

Title: The use of Mood Disorder Questionnaire, Hypomania Checklist-32 and clinical predictors for screening previously unrecognized bipolar disorder in a general psychiatric setting

Running head: Screening of bipolar disorder

Authors: Yvette Poon^a, Ka-Fai Chung^{b,*}, Kwok-Chu Tso^a, Chi-Lok Chang^a, Dorothy Tang^a

Name of department: ^a Department of Psychiatry, Queen Mary Hospital, Hong Kong SAR, China; ^b Department of Psychiatry, University of Hong Kong, Hong Kong SAR, China

***Correspondence:** Dr. KF Chung, Clinical Associate Professor, Department of Psychiatry, Queen Mary Hospital, 102 Pokfulam Road, Hong Kong. Tel: +852 2255 4488, Fax: +852 2855 1345. E-mail: kfchung@hkucc.hku.hk.

Word count: 4340 (excluding cover page, abstract, references, tables and figures);

2 tables; 2 figures.

Abstract

Bipolar disorder is often unrecognized and misdiagnosed in the general psychiatric setting. This study compared the psychometric properties of Mood Disorder Questionnaire (MDQ) and Hypomania Checklist-32 (HCL-32), examined the clinical predictors of bipolar disorder, and determined the best approach for screening previously unrecognized bipolar disorder in a general psychiatric clinic. A random sample of 340 non-psychotic outpatients with no previous diagnosis of bipolar disorder completed the MDQ and HCL-32 during their scheduled clinic visits. Mood and alcohol/substance use disorders were re-assessed using a telephone-based Structured Clinical Interview for DSM-IV. We found that the HCL-32 had better psychometric performance and discriminatory capacity than the MDQ. The HCL-32's internal consistency and 4-week test-retest reliability were higher. The area under curve was also greater than those of MDQ at various clustering and impairment criteria. The optimal cut-off of MDQ was co-occurrence of 4 symptoms with omission of impairment criterion; for HCL-32, it was 11 affirmative responses. Multivariable logistic regression found that bipolar family history was associated with an increased risk of bipolar disorder (odds ratio = 4.93). The study showed that simultaneous use of HCL-32 and bipolar family history was the best approach for detecting previously

unrecognized bipolar disorder.

Keywords: bipolar disorder, bipolar spectrum disorder, Chinese, detection,

Hypomania Checklist, Mood Disorder Questionnaire, screening

1. Introduction

In recent years, there has been emerging evidence supporting the dimensional concept of bipolar disorder (Akiskal and Pinto, 1999). Bipolar spectrum disorder represents a continuum of mood changes of different severities ranging between full blown mania and unipolar depression (Katzow et al., 2003). It consists of not only bipolar I and bipolar II, but also cyclothymia and bipolar disorder not otherwise specified (bipolar NOS), which includes a heterogeneous group of clinically significant bipolar conditions not meeting the DSM-IV criteria (American Psychiatric Association, 2000). Bipolar disorder, especially the milder forms (bipolar II and bipolar NOS), is often unrecognized and misdiagnosed in clinical practice (Ghaemi et al., 2002). Identification of past history of hypomania can be difficult, as a majority of patients seek treatment during their depressive rather than hypomanic episodes (Hirschfeld, 2001). Up to 69% of bipolar patients were initially misdiagnosed and most frequently as unipolar depression, followed by anxiety disorders, personality disorders and substance or alcohol use disorders, due to overlapping symptomatology (Hirschfeld et al., 2003). Correct diagnosis and treatment can be delayed by 8 to 10 years (Lish et al., 1994). Under-recognition of bipolar disorder results in substantial negative impact on individual patients and the whole society. It is associated with higher suicide rate,

poorer quality of life, greater functional impairment and increased healthcare cost (Shi et al., 2004; Matza et al., 2005; Awad et al., 2007). Inappropriate antidepressant monotherapy is less effective in treating bipolar depression, and it also increases the risk of manic switch and cycle acceleration (Dunner, 2003). The clinical significance of subthreshold bipolar conditions is increasingly recognized on the basis of higher illness severity, suicidality, disability and healthcare utilization comparable to bipolar I and II disorders (Judd and Akiskal, 2003; Merikangas et al., 2007). Hence, early detection and correct treatment of bipolar disorder is very important.

The use of clinical predictors and screening instruments can improve the recognition of bipolar disorder (Phelps and Ghaemi, 2006). Benazzi (2004) has found that bipolar family history and early age of onset are the two most significant bipolar validators; early onset has the highest sensitivity and bipolar family history has the highest specificity. The most widely used screening instruments for bipolar disorder include Mood Disorder Questionnaire (MDQ) and Hypomania Checklist-32 (HCL-32). The MDQ is a single-page self-report questionnaire consisting of three sections (symptom endorsement of 13 items, symptom clustering, and level of functional impairment). In the original validation study in a psychiatric population, the standard cut-off criterion is a clustering of at least 7 symptoms with at least moderate level of impairment,

where sensitivity is 0.73 and specificity is 0.90 (Hirschfeld et al., 2000). The HCL-32 is another self-administered questionnaire comprising of a checklist of 32 yes/no questions to screen for past hypomanic symptoms (Angst et al., 2005). The standard cut-off score is 14, yielding a sensitivity of 0.80 and specificity of 0.51. Although HCL-32 was originally developed for use in depressed patients, it could be useful in non-clinical and non-specialized psychiatric settings (Meyer et al., 2007).

Previous studies on the screening performance of MDQ and HCL-32 have focused on patients with mood disorders in specialized clinics. Little is known about the performance of these screening tools in detecting previously unrecognized bipolar disorder, of which inappropriate treatment and functional impairment are common. There has not been any study examining the clinical predictors of bipolar disorder among Chinese. In this study, firstly, we examined the psychometric properties of MDQ and HCL-32 in a representative sample of general psychiatric outpatients who had not been previously received a bipolar disorder diagnosis. Secondly, we identified the clinical predictors of bipolar disorder; and lastly, we found out the best method for screening previously unrecognized bipolar disorder by comparing the performance of MDQ, HCL-32, and a combination of MDQ and HCL-32 with clinical predictors.

2. Methods

This study was conducted in a regional psychiatric clinic in Hong Kong. It was reviewed and approved by the local institutional review board.

2.1 Participants

The sample size calculation was based on previous sensitivity and specificity values of the Chinese MDQ (Chung et al., 2008) and the Taiwanese HCL-32 (Wu et al., 2008), and the local prevalence of bipolar disorder (Mak, 2009). Setting the level of significance at 0.05 and the acceptable width of 95% confidence interval for sensitivity and specificity at 8%, calculation using the prior sensitivity of MDQ yielded the largest sample size, where the number of subjects was estimated at 330 (Buderer, 1996). Assuming an overall refusal rate of 25%, 450 subjects would be sufficient to achieve statistically significant results on the accuracy of MDQ and HCL-32.

The inclusion criteria were ethnic Chinese, aged 18 to 64 years, and no previous diagnosis of bipolar disorder, psychotic disorders, mental retardation, dementia, and

organic mental disorders. A complete list of outpatients who had visited the clinic between 1 March 2008 and 30 June 2008 was generated from the computerized patient record system. From 6108 active cases, 3534 patients satisfied the inclusion criteria. A total of 450 subjects were selected through a simple randomization process using computer-generated random numbers. Three-hundred forty subjects gave informed consent and completed the questionnaires while 110 patients did not participate (89 refused and 21 were excluded due to illiteracy). From the computerized record, the original psychiatric diagnoses of the 340 participants were as follows: 49.4% (n = 168) had major depressive disorder; 11.5% (n = 39) had generalized anxiety disorder; 9.4% (n = 32) had mixed anxiety and depressive disorder; 8.8% (n = 30) had adjustment disorder; 7.1% (n = 24) had panic or phobic disorder; 5.9% (n = 20) had dysthymic disorder, 3.2% (n = 11) had obsessive-compulsive disorder; 4.7% (n = 16) had alcohol or substance use disorder; 2.1% (n = 7) had post-traumatic stress or acute stress disorder; 1.8% (n = 6) had primary insomnia; and 1.5% (n = 5) had personality disorder. Only 16 (4.7%) of the 340 participants were given more than 1 psychiatric diagnosis. Major depressive disorder comorbid alcohol or substance use disorder, which was diagnosed in 6 subjects, was the most common form of psychiatric comorbidity. There was no significant difference in age, gender, marital status, and psychiatric diagnosis by

medical record between the 340 participants and the 110 non-participants.

2.2 Measures

The Chinese version of MDQ was used. The process of translation from English into Chinese was reported in details in a previous study (Chung et al., 2008). The Taiwanese version of HCL-32 was obtained with approval from one of the authors of the Taiwanese study (RBL) (Wu et al., 2008). Due to differences of language and terminology use in Cantonese (Hong Kong) and Mandarin (Taiwan), some items in the Taiwanese version were modified. The comprehension of each item was reviewed by an expert panel consisting of four bilingual psychiatrists (YP, KFC, KCT and CLC) and amended accordingly. It was tested in a pilot sample of 16 clinically stable patients. Four items (question 2, items 6 and 20 of question 3, and question 5) required further modifications. The final Chinese (Hong Kong) version of HCL-32 was then re-edited and approved by the expert panel for use in this study.

The subjects' psychiatric diagnoses were re-assessed using a telephone-based Chinese version of the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID) (First et al., 2002; So et al., 2003). Only the modules of mood and substance use

disorders were used. To improve the assessment of past hypomania, we ignored the skip-out instruction of the screening question on mood (Benazzi and Akiskal, 2003). The diagnosis of all bipolar subtypes was based on the DSM-IV-TR criteria. Bipolar NOS refers to brief hypomania meeting the DSM-IV-TR symptom criteria, lasting for 2 to 4 days, and having at least 1 major depressive episode. Our definition of bipolar NOS has been validated and used in previous studies (Akiskal and Benazzi, 2005; Benazzi and Akiskal, 2006; Kim et al., 2008; Lee et al., 2009).

2.3 Procedure

The authors contacted the participants during their scheduled clinic visits. Written informed consent was obtained from all subjects and they were asked to complete a demographic and clinical information sheet and the Chinese version of MDQ and HCL-32. Clinical information included family history of depression and bipolar disorder in first-degree relatives and age of onset of illness, defined as age when patients first had mood symptoms that caused clinically significant distress or functional impairment. Early onset was defined as onset of illness <21 years, which was the most validated and commonly used cut-off (Benazzi and Akiskal, 2008; Mak, 2009).

The subjects were then contacted by telephone within 2 weeks upon questionnaire completion. The author (YP) who was trained and experienced in using SCID and was blind to the patients' MDQ and HCL-32 results and their original psychiatric diagnoses conducted the diagnostic interview and further verified with patients the clinical information reported on the questionnaire.

We assessed the inter-rater reliability of the SCID-derived lifetime diagnosis in 20 consecutive patients. Three psychiatrists (YP, KFC and KCT) referred to audio-recorded interviews and independently rated whether the patients had a lifetime diagnosis of bipolar disorder, major depressive disorder, or other diagnoses. The kappa and Yule's coefficients (Helzer et al., 1985) for all diagnostic categories were 1.00, suggesting excellent agreement among the raters. Previous studies have shown that telephone SCID interview is comparable to face-to-face interview in diagnostic assessments for lifetime psychiatric diagnoses (Cacciola et al., 1999; Crippa et al., 2008), and it can increase participation rate when face-to-face interview is not feasible (Allen et al., 2003). We examined the level of agreement in the SCID-derived lifetime diagnosis between telephone and face-to-face interview in a convenient sample of 20 patients; both the kappa and Yule's coefficients were 1.00, suggesting that the two

methods were highly comparable.

We assessed the test-retest reliability of MDQ and HCL-32 by asking a consecutive sample of 180 patients to complete the scales twice over 4 weeks. The second set of scales was posted to the subjects about 3 weeks after the day of recruitment. Only 109 patients returned the questionnaires, and 2 were excluded as the scales were completed later than 6 weeks after the first test. The test-retest reliability was examined in 107 patients. The mean time between the first and second questionnaire administration was 29.7 days (range = 20-42 days).

2.4 Data analysis

All statistical analysis was done by SPSS version 15.0 for Windows (SPSS, Chicago, USA). Categorical variables were analyzed by Chi-square or Fisher exact test. Due to the lack of normal distribution, continuous variables were analyzed by Mann-Whitney U test. P value <0.05 was considered statistically significant.

The internal consistency of MDQ and HCL-32 was evaluated by Cronbach's alpha. Test-retest reliability was assessed by intraclass correlation coefficient (ICC). The

discriminatory capacity was evaluated in terms of sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and likelihood ratio for positive and negative tests (LR(+) and LR(-)). Receiver-operating-characteristic (ROC) analysis was used to compare the discriminatory capacity of MDQ and HCL-32. The optimal cut-off was the point closest to the left upper corner of the ROC curve and the discriminatory power was measured by the area under curve (AUC) (Chu, 1999).

Based on the SCID-derived diagnosis, subjects were divided into bipolar and non-bipolar groups. The two groups were compared on demographic and clinical variables. Univariate and multivariable logistic regression were used to identify the clinical correlates indicative of bipolar disorder by controlling for possible demographic confounders. The most validated bipolar predictors found in previous studies, including bipolar family history in first-degree relatives and age of onset of illness <21 years were entered as independent variables; while diagnosis of bipolar disorder was the dependent variable.

3. Results

Of the 340 subjects who completed the questionnaires, 16 refused the telephone-based SCID interview and we were unable to contact 19 subjects by telephone. There was no significant difference in demographics, clinical variables, psychiatric diagnosis by medical record, and median MDQ and HCL-32 scores between the 305 respondents and 35 non-respondents of SCID.

3.1 Sample description

The median age of the 340 participants was 50.0 years (range = 18-64); 66.8% were female. About two-thirds were married or cohabited and 65.0% had secondary educational level or above. Ninety-four subjects (27.6%) were professional, semi-professional or clerical worker; 74 (21.8%) were manual worker, 88 (25.9%) were homemaker, and 84 (24.7%) were unemployed. The median age of onset of illness was 37.5 years (range = 10-62) and the median duration of illness was 9.0 years (range = 1-44). Sixty-eight participants (20.0%) had family history of depressive disorder in their first-degree relatives; while 27 (7.9%) had bipolar family history.

Based on the telephone-based SCID, 31 (10.2%) of the 305 subjects were re-diagnosed to have a lifetime DSM-IV bipolar disorder, in which 14 (4.6%) were bipolar II and 17 (5.6%) were bipolar NOS. All bipolar NOS subjects reported to have a history of brief hypomania lasting for 2 to 3 days. Participants who were re-diagnosed with bipolar disorder included 20 patients with an initial diagnosis of major depressive disorder, 5 with anxiety disorders, 3 with mixed anxiety and depressive disorder, 1 with primary insomnia, 1 with major depressive disorder comorbid generalized anxiety disorder, and 1 with adjustment disorder comorbid personality disorder.

3.2 Psychometric properties of the Chinese MDQ and HCL-32

The Cronbach's alpha of MDQ and HCL-32 were 0.75 and 0.89, respectively. In both questionnaires, elimination of each item did not result in a substantial increase in their internal consistency. The test-retest reliability (ICC) of MDQ and HCL-32 ($n = 107$) were 0.74 (95% CI: 0.64-0.81) and 0.81 (95% CI: 0.73-0.87), respectively.

The ROC curves of the Chinese MDQ were compared at 6 threshold levels made up

of different combination of sections 2 and 3 results (Figure 1). Using the standard cut-off criterion, sensitivity was only 0.16 (95% CI: 0.03-0.29) and specificity was 0.99 (95% CI: 0.98-1.00). Lowering and omission of impairment criterion increased the sensitivity to 0.32 and 0.39, respectively, while specificity only decreased to 0.97. Omission of both sections 2 and 3 increased the sensitivity to 0.42 and decreased the specificity to 0.93. The cut-off level that included section 2 and removed section 3 had the highest AUC of 0.77 (95% CI: 0.67-0.87). The optimal cut-off point in our sample was a clustering of 4 or more positive symptoms with omission of impairment criterion, providing the best balance of sensitivity of 0.65 (95% CI: 0.48-0.82) and specificity of 0.77 (95% CI: 0.72-0.82). The PPV and NPV was 0.24 and 0.95, respectively. The LR(+) and LR(-) was 2.83 and 0.45, respectively.

The AUC of the Chinese HCL-32 was 0.80 (95% CI: 0.72-0.88), indicating good discriminatory power (Figure 2). The original cut-off score of 14 only obtained a sensitivity of 0.68 (95% CI: 0.52-0.84) and specificity of 0.80 (95% CI: 0.75-0.85). From the ROC curve, the optimal cut-off score in our study was 11, yielding a sensitivity of 0.84 (95% CI: 0.71-0.97) and specificity of 0.70 (95% CI: 0.65-0.75), while the PPV was only 0.24 and the NPV was 0.98. The LR(+) and LR(-) was 2.80 and 0.23, respectively.

3.3 Comparing the psychometric performance of the Chinese MDQ and HCL-32

The Chinese HCL-32 had higher internal consistency (0.89 vs. 0.75) and 4-week test-retest reliability (0.81 vs. 0.74) than the Chinese MDQ. The AUC of HCL-32 (0.80) was higher than those of MDQ at all threshold levels (0.53-0.77). At the original cut-offs of both questionnaires, HCL-32's sensitivity (0.68) and specificity (0.80) fell outside the 95% confidence intervals for MDQ's sensitivity (0.03-0.29) and specificity (0.98-1.00). At their optimal cut-offs, HCL-32's sensitivity (0.84) and specificity (0.70) also fell outside the 95% confidence intervals for MDQ's sensitivity (0.48-0.82) and specificity (0.72-0.82). Hence, HCL-32 had higher sensitivity but lower specificity than MDQ at both cut-off levels.

3.4 Clinical predictors of bipolar disorder

Subjects with SCID-derived bipolar disorder were younger and more likely to have educational level at secondary or above and positive bipolar family history in first-degree relatives than non-bipolar subjects; however, psychiatric comorbidity was not more common in subjects with bipolar disorder (Table 1). Univariate logistic

regression showed that bipolar disorder was associated with positive bipolar family history (OR = 4.15, 95% CI: 1.58-10.92, P = 0.004); while age of onset younger than 21 years was not significantly associated with bipolar disorder (OR = 1.79, 95% CI: 0.57-5.58, P = 0.32). Multivariable logistic regression showed that bipolar family history (OR = 4.93, 95% CI: 1.73-14.02, P = 0.003) was the only independent clinical factor associated with bipolar disorder after controlling for the demographic confounders (age and educational level).

3.5 Screening of bipolar disorder using combinations of bipolar family history and MDQ and HCL-32 scores

Table 2 presents the screening performance of different combinations of bipolar family history and MDQ and HCL-32 scores. Adding bipolar family history to MDQ increased the sensitivity from 0.65 to 0.71 but decreased the specificity from 0.77 to 0.72, compared to MDQ alone. Adding bipolar family history to HCL-32 increased the sensitivity from 0.84 to 0.90 and slightly decreased the specificity from 0.70 to 0.67, compared to HCL-32 alone.

4. Discussion

This was the first systematic study comparing the psychometric properties of MDQ, HCL-32, clinical predictors and their combination in a general psychiatric outpatient setting. We found that both MDQ and HCL-32 were valid and reliable screening instruments for previously unrecognized bipolar disorder; however, the optimal cut-offs were different from the original criteria obtained in specialized mood disorder clinics. The HCL-32 was more sensitive than the MDQ in detecting hypomanic conditions; in addition, bipolar family history was a useful clinical predictor with high specificity. Compared to using MDQ, HCL-32, or bipolar family history alone, simultaneous use of HCL-32 and family history could achieve better sensitivity for detecting bipolar disorder that was previously undiagnosed.

The Chinese MDQ had an internal consistency of 0.75, which was comparable to 2 previous local studies (Chung et al., 2008; Chung et al., 2009) and a Finnish study (Isometsa et al., 2003). The short-term test-retest reliability was satisfactory, but lower than that of the Spanish MDQ (Vieta et al., 2007; Sanchez-Moreno et al., 2008). The recall of hypomania and reproducibility of MDQ could be influenced by the severity of past mood symptoms (Gervasoni et al., 2009). It was possible that the bipolar I

subjects in the Spanish study might have recalled the past manic episode more reliably than the current sample with mild hypomanic conditions.

In line with previous studies (Benazzi, 2003; Miller et al., 2004; Twiss et al., 2008), lowering or omission of the impairment criterion of MDQ increased the sensitivity without significantly sacrificing the specificity. At any cut-off level, the sensitivity of the Chinese MDQ for bipolar II/NOS was lower than those reported in most previous studies (Hirschfeld et al., 2000; Isometsa et al., 2003; Weber Rouget et al., 2005; Chung et al., 2008; Sanchez-Moreno et al., 2008; Twiss et al., 2008). The finding was possibly due to differences in study population. Our subjects were selected from a non-specialized psychiatric setting among patients without previous diagnosis of bipolar disorder; while most previous studies were conducted in specialized mood disorder clinics and included known bipolar patients. Patients in specialized psychiatric settings had more prototypical and severe illness regardless of the bipolar subtype and better insight and knowledge about their bipolar diagnosis as a result of psychoeducation. Hence, they could recognize their past hypomania better than the previously undiagnosed bipolar II/NOS patients in our study. The optimal cut-off in this study was also lower compared to most previous studies, but it was closest to an Italian study conducted in a similar general psychiatric outpatient setting (Hardoy et

al., 2005). Cultural factors may also influence the MDQ's screening performance. The Chinese tend to have stigma, negative attitudes, and misconception toward mental illness and low perceived need for psychiatric treatment (Ng, 1997; Lee et al., 2007). This may possibly result in denial of hypomanic symptoms among Chinese subjects, leading to underreporting during questionnaire completion. A recent study in Germany found that young people's attitudes toward mania were more negative than for depression (Wolkenstein and Meyer, 2008); however, no study has been conducted in the Chinese population. Future cross-cultural studies on the attitudes and knowledge toward bipolar disorder are needed. The finding that the psychometric performance of the Chinese MDQ being similar to that of the Korean MDQ in depressed outpatients without previous bipolar diagnosis (Kim et al., 2008) supported our hypothesis that the sensitivity of MDQ was dependent on the study population and cultural factor.

The Chinese HCL-32 had a high internal consistency that was comparable to previous reports (Angst et al., 2005; Wu et al., 2008). The test-retest reliability was similar to that of the Spanish HCL-32 (Vieta et al., 2007). The sensitivity of Chinese HCL-32 for bipolar II/NOS at the original cut-off score of 14 was slightly lower than the Taiwanese and Spanish HCL-32 (Vieta et al., 2007; Wu et al., 2008). The optimal

cut-off score in our study was also lower than those obtained in previous studies (Angst et al., 2005; Vieta et al., 2007; Wu et al., 2008; Forty et al., 2009). The lower sensitivity and optimal cut-off in our study relative to other versions could also be due to differences in study population. Our optimal cut-off score of HCL-32 was 11 (sensitivity 0.84; specificity 0.70), which was most comparable to the optimal cut-off score of 12 for bipolar II in a non-specialized psychiatric setting (sensitivity 0.80; specificity 0.54) (Carta et al., 2006).

In line with previous studies (Carta et al., 2006; Vieta et al., 2007), we found that the Chinese HCL-32 had better discriminatory power and was more sensitive in detecting hypomanic conditions compared to MDQ. There are two possible reasons for the superiority of HCL-32 over MDQ in screening for the milder forms of bipolar disorder. Firstly, the development of MDQ and HCL-32 is based on different concepts. The MDQ is built on the 'categorical' concept of DSM-IV; while the HCL-32 is based on the 'dimensional' approach of bipolarity and has a wider range of hypomanic symptoms (Phelps and Ghaemi, 2006). Secondly, the questions of HCL-32 are descriptive and non-stigmatizing; while some items in MDQ tend to portray severe psychopathology and can be perceived by patients as signs of severe mental illness (Angst, 2008).

Our study replicated the findings in previous Western and Asian studies that bipolar family history in first-degree relatives was a clinical indicator of bipolarity (Benazzi, 2007; Kim et al., 2008; Mak, 2009). A previous study found that bipolar family history was the strongest validator of bipolar II disorder (Benazzi and Akiskal, 2008). We showed that the absence of a family history of bipolar disorder had a high specificity, supporting its usefulness in ruling out bipolarity (Benazzi, 2004). Unlike most Western studies, we found that early onset of mood symptoms was not associated with bipolar disorder. The Chinese might have later age of onset of mood disorders (Lee et al., 2007); hence the definition of early-onset could be different between Chinese and Western populations. Further research is needed to compare the age of onset of mood symptoms between the Chinese and Western patients with bipolar disorder. Our finding that bipolar disorder was more common among younger subjects was in line with previous epidemiological and clinical studies (Merikangas et al., 2007; Lee et al., 2009; Mak, 2009). However, the association between bipolar disorder and educational level was still inconclusive according to a systematic review (Tsuchiya et al., 2003).

Our study has a number of strengths as well as several methodological limitations. We

used a large and representative sample of general psychiatric outpatients and included patients with common comorbidities of bipolar disorder; hence our sample is representative of the real-life setting where bipolar disorder is commonly unrecognized. The major limitation was the use of SCID alone for diagnosis of bipolar disorder. Although the same methodology was adopted in previous MDQ and HCL-32 validation studies, future studies utilizing collateral information would improve our understanding of the scales' actual performance. Another limitation was that psychiatric diagnoses other than mood disorders and alcohol or substance use disorder were derived only from medical record. Although DSM-IV Axis I and II comorbidity may not influence the screening performance of MDQ and HCL-32 (Meyer et al., 2011), the under-recognition of psychiatric comorbidity in this study undermined our finding that psychiatric comorbidity was not a clinical predictor of bipolar disorder. Standardized instruments for assessing age of onset of illness and family history of mood disorders were not used, although efforts had been made to verify the clinical information during telephone interview. In addition, we had not examined the family history of other psychiatric disorders, which may have an association with bipolar disorder. The subjects were considered clinically stable at the time of recruitment; however, standardized rating scales were not used to quantify their mood during the administration of questionnaires. Although some researchers

have cautioned that moderately or severely depressed patients may underreport their previous hypomanic symptoms, a recent study showed that in subjects who had remitted from a severe depressive episode the MDQ total score did not significantly change over time (Gervasoni et al., 2009). Lastly, a single SCID rater might have introduced diagnostic bias; however, the semi-structured format and excellent inter-rater reliability of the SCID supported the diagnostic accuracy in this study.

In conclusion, compared to using MDQ, HCL-32 and bipolar family history alone, simultaneous use of HCL-32 and family history was the best approach for screening bipolar disorder in a general psychiatric setting in Hong Kong. This screening method could detect most of the previously unrecognized bipolar disorder with satisfactory specificity and low false-negative rate. This approach only requires around 10 minutes to complete, so it is potentially useful in busy outpatient settings. Future studies should evaluate the screening of bipolar disorder in non-psychiatric populations. This is particularly relevant for HCL-32, which has not been validated in community and family medicine settings.

Acknowledgements

We thank our colleagues in Queen Mary Hospital for their help in patient recruitment.

We also thank Dr. Ru-Band Lu for his permission in our use of the Chinese version of

HCL-32. The participation of all patients in our study is deeply appreciated.

Table 1. Comparison between bipolar and non-bipolar subjects on demographic and clinical variables

Variables	Total (n = 305)	Bipolar (n = 31)	Non-bipolar (n = 274)	P value ^a
	n (%)	n (%)	n (%)	
Female gender	204 (66.9)	17 (54.8)	187 (68.2)	0.13
Educational level				0.002
Below secondary	108 (35.4)	3 (9.7)	105 (38.3)	
Secondary or above	197 (64.6)	28 (90.3)	169 (61.7)	
Marital status				0.83
Single/divorced/widowed	103 (33.8)	11 (35.5)	92 (33.6)	
Married/cohabited	202 (66.2)	20 (64.5)	182 (66.4)	
Occupation				0.98
Unemployed	74 (24.3)	7 (22.6)	67 (24.5)	
Homemaker	83 (27.2)	8 (25.8)	75 (27.4)	
Labour worker	65 (21.3)	6 (19.4)	59 (21.5)	
Business/clerk	57 (18.7)	7 (22.6)	50 (18.2)	
Semi-/professional	26 (8.5)	3 (9.7)	23 (8.4)	
Monthly household income ^b				0.10
On government subsidies	67 (22.3)	5 (16.7)	62 (23.0)	
Low income group (≤ HK\$20000)	169 (56.3)	14 (46.7)	155 (57.4)	
High income group (> HK\$20000)	64 (21.3)	11 (36.7)	53 (19.6)	
Positive family history in first-degree relatives				
Bipolar disorder	25 (8.2)	7 (22.6)	18 (6.6)	0.007
Depressive disorder	64 (21.0)	8 (25.8)	56 (20.4)	0.49
Psychiatric comorbidity				
Axis I and II comorbidity by medical record	16 (5.2)	2 (6.5)	14 (5.1)	0.67
Substance/alcohol use disorder by SCID	23 (7.5)	2 (6.5)	21 (7.7)	1.00
	median (range) (mean ± SD)	median (range) (mean ± SD)	median (range) (mean ± SD)	P value ^c
Age	50.0 (18-64) (48.9 ± 9.6)	44.0 (18-62) (44.1 ± 9.2)	50.5 (23-64) (49.4 ± 9.5)	0.002
Age of onset of illness	37.0 (10-60) (37.4 ± 10.7)	34.0 (10-60) (33.9 ± 10.9)	38.5 (12-60) (37.8 ± 10.6)	0.057
MDQ total score	3.0 (0-13) (3.2 ± 2.5)	5.0 (1-13) (5.8 ± 3.2)	3.0 (0-10) (2.9 ± 2.2)	<0.001
HCL-32 total score	8.0 (0-27) (8.9 ± 6.6)	17.0 (2-25) (15.7 ± 6.2)	7.0 (0-27) (8.1 ± 6.2)	<0.001

^a P value by Pearson chi-square or Fisher's exact test.

^b n = 300 due to missing data (bipolar group, n = 30; non-bipolar group, n = 270).

^c P value by Mann-Whitney U test.

Table 2. Comparison of screening performance for combination of Chinese MDQ^a and HCL-32^b at optimal cut-offs and bipolar family history

Screening method	Sensitivity	Specificity	PPV	NPV	LR(+)	LR(-)
MDQ alone	0.65	0.77	0.24	0.95	2.83	0.45
HCL-32 alone	0.84	0.70	0.24	0.98	2.80	0.23
Bipolar family history alone	0.23	0.93	0.28	0.91	3.29	0.83
MDQ and bipolar family history (either one positive)	0.71	0.72	0.22	0.96	2.54	0.40
HCL-32 and bipolar family history (either one positive)	0.90	0.67	0.24	0.98	2.73	0.15

^aOptimal cut-off criteria of MDQ in this population is ≥ 4 positive items, symptom clustering and omission of impairment criterion.

^bOptimal cut-off criteria of HCL-32 in this population is ≥ 11 positive items.

PPV = Positive predictive value; NPV = Negative predictive value; LR(+) = Positive likelihood ratio; LR(-) = Negative likelihood ratio.

Figure 1. Receiver operating characteristic curves of Chinese MDQ at six different cut-off levels.

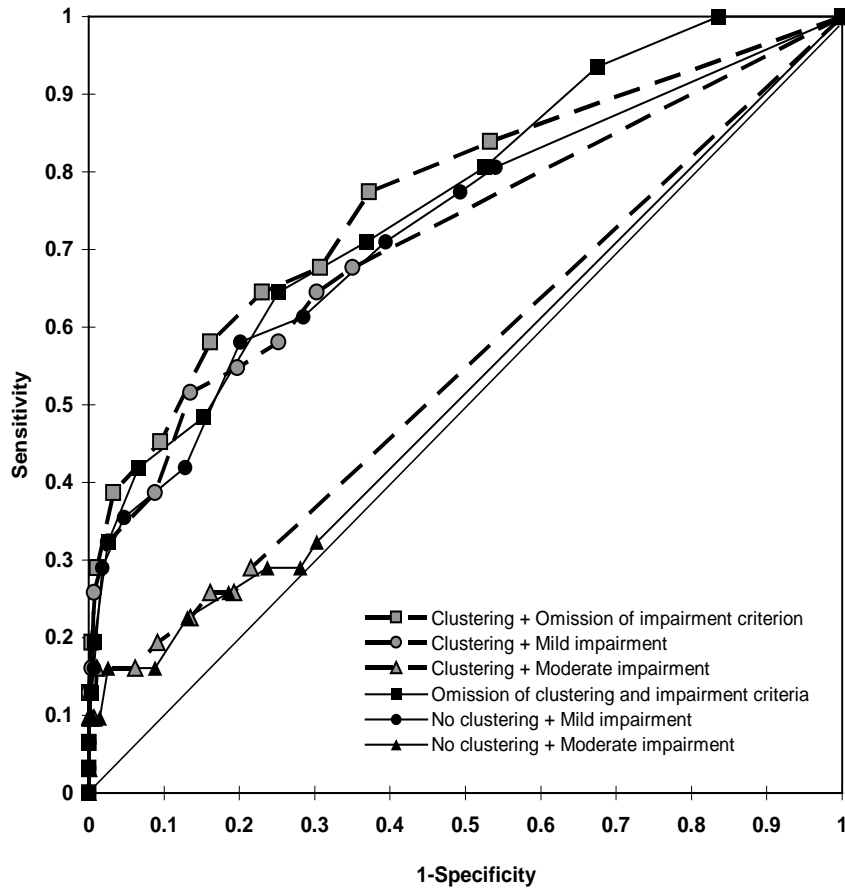
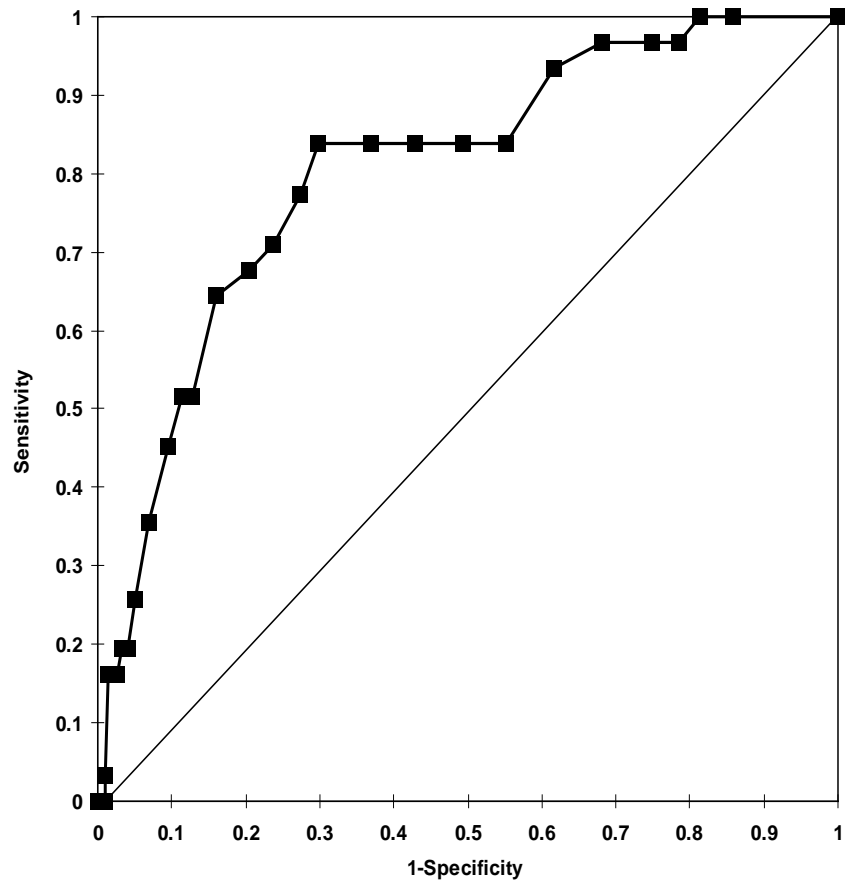


Figure 2. Receiver operating characteristic curve of Chinese HCL-32



References

- Akiskal, H.S., Benazzi, F., 2005. Optimizing the detection of bipolar II disorder in outpatient private practice: toward a systematization of clinical diagnostic wisdom. *Journal of Clinical Psychiatry* 66, 914-921.
- Akiskal, H.S., Pinto, O., 1999. The evolving bipolar spectrum. Prototypes I, II, III, and IV. *Psychiatric Clinics of North America* 22, 517-534, vii.
- Allen, K., Cull, A., Sharpe, M., 2003. Diagnosing major depression in medical outpatients: acceptability of telephone interviews. *Journal of Psychosomatic Research* 55, 385-387.
- American Psychiatric Association, 2000. *Diagnostic and Statistical Manual of Mental Disorders Fourth Edition*. American Psychiatric Association, Washington DC. Text Revision.
- Angst, J., 2008. Bipolar disorder - methodological problems and future perspectives. *Dialogues in Clinical Neuroscience* 10, 129-139.
- Angst, J., Adolfsson, R., Benazzi, F., Gamma, A., Hantouche, E., Meyer, T.D., Skeppar, P., Vieta, E., Scott, J., 2005. The HCL-32: towards a self-assessment tool for hypomanic symptoms in outpatients. *Journal of Affective Disorders* 88, 217-233.

- Awad, A.G., Rajagopalan, K., Bolge, S.C., McDonnell, D.D., 2007. Quality of life among bipolar disorder patients misdiagnosed with major depressive disorder. *Primary Care Companion to the Journal of Clinical Psychiatry* 9, 195-202.
- Benazzi, F., 2003. Improving the Mood Disorder Questionnaire to detect bipolar II disorder. *Canadian Journal of Psychiatry* 48, 770-771.
- Benazzi, F., 2004. Bipolar II disorder family history using the family history screen: findings and clinical implications. *Comprehensive Psychiatry* 45, 77-82.
- Benazzi, F., 2007. Testing predictors of bipolar-II disorder with a 2-day minimum duration of hypomania. *Psychiatry Research* 153, 153-162.
- Benazzi, F., Akiskal, H., 2006. The duration of hypomania in bipolar-II disorder in private practice: methodology and validation. *Journal of Affective Disorders* 96, 189-196.
- Benazzi, F., Akiskal, H.S., 2003. Refining the evaluation of bipolar II: beyond the strict SCID-CV guidelines for hypomania. *Journal of Affective Disorders* 73, 33-38.
- Benazzi, F., Akiskal, H.S., 2008. How best to identify a bipolar-related subtype among major depressive patients without spontaneous hypomania: superiority of age at onset criterion over recurrence and polarity? *Journal of Affective Disorders* 107, 77-88.

- Buderer, N.M., 1996. Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. *Academic Emergency Medicine* 3, 895-900.
- Cacciola, J.S., Alterman, A.I., Rutherford, M.J., McKay, J.R., May, D.J., 1999. Comparability of telephone and In-person structured clinical interview for DSM-III-R (SCID) diagnoses. *Assessment* 6, 235-242.
- Carta, M.G., Hardoy, M.C., Cadeddu, M., Murru, A., Campus, A., Morosini, P.L., Gamma, A., Angst, J., 2006. 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. *Clinical Practice and Epidemiology in Mental Health* 2, 2.
- Chu, K., 1999. An introduction to sensitivity, specificity, predictive values and likelihood ratios. *Emergency Medicine* 11, 175-181.
- Chung, K.F., Tso, K.C., 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, K.F., Tso, K.C., Chung, R.T., 2009. Validation of the Mood Disorder Questionnaire in the general population in Hong Kong. *Comprehensive Psychiatry* 50, 471-476.

- Crippa, J.A., de Lima Osorio, F., Del-Ben, C.M., Filho, A.S., da Silva Freitas, M.C., Loureiro, S.R., 2008. Comparability between telephone and face-to-face structured clinical interview for DSM-IV in assessing social anxiety disorder. *Perspectives in Psychiatric Care* 44, 241-247.
- Dunner, D.L., 2003. Clinical consequences of under-recognized bipolar spectrum disorder. *Bipolar Disorders* 5, 456-463.
- First, M.B., Spitzer, R.L., Gibbon, M., Williams, J.B.W., 2002. Structured Clinical Interview for DSM-IV-TR Axis I Disorders - Research Version, Patient Edition (SCID-I/P). Biometrics Research, New York State Psychiatric Institute, New York.
- Forty, L., Smith, D., Jones, L., Jones, I., Caesar, S., Fraser, C., Gordon-Smith, K., Craddock, N., 2009. Identifying hypomanic features in major depressive disorder using the hypomania checklist (HCL-32). *Journal of Affective Disorders* 114, 68-73.
- Gervasoni, N., Weber Rouget, B., Miguez, M., Dubuis, V., Bizzini, V., Gex-Fabry, M., Bondolfi, G., Aubry, J.M., 2009. Performance of the Mood Disorder Questionnaire (MDQ) according to bipolar subtype and symptom severity. *European Psychiatry* 24, 341-344.
- Ghaemi, S.N., Ko, J.Y., Goodwin, F.K., 2002. "Cade's disease" and beyond:

misdiagnosis, antidepressant use, and a proposed definition for bipolar spectrum disorder. *Canadian Journal of Psychiatry* 47, 125-134.

Hardoy, M.C., Cadeddu, M., Murru, A., Dell'Osso, B., Carpiniello, B., Morosini, P.L., Calabrese, J.R., Carta, M.G., 2005. Validation of the Italian version of the "Mood Disorder Questionnaire" for the screening of bipolar disorders. *Clinical Practice and Epidemiology in Mental Health* 1, 8.

Helzer, J.E., Robins, L.N., McEvoy, L.T., Spitznagel, E.L., Stoltzman, R.K., Farmer, A., Brockington, I.F., 1985. A comparison of clinical and diagnostic interview schedule diagnoses. Physician reexamination of lay-interviewed cases in the general population. *Archives of General Psychiatry* 42, 657-666.

Hirschfeld, R.M., 2001. Bipolar spectrum disorder: improving its recognition and diagnosis. *Journal of Clinical Psychiatry* 62 Suppl 14, 5-9.

Hirschfeld, R.M., Lewis, L., Vornik, L.A., 2003. Perceptions and impact of bipolar disorder: how far have we really come? Results of the national depressive and manic-depressive association 2000 survey of individuals with bipolar disorder. *Journal of Clinical Psychiatry* 64, 161-174.

Hirschfeld, R.M., Williams, J.B., Spitzer, R.L., Calabrese, J.R., Flynn, L., Keck, P.E., Jr., Lewis, L., McElroy, S.L., Post, R.M., Rapport, D.J., Russell, J.M., Sachs, G.S., 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.
- Isometsa, E., Suominen, K., Mantere, O., Valtonen, H., Leppamaki, S., Pippingskold, M., Arvilommi, P., 2003. The mood disorder questionnaire improves recognition of bipolar disorder in psychiatric care. *BMC Psychiatry* 3, 8.
- Judd, L.L., Akiskal, H.S., 2003. The prevalence and disability of bipolar spectrum disorders in the US population: re-analysis of the ECA database taking into account subthreshold cases. *Journal of Affective Disorders* 73, 123-131.
- Katzow, J.J., Hsu, D.J., Nassir Ghaemi, S., 2003. The bipolar spectrum: a clinical perspective. *Bipolar Disorders* 5, 436-442.
- Kim, B., Wang, H.R., Son, J.I., Kim, C.Y., Joo, Y.H., 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, K.L., Tsang, A., 2009. A community survey of the twelve-month prevalence and correlates of bipolar spectrum disorder in Hong Kong. *Journal of Affective Disorders* 117, 79-86.
- Lee, S., Tsang, A., Zhang, M.Y., Huang, Y.Q., He, Y.L., Liu, Z.R., Shen, Y.C., Kessler, R.C., 2007. Lifetime prevalence and inter-cohort variation in DSM-IV disorders

in metropolitan China. *Psychological Medicine* 37, 61-71.

Lish, J.D., Dime-Meenan, S., Whybrow, P.C., Price, R.A., Hirschfeld, R.M., 1994.

The National Depressive and Manic-depressive Association (DMDA) survey of bipolar members. *Journal of Affective Disorders* 31, 281-294.

Mak, A.D., 2009. Prevalence and correlates of bipolar II disorder in major depressive

patients at a psychiatric outpatient clinic in Hong Kong. *Journal of Affective Disorders* 112, 201-205.

Matza, L.S., Rajagopalan, K.S., Thompson, C.L., de Lissovoy, G., 2005.

Misdiagnosed patients with bipolar disorder: comorbidities, treatment patterns, and direct treatment costs. *Journal of Clinical Psychiatry* 66, 1432-1440.

Merikangas, K.R., Akiskal, H.S., Angst, J., Greenberg, P.E., Hirschfeld, R.M.,

Petukhova, M., Kessler, R.C., 2007. Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey replication.

Archives of General Psychiatry 64, 543-552.

Meyer, T.D., Bernhard, B., Born, C., Fuhr, K., Gerber, S., Schaerer, L., Langosch,

J.M., Pfennig, A., Sasse, J., Scheiter, S., Schottle, D., van Calker, D.,

Wolkenstein, L., Bauer, M., 2011. The Hypomania Checklist-32 and the Mood

Disorder Questionnaire as screening tools – going beyond samples of purely mood-disordered patients. *Journal of Affective Disorders* 128. 291-298.

- Meyer, T.D., Hammelstein, P., Nilsson, L.G., Skeppar, P., Adolfsson, R., Angst, J., 2007. The Hypomania Checklist (HCL-32): its factorial structure and association to indices of impairment in German and Swedish nonclinical samples. *Comprehensive Psychiatry* 48, 79-87.
- Miller, C.J., Klugman, J., Berv, D.A., Rosenquist, K.J., Ghaemi, S.N., 2004. Sensitivity and specificity of the Mood Disorder Questionnaire for detecting bipolar disorder. *Journal of Affective Disorders* 81, 167-171.
- Ng, C.H., 1997. The stigma of mental illness in Asian cultures. *Australian and New Zealand Journal of Psychiatry* 31, 382-390.
- Rybakowski, J.K., Angst, J., Dudek, D., Pawlowski, T., Lojko, D., Siwek, M., Kiejna, A., 2010. Polish version of the Hypomania Checklist (HCL-32) scale: the results in treatment-resistant depression. *European Archives of Psychiatry and Clinical Neuroscience* 260, 139-144.
- Sanchez-Moreno, J., Villagran, J.M., Gutierrez, J.R., Camacho, M., Ocio, S., Palao, D., Querejeta, I., Gascon, J., Sanchez, G., Vieta, E., 2008. Adaptation and validation of the Spanish version of the Mood Disorder Questionnaire for the detection of bipolar disorder. *Bipolar Disorders* 10, 400-412.
- Shi, L., Thiebaud, P., McCombs, J.S., 2004. The impact of unrecognized bipolar disorders for patients treated for depression with antidepressants in the

Formatted: German (Germany)

- fee-for-services California Medicaid (Medi-Cal) program. *Journal of Affective Disorders* 82, 373-383.
- So, E., Kam, I. Leung, C.M., 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.
- Tsuchiya, K.J., Byrne, M., Mortensen, P.B., 2003. Risk factors in relation to an emergence of bipolar disorder: a systematic review. *Bipolar Disorders* 5, 231-242.
- Twiss, J., Jones, S., Anderson, I., 2008. Validation of the Mood Disorder Questionnaire for screening for bipolar disorder in a UK sample. *Journal of Affective Disorders* 110, 180-184.
- Vieta, E., Sanchez-Moreno, J., Bulbena, A., Chamorro, L., Ramos, J.L., Artal, J., Perez, F., Oliveras, M.A., Valle, J., Lahuerta, J., Angst, J., 2007. Cross validation with the mood disorder questionnaire (MDQ) of an instrument for the detection of hypomania in Spanish: the 32 item hypomania symptom check list (HCL-32). *Journal of Affective Disorders* 101, 43-55.
- Weber Rouget, B., Gervasoni, N., Dubuis, V., Gex-Fabry, M., Bondolfi, G., Aubry, J.M., 2005. Screening for bipolar disorders using a French version of the Mood Disorder Questionnaire (MDQ). *Journal of Affective Disorders* 88, 103-108.

Wolkenstein, L., Meyer, T.D., 2008. Attitudes of young people towards depression and mania. *Psychology and Psychotherapy* 81, 15-31.

Wu, Y.S., Angst, J., Ou, C.S., Chen, H.C., Lu, R.B., 2008. Validation of the Chinese version of the hypomania checklist (HCL-32) as an instrument for detecting hypo(mania) in patients with mood disorders. *Journal of Affective Disorders* 106, 133-143.