Comparative validity of inventories and checklists for identifying depressed patients with hidden bipolarity

Per Bech\textsuperscript{*1}, Rikke Engell\textsuperscript{1} & Stine Bjerrum Møller\textsuperscript{1}

Practice points

- A high proportion of treatment-resistant depressed patients may be bipolar patients, and clinicians should screen for signs of bipolarity when a patient presents with major depressive disorder.
- Clinical validity and measures assessing unidimensional constructs should be prioritized over lengthy multidimensional measures.
- The symptom profiles in unipolar depression and bipolar depression are rather similar in the acute phase.
- The duration criterion of bipolar phases is not clinically relevant and should be omitted.
- Overactivity is equally as important as elevated mood when searching for 'hidden' bipolarity in depressed patients and the dialog could be opened with Eysenck's neuroticism question 'are you sometimes bubbling over with energy and sometimes very sluggish?'
- Fluctuating mood as a risk factor for developing bipolar depression is conceptualized in contemporary cognitive models of bipolar depression. Individuals experiencing fluctuating moods in combination with behavior change (ascent behaviors) are at risk of exacerbating their mood and fueling the spiraling process towards mania. Differentiating between ordinary mood swings and pathological mood swings is an important future task.
- The ability to detect hidden bipolarity in a population of patients with major depressive disorder is equal for the Hypomania Checklist-32 (HCL-32) and the Mood Disorder Questionnaire (MDQ). These tools might have high applicability in the daily clinic when supplemented with the clinician-rated bipolarity instrument made by Parker, and modified by items from the HCL-32, MDQ or Eysenck Personality Questionnaire as suggested.
- The great clinical impact of identifying hidden bipolarity is in depressed patients who are resistant to antidepressants.
SUMMARY

Current diagnostic criteria favor overdiagnosis of major depression disorder, and evidence suggests that bipolarity is underdiagnosed in depressed samples. In this review, we focus on the comparative validity of questionnaires to identify depressed patients at risk for bipolar disorder. The review focuses on clinical validity as expressed by experienced psychiatrists as opposed to focusing on the psychometric validity of the individual items in a questionnaire. Compared with the Mood Disorder Questionnaire (MDQ), the Hypomania Checklist-32 (HCL-32) has been the subject of many more psychometric analyses. It is unfortunate that the classical methods (factor analysis and Cronbach’s coefficient \( \alpha \)) have been used to evaluate the psychometric validity of the HCL-32, because they favor scales with many items. In this review, we found no differences in the two measures regarding their clinical validity.

The past decade has seen an intense focus on ‘hidden’ bipolarity in patients with major depression because bipolar affective disorder is now considered to be a more severe disorder than major depression \([1]\). Correctly diagnosing bipolar risk is, as stated by Angst \([1]\), essential for appropriate treatment. Therefore, the lack of response to treatment in patients who have been prescribed several antidepressant medications within their major depressive episode might be due to a hidden bipolarity \([2]\). The diagnostic criteria favor overdiagnosis of major depressive disorder, and evidence suggests that bipolarity is underdiagnosed in depressed groups \([3]\).

This review on the comparative validity of questionnaires to identify depressed patients with hidden bipolarity is based on the cliniometric approach \([4,5]\), in which clinical validity has a higher priority than psychometric validation. Clinical validity, as expressed by experienced psychiatrists (macroanalysis or clinical judgement), captures the clinical reality, often with a rather limited number of items, whereas psychometric validation (microanalysis) tends to increase the number of items in an attempt to identify factors. However, in this review we have favored the modern psychometric method (item response theory [IRT] analysis) to classical factor analysis. In Box 1 we have shown how the classical analysis is an attempt to identify the core structure of factors, whereas the IRT is a test for measurement, in other words, to what extent the summed total score of a scale is a sufficient statistic. Therefore, the tendency to increase the number of items in a questionnaire to satisfy the microanalytic tests for homogeneity has been discarded in favor of the clinical content judgment made by the experienced psychiatrist in accordance with the macroanalytic approach.

Box 1. Psychometric analysis of severity questionnaires.

**Classical method: correlation analysis**

- Factor analysis is a statistical model to capture the core structure of factors among the items of a scale. The procedure of rotation (oblique or varimax) identifies a range of factors that are distinct from each other (e.g., somatic vs psychic anxiety).
- Principal component analysis without rotation is a mathematical model in which the first component extracts all the positive correlations between items of a scale. The second component identifies bidirectional factors by negative vs positive loadings (e.g., somatic vs psychic anxiety).
- Cronbach’s coefficient \( \alpha \) identifies the mean correlation between items. A high \( \alpha \) coefficient is, however, not an argument for the unidimensionality of a scale.

**Modern method: item response theory analysis**

- Item response theory analysis is a model for the evaluation of the extent to which the items of a scale are located on a dimension for measurement, for example, the degree of mood or anxiety states.
- The basic principle of this item response theory analysis is the extent to which each item of a scale covers its own specific location on the dimension being measured. An item with lower prevalence is located on the severe part of this dimension, whereas an item with higher prevalence is located on the mild-to-moderate part of the dimension. In the Mokken analysis, it is the mean score of the individual items that identifies their locations. Therefore, the answer categories of the different items have to be similar.
- If the scale fulfills the Mokken analysis, the scale is unidimensional and the total summed score is a sufficient statistic; in other words, no profile scores are necessary.
of their investigations [8], which appeared to be rather consistent in most respects. Their distinction between bipolar (manic-depressive illness) and unipolar (either depressive or manic episodes) disorders within recurrent mood disorders was subsequently included in international diagnostic classification systems, such as the International Classification of Diseases (ICD)-9 [9] or ICD-10 [10], and the DSM-III [11] or DSM-IV [12].

In 1992, Perris reviewed clinical investigations from 1966 to 1990, focusing on the clinical discrimination between unipolar and bipolar illness, taking into account many of the elements within a global clinical analysis (the four time frames in Figure 1) [13]. First, he demonstrated that patients with unipolar mania are extremely rare (less than the 5% originally identified by Kraepelin) [14]. Therefore, unipolar affective illness should be considered in patients with recurrent depressive episodes without mania. In his review, Perris demonstrated that approximately 20% of patients with unipolar depression might later become bipolar patients, using the time frames indicated in Figure 1 [13]. Perris especially focused on time frames 1 and 4. Concerning the clinical symptoms of depression (e.g., according to the Hamilton Depression Scale [HAM-D]), Perris concluded that no consistent symptomatological difference has been found between unipolar depressed patients and bipolar depressed patients in the acute depressive phase [13]. Concerning time frame 4, Perris focused on such personality inventories as the Eysenck Personality Questionnaire (EPQ) (Table 1); here he also had difficulties in finding clinically significant differences, although unipolar depressed patients in euthymic phases seem to score higher on the Eysenck Neuroticism Scale and lower on the Eysenck Extroversion Scale than bipolar depressed patients in euthymic phases [13]. Recent research on personality differences in unipolar and bipolar depression has also been contradictory [15]. However, when reanalyzing our data set from the first three Danish University Antidepressant Group (DUAG) trials [16–18] using a follow-up study of these patients for the classification of bipolar (n = 77) and unipolar (n = 225) patients [Kristoffersen JH, Unpublished Data], it was shown that, on the Newcastle Diagnostic Depression Scale [19], unipolar individuals score higher than bipolar individuals on the item of character neurosis (p = 0.04) [Kristoffersen JH, Unpublished Data].

From a 20-year prospective cohort of community young adults, Angst et al. compared subjects that met the DSM-IV stem criterion A (mood changes, euphoria and irritability) and manifested overactivity with subjects only manifesting overactivity and found that subjects reporting overactivity but no mood changes did not significantly differ from those with mood changes with regard to the clinical validators [20]. Angst et al. concluded that overactivity should be added as a stem criterion for hypomania, and that the duration of the hypomanic episodes had no clinical relevance [20]. Other researchers have also acknowledged overactivity as the most sensitive screening variable among depressed patients [21].

Angst has reviewed the assessment of the risk of bipolar disorder in patients with major depression from 1966 to 2008 [3,22]. He found the DSM-IV criteria for hypomanic episodes problematic. Thus, Angst recommended to include the symptoms of increased activity among the symptomatic criteria and to accept the duration criterion to cover a duration of less than 3 days.
of a temporal criterion rather than the DSM-IV 4-day duration of a hypomanic state [3]. Use of the symptom of irritability as a mandatory symptom as in the DSM-IV approach was also questioned by Angst, who called for a strict unidimensional measure of hypomanic states rather than the trait approach of the EPQ [3].

With reference to Figure 1, Angst [3,22] and Perris [13] found that the symptom profile of bipolar depression does not differ from that found in unipolar depression. According to Angst, the most valid measure of the risk of bipolar illness in patients with major depression is the measure of lifetime episodes of hypomanic states (time frame 4; Figure 1) [3,22].

Concerning a genetic factor, we are still awaiting a valid genetic marker [23]. Since its commencement in 1981, the Zürich community cohort study performed by Angst and his group [3] has undertaken a prospective investigation of hypomanic states by asking patients about episodes of increased enterprise, increased activity, lower fatigability, less need for sleep than usual, talking more, traveling more and doing more activities. A list of 20 hypomanic symptoms was developed for a Hypomania Checklist (HCL) (Table 2) [24]. In a nationwide French investigation, including 994 patients from the primary care setting and 772 from a psychiatric outpatient setting, the
HCL-20 questionnaire identified approximately 60% of the patients with major depression as being at risk of bipolar illness [25]. Subthreshold bipolarity has been identified at a level of approximately 40% in depressed patients [26,27].

In accordance with ICD-10 or DSM-IV, time frame 2 (Figure 1) is used for the consideration of the course of depressive symptoms over the past 2 weeks in order to confirm the diagnosis of depressive episodes. Thus, the symptoms must have persisted for most days; if there is any fluctuation in the course of symptoms then diurnal variance (with the morning being the worst time of the day) is indicative of a diagnosis of depressive episode. When comparing bipolar depression and unipolar recurrent depression, no difference in the persistence or diurnal variation of symptoms has been found.

The Newcastle Diagnostic Depression Scales cover the current depressive episode under investigation and include the symptoms of sleep and appetite, as measured by any change at the time of evaluation compared with the pattern in the euthymic phase before the episode [19]. However, the Newcastle scales only focus on too little sleep or too small an appetite. When bipolar patients are compared with unipolar patients there seems to be some evidence of an increased frequency of hypersomnia and hyperphagia symptoms in bipolar depression [28].

The mandatory clinical assessment when identifying hidden bipolarity in patients with major depressive episodes is, therefore, the lifetime frame of reference as indicated by time frame 4 in Figure 1.

It has been suggested that individuals with high scores on the Hypomanic Personality Scale [29] may have an increased risk of developing bipolar disorder [30]. The Hypomanic Personality Scale contains 48 items. An attempt

| Table 2. The original Hypomania Checklist-20 and the final Hypomania Checklist-32 version with corresponding items in the Mood Disorder Questionnaire. |
|-------------------------------------------------|-----------------|-----------------|-----------------|
| Item number | Item content | Item present in checklist? | Item content | Item present in MDQ? |
|-------------|--------------|-----------------|-----------------|
| 1           | Less sleep  | Y               | –               | Y               |
| 2           | Energy       | Y               | –               | Y               |
| 3           | Self-confidence | Y | –               | Y               |
| 4           | Work         | Y               | –               | N               |
| 5           | Sociability  | Y               | –               | Y               |
| 6           | Travel       | Y               | –               | N               |
| 7           | Excessive shopping | Y | Drive faster | Y               |
| 8           | Risks        | Y               | –               | Y               |
| 9           | Increased physical activity | Y | –               | Y               |
| 10          | Plans/ideas | Y               | More creative   | N               |
| 11          | Less shy     | Y               | –               | Y               |
| 12          | Talkative    | Y               | Colored clothes, meet people | Y               |
| 13          | Irritable    | Y               | Bug other people, gets into quarrels | Y               |
| 14          | Disability   | Y               | Lots of new things | Y               |
| 15          | Sex          | Y               | Flirts more     | Y               |
| 16          | Coffee/cigarettes | Y | Separate coffee vs cigarettes | Y               |
| 17          | Alcohol      | Y               | Drugs           | N               |
| 18          | Euphoria     | Y               | –               | N               |
| 19          | Laughing     | Y               | –               | N               |
| 20          | Thinks faster | Y | Does things more quickly | N               |

Patients were asked the following questions: please try to remember a period in your life when swings in energy, activity and mood were in a ‘high’ state; how did you feel then; and please answer all these statements independently of your present condition with a Y or N.

HCL: Hypomania Checklist; MDQ: Mood Disorder Questionnaire; N: No; Y: Yes.

HCL-20 data taken from [24].
HCL-32 data taken from [32].
MDQ data taken from [33].
has been made to reduce the number to 20 items by the use of IRT models [31]. However, such microanalytic studies lack reference to the clinical content as found by experienced psychiatrists (macroanalysis). There is still no landmark study on the clinical validity of the Hypomanic Personality Scale for detecting hidden bipolarity in patients with major depression. In Table 1 we have selected the clinically most relevant items for hidden bipolarity in the EPQ from a macroanalytic point of view with reference to the dimension of extraversion and neuroticism. We have then selected the corresponding items in the Hypomanic Personality Scale.

**Psychometric validity of the HCL-32 versus the Mood Disorder Questionaire**

When undertaking this review we decided to use the HCL-32 (Table 2) as a starting point because this is the most widely used instrument in this connection. We performed a MEDLINE and PsycINFO search using ‘Hypomania Checklist’ and ‘HCL-32’ as search terms. The HCL-32 was originally published in 2005 [32]. At that point in time, the Mood Disorder Questionaire (MDQ) had been available for 5 years [33]. The MDQ is a brief 13-item self-report questionnaire designed to screen for hidden bipolarity and is an easy-to-administer screening test. It soon reached a rather high acceptability and was translated into several languages. In our search for questionnaires with which to compare the validity of the HCL-32 it was, therefore, obvious that the MDQ was the candidate. Another screening inventory, the Bipolar Spectrum Diagnostic Scale (BSDS), which was developed to be sensitive in milder variants of bipolar disorder [34], has only been very infrequently compared with the HCL-32. In a review by Zimmerman its value as a screening measure in routine clinical practice was questioned [35].

Box 1 shows the two major methods found in psychometric literature for the evaluation of the psychometric (microanalytic) validity of a questionnaire or an inventory. Compared with the MDQ, the HCL-32 has undergone much more psychometric analysis. Unfortunately, the classical methods (factor analysis and Cronbach’s coefficient $\alpha$) have been used to evaluate the psychometric validity of the HCL-32 (Table 3). When evaluating an $\alpha$ coefficient, it is very important to take the number of items in a questionnaire into account because a higher number will give a higher coefficient. Typically, a questionnaire with eight items will produce an $\alpha$ coefficient of approximately 0.60, whereas a questionnaire with 20 items will produce an $\alpha$ coefficient of approximately 0.80. In Table 3, the MDQ with its 13 items has produced $\alpha$ coefficients of between 0.65 and 0.76 as expected, while the HCL-32 has produced $\alpha$ coefficients of between 0.77 and 0.94, as to be expected with its 32 items.

### Table 3. Psychometric validity of Hypomania Checklist-32 versus the Mood Disorder Questionaire.

<table>
<thead>
<tr>
<th>Study (year)</th>
<th>Patients (n)</th>
<th>Cronbach's $\alpha$</th>
<th>Factor analysis</th>
<th>ROC</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCL-32 MDQ</td>
<td>HCL-32 MDQ</td>
<td>HCL-32 MDQ</td>
<td>HCL-32 MDQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>Specificity</td>
<td>Sensitivity</td>
<td>Specificity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>An et al. (2011)</td>
<td>608</td>
<td>0.89</td>
<td>–</td>
<td>Oblique</td>
<td>–</td>
</tr>
<tr>
<td>Carta et al. (2006)</td>
<td>123</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.76</td>
</tr>
<tr>
<td>Chou et al. (2012)</td>
<td>59</td>
<td>0.91</td>
<td>0.76</td>
<td>–</td>
<td>1.00</td>
</tr>
<tr>
<td>Forty et al. (2009, 2010)</td>
<td>582</td>
<td>0.77</td>
<td>0.65</td>
<td>–</td>
<td>0.65</td>
</tr>
<tr>
<td>Leao and Del Porto (2012)</td>
<td>200</td>
<td>0.83</td>
<td>0.76</td>
<td>Varimax</td>
<td>0.80</td>
</tr>
<tr>
<td>Meyer et al. (2011)</td>
<td>288</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.88</td>
</tr>
<tr>
<td>Perugi et al. (2012)</td>
<td>563</td>
<td>0.94</td>
<td>–</td>
<td>–</td>
<td>0.85</td>
</tr>
<tr>
<td>Poon et al. (2012)</td>
<td>340</td>
<td>0.89</td>
<td>0.75</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Rybakowski (2012)</td>
<td>1051</td>
<td>0.93</td>
<td>–</td>
<td>Oblique</td>
<td>–</td>
</tr>
<tr>
<td>Smith et al. (2011)</td>
<td>576</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.79</td>
</tr>
<tr>
<td>Soares et al. (2010)</td>
<td>123</td>
<td>0.86</td>
<td>–</td>
<td>Varimax</td>
<td>0.75</td>
</tr>
<tr>
<td>Wu et al. (2008)</td>
<td>199</td>
<td>0.88</td>
<td>–</td>
<td>–</td>
<td>0.82</td>
</tr>
<tr>
<td>Yang et al. (2012)</td>
<td>1487</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.79</td>
</tr>
<tr>
<td>Yang et al. (2011)</td>
<td>456</td>
<td>0.88</td>
<td>–</td>
<td>–</td>
<td>0.74</td>
</tr>
<tr>
<td>Bech et al. (2011)</td>
<td>122</td>
<td>–</td>
<td>–</td>
<td>Unrotated</td>
<td>0.90</td>
</tr>
</tbody>
</table>

HCL: Hypomania Checklist; MDQ: Mood Disorder Questionaire; ROC: Receiver operating characteristic.
It is, however, always possible to obtain the α coefficient levels found in the MDQ and HCL-32 by simply using questions that are mere variations of one very simple and very restricted theme [36]. In conclusion, it is not correct, as pointed out by Leao and Del Porto [37], that the HCL-32 is superior to the MDQ because of a coefficient α of 0.85 versus 0.76.

In the field of factor analysis we often refer to the British versus the American approach [4]. In the original British tradition of factor analysis going back to Spearman [38], factor analysis is used to interpret the first two factors of a questionnaire, namely the nature of the first general factor versus that of the second bidirectional, or dual, factor (Box 1). By contrast, an American tradition emerged in which the goal was to identify as many factors as possible [4]. The American tradition advocated rotation of the various factors. Therefore, Guilford recommended an orthogonal rotation (varimax), that is, that factors may not intercorrelate [39], whereas Cartell recommended oblique rotation, in other words, permitting a certain degree of intercorrelation [40].

Table 3 shows four studies using varimax or oblique rotation on the HCL-32. In the Korean study [41], oblique rotation identified the highest number of factors, namely seven, while in the Brazilian study [42], varimax rotation identified the highest number of factors, namely nine. However, in most of these studies the scree test was used [43]. This test performs plots of eigenvalues against factor number. The resulting curve is then used to judge the cutoff point.

Out of the nine HCL-32 factors identified by Soares et al., the scree test only accepted a two-factor solution with 20 items in the first factor and seven items in the second [42]. In the Leao and Del Porto study, the scree test also accepted a two-factor solution with 16 items in the first factor and six items in the second [37]. In the oblique rotation study, the scree test accepted three of the seven factors, where the third factor only included the two items on sex-related behavior [41]. In the other oblique rotation study, the second of the three factors included the sex-related items.

The Danish study is the only HCL-32 study to use the unrotated principal component analysis [45]. However, in this study the MDQ was also tested by a principal component analysis. In the MDQ, two components were identified, namely a first principal component with positive loadings on all items and a second principal component that was bidirectional: six items had negative loadings (covering the sunny side of hypomania or active/elevated behavior) and the remaining seven items had positive loadings (covering the dark side of hypomania or irritable/risk-taking behavior). In the HCL-32 the first principal component was a general factor with positive loadings on all items, while a second principal factor had negative versus positive loadings. When compared with the Bech–Rafaelsen Mania Scale [4], the HCL-32 items could be reduced to a HCL-20 version in which ten items covered the sunny side of hypomania and the remaining ten items covered the dark side of hypomania (Table 4).

Use of IRT models (the nonparametric Mokken analysis [4]) demonstrated that the coefficient of homogeneity was 0.60 for the ten items in the sunny side of hypomania and 0.49 for the ten items in the dark side of hypomania (Table 4) [45]. However, when considering all
20 items, the coefficient of homogeneity was 0.54. The level of unidimensionality (total score a sufficient statistic, i.e., profile scores not necessary) on the coefficient of homogeneity was 0.40. Moreover, the coefficient of homogeneity for the HCL-32 was 0.53 and 0.60 for the MDQ. It was, therefore, concluded that although the HCL-32 and the MDQ contain two factors (the sunny vs the dark side of hypomania), they refer to a latent dimension of hypomania severity, implying that hypomania, as measured by the HCL-32 or MDQ, is a dimension of severity.

**The receiver operating characteristic analyses of diagnostic sensitivity & specificity**

Both the HCL-32 and the MDQ have been evaluated by use of a receiver operating characteristic analysis. In some of these studies the diagnosis against which the HCL-32 or MDQ were tested was bipolar I, while in other studies both bipolar I and II were used. If both diagnostic categories were used, they have been combined in the results in Table 3. The sensitivity and specificity calculations are often considered to be an additional type of psychometric validity [46]. Sensitivity is the proportion of cases that have a positive score and specificity is the proportion of noncases that score negatively. Most of the HCL-32 studies in Table 3 recommend a cutoff score of 14. On the MDQ the recommended cutoff score of 7 has been used. Table 3 demonstrates that HCL-32 sensitivity ranged from 0.65 to 1.00, whereas its specificity ranged from 0.41 to 0.83. The MDQ sensitivity ranged from 0.68 to 0.88, whereas its specificity ranged from 0.58 to 0.89. However, the specificity of the MDQ was considerably higher than that of the HCL-32, whereas HCL-32 sensitivity was considerably higher than that of the MDQ (Table 3).

**Clinical validity of HCL-32 & MDQ in treatment-resistant depression**

When using the MDQ as a screening test for hidden bipolarity, Calabrese et al. found that approximately 18% of patients with treatment-resistant depression were MDQ positive [47]. When comparing the validity of the HCL-32 versus the MDQ for screening of hidden bipolarity in patients with treatment-resistant depression, Dudek et al. found that even in patients who had no DSM-IV criteria of bipolarity at baseline and had not previously been treated with mood-stabilizing drugs, 14% were MDQ positive for bipolarity and 44% were HCL-32 positive for bipolarity [48]. However, the MDQ cutoff score was 6 and not the conventional 7, whereas the HCL-32 cutoff score was the conventional 14. In a recent study, Rybakowski et al. found that, among unipolar patients resistant to antidepressant treatment, approximately 44% were HCL-32 positive for bipolarity (cutoff score of 14 or more) and 26% were MDQ positive for bipolarity (cutoff score of 7 or more) [49].

In a group of unipolar depressed patients not receiving mood stabilizers [45], 18% were MDQ positive for bipolarity using the conventional cutoff of seven (similar to Calabrese et al. [47]) and 36% were HCL-32 positive for bipolarity using a cutoff score of 18.

Placebo-controlled trials have found lithium augmentation in patients with treatment-resistant depression very effective, with a cumulative response rate of 56% [50]. Therefore, there is a need for prospective trials with lithium augmentation in unipolar depressed patients with hidden bipolarity resistant to antidepressive medication [27,51].

**Clinical judgment versus questionnaires in the diagnosis of hidden bipolarity**

In its original form, the Beck Depression Inventory [52] was actually administered by a trained interviewer who read each statement aloud and asked the patient to select the statement that seemed to best fit him or her. However, this structured clinical interview soon became a patient-administered scale. As stated by Kendell, a self-administered scale eliminates many of the problems with interobserver reliability of clinical interviews [53]; however, in doing so it also creates several other problems. Therefore, there is no guarantee that different patients all associate the same meaning with the particular words they use to describe the sunny side or dark side of hypomania. Retrospectively, many patients who have been through severe manic episodes recall them as dark, depressive episodes rather than sunny highs. Others, again, do not perceive hypomania as pathological and, as such, may fail to report it [32]. Furthermore, depressed patients may have difficulty reporting hypomania owing to being unaware of mood changes [20].

Parker and his study group have declared that bipolar disorder is often undetected at the clinical interview, most commonly because the patient fails to describe any thoughts of ‘highs’,
either because of not being asked or because the interviewer sets thresholds (based on the number of symptoms) too high for a positive diagnosis of bipolarity. Parker, therefore, recommends the following items as screening questions for bipolarity in patients with major depression at the interview, in order to capture the sunny side of previous hypomanic episodes:

- Do you have times when you feel ‘wired’ or energized more than usual?
- Do you talk more and talk over people?
- Do you feel the need for less sleep but not feel tired?

And the following to capture the dark side of previous hypomanic episodes:

- Do you then spend more money and buy things you do not really need?
- Do you then say or do things that you later regret?
- If the latter question is acknowledged then the interviewer should ask about being impatient and irritated, or more disinhibited.

As irritability has been associated with unipolar depression as well as bipolar depression, Perlis et al. hypothesized that anger attacks (sudden outbursts of anger) would also be common. They found, in a sample of outpatients in a current depressive state, that bipolar patients had a significantly higher number of anger attack episodes than unipolar patients. This finding supports the benefit of inquiring about irritability and disinhibition, resulting in anger attacks among depressed patients when screening for bipolarity.

In Table 4, the Parker suggestions for a clinical judgment interview are listed with the corresponding HCL-32 or MDQ items.

In both the MDQ and HCL-32 it is the severity of the hypomanic state that is measured, and the total scores of these scales are sufficient statistics.

Clinical validity of traits predicting bipolarity

Kraepelin considered hypomanic states to be a variation of the cyclothymic predisposition to bipolarity. In his core description of cyclothymia, Kraepelin focused on patients who would constantly swing back and forth between the two opposite poles of emotion: ‘now shouting with joy to heaven, now grieved to death’. The clinically valid trait predicting schizophrenia was for Kraepelin the ‘autistic’ trait, which he placed within his concept of negativism. This autistic trait was a contrast to cyclothymia with such elements as a lack of sociability, retiring and resistance to influence.

The Hypomanic Personality Scale (Table 1) is based on the premorbid personality style with a focus on patients with predominantly hypomanic tendencies. In Table 1 the most clinically valid items have been selected.

For personality questionnaires such as the EPQ or the Hypomanic Personality Scale (Table 2) the responses to the statements have to be independent of the patient’s current mood state. Kendell and DiScipio have demonstrated that if the patients are instructed to describe their personality before the onset of their illness, they are able to disregard their current symptomatology; this is also the case with the HCL-32. However, the clinician assessing a currently depressed patient should be careful to remind the patient about time frame 4 in order to not contaminate state and trait variables.

A study on Eysenck’s trait markers (neuroticism and extraversion) found that the item ‘when considering how you would describe yourself in general, does your mood often go up and down?’ had a significant predictive validity for depression in patients with chronic idiopathic pain disorder.

The very important clinimetric issue of how to open the dialog with a patient when measuring hidden bipolarity might be all three ‘neuroticism’ items in Table 1. They all have a high degree of applicability. From the results found by Bech et al., the most appropriate opening item is: ‘are you sometimes bubbling over with energy and sometimes very sluggish?’ This
can then be followed, as in the HCL-32 or the Parker approach, by ‘let us now consider a period when your energy is bubbling over’ [54].

As Angst advocates, caution should be taken regarding normal and pathological mood swings [3]. As temperament characterizes a normal state, its elements cannot serve as diagnostic criteria for psychopathology. That said, as shown above, traits in bipolar depression may very well differ from traits in unipolar depression, and could be promising markers with which to differentiate between unipolar depression and bipolar depression in an individual already presenting with symptoms of depression.

Conclusion & future perspective
Clinicians should be warned not to set thresholds too high for a positive diagnosis of bipolarity, as evidence concludes that a high proportion of treatment-resistant depressed patients may be bipolar patients.

The clinical content judgment of the experienced psychiatrist/psychologist in tandem with the assessment of clinically validated unidimensional constructs is considered good clinical practice. Inaccurate and lengthy tests (to boost psychometric performance) with questionable or no clinical validity should be avoided.

When assessing a depressed individual, focusing on the current state (time frame 1) may not facilitate screening for bipolar depression, as no consistent differences have been found in the acute phase.

Evidence points to the benefit of focusing on the time frame between episodes (time frame 4) in this screening. When inquiring about hypomanic episodes in time frame 4, the duration criterion does not display clinical relevance and we suggest, as others have done, disregarding this criterion.

As Angst argues, inquiring about mood changes necessitates the subject’s awareness of a mood change, which is not always the case [3]. He adds that overactivity seems to be just as endorsing as an elevated mood in regard to the underlying concept of bipolarity, and recommends to include ‘increased activity’ when assessing for hypomanic episodes because people are more aware of their behavioral changes than their mood changes. Therefore, we suggest searching for hidden bipolarity in depressed patients by opening with Eysenck’s neuroticism question ‘are you sometimes bubbling over with energy and sometimes very sluggish?’

In addition to considering time frame 4, fluctuating mood has drawn attention as a risk factor for developing bipolar depression; this is echoed in a contemporary cognitive model on bipolar depression [62]. In this model, the individual changes in internal state are involved in triggering a further escalation of mood change through the attempt to control mood using different descent or ascent behaviors. As such, as long as a person can remain within a relatively moderate range of mood change, there is no risk of triggering an extreme mood swing. However, individuals experiencing fluctuating moods in combination with behavior change (ascent behaviors) are at risk of exacerbating their mood and fueling the spiraling process towards mania.

As Angst et al. have argued, evidence supports a dimensional view of the affective spectrum from ‘normal’ mood swings to ‘abnormal’ mood swings [20,32]. For future development of the treatment of depression, assessing the tendency of mood swings as a trait (time frame 4), as well as the typically chosen mood control behaviors, may be of value. Future research may even find trait differences between unipolar and bipolar depression in regard to the intensity with which individuals respond behaviorally and affectively to different classes of stimuli and life events [63].

In identifying mood swings as a risk factor for developing bipolar disorder, differentiating between ordinary mood swings and pathological mood swings becomes important. Scales measuring trait mood swings should, therefore, collect and provide norms for relevant sample types.

In this review we have considered the HCL-32, which is based on Angst’s vast experience with bipolar disorder, as an excellent model for identifying hidden bipolarity. However, in our clinical analysis of trials in which the HCL-32 was compared with the MDQ we have found no major differences when used in a population of patients with major depressive disorder. In daily clinical practice with these patients the clinician-rated version made by Parker might have high applicability, especially when modified by items from the HCL-32, MDQ or EPQ as suggested [54,55].

The great therapeutic impact of identifying hidden bipolarity is in depressed patients who are resistant to antidepressants by the addition of a mood stabilizer. Lithium is the most specific drug treatment in psychiatry, selectively acting in patients with bipolar depression.
Comparative validity of inventories & checklists for identifying depressed patients with hidden bipolarity

**Financial & competing interests disclosure**

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

**References**

Papers of special note have been highlighted as:
- of interest
- of considerable interest


The final version of the Hypomania Checklist (HCL-32) is released in this article, and results indicate the high applicability of this rather extensive questionnaire for bipolarity in diverse study centers around the world.

The Mood Disorder Questionnaire (MDQ) is presented as a brief, easy-to-administer screening questionnaire for bipolar disorder. Time has shown its high clinical utility.
Both the HCL-32 and the MDQ were psychometrically analyzed by principal component analysis (unrotated) and an item response theory model. Results indicate that both questionnaires are clinically valid.

Both the HCL-32 and the MDQ were psychometrically analyzed by principal component analysis (unrotated) and an item response theory model. Results indicate that both questionnaires are clinically valid.

Both the HCL-32 and the MDQ were psychometrically analyzed by principal component analysis (unrotated) and an item response theory model. Results indicate that both questionnaires are clinically valid.
Comparative validity of inventories & checklists for identifying depressed patients with hidden bipolarity


