Validation of the Chinese version of the "Mood Disorder Questionnaire" for the screening bipolar disorder among patients with current depressive episode in Chinese mainland

Zhaoyu Gan¹, Zili Han¹, Kanglai Li², Feici Diao¹, Nianhong Guan¹, Xiaoli Wu¹, Jinbei Zhang¹*

¹Psychiatry Department, the Third Affiliated Hospital, Sun Yat-sen University, Guangzhou 510630, China

²VIP Department, the Third Affiliated Hospital, Sun Yat-sen University, Guangzhou 510630, China.

Zhaoyu Gan and Zili Han contributed equally to this work.

*Corresponding author

E-mail address:

GZY: moyanchou@gmail.com

HZL: 13312883609@189.cn

LKL: likanglai1979@163.com

DFC: feici@126.com

GNH: kuwanhong@yahoo.com.cn

WXL: wxhzp2006@126.com

ZJB: jinbeizhang@gmail.com
Abstract

Background
Mood Disorder Questionnaire (MDQ) is a well-recognized screening tool for bipolar disorder. But its Chinese version needs further validation, especially among populations with bipolar disorder type II treated for major depressive episode. This study aims to measure the accuracy of the Chinese version of the MDQ as a screening instrument for bipolar disorder (BPD) in a group of patients with current major depressive episode.

Methods
One hundred and forty two consecutive patients with an initial DSM-IV diagnosis of present depressive episode were screened for BPD using the Chinese translation of the MDQ over 1-year follow-up. The 1-year final diagnosis, determined by a special committee consisting of three trained senior psychiatrists, was used as a ‘gold standard’ and ROC was plotted to evaluate the performance of the MDQ. The optimal cut-off was chosen by maximizing the Youden’s index.

Results
Of the 142 patients, 122 (85.9%) finished 1-year follow-up. On the basis of Semi-structured Clinical Interview: 48.4% (59/122) received a diagnosis of unipolar depression (UPD), 36.9% (45/122) BPD and 14.8% (18/122) BPD. At the end of 1-year follow-up, 9 moved from UPD to BPD, 2 from BPD to UPD, 1 BPD to BPD, the overall rate of initial misdiagnosis was 16.4%. MDQ showed a good accuracy for BPD: the optimal cut-off was 4, with a sensitivity of 0.72 and a specificity of 0.73. When BPD and BPD were calculated independently, the optimal cut-off for BPD was 4, with a sensitivity of 0.70 and a specificity of 0.73; while the optimal cut-off for BPD was 5, with a sensitivity of 0.67 and a specificity of 0.86. The section 2 and section 3 showed no significant improvement on the performance of MDQ.

Conclusions
Our results show that the Chinese version of MDQ seems to be a valid tool for
screening for BPD in a group of patients with current depressive episode in Chinese mainland.
**Background**

Substantial studies have reported that patients with BPD are frequently misdiagnosed with other disorders. The frequency of initial misdiagnosis was reported to reach as high as 69%, with more than one third patients with BPD experiencing a delayed correct diagnosis as long as 10 years or even more[1]. At the same time, over-diagnosis of BPD is also reported to be highly frequent. A study by Lopez showed that frequency of over-diagnosis was 47.2% in 1153 subjects diagnosed as BPD[2]. The incorrectness and delay in diagnosis often lead to inappropriate treatment, which in turn results in poor outcome[3]. Over these years, a number of strategies have been proposed to facilitate the detection of BPD in real clinical practice. Using the Mood Disorder Questionnaire (MDQ) [4] is one of the common strategies.

As a tool created to aid diagnosis of bipolar spectrum disorder, MDQ has been translated into many languages and proved to be a helpful tool in screening BPD[5-8]. A Chinese version of MDQ had been proven a valid screening tool for BPD in a psychiatric population but not in the general population in Hong Kong[8-9]. However, psychometric properties of MDQ were found to differ slightly under different language settings and among different sample population [5-8]. Especially its poor performance in identifying mild bipolar spectrum, such as BPD II [10], greatly reduces its value in detecting BPD, since BPD II accounts for the majority of misdiagnosis among patients with BPD in clinical reality[1, 11].The section 2 and section 3 of MDQ might partly contribute to this [10, 12-13]. Another potential reason might the gold standard used as a reference. Over the past decades, most studies applied diagnosis based on a single structural clinical interview as the gold standard for evaluation. However, a number of studies have shown that a single structural clinical interview based on DSM-IV criteria is far from enough to achieve an accurate diagnosis of BPD in real clinical practice, especially for those who lack insight or those with BPD II[14]. For instance, according to the variation of observation length, approximate 12.5%-30% patients with an initial diagnosis of UPD eventually receive a diagnosis of BPD [15-17].
In the present study, we hypothesize that, among patients with current depressive episode, it might be reasonable and valuable to use MDQ as a screening tool for BPD, since most of those patients with bipolar spectrum disorder, especially BPD II pay visits to doctors when they are depressive, which makes them more likely to be misdiagnosed as UPD [16]. At the same time, one year follow-up will be administrated to justify the initial diagnosis based on SCID interview in an attempt to improve the accuracy of the gold standard. Finally, the performance of MDQ without the section 2 and section 3 was also accessed.

Methods

Subjects
This study sample consisted of 142 eligible subjects that were treated currently for major depressive episode (MDE) based on the criteria of DSM-IV-R in the psychiatric department, the 3rd Affiliated Hospital of Sun Yat-sen University between July 2006 and July 2007. Written informed consent was obtained from all participants, and all procedures used in the present study were reviewed and approved by the local institutional review board. Patients with a psychiatric or physical disorder that prevented them from being interviewed or undermined their ability to provide accurate information, and those who declined participation in the study or refused to provide informed consent were excluded.

Instruments
The Chinese version of SCID: Chinese version of the Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR) Axis 1 Disorders (SCID-I) was used for diagnostic interview.

MDQ: The translation of MDQ into Chinese was approved by one of the developers of the original version. The Chinese version was translated back into English and re-edited to make it comparable to the original version.

Procedure
Prior to the start of this study, 3 senior psychiatrists (GNH, GNH and WXL) attended a training program focused on SCID, self-compiled questionnaire and the detection of
switch. At the end of the program, their kappa coefficient reached 0.93 in terms of interrater reliability. Throughout the whole study period, all diagnostic interview and assessments were performed by these three psychiatrists, who already constituted a special committee responsible for these tasks and keep blind to the result of MDQ. Potential participants for this study were found by a study nurse (LKL) through reviewing the archive records and clinical outpatient files. The cases were included if they had been or would like to be followed up by the psychiatrists of our department. All the participants submitted written inform consent. At the study entry, participants were invited (by LKL) to fill in the Chinese version of MDQ. Then SCID-I was performed for each participant to establish an initial diagnosis meeting with the criteria of DSM-IV-TR. Demographic and clinical characteristics and features of the current depressive episode were collected using the self-compiled questionnaire. The participants were then followed up for 1 year, being interviewed by one of the three senior psychiatrists for at least six times with a flexible interval of 1–2 months via telephone or face to face talking. In each interview, if suspected switch was detected, the patients' relatives or friends who knew well about them were asked to provide additional information, and then all the data about this patient was submitted to the committee, who would decide whether the patient had experienced a switch according to the criteria of DSM-IV-TR. To insure the quality and objectivity of switch detection, those who did not complete 1 year follow-up or did not regularly follow up for more than 6 times within the year were excluded. At the end of study, the committee reviewed the 1-year medical records and came up with a final diagnosis. During the study period, all treatment decisions or changes in treatment medications such as dose reduction, dose augmentation, or switch strategies were made by their treating psychiatrists. This study was carried out under naturalistic clinical settings and no treatment information was obtained.

Data analysis

All statistical analysis was performed using commercial statistical package SPSS 13.0(SPSS Inc., Chicago). The Mann-Whitney U test was used to compare numerical variables and the chi-square test was used to compare categorical variables. Cronbach
alpha was used to access the internal consistency of the scale. The receiver operating characteristic (ROC) curve was plotted to assess the screening performance of the questionnaire. Its accuracy was calculated in terms of sensibility and specificity for each theoretically possible cut-off, and then the method of linear interpolation was used to calculate the sensibility and specificity for each actually possible cut-off (number of positive answers). The cut-off point was determined by maximizing the Youden's index.

**Results**

**Comparison of the dropout group and the rest**

At the study entry, 102 subjects (71.8%) were inpatients and 40 (28.2%) were outpatients. Of the 142 subjects, 122 (85.9%) completed 1 year follow-up receiving 6-12(7±2 ) visits. The dropout reasons included transferring to other psychiatric institution (9 subjects, 6.33%) and refusal to continue the study (11 subjects, 7.74%). No difference was found between the dropout group and the rest with regard to demographic and clinical features. Given there is no eminent features in dropout group, this group is not involved in the subsequent statistical analysis.

**Comparison of initial diagnosis and final diagnosis**

As the table2 showed, both underdiagnosis and overdiagnosis of BPD existed in this study, but the underdiagnosis seemed to be more prominent than overdiagnosis.

**The internal consistency of the Chinese version of the MDQ**

In this sample, the Cronbach coefficient for the 13-item symptom scale was 0.735, the item-total scale correlations ranged from 0.195 (less sleep) to 0.597 (more active). The elimination of each item did not impose great impact on the scale’s internal consistency.

**Scores of MDQ in each section**

The table 3 showed the summary scores of all subjects in each section of MDQ. The high proportion (34.7%) of subjects with UPD scoring moderate or severe in section 3 and the high prevalence of missing values in the section2 and section3 seemed to implicate that there might be some misunderstanding in these sections. While the percentage of scoring moderate or severe among subjects with BP was less than
expected, if combined with the number of symptoms, this percentage would be lower. Therefore, the section 2 and section 3 makes no sense in distinguishing BPD from UPD.

**ROC analysis of section 1**

Based on the abovementioned reasons, the section 2 and section 3 were directly ignored in the following statistic analysis. ROC was plotted according to the scores in the section 1. The corresponding sensibility and specificity for each actual possible cut-off (number of positive answers) were calculated by linear interpolation based on the sensibility and specificity of the corresponding theoretical cut-off point in ROC. The results and the corresponding area under curve (AUC) and p value were listed in table 5. By maximizing the Youden’s index, 4 was selected as the optimal cut-off point for patients with BP or BP II, with a sensibility of 0.72 or 0.70 respectively and a specificity of 0.73. If BP I was separately calculated, 5 was considered the optimal cut-off point, with a sensibility of 0.67 and a specificity of 0.86.

**Discussion**

As this study and our previous report [18] have showed, the diagnosis of UPD and BPD based on a single interview was unstable over time, with a rate of diagnosis transition ranging from 16.4% to 19.4%, which was consistent with other reports [15, 19-21]. Therefore, using diagnosis confirmed by one year follow-up visits as ‘gold standard’, will help make the conclusion more persuasive.

Compared to the optimal cut-off of 7 reported by studies from western countries [4, 6] and Hongkong [8], this study showed a smaller optimal cut-off, which was similar to findings from Chinese mainland [22]. This might partly due to the culture difference, since Hongkong is a very westernized city in China, which makes its culture and language greatly different from Chinese mainland.

In line with previous studies [4, 6-7], MDQ is more sensitive in detecting BP than detecting BPD in this study, given the same cut-off point. Although the originator of the MDQ did not specially access patients in remission from a mood episode, whether the patient’s symptomatology at the time of screening will affect the
MDQ performance is interesting topic. A previous study with subsample of small size [23] showed the performance of MDQ seemed to be independent of depressive symptoms, but the relatively low test-retest reliability (kappa coefficient 0.64) with the whole sample implicates the possible influence of clinically relevant factors, such as patient’s mood states at time of completion. While compared to report which sampled patients treated for depression[14], the performance of MDQ in detecting BPD in this study was quite close (sensibility: 0.706 vs. 0.70), in spite of the different cut-off point (7 vs. 4).

According to the initial conception of MDQ’s developers, a subject who will be screened positive has to meet the DSM-IV-TR criteria of manic or hypomanic episode, including symptom criteria, severity criteria. However, the poor performance of the section 2 and section 3 in this study and other reports [5-6, 14] proves this kind of conception doesn’t work, especially in screening patients with BPD. In this study, we went further by adding a question to ask subjects how long the positive symptoms last. It turned out to be: 16(32%) subjects with BPD did not meet the DSM-IV-TR duration criteria of hypomanic episode (lasting at least 4 days). That means it is unrealistic to expect a self-rated questionnaire to help improve recognition of a past criteria-met hypomanic episode among patients with BPD. In addition, this kind of questionnaire makes no sense to those who experience their first visit as a depressive episode.

However, MDQ without section 2 and section 3 has been proved to be a valid screening tool for BPD, and even for previously unrecognized bipolar disorder[24]. Possible explanation for this might be: MDQ without section 2 and section 3 helps recognize the opposite polarity-manic or hypomanic symptoms of BPD, which help improve the recognition of BPD [25-26]. Recently, convergent evidences have shown that bipolarity is a sensitive and characteristic feature of BPD [25, 27-28]. For instance, a cross-sectional study [29] found that clinically significant depressive symptoms occurred in 94.1% of those with (hypo) mania, while 70.1% in a depressive episode had clinically significant manic Symptoms. In this study, manic symptoms are also found to be more likely to occur in patients with BPD than those with UPD.
Under this context, it is not difficult to understand why MDQ without section 2 and section 3 can be used as a screening tool to detect bipolar diathesis in depression [27, 30].

Generally speaking, this study shows that the Chinese version of MDQ without the section 2 and section 3 is a valid, brief and feasible tool for screening BPD from patients with current depressive episode in Chinese mainland, although the psychometric properties in terms of internal consistency is not as excellent as reports in western countries[5-6], which means some modification should be done.

**Competing interests**
The authors declare that they have no competing interests.

**Authors' contributions**
ZJB designed and organized the study, trained counselors conducting structured interviewed and supervised the quality of research. GZY designed the study, translated the MDQ and drafted the manuscript. HZL, as the chief of the special committee, conducted structured interview, collected other clinical information, follow up each subject and make final diagnostic decision. LKL recruited subjects, surveyed with MDQ, arranged follow-up visits and entered data. DFC translated the MDQ and conducted statistical analyses. GNH and WXL, both as member of the special committee, conducted structured interview, follow up each subject and make final diagnostic decision. All authors read and approved the final manuscript.

**Acknowledgements**
Funding for this study was provided by Natural Science Foundation of Guangdong Province, China, (101510089010000214). The Natural Science Foundation of Guangdong Province had no further role in study design, in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication. The authors would like to gratefully acknowledge the contributions of all of the nurses, technicians, and patients that participated in this study.
Reference


<table>
<thead>
<tr>
<th>Features</th>
<th>BP&lt;sup&gt;a&lt;/sup&gt;</th>
<th>DEP</th>
<th>P&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=20</td>
<td>N=52</td>
<td>N=70</td>
</tr>
<tr>
<td>Age(χ±s)</td>
<td>28.0±8.7</td>
<td>29.2±8.6</td>
<td>28.8±8.6</td>
</tr>
<tr>
<td>Female (%)</td>
<td>5 (25.0)</td>
<td>27 (54.0)</td>
<td>32 (45.7)</td>
</tr>
<tr>
<td>Atypical features&lt;sup&gt;c&lt;/sup&gt; (%)</td>
<td>6 (30.0)</td>
<td>8 (16.0)</td>
<td>14 (20.0)</td>
</tr>
<tr>
<td>Manic symptoms&lt;sup&gt;d&lt;/sup&gt;(%)</td>
<td>6 (30.0)</td>
<td>16 (32.0)</td>
<td>22 (31.4)</td>
</tr>
<tr>
<td>Comorbidity of anxiety disorder&lt;sup&gt;e&lt;/sup&gt;(%)</td>
<td>7 (35.0)</td>
<td>14 (28.0)</td>
<td>21 (30.0)</td>
</tr>
<tr>
<td>Comorbidity of psychoactive drug abuse (%)</td>
<td>1 (5.0)</td>
<td>1 (2.0)</td>
<td>2 (2.9)</td>
</tr>
<tr>
<td>Recurrent depression (%)</td>
<td>15 (75.0)</td>
<td>27 (54.0)</td>
<td>42 (60.0)</td>
</tr>
</tbody>
</table>

**Note:**

a. The categorization was based on the final diagnosis.

b. all compared with UPD.

c. Atypical features including mood reactivity, overeating or weight gain, oversleeping, leaden paralysis and interpersonal rejection sensitivity()

d. Manic symptoms were found by psychiatric interview at the entry of this study.

e. Anxiety disorder here consisted of generalized anxiety disorder, panic disorder, obsessive-compulsive disorder, phobia, somatization disorder

<table>
<thead>
<tr>
<th>Initial diagnosis</th>
<th>Final diagnosis</th>
<th>N of cases</th>
<th>Percentage(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UPD(59)</td>
<td>UPD</td>
<td>50</td>
<td>84.7</td>
</tr>
<tr>
<td>BP&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
<td>8</td>
<td>13.6</td>
</tr>
<tr>
<td>BP&lt;sup&gt;−&lt;/sup&gt;</td>
<td></td>
<td>1</td>
<td>1.6</td>
</tr>
<tr>
<td>BP&lt;sup&gt;+&lt;/sup&gt;(45)</td>
<td>UPD</td>
<td>2</td>
<td>4.4</td>
</tr>
<tr>
<td>BP&lt;sup&gt;−&lt;/sup&gt;</td>
<td></td>
<td>42</td>
<td>93.3</td>
</tr>
<tr>
<td>BP&lt;sup&gt;−&lt;/sup&gt;</td>
<td></td>
<td>1</td>
<td>2.2</td>
</tr>
<tr>
<td>BP&lt;sup&gt;+&lt;/sup&gt;(18)</td>
<td>UPD</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BP&lt;sup&gt;−&lt;/sup&gt;</td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BP&lt;sup&gt;−&lt;/sup&gt;</td>
<td></td>
<td>18</td>
<td>100%</td>
</tr>
<tr>
<td>Total(122)</td>
<td>Agreed with initial diagnosis</td>
<td>110</td>
<td>83.6%</td>
</tr>
</tbody>
</table>
Table 3. Scores of the subjects in each section of MDQ.

<table>
<thead>
<tr>
<th>MDQ</th>
<th>BP I</th>
<th>BP II</th>
<th>UDP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(N=20)</td>
<td>(N=50)</td>
<td>(N=52)</td>
</tr>
<tr>
<td>Section 1</td>
<td>6.85±3.33</td>
<td>5.76±2.73</td>
<td>3.02±2.00*</td>
</tr>
<tr>
<td>Section 2</td>
<td>Yes</td>
<td>12(60.0%)</td>
<td>26(50%)</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>6(30.0%)</td>
<td>16(34.6%)</td>
</tr>
<tr>
<td></td>
<td>Missing</td>
<td>2(10.0%)</td>
<td>8(15.4%)</td>
</tr>
<tr>
<td>Section 3</td>
<td>unafflicted</td>
<td>4 (21.1%)</td>
<td>6 (12.0%)</td>
</tr>
<tr>
<td></td>
<td>mild</td>
<td>7 (36.8%)</td>
<td>12 (24.0%)</td>
</tr>
<tr>
<td></td>
<td>moderate</td>
<td>3 (15.8%)</td>
<td>16 (32.0%)</td>
</tr>
<tr>
<td></td>
<td>severe</td>
<td>3 (15.8%)</td>
<td>13 (26.0%)</td>
</tr>
<tr>
<td></td>
<td>missing</td>
<td>2 (10.5%)</td>
<td>3 (6.0%)</td>
</tr>
</tbody>
</table>

Note: a. The final diagnosis after one-year follow up.

b. compared to subjects with BPD, subjects with UPD scored significantly lower (P<0.01) in section 1 of MDQ.

Table 4. ROC analysis of section 1 for BP, BP II and BP III

<table>
<thead>
<tr>
<th></th>
<th>AUC</th>
<th>P</th>
<th>Cut-off</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP</td>
<td>0.803</td>
<td>&lt;0.001</td>
<td>sen</td>
<td>0.98</td>
<td>0.93</td>
<td>0.85</td>
<td>0.72</td>
<td>0.59</td>
<td>0.42</td>
<td>0.35</td>
<td>0.26</td>
<td>0.19</td>
<td>0.14</td>
<td>0.06</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>sep</td>
<td>0.15</td>
<td>0.36</td>
<td>0.55</td>
<td>0.73</td>
<td>0.86</td>
<td>0.90</td>
<td>0.94</td>
<td>0.97</td>
<td>0.99</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>BP</td>
<td>0.794</td>
<td>&lt;0.001</td>
<td>sen</td>
<td>0.99</td>
<td>0.94</td>
<td>0.84</td>
<td>0.70</td>
<td>0.55</td>
<td>0.43</td>
<td>0.31</td>
<td>0.20</td>
<td>0.13</td>
<td>0.09</td>
<td>0.05</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>sep</td>
<td>0.15</td>
<td>0.36</td>
<td>0.55</td>
<td>0.73</td>
<td>0.86</td>
<td>0.90</td>
<td>0.94</td>
<td>0.97</td>
<td>0.99</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>BP</td>
<td>0.826</td>
<td>&lt;0.001</td>
<td>sen</td>
<td>0.95</td>
<td>0.90</td>
<td>0.88</td>
<td>0.78</td>
<td>0.67</td>
<td>0.58</td>
<td>0.50</td>
<td>0.43</td>
<td>0.33</td>
<td>0.25</td>
<td>0.13</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>sep</td>
<td>0.15</td>
<td>0.36</td>
<td>0.55</td>
<td>0.73</td>
<td>0.86</td>
<td>0.90</td>
<td>0.94</td>
<td>0.97</td>
<td>0.99</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Note: a. The sensibility and specificity were calculated by linear interpolation based on the sensibility and specificity of the corresponding theoretical cut-off point in ROC.

b. the sensibility and specificity in boldface maximized the Youden’s index and the corresponding cut-off point was considered the optimal one.