.6)	62.3	(4.6)	56.6	(5.8)	45.6	(6.1)
8	68.6	(6.5)	72.1	(11.9)	67.4	(11.8)	77.4	(9.0)	65.3	(8.3)	59.8	(11.7)
9	68.6	(6.5)	72.1	(11.9)	67.4	(11.8)	77.4	(9.0)	65.3	(8.3)	59.8	(11.7)
AUC	.765		.746		.755		.738		.750		.739	
(n)	(9282)		(578)		(5225)		(821)		(6314)		(695)	

[1]High PCP = respondents who saw a primary care physician at least 12 times in the year before interview; WT PCP = respondents who was a primary care physician at least ones in the year before interview weighted for number of visits; SMH = respondents who received specialty mental health treatment in the year before interview; LOW INC = respondents with family incomes at or below the federal poverty line.

[2]Negative stem = respondents who either failed to endorse either of the two CIDI BPD diagnostic stem questions or endorsed a diagnostic stem question but failed to endorse the Criterion B screening question; Number endorsed/Positive Criterion B screen = number of Criterion B symptom questions endorsed out of a maximum of nine among respondents who endorsed the Criterion B screening question. Equivalence of PPV for some numbers is due to these sub-samples being collapsed because of insignificant differences in PPV.