

# Test performance of the Cantonese Chinese Mood Disorder Questionnaire for detecting bipolar spectrum disorder in the community of Hong Kong

Received 31 May 2011; Revised 20 July 2011; Accepted 22 July 2011

**Key words:** Bipolar spectrum disorder, Mood Disorder Questionnaire, Reliability, Concordance.

Dear Editor

Bipolar spectrum disorder (BSD) is common in the general population with a lifetime prevalence of 2.4–5%. A recent cross-national community epidemiological study confirmed that it is a common and valid illness entity across 11 countries (Merikangas *et al.*, 2011). Although BSD has been uncommonly studied in community settings (Benazzi, 2007; Lee *et al.*, 2009), its early age of onset, high prevalence, typically late recognition and impairing nature (Merikangas *et al.*, 2011) have prompted recent interest in early detection in both clinical and community settings. The under-recognition of BSD (Akiskal *et al.*, 2000) may be improved by enhancing the reliability and validity of screening instruments (Young & MacPherson, 2011).

The Mood Disorder Questionnaire (MDQ) is commonly used to screen for lifetime manic or hypomanic syndromes (Hirschfeld, 2002). Validation studies across different settings have not produced consistent results. Overall, they found its English and several non-English versions to exhibit moderate sensitivity and high specificity for assessing bipolar disorder among clinical samples (Hirschfeld *et al.*, 2003). However, it was less commonly examined in the community and was usually used for the screening of bipolar I (BP-I) and bipolar II (BP-II) disorders rather than BSD. Available studies suggested that it exhibited a much lower sensitivity for bipolar disorder in community studies than in clinical studies (Hirschfeld *et al.*, 2003), and that seemed to be especially so in a Chinese setting (Chung *et al.*, 2009). The restrictive criteria of bipolar disorder and the telephone-based mode of clinical reappraisal interviews adopted in these studies might have contributed to the finding of low sensitivity. Moreover, these studies used the single four-level impairment item in the original MDQ

which validation studies had found to adversely affect sensitivity (Weber Rouget *et al.*, 2005; Chung *et al.*, 2008; Kim *et al.*, 2008).

The role of the MDQ in screening for BSD has not been studied in Chinese populations before. The present study examined the concordance of the Chinese MDQ with face-to-face clinical diagnostic interviews in a general population setting. A clinical diagnosis of BSD referred to the Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV) diagnoses of BP-I and BP-II disorder, as well as bipolar disorder not otherwise specified (NOS) which consists of major depressive episode accompanied by sub-threshold hypomania lasting 2–3 days. We attempted to improve on the previous community studies in several ways. We replaced the single four-level impairment item of the MDQ with the multi-domain Sheehan Disability Scale (SDS) (Leon *et al.*, 1997). In translating the MDQ, we paid attention to the contextual meanings of the items with a view to enhancing their sensitivity without unduly changing the original meanings. Finally, we conducted detailed face-to-face interviews using an enhanced version of the Structured Clinical Interview for DSM-IV (SCID) that assesses a spectrum of hypomania beyond conventionally recognized bipolar disorder (Benazzi & Akiskal, 2003).

## Methods

An independent survey research organization, the Hong Kong Institute of Asia-Pacific Studies of The Chinese University of Hong Kong, was commissioned to conduct the telephone survey from January to February 2007. Trained interviewers obtained verbal consent from respondents prior to each successfully completed interview that lasted 7.3 min on average (s.d. = 3.1). Three hundred and eighty of 3016 successfully interviewed respondents expressed an interest to participate in a subsequent face-to-face interview. Among these respondents, a research assistant

---

\* Address for correspondence: Professor Sing Lee, Director, Hong Kong Mood Disorders Center, 7A, Block E, Staff Quarters, Prince of Wales Hospital, Shatin, N.T., Hong Kong.  
(E-mail: singlee@cuhk.edu.hk)

identified 87 who fulfilled the DSM-IV criteria of 1-year major depressive episode and any lifetime hypomanic/manic symptoms as assessed in the telephone survey (Lee *et al.*, 2009). Thirty-seven of these 87 respondents took part in the re-interview. The rest of the re-interview sample ( $n=68$ ) was randomly selected from respondents who did not fulfill the above criteria. From March 2007 to January 2008, 105 respondents were re-interviewed. This sample size was larger than what would be required (82) for the anticipated area under curve (AUC, 0.8) and its S.D. (0.05).

The research assistant assigned re-contacted telephone survey respondents to six clinical interviewers who were blind to respondents' result in the phone survey. Written informed consent was obtained prior to these interviews that lasted 2 hours on average. The clinical interviewers consisted of four practicing psychiatrists, one clinical psychologist and one senior research assistant with clinical training. They all had previous research experience in using the Chinese SCID and went through three 3-h training and consensus-building meetings with three patients diagnosed as having DSM-IV bipolar disorder. The ethics review board of The Chinese University of Hong Kong approved the above procedure of the study.

### Instrument

The telephone survey instrument was composed of the Cantonese Chinese MDQ, SDS, questions for the assessment of 12-month DSM-IV major depressive episode and mania/hypomania, help-seeking behavior and socio-demographic information. The MDQ is a self-report inventory of 13 yes/no questions about any lifetime history of manic or hypomanic syndrome(s). It has another binary item asking whether several of the endorsed symptoms occur during the same period of time, and a four-point scale of functional impairment. Endorsing seven items or more was previously chosen as an optimal cut-off (Hirschfeld *et al.*, 2000). Instead of the single four-level impairment item, we used the SDS that assesses in greater detail how manic/hypomanic symptoms interfered with functioning in four domains of life, namely, work, housework, close relationship and social roles. Responses were scored with a 0–10 scale and severity was classified as, none (0), mild (1–3), moderate (4–6), severe (7–9) and very severe (10). The Chinese version of the SDS was widely used in community surveys (Lee *et al.*, 2008). Scores in the moderate or high range were taken to indicate impairment.

Translation of the MDQ items was performed by experienced bilingual investigators (S. Lee and A. Tsang) and adopted a collaborative and iterative

approach. For example, the literal Chinese translation of MDQ8 ‘...you had much more energy than usual?’ could be misunderstood as ‘being more sexually active than usual’. When the item was understood behaviorally, it could also be taken to mean ‘being more active and doing more things than usual’, which was covered by the item MDQ9. Therefore, the translation we used emphasized the feeling of being eager to do more and it became ‘... more eager to do more or having more plans than usual’. The translated MDQ was pilot-tested face-to-face with three outpatients with a history of DSM-IV bipolar disorder and through telephone with 24 non-patients for further linguistic adaptation. In order to screen positively for BSD, in addition to a threshold number of items, the respondent had to report that the symptoms clustered in the same time period and caused moderate or more range of impairment as assessed by the SDS.

We used the non-patient Chinese version of the SCID (First *et al.*, 2002; So *et al.*, 2003). To make the SCID less stringent for detecting hypomania, we removed all skip-out instructions in the lifetime manic/hypomanic episode sections so that all lifetime manic/hypomanic symptoms and behaviors were assessed. Diagnoses of hypomanic and manic episode followed the DSM-IV. Moreover, in the enhanced version of the SCID used in this study, hypomania lasting 2–3 days was classified as a valid (sub-threshold) episode. Accordingly, a diagnosis of major depressive disorder with sub-threshold hypomania would be classified as bipolar disorder NOS (Benazzi & Akiskal, 2003). BSD in the present study thus included those with BP-I, BP-II or BP-NOS.

### Analysis

The telephone survey responses of the re-interviewed participants were retrieved from the survey data pool and combined into the data file containing their responses to SCID questions asking about mood disorders. We assessed overall diagnostic efficiency by estimating non-parametrically the AUC ( $AUC = [\text{sensitivity} + \text{specificity}]/2$ ) from receiver operating characteristic analyses. This was done in terms of the concordance between MDQ dichotomous classification and the primary criterion measure of SCID diagnosis of any BSD. While calibrating the cut-off, we examined variations in sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). Since these values are affected by the base rate of a disorder and may not accurately reflect the performance of an instrument, we also computed the positive and negative diagnostic likelihood ratios (DLR+ and DLR-) (Pepe, 2004). Analysis was performed with SPSS and Excel.

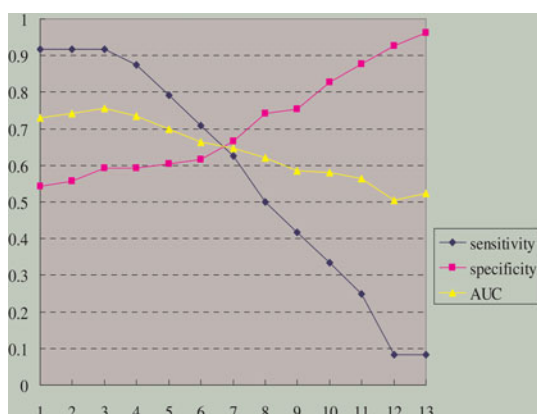
## Result

### Participants

With a participation rate of 63.6%, a total of 3016 respondents were interviewed by telephone (1414 males and 1601 females; age 18–24 (13.2%), age 25–34 (21.5%), age 35–44 (25.5%), age 45–54 (24.7%), age 55–65 (15.1%); 62.8% married/cohabited, 33% single, 3.2% previously married; 87.4% had high-school education or above). One hundred and five respondents were re-interviewed with the SCID (44 males and 61 females). Comparison between the clinically interviewed group and the telephone survey group showed that they did not differ significantly in terms of gender ( $\chi^2 = 0.31, p = 0.58$ ), age group ( $\chi^2 = 1.78, p = 0.78$ ) and work status ( $\chi^2 = 7.4, p = 0.12$ ).

### Concordance between MDQ classification and SCID diagnosis with changing cut-offs

The 13-item MDQ exhibited satisfactory internal consistency (Cronbach's alpha = 0.86). Among the 105 re-interviewed respondents, 24 (22.9%) received diagnoses of BSD (BP-I = 2, BP-II = 10, BP-NOS = 12). Their endorsement of MDQ items ranged from 24.8% to 75.2%. The three most frequently endorsed items were 'easily distracted' (75.2%), 'so irritable' (67.6%) and 'racing thought' (62.9%). Fig. 1 shows the operating characteristics of the MDQ with changing cut-offs. The sensitivity decreased and specificity increased when the cut-off for classifying BSD was increased. In accordance with the usual practice of setting the cut-off at seven items (Hirschfeld *et al.*, 2000), the sensitivity (0.64), specificity (0.68) and AUC (0.66) were moderate. Sensitivity increased from 0.64 to 0.92 when the cut-off decreased from seven items to three



**Fig. 1.** Operating characteristics of the Chinese MDQ for various threshold scores. Variations of cut-off in terms of endorsed number of items. (A colour version of this figure is available online at <http://journals.cambridge.org/eps>)

items, while specificity only dropped slightly (0.68–0.6). Thus, a lower cut-off could improve sensitivity substantially with only slight impact on specificity.

### Other accuracy indicators upon different cut-offs

Regarding the DLRs, being classified as BSD by the MDQ with a cut-off of seven could only increase the odds of BSD diagnosed by the SCID by 1.97 times, while being classified as non-BSD by the MDQ with the same cut-off could only decrease the odds of non-BSD by the SCID by 0.53 times (Table 1). The PPV showed that only 38% of the positive cases found by the MDQ were diagnosed as BSD by the SCID, while the NPV showed that 86% of the negative cases found by the MDQ were also not diagnosed by the SCID. These indicators showed that the PPV did not change much, but NPV decreased as the cut-off increased. The DLR+ was the highest and the DLR– was the lowest when the cut-off was set at three. Using a three-item cut-off could maximize the AUC, DLR+ and sensitivity substantially (0.76, 2.3 and 0.92, respectively), and minimize the DLR– (0.13), although specificity was compromised (0.6). At this cut-off for the community sample, being classified as BSD by the MDQ could increase the odds of BSD diagnosed by the SCID by 2.3 times, while being classified as non-BSD by the MDQ with the same cut-off could decrease the odds of non-BSD by the SCID by 0.13 times. The receiver operating characteristic (ROC) curve also showed a satisfactory AUC (Fig. 2).

## Discussion

Using several methodological enhancement measures, the present study showed that when the cut-off of our Cantonese Chinese version of the MDQ was lowered, it performed moderately for the screening of BSD in a community sample. At the original cut-off of seven items, its sensitivity was higher than the English version for detecting SCID diagnosis of bipolar disorder in the community (28.1%) (Hirschfeld *et al.*, 2003), but its specificity was lower. It also demonstrated a higher sensitivity than in a previous community study that, with regard to DSM-IV bipolar disorder, found the MDQ to have zero sensitivity (0) and high specificity (0.95) (Chung *et al.*, 2009). The authors of that study suggested deleting the impairment criterion to achieve better sensitivity (0.5) without compromising specificity (0.92) excessively. Our use of the SDS could partly solve the low sensitivity problem of the MDQ. This is also supported by the finding (analysis not shown but available on request) that if we excluded bipolar disorder NOS and only focused on

**Table 1.** Performance of the Chinese MDQ with respect to SCID diagnostic interview (n = 105)

Cut-off*	TP	TN	FP	FN	Sensitivity	Specificity	PPV	NPV	DLR+	DLR-	AUC (95% CI)
1	23	44	36	2	0.92	0.55	0.39	0.96	2.04	0.15	0.74 (0.63–0.84)
2	23	46	34	2	0.92	0.58	0.40	0.96	2.16	0.14	0.75 (0.65–0.85)
3	23	48	32	2	0.92	0.6	0.42	0.96	2.3	0.13	0.76 (0.66–0.86)
4	22	48	32	3	0.88	0.6	0.41	0.94	2.2	0.2	0.74 (0.64–0.84)
5	20	49	31	5	0.80	0.61	0.39	0.91	2.06	0.33	0.71 (0.59–0.82)
6	18	50	30	7	0.72	0.63	0.38	0.88	1.92	0.45	0.67 (0.55–0.78)
7	16	54	26	9	0.64	0.68	0.38	0.86	1.97	0.53	0.66 (0.53–0.78)
8	13	60	20	12	0.52	0.75	0.39	0.83	2.08	0.64	0.64 (0.51–0.76)
9	11	61	19	14	0.44	0.76	0.37	0.81	1.85	0.73	0.60 (0.47–0.73)
10	8	66	14	17	0.32	0.83	0.36	0.80	1.83	0.82	0.58 (0.44–0.71)
11	6	70	10	19	0.24	0.88	0.38	0.79	1.92	0.87	0.56 (0.42–0.69)
12	2	74	6	23	0.08	0.93	0.25	0.76	1.07	0.99	0.50 (0.37–0.63)
13	2	77	3	23	0.08	0.96	0.4	0.77	2.13	0.96	0.52 (0.39–0.65)

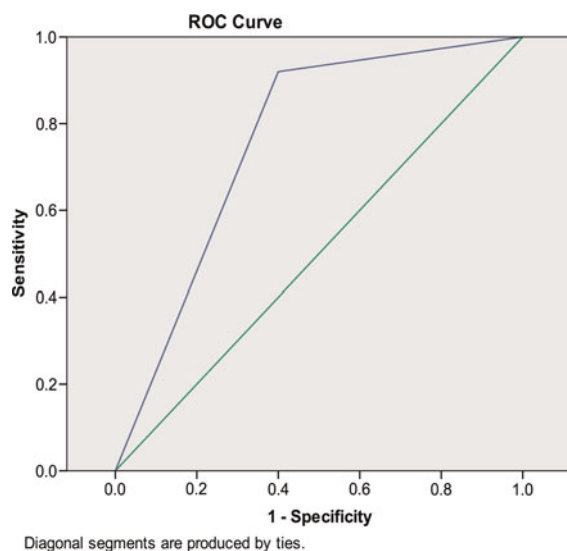
Cut-off\*: The interviewee was classified as having BSD when (i) endorsing the number of item or more, (ii) scoring 4 or more (moderate or more severe) about mania/hypomania related impairment in any area of living in the SDS and (iii) endorsing the item asking about whether any two of the symptoms in MDQ occurred at the same period of time

TP: true-positive; TN: true-negative; FP: false-positive; FN: false-negative.

PPV: positive predictive value; NPV: negative predictive value; DLR+: positive diagnostic likelihood ratio; DLR-: negative diagnostic likelihood ratio; AUC: area under curve; CI: confidence interval.

BP-I and BP-II disorders, the sensitivity of the MDQ using a seven-item cut-off and the SDS significantly improved to 0.69 (specificity 0.64).

Our findings suggested that empirically supported adaptations could enhance the performance of the MDQ in a community. To detect BSD in the general population of Hong Kong, using an enhanced multi-domain impairment measure (SDS) and lowering the item threshold to three could greatly increase the sensitivity of the MDQ without significant compromise of



**Fig. 2.** ROC curve when cut-off is set at 3 items or more (AUC = 0.76, s.e. = 0.05). (A colour version of this figure is available online at <http://journals.cambridge.org/eps>)

the specificity. Our findings were strengthened by the use of DLR that was not confounded by the low base rate of bipolar disorder. This indicated that a lower cut-off could increase the odds of predicting BSD from a positive MDQ result and decrease the odds of BSD from a negative MDQ result. Given that considerable skills and time are needed for the administration of clinical diagnostic interviews like the SCID, the brevity of the MDQ and its moderate accuracy as a screen-out tool make it a potentially valuable tool in the epidemiological study of BSD. The still low sensitivity of the MDQ we found might partly be related to respondents' tendency to minimize the report of impairment since many items of the MDQ tap apparently 'normal' symptoms that may even be positively experienced by those who endorse them. Accordingly, the original Likert-style single item of impairment could be more likely to create false negatives. By using the SDS that consists of a larger number of similar Likert-scale impairment items in four domains of life, sensitivity was improved. The high NPV and low PPV also showed that the MDQ could be a more useful tool in screening out than screening in BSD. Regardless of what cut-off was set, less than 40% of the positive cases found by the MDQ could have BSD, but more than 80% of the negative cases would not be diagnosed as having BSD. The DLR ratio also indicated that at the cut-off of three, a positive finding of MDQ could double the odds of BSD, while a negative finding could reduce the odds by nearly one-seventh.



One methodological issue of note is that owing to the practical difficulty of obtaining the contacts of all the respondents who took part in the telephone survey and the low base rate of bipolar disorder, our selection of respondents for SCID interviews was not random. It remains possible that the respondents who volunteered for face-to-face assessment could be a biased group. How this might have affected our findings remains to be clarified. Besides, although our SCID interviewers were clinicians, we did not assess the inter-rater reliability of the SCID enhanced for the diagnosis of BSD. Although the SDS covers a multi-domain spectrum of impairment that may occur in BSD with differing severity of manic or hypomanic symptoms, it was not externally validated in this study.

#### Declaration of interest

None.

#### Conflict of interest

None.

S. Lee<sup>1,2,3,\*</sup>, A. Tsang<sup>3</sup>, Y. L. Ma<sup>1</sup> and K. L. Ng<sup>1</sup>

<sup>1</sup>Department of Psychiatry, The Chinese University of Hong Kong, Hong Kong, People's Republic of China

<sup>2</sup>Department of Global Health and Social Medicine, Harvard Medical School, Boston, MA, USA

<sup>3</sup>Hong Kong Mood Disorders Center, The Chinese University of Hong Kong, Hong Kong, People's Republic of China

#### References

- Akiskal HS, Bourgeois ML, Angst J, Post R, Moller H, Hirschfeld R (2000). Re-evaluating the prevalence of and diagnostic composition within the broad clinical spectrum of bipolar disorders. *Journal of Affective Disorders* **59**, S5–S30.
- Benazzi F (2007). Bipolar II disorder: epidemiology, diagnosis and management. *CNS Drugs* **21**, 727–740.
- Benazzi F, Akiskal HS (2003). Refining the evaluation of bipolar II: beyond the strict SCID-CV guidelines for hypomania. *Journal of Affective Disorders* **73**, 33–38.
- Chung KF, Tso KC, Cheung E, Wong M (2008). Validation of the Chinese version of the Mood Disorder Questionnaire in a psychiatric population in Hong Kong. *Psychiatry and Clinical Neurosciences* **62**, 464–471.
- Chung KF, Tso KC, Chung TY (2009). Validation of the Mood Disorder Questionnaire in the general population in Hong Kong. *Comprehensive Psychiatry* **50**, 471–476.
- First MB, Spitzer RL, Gibbon M, Williams JBW (2002). *Structured clinical interview for DSM-IV axis I disorders, research version, non-patient edition (SCID-I/NP)*. Biometrics Research, New York State Psychiatric Institute: New York.
- Hirschfeld RM (2002). The Mood Disorder Questionnaire: a simple, patient-rated screening instrument for bipolar disorder. *Primary Care Companion to the Journal of Clinical Psychiatry* **4**, 9–11.
- Hirschfeld RM, Williams JBW, Spitzer RL, Calabrese JR, Flynn L, Keck PE Jr, Lewis L, McElroy SL, Post RM, Rapport DJ, Russell JM, Sachs GS, Zajecka J (2000). Development and validation of a screening instrument for bipolar spectrum disorder: the Mood Disorder Questionnaire. *American Journal of Psychiatry* **157**, 1873–1875.
- 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 (2003). Validity of the mood disorder questionnaire: a general population study. *American Journal of Psychiatry* **160**, 178–180.
- Kim B, Wang HR, Son JI, Kim CY, Joo YH (2008). Bipolarity in depressive patients without histories of diagnosis of bipolar disorder and the use of the mood disorder questionnaire for detecting bipolarity. *Comprehensive Psychiatry* **49**, 469–475.
- Lee S, Ng KL, Tsang A (2009). A community survey of the twelve-month prevalence and correlates of bipolar spectrum disorder in Hong Kong. *Journal of Affective Disorders* **119**, 79–86.
- Lee S, Tsang A, Huang YQ, He YL, Liu ZR, Zhang MY, Shen YC, Kessler RC (2008). The epidemiology of depression in metropolitan China. *Psychological Medicine* **39**, 1–13.
- Leon AC, Olfson M, Portera L, Farber L, Sheehan DV (1997). Assessing psychiatric impairment in primary care with the Sheehan Disability Scale. *International Journal of Psychiatry in Medicine* **27**, 93–105.
- Merikangas KR, Jin R, He JP, Kessler RC, Lee S, Sampson NA, Viana MC, Andrade LH, Hu CY, Karam EG, Ladea M, Medina-Mora ME, Ono Y, Posada-Villa J, Sagar R, Wells JE, Zarkov Z (2011). Prevalence and correlates of bipolar spectrum disorder in the World Mental Health Survey Initiative. *Archives of General Psychiatry* **68**, 241–251.
- Pepe MS (2004). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press: Oxford.
- So E, Kam I, Leung CM, Chung D, Liu Z, Fong S (2003). The Chinese-bilingual SCID-I/P project: stage 1 – reliability for mood disorders and schizophrenia. *Hong Kong Journal of Psychiatry* **13**, 7–18.
- Weber Rouget B, Gervasoni N, Dubuis V, Gex-Fabry M, Bondolfi G, Aubry JM (2005). Screening for bipolar disorders using a French version of the Mood Disorder Questionnaire (MDQ). *Journal of Affective Disorders* **88**, 103–108.
- Young AH, MacPherson H (2011). Detection of bipolar disorder. *British Journal of Psychiatry* **199**, 3–4.