Research Methods

Diagnosis of depressed young people—criterion validity of WHO-5 and HSCL-6 in Denmark and Norway

Kaj S Christensen*a,*, Wenche Haugen*a, Manjit K Sirpal*b and Ole Rikard Haavetb

*aResearch Unit for General Practice, Institute for Public Health, Aarhus University, Aarhus, Denmark and
bDepartment of General Practice and Community Medicine, University of Oslo, Oslo, Norway.

*Correspondence to Kaj S Christensen, Research Unit for General Practice, Institute for Public Health, Aarhus University, Bartholins Allé 2, DK-8000 Aarhus C, Denmark; E-mail: kasc@alm.au.dk

Abstract

Background. According to the World Health Organization, depression ranks as a major contributor to the global burden of disease. A large proportion of adult depressions had their first appearance in adolescence. Brief and reliable rating scales are needed for early detection.

Objective. The aim of this study is to validate the WHO-5 and the HSCL-6 for detection of depression among adolescents.

Methods. The project is a GP multicentre study conducted in Denmark and Norway. A total of 294 adolescents (14–16 years) responded by answering a paper- or web-based questionnaire and later completed a Composite International Diagnostic Interview, which was used as the gold standard. Depression was defined by ICD-10 criteria. Criterion validity was examined, the likelihood ratios measured and cut-offs for depression were calculated using the Youden index.

Results. The prevalence of depression was 11.8% in our sample. The optimal cut-off point was 11 for the WHO-5 test and 9 for the HSCL-6 test. There were no gender differences. For the WHO-5 and the HSCL-6, respectively, the sensitivity was 0.88 [95% confidence interval (CI): 0.74–0.96] and 0.85 (95% CI: 0.70–0.94), the specificity was 0.80 (95% CI: 0.75–0.85) and 0.79 (95% CI: 0.74–0.84) and the likelihood ratio was 4.5 and 3.8 (P = 0.39). We observed no statistically significant diagnostic differences with respect to nationality or administration procedure.

Conclusions. The WHO-5 and the HSCL-6 may both serve as valid case finding instruments for depression in young people in primary care.

Key words: Adolescent, brief psychiatric rating scale, depression, mass screening, primary health care, sensitivity and specificity.

Introduction

Unipolar depression is a prevalent disorder among adolescents with an estimated 1-year prevalence in the general population of 4–5% (1). According to the World Health Organization, depression ranks as a major contributor to the global burden of disease measured in disability-adjusted life years (2). Population studies have consistently shown that depression is about twice as common in women as in men. In puberty, the relative increase in its prevalence is related to pubertal development rather than to chronological age, and it reaches adult ratios between the ages of 15 and 18 years (3). Depressive disorders are more common in urban than in rural populations and are strongly associated with low socio-economic status (4). Primary health care professionals face a shortage of valid instruments for the identification of depression in adolescents (5–7). The lack of detection is particularly problematic for young people because early recognition is pivotal in avoiding negative short- and long-term effects on social, mental and academic functioning (8–10).
Moreover, early diagnosis is crucial for prevention of acute and chronic morbidity and mortality (6,11).

Mauerhofer et al. (12) report that only 13% of 16- to 20-year olds needing help for psychological problems consulted a health professional; hereof 45% visited a GP, 41% a psychologist or psychiatrist and 14% another health care professional. Yet 78% of the age group had seen their GP at least once during the previous year for reasons other than depression, which may suggest that current diagnostic routines in primary care services are inadequate.

Between one in five (13,14) and one in three (6) depressed youth are correctly diagnosed, and ≤50% of identified cases (15) receive any type of treatment before adulthood.

Depression screening instruments for young people are mainly developed and endorsed in the specialist health care (16). The sensitivity, specificity and cut-off points of these instruments may differ between depressed young hospital inpatients and adolescents meeting health care professionals in primary care, e.g. secondary school students (17). A call for depression screening instruments in primary care is therefore renewed, and also Internet-based alternatives are requested (1).

The National Institute for Health and Clinical Excellence has released guidelines and recommendations for the diagnosis and management of depression in primary care (www.nice.org.uk/pdf/CG023quickrefguide.pdf). The recommendations conclude that better diagnosis may be accomplished through short questionnaires or a few screening questions. These recommendations are consistent with the fact that GPs having to consider many possible diagnoses during short appointments. In the Danish-Norwegian validation study, we are investigating the use of different diagnostic tools of different length in adolescents (18). The HSCL-10, a 10-item questionnaire, is already found to be suitable and valid (19). However, even shorter instruments like the HSCL-6 and the WHO-5 may prove valid. The HSCL-6 uses a set of negatively worded questions and the WHO-5 uses a set of positively worded phrasings (Box 1). So far, these two instruments have only been validated for use among adults (20,21) and within single-country settings. The present study adopts a broader perspective in two ways: first, by using existing screening instruments developed for adults in adolescents; second, by applying the same validation procedure in two different countries to explore the generalizability of these instruments across country settings. The aim of the present paper is accordingly to assess the criterion valid and the WHO-5 may prove valid. The HSCL-6 uses a set of negatively worded questions and the WHO-5 uses a set of positively worded phrasings (Box 1). So far, these two instruments have only been validated for use among adults (20,21) and within single-country settings. The present study adopts a broader perspective in two ways: first, by using existing screening instruments developed for adults in adolescents; second, by applying the same validation procedure in two different countries to explore the generalizability of these instruments across country settings. The aim of the present paper is accordingly to assess the criterion valid and the WHO-5 may prove valid. However, even shorter instruments like the HSCL-6 and the WHO-5 may prove valid.

Methods

The project is a collaboration between the Research Unit for General Practice, University of Aarhus, Denmark and the Institute of General Practice and Community Medicine, University of Oslo, Norway. The primary objective of this collaborative project was to assess screening methods for improving recognition of depression among adolescents.

General practitioners

A random sample of GPs listed in Norway and Denmark was included. We invited 25 GPs in Norway and 18 GPs in Denmark to extract the files of adolescents, 14–16 years of age, from their medical records. On the basis of these files, a list was made that included the name and the date of birth of each adolescent. A participant code number was added to the list for identification and to serve as a login code for a web questionnaire.

Study group

A standard letter of invitation to participate was sent to the adolescents on the GPs’ lists; it contained standardized information, a questionnaire and a pre-stamped envelope. In Norway, a similar letter was also sent to the parents of the 14- and 15-year olds in the same envelope, as required by the ethical committees in that country.

The invitation letter contained information about the goal of the study and its procedure. In Denmark, the adolescents’ returning of the questionnaire was interpreted as tantamount to agreement to participate in the study. In Norway, the 14- and 15-year olds and their parents were required to sign a form to be returned to their GP to indicate their agreement to participate. The adolescent could respond by using the forms sent by post or by using the website: www.ungdep.au.dk. Responses were identified by the login code and a unique participation number.

All participating adolescents who filled in a questionnaire also provided a telephone number by which they wished to be contacted (but not their names), and a member of the study group called the adolescent by phone. The date when the questionnaire and consent form were received and the date of the Composite International Diagnostic Interview (CIDI) telephone interview were noted.

Rating scales

The HSCL-6 is a depression subscale derived from the Symptom Checklist-90 (SCL-90) (20). The WHO-5 (22) is a subscale developed from the Short Form 36 (SF-36). The WHO-5 is a widely used instrument to measure well-being and it is available in 28 languages. The SCL-90 and the WHO-5 have both been translated and culturally adapted by the copyright holders of the Danish and Norwegian versions. Both instruments have been validated as screening instruments for detection of depression among adults (20,23).

Box 1. Wording of WHO-5 and HSCL-6

WHO-5

Over the last 2 weeks:

1. I have felt cheerful and in good spirits
2. I have felt calm and relaxed
3. I have felt active and vigorous
4. I woke up feeling fresh and rested
5. My daily life has been filled with things that interest me

Response categories: all of the time (5), most of the time (4), more than half of the time (3), less than half of the time (2), some of the time (1), at no time (0).

WHO-5 is usually converted to a scale of 0–100 by multiplying the sum score by 4.

HSCL-6

In the course of the past 2 weeks, have you been troubled by:

1. Feeling blue? [HSCL item 30]
2. Feelings of worthlessness? [HSCL item 79]
3. Thoughts of ending your life? [HSCL item 15]
4. Feelings of being trapped or caught? [HSCL item 22]
5. Feeling lonely? [HSCL item 29]
6. Blaming yourself for things? [HSCL item 26]

Response categories: no (1), slightly (2), much (3), very much (4).
Diagnostic interview
The CIDI, a well-established clinical instrument for measuring depression based upon the DSM-IV and the ICD-10 (9), was used as the gold standard interview. Two of the authors (WH and MKS) were both certified as CIDI interviewers and conducted all the interviews. Interviewers were blinded to WHO-5 and HSCL-6 ratings.

Composite International Diagnostic Interview
The CIDI version 2.1 was used as the gold standard interview. The module for depressive disorders can identify seven different depressive disorders based upon the DSM-IV and the ICD-10 (F31.1–F34.1). A diagnosis within the past 2 weeks was defined as ongoing; diagnosis before this was defined as previous. The CIDI is a fully structured, standardized diagnostic interview that has shown good feasibility, high inter-rater reliability and has been subjected to tests of its reliability and validity (24,25). Because the interview is fully structured, it is suitable for use as a telephone interview (26,27).

Statistical analysis
The data collected from the website were recorded directly into a database. All the data collected in the CIDIs and the paper versions of the questionnaire were fed into the same database. The data files were analysed using STATA version 11; only complete data sets were used.

Criterion validity was analysed by means of the receiver operating characteristic (ROC) curve with specific measurements for sensitivity and specificity. In order to match the scoring of the HSCL-6, the scoring of the WHO-5 was reversed. Using the Youden index \( J = \max \{ \text{Sensitivity}(c) + \text{Specificity}(c) - 1 \} \), we calculated the optimal cut-off point for depression among girls and boys and measured the likelihood ratios for positive and negative test results. When calculating the cut-off point, we used the ICD-10 diagnoses for Mild depressive episode (F32.0), Moderate depressive episode (F32.1), Severe depressive episode without psychotic symptoms (F32.2), Recurrent depressive disorder, current episode without psychotic symptoms (F 33.2) and Dysthymia (F34.1).

Sample size
The optimal number of adolescents needed for the study was calculated using the formula \( N = 4 \times \text{Zcrit}^2 \times P \times (1-P)/D^2 \). \( N \) is the sample size, the Zcrit is 1.96, \( P \) is the pre-study estimate of sensitivity of 0.70 and \( D \) is the total width of the confidence interval (CI) of 0.1. The number of adolescents needed is 323.

Results
Study population
A total of 2374 adolescents (1167 boys and 1207 girls) were invited to participate by mail, and 373 questionnaires were returned, either by mail or by web, representing a response rate of 16%. Hereof 294 (77%) completed the CIDI (Table 1). The average time between completion of the questionnaire and the CIDI was 20.0 days. Among the CIDI-interviewed youth, 33 (11%) were diagnosed with a current depressive episode, hereof 23 girls and 8 boys (\( P = 0.02 \)). Twelve girls and 1 boy (4%) met the ICD-10 criteria of moderate to severe depression.

Criterion validity
The diagnostic accuracy of the HSCL-6 and the WHO-5 demonstrated no significant differences (Table 2). ROC areas varied between 0.81 and 0.94. No significant differences were observed between Denmark and Norway or between responses on paper and web. The mean score for non-depressed youth was HSCL6 7.65 (SD 2.53) and WHO-5 7.34 (SD 4.37), whereas the mean score for depressed youth was HSCL6 12.64 (SD 4.76) and WHO-5 15.18 (SD 4.65).

Recommended cut-off point
The optimal cut-off point for the WHO-5 and the HSCL-6 test was 11 and 9, respectively, with no gender differences. For WHO-5 and HSCL-6, respectively, the sensitivity was 0.88 (95% CI 0.74–0.96) and 0.85 (95% CI 0.70–0.94), the specificity was 0.80 (95% CI 0.74–0.84) and 0.78 (95% CI 0.72–0.83) and the likelihood ratio was 4.5 and 3.8 (\( P = 0.39 \)). Measures of sensitivity and specificity for different cut-offs are displayed in Table 3. ROC curves are illustrated in Figure 1.

The optimal cut-off points for the HSCL-6 are 9 for mild depression (sensitivity 0.85, specificity 0.78), 13 for moderate depression (sensitivity 0.83, specificity 0.78) and 14 for severe depression (sensitivity 0.61, specificity 0.96). The optimal cut-off points for the WHO-5 are 11 for mild depression (sensitivity 0.88, specificity 0.80), 13 for moderate depression (sensitivity 0.75, specificity 0.89) and 14 for severe depression (sensitivity 0.84, specificity 0.90).

Discussion
The WHO-5 and the HSCL-6 enjoyed equal and good criterion validity in the Danish and Norwegian primary care populations; we observed no significant differences in validity measures between the web-based and the paper-based questionnaire responses. Although measures of both sensitivity and specificity seem fair, systematic screening in low-prevalence populations should always be accompanied by careful diagnostic assessment; thus, at a prevalence rate of 10%, a sensitivity of 90% and with a specificity of 80%, a screening would produce 9% true-positive tests and 18% false-positive tests for depression. If the test is negative, only 1% would have a false-negative test at this prevalence rate. A systematic screening approach may therefore be a more effective means for ruling out rather than...
Table 3. Measures of sensitivity and specificity of the WHO-5 and the HSCL-6 scales using the CIDI for ICD-10 diagnosis of depression among adolescents

<table>
<thead>
<tr>
<th>Cut point</th>
<th>WHO-5</th>
<th>HSCL-6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity (95% CI)</td>
<td>Specificity (95% CI)</td>
</tr>
<tr>
<td>≥7</td>
<td>0.94 (0.82–0.99)</td>
<td>0.47 (0.41–0.54)</td>
</tr>
<tr>
<td>≥8</td>
<td>0.94 (0.82–0.99)</td>
<td>0.60 (0.54–0.66)</td>
</tr>
<tr>
<td>≥9</td>
<td>0.94 (0.82–0.99)</td>
<td>0.65 (0.59–0.71)</td>
</tr>
<tr>
<td>≥10</td>
<td>0.91 (0.78–0.97)</td>
<td>0.73 (0.67–0.78)</td>
</tr>
<tr>
<td>≥11</td>
<td>0.88 (0.74–0.96)</td>
<td>0.80 (0.75–0.85)</td>
</tr>
<tr>
<td>≥12</td>
<td>0.79 (0.63–0.90)</td>
<td>0.84 (0.80–0.89)</td>
</tr>
<tr>
<td>≥13</td>
<td>0.70 (0.53–0.83)</td>
<td>0.88 (0.84–0.92)</td>
</tr>
<tr>
<td>≥14</td>
<td>0.64 (0.47–0.78)</td>
<td>0.90 (0.86–0.93)</td>
</tr>
</tbody>
</table>

The scoring of the WHO-5 was reversed in order to match the scoring of the HSCL-6.

Figure 1. Criterion validity of HSCL-6 and WHO-5 according to CIDI diagnosis of depression in the total population of adolescents in Denmark and Norway. The scoring of the WHO-5 was reversed in order to match the scoring of the HSCL6.

ruling in depression. The HSCL-6/WHO-5 may therefore serve as good screening questionnaires in so far as we can be almost certain that patients who score negative do not have depression. If, on the other hand, they score positive, we cannot be sure that the patient is actually depressed; in this situation, it would be sensible to include a more comprehensive depression questionnaire, i.e. to investigate further.

However, in clinical situations where depression is to be ruled out or in, a test may actually be more effective. Assuming a pre-test probability of 50% and that the sensitivity and the specificity values also apply to this clinical population, a case-finding strategy would include 45% true-positive and 10% false-positive tests for depression.

To our knowing, this is the first published study jointly validating the WHO-5 and the HSCL-6 for depression screening among adolescents in a primary care sample. A recent study by Prast et al. (28) validating the WHO-5 for depression screening among adolescents in pediatric care demonstrated a similar diagnostic accuracy with an area under the curve of 0.88.

This study has some limitations. First of all, the low response rate does raise questions regarding the representativeness of the interviewed population. Still, low participation rates could be expected in this ‘hard to reach group’. Participation rates were higher in Denmark (21%) than in Norway (12%), presumably because in Norway parental consent is required for studies including the 14- and 15-year olds. In order to study the possible effect of attrition bias, we compared early and late responders (29). Groups were split on a median time between questionnaire response and interview of 15 days. The differences in mean scores for the WHO-5 and the HSCL-6 were −0.73 (P = 0.23) and −0.79 (P = 0.04), respectively. Assuming that late responders resemble non-responders, these figures indicate that non-responders may actually have more depressive symptoms than participants. This would clearly tend to underestimate the observed depression prevalence in our study. Despite differences in participation rates between the two countries, prevalence rates of depression were similar. Previous primary care studies have demonstrated prevalences of depression among adolescents similar to those reported in the present study, and our participants are therefore likely to be as representative as the populations included in other studies.

A second limitation is that we included a slightly smaller number of participants (294) than required according to the protocol calculation (323). However, during the study, the sensitivity of the instruments proved to be higher than expected, and the recruitment of participants could therefore be stopped at a smaller number.

A third limitation may concern the validity of the diagnostic interviews made by telephone rather than as face-to-face interviews. The CIDI is a fully structured interview that can be administered by trained lay interviewers (30) and independently of clinical observations. An additional concern regarding the CIDI is that it has not been formally validated for use among adolescents. Both interviewers, however, were not only trained and certified CIDI interviewers but they were also experienced GPs familiar to the populations approached.

A fourth limitation to be discussed is the context in which the adolescents completed the questionnaires. The completion of a depression questionnaire at home in a confidential setting with parents close by may have affected how the adolescents completed the questionnaire. We have no data to assess the importance of these factors. Respondents were informed that the symptoms asked all regard symptoms of depression, and we therefore have no reason to believe that conjoint use of negatively and positively worded questionsnaires should influence how they were completed.

In summary, we conclude that the HSCL-6 and the WHO-5 are both valid instruments for depression screening among adolescents. No differences in criterion validity between paper-based and web-based versions were observed. Adolescents who test positive may, however, still need careful diagnostic evaluation in order to rule out a false-positive diagnosis of depression.
Both instruments may be useful for research or regular check-ups on the mental health of young people in a primary care setting. The positively worded WHO-5 may be preferred to the negatively worded HSCL-6, especially in a non-clinical setting such as a school nursery.

Finally, it should be stated that questionnaire-based screening for depression may cause over-diagnosis in populations with a low prevalence of depression. Future research should be conducted with a focus on validating such forms in the context in which they are routinely being used to rule out or in a clinical suspicion of depression.

Acknowledgements

We are very grateful to the patients and practices who agreed to participate.

Declaration

Funding: All authors are funded through fixed pay from the institutions where they belong. The Fund for Quality Assurance of Primary Health Care by the Norwegian Medical Association also provided a grant for this research.

Ethical approval: The Ethical Committee in Denmark (2006-2.0/44); The National Committees for Research Ethics in Norway (S-06293a).

Conflicts of interest: none.

References