Title: Hypomania Checklist-32 - cross-validation of shorter versions screening for bipolar disorders in an epidemiological study.

Authors: Meyer TD, Castelao E, Gholamrezaee M, Angst J, Preisig M

Journal: Acta psychiatrica Scandinavica

Year: 2017

Volume: 135

Issue: 6

DOI: 10.1111/acps.12715

Hypomania Checklist-32 – Cross-validation of shorter versions screening for bipolar disorders in an epidemiological study

Thomas D Meyer, PhD [1], *Enrique Castelao, MSc* [2], *Mehdi* Gholamrezaee , PhD [2], *Jules Angst, MD* [3], & Martin Preisig, MD [2]

Accepted, uncorrected version (to be published in Acta Psychiatrica Scandinavica)

[1]*University of Texas Health Science Center at Houston, Department of Psychiatry and Behavioral Sciences, Houston, TX, USA*

[2]Center for Psychiatric Epidemiology and Psychopathology, Department of Psychiatry, Lausanne University Hospital, 1008 Prilly, Switzerland:

[3]Department of Psychiatry, Psychotherapy and Psychosomatics, Psychiatric Hospital, University of Zurich

Correspondence to: Thomas D Meyer, University of Texas HSC at Houston, Department of Psychiatry and Behavioral Sciences, 1941 East Rd (BBSB 3118), Houston, TX 77054, USA; Email: Thomas.d.meyer@uth.tmc.edu, phone +1 (713) 485-2759

Abstract

Objective: Self-reports such as Hypomania Checklist (HCL-32) can be used to enhance recognition of bipolar disorders but they are often too long and only validated in clinical samples. The objectives of this study are therefore to test if a) the HCL-32 can be used for screening in the community and b) whether two previously suggested shorter versions would do as well.

Method: Data stemmed from the CoLaus/PsyColaus, a prospective cohort study including randomly selected residents aged 35 to 66 years from an urban area. Participants underwent semi-structured interviews to assess DSM-IV disorders and 1712 of them completed the HCL-32).

Results: Forty individuals (2.3%) were diagnosed as having BD. Compared to others, participants with BD scored significantly higher on the HCL-32. The HCL-32 had a sensitivity of .78 and specificity of .68. Very similar figures were found for two previously proposed shorter versions with 16 and 20 items. The results of Confirmatory Factor Analysis and Item Response Theory (IRT) models supported the postulated two-factor structure for the three HCL versions.

Conclusion: Despite the low base rate of BD in this sample the screening properties of the HCL-32 remained almost as good. Importantly, two previously proposed shorter versions performed as well, suggesting that those could be used without losing essential information.

Keywords: *screening; bipolar disorders; depression; epidemiology; Hypomania Checklist*

Significant outcomes

- The HCL-32 showed almost as good screening properties in this epidemiological sample as in prior clinical studies
- Two previously suggested shorter versions with 16 and 20 items do as well as the original longer version with 32 items and therefore could be used in clinical practice.
- Future studies should examine if even shorter versions can be created or practical clinical algorithms can be developed based on the existing measure.

Limitations

- Unfortunately some participants had to be excluded from the analyses because of a printing error exposing them only to one of the two pages of the HCL-32.
- We only cross-validated two previously suggested shorter versions but there might be others out there.

While research has moved forward towards early identification and treatment of bipolar disorders (BD) (e.g. 1, 2, 3), diagnosing BD is still a complex process in individuals with fully syndromal expressions. It requires looking beyond the current presentation, which mostly is of a depressive nature, taking into account its longitudinal course (e.g. 4). In many cases the appropriate diagnosis is made with a long delay which is not just due to the first onset, often being a depressed episode (5-7). Other reasons referred to in the literature are that the patients do not spontaneously report periods which are subjectively dominated by good mood, high energy levels or decreased need for sleep (e.g. 8), or that there are no distinct boundaries between normal and abnormally elevated mood or energy levels; the latter leaves room for individual interpretation of symptoms and use of heuristics which can affect diagnostic decisions (e.g. 9, 10, 11). Last but not least, clinicians, especially under pressure of time, might not specifically probe for mania or hypomania in patients (12).

One way to address this issue of potential under-recognition of BD is the use of patient-rated screening tools that are both time- and cost-sensitive. Since this can be done between consultations or in the waiting room, it decreases the time needed during consultations especially if the result of the screening is 'negative'. Because mania and hypomania are in effect the defining features for a DSM diagnosis of BD (13), only a screening of a potential lifetime history of episodes of those symptoms is needed.

While several instruments have been developed and evaluated, such as the Mood Disorder Questionnaire (MDQ; 14) and the Bipolar Spectrum Diagnostic Scale (BSDS; 15), the Hypomania Checklist-32 (HCL-32; 16) seems to be the one that has been translated in more languages than any other (11, 17) and its popularity is still growing (e.g. 18, 19). Meyer et al. (20) published a systematic review on the screening properties of the HCL-32 and found an average sensitivity

of 80%, regardless whether the comparison group was consisting of all patients or only unipolar depressed patients. Specificity was slightly higher with an estimated average of 65% with respect to the unipolar depressed groups than when patients with all kind of diagnoses were considered (57%). Meyer et al. (2014) pointed out that while most studies screened patients with or without a confirmed prior diagnosis (e.g. 21, 22, 23), some of those studies followed a different approach by excluding all patients with an established diagnosis of BD which is a much more conservative approach (e.g. 24, 25, 26). One unanswered question is, however, whether the HCL-32 would be able to identify BD in the general population.

It is relevant to know that evidence based guidelines such as the one from the National Institute for Health and Care Excellence (27) do not recommend screening tools, such as the HCL-32, because of its length. Only very few studies have tried to reduce the length of the published HCL-32. One such study was provided by Bech et al. (28), which proposed a HCL-20 based on their item response analysis with an optimal cut-off score of 10. However, their sample was quite small with 59 patients with BD-I and 63 with MDD, and the MDD group was also significantly older than the patients with BD. Forty et al. (29) used a prior data set which led to the development of a 16-item HCL version and then evaluated this in a separate sample. From primary and secondary services, they recruited 59 patients with BD (BD-I: n=28; BD-II: n=23, Bipolar-NOS: n=8) and 76 patients with MDD who were considered in clinical recovery. In their study, a score of 8 and higher was identified as optimal for screening purposes with this 16-item version. Although such briefer screening tools would be appreciated in clinical practice, cross-validations of these shorter versions do not exist yet.

Aims of this study: a) Using an epidemiological sample, the CoLaus/PsyCoLaus study (30), we examine whether the screening properties of the original HCL-32 in terms of sensitivity,

specificity, as well as positive and negative predictive power are comparable to those of prior studies, and b) cross-validating the cut-offs of the before mentioned 16- and 20-item versions of the HCL for the very first time in an independent sample. Additionally we investigated the psychometric properties of the three HCL versions using Cronbach's α coefficient, confirmatory factor analysis and Item Response Theory.

Method

Participants

The data of the present paper stemmed from CoLaus¦PsyCoLaus study, a cohort study designed to prospectively study mental disorders and cardiovascular risk factors (CVRF) in the general population. The original sample of 6733 individuals was randomly selected from the 35 to 75-year-old residents of the city of Lausanne (Switzerland) from 2003 to 2006, according to the civil registry (30, 31). Among these subjects, 3721 in the age range of 35 to 66 years also participated in the psychiatric assessment. A first follow-up of the cohort took part between 2010 and 2013. Among the 2582 who participated in both the psychiatric baseline and follow-up investigations, 1712 completed the HCL-32 questionnaire and could be included in the present analysis. The Institutional Ethics' Committee of the University of Lausanne approved the CoLaus|PsyCoLaus study. All participants signed a written informed consent after having received a detailed description of the goal and funding of the study. Their mean age was 59.70 years (SD = 10.22, range 41-84). Women were slightly overrepresented with 57.5% (n = 1485 of 2582). More than half (53.8%, n = 1388) did not have any lifetime or current history of a mood disorder. Forty three and a half percent (n = 1122) reported a current or lifetime history of major depressive disorder (MDD), 1% received a diagnosis of bipolar I disorder (BD-I; n = 27),

0.6% of bipolar II disorder (BD-II, n =25) and 1.2 % of BD NOS (n = 30). Women were equally likely to have a history of BD, but more likely to report a history of major depression than males ($\chi^2_{(2)}$ = 79.05, p < .001). 93.2 % of the sample was euthymic at the time of assessment, 6.5% were currently in a major depression episode and 0.3 % were currently (hypo)manic.

We evaluated if this could introduce a systematic bias for our results and, therefore, compared those with and without HCl-32 data. Those individuals we were able to include in our analyses did not significantly differ from the 870 excluded cases with respect to age ($t_{(1679.63)}$ = 1.72, p= .09) or in their current mood (HCL item 1) on the day of the assessment ($t_{(1470)}$ = 0.68, n.s.). Gender distribution ($\chi^2_{(1)}$ = 1.23, n.s.), rates of alcohol use disorder ($\chi^2_{(1)}$ = 1.27, n.s.) or substance use disorder were equivalent, as well ($\chi^2_{(1)}$ = 0.03, n.s.). There was, however, a significant difference with respect to diagnostic status. While 65% of individuals without any mood disorders and 68.6% of subjects with major depression had sufficient HCL data, this rate was lower with 55.6% in the group with BD ($\chi^2_{(2)}$ = 7.51, p < .05). The final sample of 1712 consisted of 2.4 % individuals with BD (n = 40; BD-I: n = 16, BD-II: n = 8, BD-NOS: n = 16), 45% individuals with a current or former history of major depression (n = 770) and 52.6% without a history of any mood disorder (n = 902) and is described in Table 1.

Measures

Diagnostic Interview for Genetic Studies (DIGS); 32)

Diagnostic information on mental disorders was collected using the French version of the DIGS (33), a well-known and widely used semi-structured interview. The French version of the DIGS revealed excellent inter-rater reliability in terms of kappa and Yule's Y coefficients for major mood and psychotic disorders, whereas the 6-week test-retest reliability was slightly

lower (33, 34). The DIGS also assesses hyperthymic personality according to the criteria of the modified Research Diagnostic Criteria, which required periods of elation or excitement lasting most of the time (chronic form) and resulted in: 1) subject communicated with a close friend or relative on how he/she felt or 2) someone complained or commented on some manifestation of this condition (35). Trained psychologists conducted these semi-structured interviews, and psychiatric diagnoses were assigned according to the *Diagnostic and Statistical Manual of Mental Disorders, fourth edition* (DSM-IV, 36).

Hypomania Checklist 32 (HCL-32) (16)

This instrument consists of 32 items for the self-assessment of hypomania. People are asked to remember "a period when you were in a 'high' [mood]" and to indicate whether specific behaviors, thoughts, or emotions were present in such a state, e.g. "I need less sleep", "I am less shy or inhibited". The questionnaire also includes items about the duration of such "highs". Furthermore, people are asked to rate the impact of such "highs" as "positive & negative", "positive", "no impact", or "negative" for family life, social life, and work. Additionally, other people's reactions and comments on such episodes were assessed (positively, no comments, or negatively) (for review:20). The checklist was first conceptualized in German by J. Angst and T.D. Meyer, and then translated in many languages. The two shorter versions, 20-items (28) and 16-items (29), are based on the 32-item version.

Statistical analyses

For group comparisons we conducted t-tests (two tailed) or analysis of variance or cross-tabulations, where appropriate. As indicators of the screening properties, we ran ROC analyses and calculated Sensitivity (SE; i.e. the probability of a positive screening among those with a diagnosis, often interpreted as the percentage of correctly identified cases), Specificity (SP; i.e.

the probability of a negative test result among those without a diagnosis, often interpreted as the percentage of correctly identified non-cases), Positive Predictive Power (PPP; i.e. the probability that those who screen positive have a diagnosis) and Negative Predictive Power (NPP; i.e. the probability that those who screen negative do not have a diagnosis) (37).

In order to investigate the psychometric properties of the three HCL versions confirmatory factor analysis and Item Response Theory were performed using lavaan (38) (URL http://www.jstatsoft.org/v48/i02/.) and mirt packages of R, respectively (39). For confirmatory factor analysis, Comparative and Non-Normed Fit Indices (CFI and NNFI) are provided as fit index, for both a fit index close to 1 suggests a good fit.

Unidimensionality, local independence, and latent monotonicity of three HCL version scales (and their corresponding dimensions) were investigated using H coefficient of homogeneity implemented in Mokken package of R (40) [http://www.jstatsoft.org/v20/i11]. If H coefficients are between 0 and 1 (for all item-pairs, items and total score) then the three mentioned assumptions seem to hold (41). As suggested by Mokken (42) a total H of at least 0.4 has been used as an indication of strong unidimensionality.

Results

Before evaluating the screening properties we wanted to describe the sample a little more and compare groups. Comparisons were made among individuals with BD to the individuals with a current or former history of major depressive disorder (MDD) and those without a history of any mood disorder in various variables (Table 1). The three groups significantly differed in age ($F_{(2, 1711)} = 22.53$, $p < .001$) with post hoc Student Newman-Keuls tests revealing that both patient groups were significantly younger than the individuals without a mood disorder. When controlling for the unequal sex distribution in the three groups, the age

differences persisted ($F_{(2, 1711)}$ = 24.32, p< .001). The groups did not, however, differ in current mood on the day completing the HCL-32 ($F_{(2,1645)}$= 1.07, p = .34), and this was still the case after controlling for sex and age ($F_{(2,1645)}$= 1.68, p = .19). Both mood disorder groups had significantly higher rates of SUD ($\chi^2_{(2)}$ = 57.04, p < .001) and AUD ($\chi^2_{(2)}$ = 46.25, p < .001) than the non-mood disorder group (and also differed from each other) . This was also the case if analyzed separately for men and women (results available on request)

For comparison with prior studies the mean score for the HCL-32 was 10.99 (SD = 5.77, range 0 - 30). The equivalent values for Bech et al.'s (28) 20-item version were M= 7.05 (SD = 3.56), and for Forty et al.'s (29) 16-item version M = 5.09 (SD = 2.86). Age was slightly negative, but significantly related to all three HCL sum scores (n = 1712: HCL-32: r =-.15; HCL-20: r=-.13; HCL-16: r=-.15, all p <.001). Slightly positive, but significant correlations were observed for current mood as assessed by the HCL-32 (n = 1646: HCL-32: r =.11; p <.001). No differences were found for any HCL-version between the two genders (all t<|1.39|).

The HCL-32 also asks whether the reported "highs" had any positive and/or negative impact on the areas of family, social life, work, or leisure. Using these areas to determine if there was any impact on their lives, regardless if positive or negative, all of the individuals with BD (100%, n = 39) reported some form of impact, while the respective figures in the group with MDD disorder and without a mood disorder were 86.4% and 78.9% respectively ($\chi^2_{(2)}$ = 24.12, p < .001). Running these analyses separately for men and women led to the same result (available on request). If one only looks at negative impact, the numbers drop but the groups still differ with 35.9%, 19.7% and 16.7% respectively ($\chi^2_{(2)}$ = 11.82, p < .01). However, when looking at men and women separately, only the women with BD reported significantly higher rates of negative consequences (available on request). Negative reactions from others to those "highs"

were reported by almost a quarter of participants with BD (23.8%) but less than 10% of the other two groups (MDD: 9.5%, no mood disorder: 8.9%).

As expected, the three groups (MDD, BD and non-mood disorder group) significantly differed in all three versions of the HCL (HCL-32: $F_{(2, 1711)} = 76.59$, $p < .001$; HCL-20: $F_{(2, 1711)} = 70.81$, $p < .001$; HCL-16: $F_{(2, 1711)} = 80.18$, $p < .001$). Controlling for sex and age in these analyses left the results virtually unchanged (HCL-32: $F_{(2, 1711)} = 70.44$, $p < .001$; HCL-20: $F_{(2, 1711)} = 65.55$, $p < .001$; HCL-16: $F_{(2, 1711)} = 61.47$, $p < .001$). On a side note, the covariate sex was only significant for the HCL 32 ($F_{(1, 1711)} = 5.14$, $p < .05$) but not the shorter versions (HCL-20: $F_{(1, 1711)} = 3.15$, $p = .08$; HCL-16: $F_{(1, 1711)} = 0.01$, $p = .91$). Post hoc comparisons showed that all three comparisons were statistically significant with the participants with BD scoring highest and those without a mood disorder scoring lowest. We also calculated the two factor scores originally identified by Angst et al. (16) and found that individuals with BD had the highest score on both the active/elated ($F_{(2, 1711)} = 62.01$, $p < .001$) and the risk-taking/irritable factor ($F_{(2, 1710)} = 22.13$, $p < .001$), and all group means comparisons were significant, i.e. the group with MDD scored significantly lower than individuals with BD on both subscales, but also higher than the controls. Controlling for sex and age in these analyses did not change the results (available on request).

Screening properties for bipolar disorder (BD) of the HCL with 32, 20 and 16 items

Regarding the screening properties of the HCL and its shorter versions, we ran the analyses for three comparisons, i.e. BD versus all other participants, BD versus MDD and BD versus those without a mood disorder. The results are presented in Table 2 and include the optimal cut-off scores based on the CoLaus¦PsyColaus study sample as well, although the cross-validation is more focused on application of the formerly established cut-off scores. For the

HCL-32 and HCL-16 the two cut-off scores are only one point apart, but for the HCL-20 the optimal cut-off score in the present sample would be three points lower. With regards to the optimal cut-off scores, the HCL-32 marginally outperforms the other two versions (see also Figure 1 and 2). Looking at the original cut-off scores the HCL-32 and HCL-16 perform similarly with regards to SE and SP, while the HCL-20 has a lower SE but higher SP than both of the others. [1]

A proportion of 3.8% of the sample met criteria for hyperthymic personality. A hyperthymic personality was reported by 3.6% in the group without a mood disorder, 2.9% in individuals with MDD, and 21.1% in individuals with a bipolar disorder ($\chi^2_{(2)}$ =60.74, p < .001). The presence or absence of self-rated hyperthymia was not associated with the HCL-32 scores (AUC = .57, 95% CI .48-.66), i.e. suggesting that HCL-32 is not screening for a temperament factor. The same was true for the HCL-20 (AUC = .55, 95% CI .46-.64) and HCL-16 (AUC = .56, 95% CI .48-.65).

Psychometric properties of the three HCL versions

Cronbach's α coefficients as a measure of internal consistency were .86 for the HCL-32, .78 for Bech et al.'s (28) 20-item version, and .73 for Forty et al.'s 16-item version (29). Evaluating dimensionality, we first estimated the H coefficient (homogeneity coefficient of Mokken) following Bech et al approach (28). For the HCL-32 this resulted in H = 0.296 (SE=0.010). Since a coefficient higher than 0.4 shows strong unidimensionality, we did not find evidence for strong unidimensionality for the original full length HCL. The same was true for versions with HCL-20 (H = 0.272 [SE = 0.011]) and HCL-16 (H = 0.281 [SE = 0.011]).

Exploratory and confirmatory factor analysis of the HCL-32 supported, however, the two-factor structure postulated by Angst et al. (16) with an 'active/elated' and a 'risk-taking/irritable' factor (CFI = 0.94 , NNFI = 0.94). Similarly, the two-factor structure could be confirmed for the 20-item version as suggested by Bech et al. (28) (CFI = 0.92 , NNFI = 0.91) and for the 16-item version (CFI = 0.96 , NNFI = 0.95). For the latter version we assumed that the remaining items would load on the same factors as in the original 32-item version given that no specific latent structure was proposed in the literature for this version.

The multidimensional two Parameter Logistic (2PL) IRT models were employed to further investigate the dimensionality of the three HCL versions. Based on results of the fitted multidimensional IRT models, the estimated difficulty and discrimination parameters from all three models were very similar which suggests that the three versions of HCL are similar in representing the latent construct of the questionnaire (analyses available upon request).

Discussion

While almost all prior studies have used clinical samples to look at the screening properties of the HCL-32 (20), we used, for the very first time, an epidemiological sample. First, although the HCL-32 (16) was developed to identify potential bipolarity in depressed patients, it is actually of high clinical interest to know whether the HCL-32 could be used more routinely in a sample outside of mood disorder clinics. The latter is especially informative since there is already some data out there that none of the existing screening measures for BD do very well in samples abusing substances (43). Second, we wanted to cross-validate two shorter versions of the HCL, one with 20 (28), and one with 16 items (29) because this would increase the likelihood of it being used and recommended for clinical practice (NICE, 2014).

Comparing the results from this epidemiological study to clinical studies, there is similarity in results such as comparing means between groups or the lack of meaningful gender differences, but contrary to most other studies, we found significant associations between HCL-scores and current mood or age (cf.22, 23, 44, 45). Nevertheless, looking at the effect sizes, they are small and probably only reached significance because of the sample size. Of more importance is that the HCL-32 showed a sensitivity of .78 and specificity of .68 in this epidemiological sample using the originally published cut-off score similar to one estimated average values across clinical studies, as presented by (20). This could not necessarily be expected given the low base rate which definitely affected the low PPP (see: 46, 47).

The other goal was to cross-validate proposed shorter versions of the HCL-32 because the two selected versions have only been used in one single study. The 20-item version from Bech et al. (28) was chosen based on their item response analysis (Mokken analysis) that was able to show that unidimensionality of the HCL-20 can be assumed and that the total score is a valid statistic. Using Mokken's analysis and IRT we could not replicate their finding of strong unidimensionality of the HCL-20 and the other versions, but we were able to replicate the previously proposed two factor structure by Angst et al. (16) for all three HCL versions, and their corresponding latent structure can be modeled satisfactorily. This provides further evidence that the underlying latent structure is highly similar in clinical and community samples.

Bech et al (28) had identified 10 or more for the HCL-20 as an optimal cut-off score, but they had only included patients with BD-I (n=59) and MDD (n=63), which also significantly differed in age. The 16-item version by Forty et al. (29) was also psychometrically developed using the data from a prior study (see: 48), and then they evaluated this abbreviated scale in a separate

sample. From primary and secondary services they had recruited 59 patients with BD and 76 patients with MDD, who were considered in clinical recovery. Both studies showed rigor in developing shortened scales but both had not been cross-validated in other samples before our study. We were able to show both versions are doing equally as well as the original HCL with 32-items when it comes to sensitivity and specificity, especially the HCL-32 and HCL-16 when using the originally published cut-off scores as criterion. As an additional cross-validation we re-analysed the data from the clinical sample recruited by Meyer et al. (22), which also shows that the HCL-20 and 16 can be used to screen for BD with a sensitivity of .82 and .81 and a specificity of .85 and .84, respectively (unpublished data).

Contrary to that clinical sample from Meyer et al. (22) the data from the CoLaus¦PsyCoLaus study show the typically observed higher sensitivity than specificity for all the HCL versions. Especially in non-clinical samples this is likely to be associated with a higher rate of false positive cases but given a) the clinical significance of unrecognised and untreated BD and b) the original aim of the HCL-32 to be especially sensitive to bipolar II disorder (16), this seems justified. Furthermore, while none of the HCL versions should be used, in itself, to diagnose a lifetime history of (hypo)mania or even BD and ideally sensitivity should be above 90% (49), the use of such a tool will probe further questions from clinicians making it more likely to recognize potential bipolar spectrum disorders in a timely manner. However, we also acknowledge that the present study has some limitations. First of all, only for a subsample, complete HCL-32 data were available due partially to a printing error; i.e. a substantial proportion of the participants only received the first page of the HCL, making them ineligible to be included in the analyses. While this subsample did not differ from the main sample with regards to most variables such as age, gender or comorbid SUD, this subsample had a lower

rate of individuals diagnosed with BD. This is unfortunate but the lower base rate of BD did not result in a drop in SE or SP, which supports the validity of the scale. Second, the representativeness of the sample was limited with regards to younger people because of the way the original study was planned. Therefore it could be that certain characteristics such as substance use disorders are less prevalent in this epidemiological sample. Third, unfortunately we do not know how many of the participants had been identified and treated as having BD prior to the study or were identified during the diagnostic interview but this can affect sensitivity and specificity and potentially induce a bias. However, although this does not mean that they have been diagnosed with BD, we know that 86.8% of the patients with BD-I and 92.9% of those with BD-II had sought help for their problems before study entry (50). Last but not least, as reported above we were not able to show unidimsionality for the HCL-32, HCL-20 or HCL-16, and our results suggest that two factors better describe their latent structure. On the one hand this questions the use of a single composite score for screening purposes, but on the other hand all prior studies relied on that single sum score to determine sensitivity and specificity of the cut-off scores. While exploratory analyses of this data set suggested for the original HCL-32 that the 'active/elated' factor does better with respect to AUC (see Footnote 1), we do not know if this would apply to clinical or help-seeking samples as well. Future studies should evaluate if the subscales on their own or in combination with additional information from the HCL (e.g. comments or negative reactions from others, self-reported impacts of the 'highs' on life) improve screening properties and comply with psychometric scale properties.

In summary, the HCL proved to be a sufficiently sensitive and specific tool to screen for BD in this epidemiological and essentially non-clinical study. This is important because almost all

prior studies recruited clinical samples, especially from mood disorders specialty settings (20). We conclude that the HCL-32, HCL-20 and HCL-16 are fairly equivalent in screening for BD. It is probably premature to recommend one specific shorter version of the HCL because even briefer untested ones are an option, for example the 10 overlapping items in the HCL-20 and HCL-16, but, in general, we think it would be good if researchers reanalyze their existing data to identify the optimal, cross-culturally valid shortened version. Only by providing more evidence that a briefer version of the HCL works as well as the original one (16) will increase the likelihood that it will be considered useful in clinical practice.

References

1. Correll CU, Olvet DM, Auther AM, Hauser M, Kishimoto T, Carrión RE, et al. The Bipolar Prodrome Symptom Interview and Scale–Prospective (BPSS-P): description and validation in a psychiatric sample and healthy controls. Bipolar disorders. 2014;16(5):505-22.
2. Duffy A, Jones S, Goodday S, Bentall R. Candidate risk indicators for bipolar disorder: Early intervention opportunities in high-risk youth. International Journal of Neuropsychopharmacology. 2015:pyv071.
3. Pfennig A, Correll CU, Marx C, Rottmann-Wolf M, Meyer TD, Bauer M, et al. Psychotherapeutic interventions in individuals at risk of developing bipolar disorder: a systematic review. Early intervention in psychiatry. 2014;8(1):3-11.
4. Goodwin FK, Jamison KR. Manic-depressive illness: bipolar disorders and recurrent depression: Oxford University Press; 2007.
5. Drancourt N, Etain B, Lajnef M, Henry C, Raust A, Cochet B, et al. Duration of untreated bipolar disorder: missed opportunities on the long road to optimal treatment. Acta Psychiatrica Scandinavica. 2013;127(2):136-44.
6. Scott J, Leboyer M. Consequences of delayed diagnosis of bipolar disorders. L'Encéphale. 2011;37:S173-S5.
7. Lish JD, Dime-Meenan S, Whybrow PC, Price RA, Hirschfeld RM. The National Depressive and Manic-depressive Association (DMDA) survey of bipolar members. Journal of affective disorders. 1994;31(4):281-94.
8. Hirschfield R, Vornik LA. Recognition and diagnosis of bipolar disorder. The Journal of clinical psychiatry Supplement. 2004;65(15):5-9.
9. Wolkenstein L, Bruchmüller K, Schmid P, Meyer TD. Misdiagnosing bipolar disorder—do clinicians show heuristic biases? Journal of affective disorders. 2011;130(3):405-12.
10. Meyer F, Meyer TD. The misdiagnosis of bipolar disorder as a psychotic disorder: some of its causes and their influence on therapy. Journal of affective disorders. 2009;112(1):174-83.
11. Angst J, Meyer TD, Adolfsson R, Skeppar P, Carta M, Benazzi F, et al. Hypomania: a transcultural perspective. World Psychiatry. 2010;9(1):41-9.
12. Hutton C, Gunn J. Do longer consultations improve the management of psychological problems in general practice? A systematic literature review. BMC health services research. 2007;7(1):1.
13. Association AP, Association AP. Diagnostic and statistical manual of mental disorders (DSM). Washington, DC: American psychiatric association. 1994:143-7.
14. Hirschfeld RM, Williams JB, Spitzer RL, Calabrese JR, Flynn L, Keck Jr PE, et al. Development and validation of a screening instrument for bipolar spectrum disorder: the Mood Disorder Questionnaire. American Journal of Psychiatry. 2000;157(11):1873-5.
15. Ghaemi SN, Miller CJ, Berv DA, Klugman J, Rosenquist KJ, Pies RW. Sensitivity and specificity of a new bipolar spectrum diagnostic scale. Journal of affective disorders. 2005;84(2):273-7.
16. Angst J, Adolfsson R, Benazzi F, Gamma A, Hantouche E, Meyer TD, et al. The HCL-32: towards a self-assessment tool for hypomanic symptoms in outpatients. Journal of affective disorders. 2005;88(2):217-33.
17. Gamma A, Angst J, Azorin JM, Bowden CL, Perugi G, Vieta E, et al. Transcultural validity of the Hypomania Checklist–32 (HCL-32) in patients with major depressive episodes. Bipolar disorders. 2013;15(6):701-12.
18. Lee K, Oh H, Lee E-H, Kim JH, Kim J-H, Hong KS. Investigation of the clinical utility of the hypomania checklist 32 (HCL-32) for the screening of bipolar disorders in the non-clinical adult population. BMC psychiatry. 2016;16(1):1.

19. Mosolov S, Ushkalova A, Kostukova E, Shafarenko A, Alfimov P, Kostyukova A, et al. Validation of the Russian version of the Hypomania Checklist (HCL-32) for the detection of Bipolar II disorder in patients with a current diagnosis of recurrent depression. Journal of affective disorders. 2014;155:90-5.
20. Meyer TD, Schrader J, Ridley M, Lex C. The Hypomania Checklist (HCL)—Systematic review of its properties to screen for bipolar disorders. Comprehensive psychiatry. 2014;55(5):1310-21.
21. Bschor T, Angst J, Azorin J, Bowden C, Perugi G, Vieta E, et al. Are bipolar disorders underdiagnosed in patients with depressive episodes? Results of the multicenter BRIDGE screening study in Germany. Journal of affective disorders. 2012;142(1):45-52.
22. Meyer TD, Bernhard B, Born C, Fuhr K, Gerber S, Schaerer L, et al. The Hypomania Checklist-32 and the Mood Disorder Questionnaire as screening tools—going beyond samples of purely mood-disordered patients. Journal of Affective Disorders. 2011;128(3):291-8.
23. Vieta E, Sanchez-Moreno J, Bulbena A, Chamorro L, Ramos J, Artal J, et al. 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. 2007;101(1):43-55.
24. Chou CC, Lee IH, Yeh TL, Chen KC, Chen PS, Chen WT, et al. Comparison of the validity of the Chinese versions of the Hypomania Symptom Checklist-32 (HCL-32) and Mood Disorder Questionnaire (MDQ) for the detection of bipolar disorder in medicated patients with major depressive disorder. International journal of psychiatry in clinical practice. 2012;16(2):132-7.
25. Hu C, Xiang Y-T, Ungvari GS, Dickerson FB, Kilbourne AM, Si T-M, et al. Undiagnosed bipolar disorder in patients treated for major depression in China. Journal of affective disorders. 2012;140(2):181-6.
26. Poon Y, Chung K-F, Tso K-C, Chang C-L, Tang D. The use of Mood Disorder Questionnaire, Hypomania Checklist-32 and clinical predictors for screening previously unrecognised bipolar disorder in a general psychiatric setting. Psychiatry research. 2012;195(3):111-7.
27. Excellence NIfHaC. Bipolar disorder: assessment and management (CG185). https://wwwniceorguk/Guidance/CG185. 2014.
28. Bech P, Christensen E, Vinberg M, Bech-Andersen G, Kessing L. From items to syndromes in the Hypomania Checklist (HCL-32): psychometric validation and clinical validity analysis. Journal of affective disorders. 2011;132(1):48-54.
29. Forty L, Kelly M, Jones L, Jones I, Barnes E, Caesar S, et al. Reducing the Hypomania Checklist (HCL-32) to a 16-item version. Journal of affective disorders. 2010;124(3):351-6.
30. Preisig, Waeber G, Vollenweider P, Bovet P, Rothen S, Vandeleur C, et al. The PsyCoLaus study: methodology and characteristics of the sample of a population-based survey on psychiatric disorders and their association with genetic and cardiovascular risk factors. BMC psychiatry. 2009;9(1):1.
31. Firmann M, Mayor V, Vidal PM, Bochud M, Pécoud A, Hayoz D, et al. The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. BMC cardiovascular disorders. 2008;8(1):1.
32. Nurnberger JI, Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J, et al. Diagnostic interview for genetic studies: rationale, unique features, and training. Archives of general psychiatry. 1994;51(11):849-59.

33. Preisig M, Fenton BT, Matthey ML, Berney A, Ferrero F. Diagnostic interview for genetic studies (DIGS): inter-rater and test-retest reliability of the French version. Eur Arch Psychiatry Clin Neurosci. 1999;249(4):174-9.
34. Berney A, Preisig M, Matthey ML, Ferrero F, Fenton BT. Diagnostic interview for genetic studies (DIGS): inter-rater and test-retest reliability of alcohol and drug diagnoses. Drug Alcohol Depend. 2002;65(2):149-58.
35. Gershon ES, Hamovit J, Guroff JJ, Dibble E, Leckman JF, Sceery W, et al. A family study of schizoaffective, bipolar I, bipolar II, unipolar, and normal control probands. Arch Gen Psychiatry. 1982;39(10):1157-67.
36. Association AP. Diagnostic and statistical manual of mental disorders (text rev.) APA. Washington, DC. 2000.
37. Kraemer HC. Evaluating medical tests: Objective and quantitative guidelines: Sage publications Newbury Park, CA:; 1992.
38. Rosseel Y. lavaan: An R package for structural equation modeling. Journal of Statistical Software. 2012;48(2):1-36.
39. Chalmers RP. mirt: A multidimensional item response theory package for the R environment. Journal of Statistical Software. 2012;48(6):1-29.
40. Van der Ark LA. Mokken scale analysis in R. Journal of Statistical Software. 2007;20(11):1-19.
41. Van Abswoude AA, van der Ark LA, Sijtsma K. A comparative study of test data dimensionality assessment procedures under nonparametric IRT models. Applied Psychological Measurement. 2004;28(1):3-24.
42. Mokken RJ. A theory and procedure of scale analysis: With applications in political research: Walter de Gruyter; 1971.
43. Nallet A, Weber B, Favre S, Gex-Fabry M, Voide R, Ferrero F, et al. Screening for bipolar disorder among outpatients with substance use disorders. European Psychiatry. 2013;28(3):147-53.
44. Feng Y, Xiang Y-T, Huang W, Wang G, Feng L, Tian T-F, et al. The 33-item Hypomania Checklist (HCL-33): A new self-completed screening instrument for bipolar disorder. Journal of affective disorders. 2016;190:214-20.
45. Perugi G, Fornaro M, Maremmani I, Canonico PL, Carbonatto P, Mencacci C, et al. Discriminative hypomania checklist-32 factors in unipolar and bipolar major depressive patients. Psychopathology. 2012;45(6):390-8.
46. Zimmerman M. Misuse of the Mood Disorders Questionnaire as a case-finding measure and a critique of the concept of using a screening scale for bipolar disorder in psychiatric practice. Bipolar disorders. 2012;14(2):127-34.
47. Phelps JR, Ghaemi SN. Improving the diagnosis of bipolar disorder: predictive value of screening tests. Journal of affective disorders. 2006;92(2):141-8.
48. Forty L, Smith D, Jones L, Jones I, Caesar S, Fraser C, et al. Identifying hypomanic features in major depressive disorder using the hypomania checklist (HCL-32). Journal of affective disorders. 2009;114(1):68-73.
49. Zimmerman M, Galione JN, Ruggero CJ, Chelminski I, Dalrymple K, Young D. Are screening scales for bipolar disorder good enough to be used in clinical practice? Comprehensive psychiatry. 2011;52(6):600-6.
50. Fassassi S, Vandeleur C, Aubry J-M, Castelao E, Preisig M. Prevalence and correlates of DSM-5 bipolar and related disorders and hyperthymic personality in the community. Journal of affective disorders. 2014;167:198-205.

Foot note

[1] Following a reviewer's suggestion we also analyzed whether the active/elated versus irritable/risk-taking factor resulted in different screening properties for the HCL-32. The ROC suggested better screening properties with respect to AUC for the active/elated factor with .80 [95% CI .73-.87] than the irritable/risk-taking factor with .72 [95% CI .64-.80] in this sample.

Funding source

The PsyCoLaus study was and is supported by research grants from GlaxoSmithKline, the Faculty of Biology and Medicine of Lausanne, and the Swiss National Science Foundation (grants 3200B0–105993, 3200B0-118308, 33CSCO-122661, 33CS30-139468 and 33CS30-148401).

Acknowledgements

The authors would also like to express their gratitude to the investigators who recruited the CoLaus sample, in particular to Prof. Peter Vollenweider and Prof Gérard Waeber.

Declaration of interests/Disclosures

TDM has received speaker honoraria in the past but has no conflict of interest with respect to the current paper. EC, MG, JA and MP have nothing to declare and report no conflict of interest.

Table 1

Description of the final sample (n = 1712) and the diagnostic groups

	No mood disorder (n = 902)	Major depressive disorder (n = 770)	Bipolar disorder (n = 40)	p-value	Group comparison [a]
% women (n)	46.8% (422)	68.3% (526)	50% (20)	p< .001	1 < 2 > 3
Age	61.07 (10.42)	57.89 (9.42)	57.09 (8.55)	p< .001	1 < 2 = 3
current mood*	3.15 (0.75)	3.21 (0.98)	3.29 (1.37)	n.s.	1 = 2 = 3
AUD	19 (2.1%)	40 (5.2 %)	9 (22.5%)	p< .001	1 < 2 < 3
SUD	24 (2.7%)	51 (6.6%)	11 (27.5%)	p< .001	1 < 2 < 3
HCL 32	9.60 (5.57)	12.03 (5.48)	18.24 (5.37)	p< .001	1 < 2 < 3
- Active/elated factor	7.29 (4.07)	9.04 (3.94)	12.37 (3.37)	p< .001	1 < 2 < 3
- Irritable/risk-taking factor	0.99 (1.51)	1.25 (1.63)	2.56 (2.11)	p< .001	1 = 2 < 3
HCL 20	5.61(3.23)	6.98 (3.18)	10.43 (3.65)	p< .001	1 < 2 < 3
HCL 16	4.44 (2.73)	5.67 (2.77)	8.43 (2.92)	p< .001	1 < 2 < 3

Notes: AUD – Alcohol use disorder; SUD – Substance use disorder (including AUD as well); HCL-32 – Hypomania Checklist-32 items (Angst et al., 2005); HCL-20 – Hypomania Checklist-20 items (Bech et al., 2011); HCl-16 – Hypomania Checklist 16 items (Forty et al., 2011).

*This is based on HCL-32 item 1: asking '*how are you feeling today compared to your usual state:*' on a scale from 'much worse' (0) to 'much better' (6) (n =1646)

[a] 1 = no mood disorder, 2 = major depressive disorder, 3 = bipolar disorder; post hoc tests for dimensional were Student Newman Keuls; for categorical variables , post hoc χ^2 tests for 2 x 2 comparisons.

Table 2

Screening properties for bipolar disorder (BD) of the Hypomania Checklist (HCL) with 32, 20 and 16 items

	HCL-32 (Angst et al., 2005)	HCL-20 (Bech et al., 2011)	HCL -16 (Forty et al., 2010)
AUC (95 % CI)			
BD vs all other participants	.83 (SE .03) .77-.89	.80 (SE .04) .73-.87	.80 (SE .04) .73-.87
BD vs MDD	.79 (SE .04) .72-.86	.76 (SE .04) .68-.84	.76 (SE .04) .68-.84
BD vs no mood disorder	.87 (SE .03) .81-.92	.84 (SE .03) .77-.90	.85 (SE .03) .77-.91
Optimal cut-off (PsyColaus)	**≥ 13**	**≥ 7**	**≥ 6**
SE (optimal cut-off) [a]	.85	.83	.80
SP:			
BD vs all other participants	.63	.55	.58
BD vs MDD	.54	.44	.48
BD vs no mood disorder	.71	.64	.66
Published cut-off	**≥14**	**≥10**	**≥7**
SE (published cut-off) [a]	.78	.63	.78
SP:			
BD vs all other participants	.68	.84	.70
BD vs MDD	.60	.78	.60
BD vs no mood disorder	.76	.89	.78
PPP:			
BD vs all other participants	.06	.09	.06
BD vs MDD	.09	.14	.09
BD vs no mood disorder	.12	.20	.14
NPP:			
BD vs all other participants	.99	.99	.99
BD vs MDD	.98	.98	.98
BD vs no mood disorder	.99	.98	.99

Notes: AUC – Area under the curve (ROC); 95 % CI – 95 % confidence interval; MDD = Major depressive disorder; NPP – negative predictive power; PPP – positive predictive power; SE = sensitivity, SP – specificity;

[a] SE for all comparisons will be the same since based on cases with BD

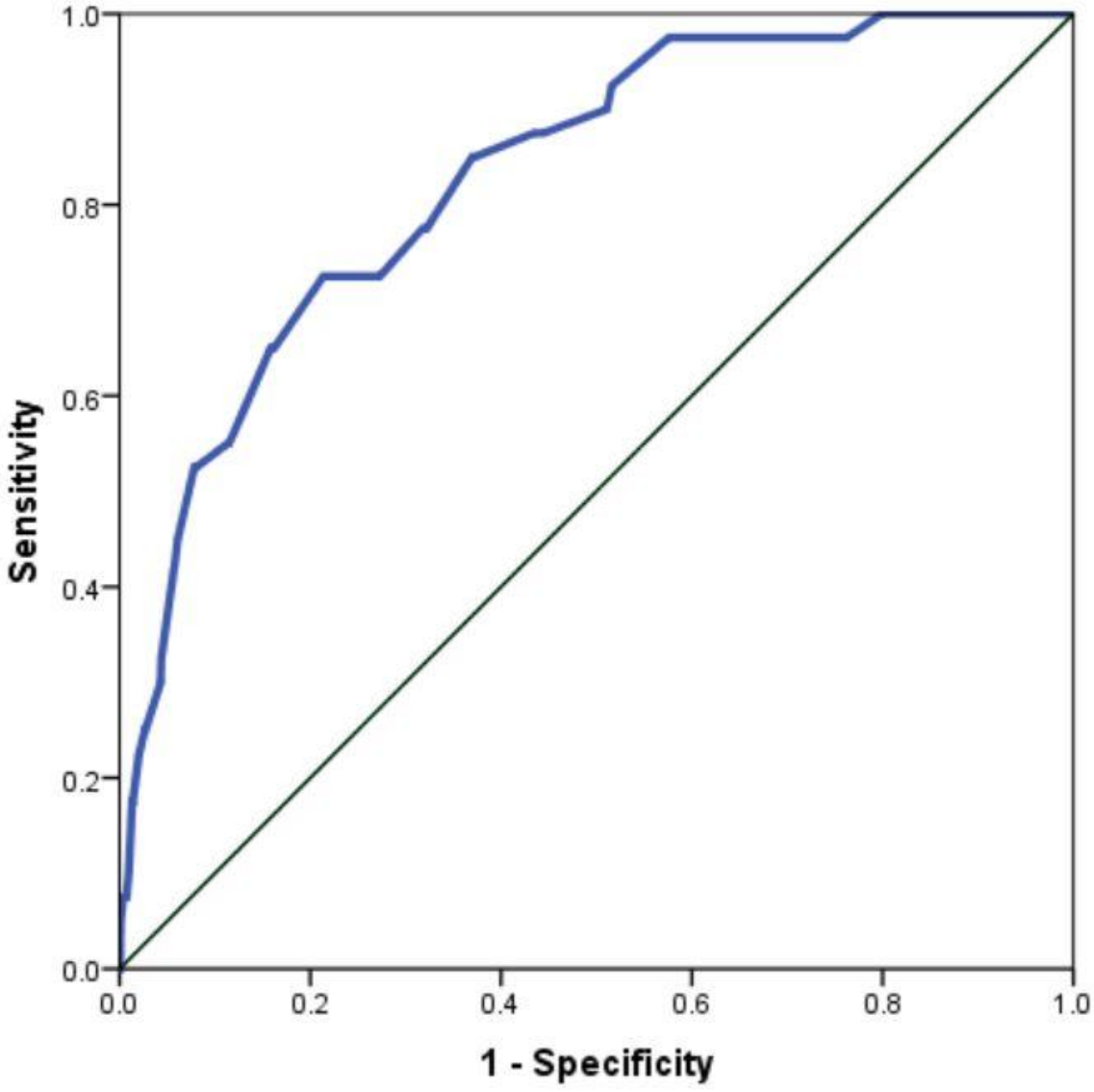

Figure 1:
ROC curve for the HCL-32 comparing patients with BD with all other participants in the PsyCoLaus study

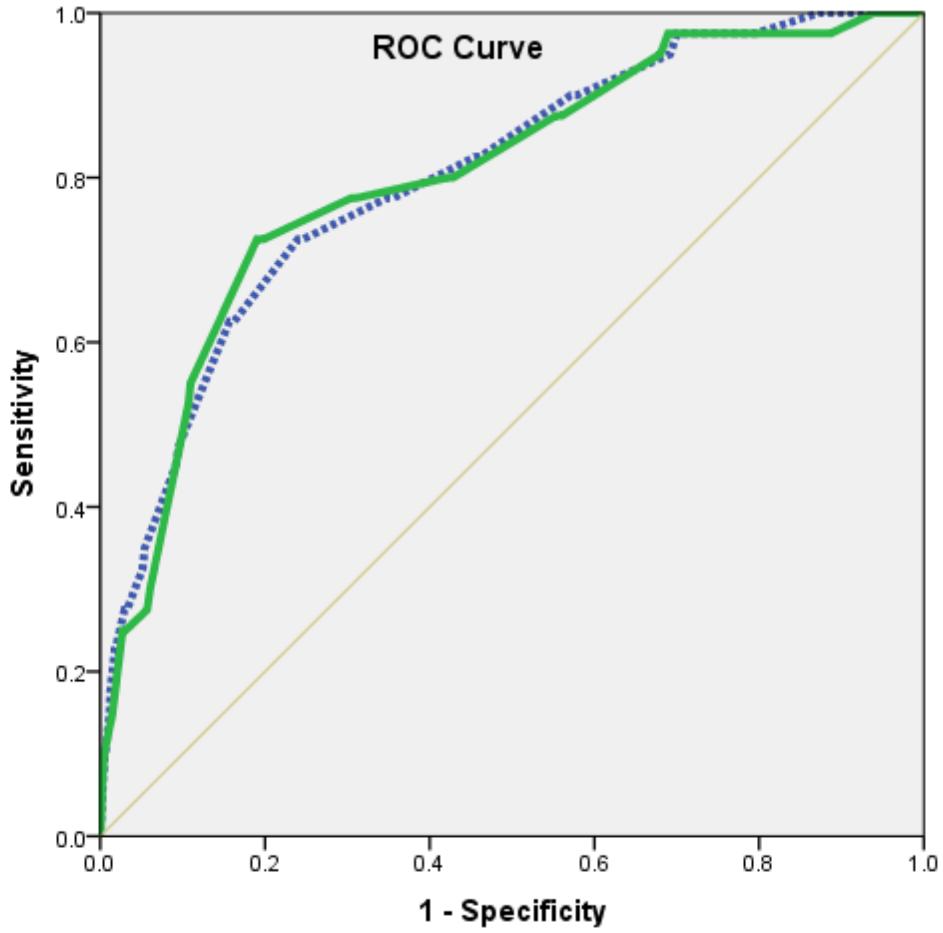

Figure 2:
ROC curves for the two briefer HCL comparing patients with BD with all other participants [HCL-20 (Bech et al) = dotted line; HCL-16 (Forty et al) = solid line]