Shortened questionnaires to assess anxiety and depression during in-hospital rehabilitation: clinical validation and cutoff scores

Giorgio Bertolotti¹
Loretta Moroni¹
Roberto Burro²
Antonio Spanevello³
Roberto FE Pedretti⁴
Giandomenico Giorgetti⁵

¹Psychology Unit, Salvatore Maugeri Foundation, IRCCS, Scientific Institute, Tradate, ²Department of Human Sciences – University of Verona, Verona, ³Department of Cardiology, ⁴Division of Pulmonary Disease, ⁵Department of Neuromotor Rehabilitation, Salvatore Maugeri Foundation, IRCCS, Scientific Institute of Tradate, Italy

Background: A postacute phase needs reliable routine screening instruments in order to identify the patients to be referred for a clinical interview with a psychologist. The aim of this study was to estimate the clinical cutoff scores of the anxiety and depression questionnaires and their clinical validity using a gold standard.

Methods: The study involved 177 patients with pulmonary, cardiac, or neurological disease undergoing in-hospital rehabilitation. Receiver operating characteristic curves were used to determine the best concordance between questionnaire’s scores and the gold standards.

Results: There was a significant difference ($P<0.001$) between clinically anxious and depressed patients and nonclinical subjects. The receiver operating characteristic curve for anxiety indicated that the best area under the curve for State Anxiety Inventory is obtained with a cutoff point of 21 for males and 25 for females; for depression scores, the highest area under the curve for Depression Questionnaire-Reduced Form is obtained with a cutoff point of six for males and eight for females.

Conclusion: Using appropriate cutoff values, the State Anxiety Inventory and Depression Questionnaire-Reduced Form allow psychologists to optimize early clinical intervention strategies selecting patients with significant needs.

Keywords: anxiety, depression, questionnaire, sensitivity and specificity, rehabilitation

Introduction

Many studies have established the presence of a high rate of psychological complaints among nonpsychiatric hospital patients.¹ Symptoms of anxiety and depression may confuse a patient’s clinical image, reduce compliance with therapeutic programs and affect the medium- or long-term outcomes pursued during the course of hospitalization,²–⁴ predict health-related quality of life,⁵ and predict the influence of symptoms of anxiety and depression on medication noncompliance.⁶–⁷

The American Heart Association recently published a Science Advisory with the recommendation that patients with coronary heart disease (CHD) should be screened for depressive symptoms.⁸–¹⁰ Ziegelstein et al¹¹ maintain that for routine screening of CHD patients for depression to be recommended, screening tests must be sufficiently sensitive, specific, and validated, because cutoff scores used in primary care may not work equivalently in patients with CHD.¹² In a very recent review,¹³ it is noted that there are few examples of screening tools with high sensitivity and specificity using an a priori defined cutoff score in >1 CHD sample.

Mild-to-moderate symptoms of anxiety and/or depression have also been observed in patients with chronic obstructive pulmonary disease (COPD) and current recommendations indicate that they should not be ignored. Appropriate outcome measures...
for mental health are needed for this patient population. Similarly, depression and anxiety were significant for the outcomes regarding readmissions to hospital or death 6 months after a stroke. These are the reasons why clinicians need to identify specific patients with stroke with preexisting mental health conditions for which additional psychotherapy treatment may result in improved stroke outcomes.15

A Cochrane review16 indicated that psychological intervention in CHD patients did produce small to moderate improvements in depression and anxiety but there was no consistent evidence of a positive effect on health-related quality-of-life (HRQOL) or other psychological outcomes, including perceived stress, Type-A behavior, anger, and perceived exhaustion or Vital Exhaustion. All the aforementioned reasons and results support our search for clinical level of anxiety or depression in a rehabilitation setting and that both will be specific to the medical conditions of the patients concerned.

The aim of this study was to use receiver operating characteristic (ROC) curves to determine the best concordance between STAI-X3 and Depression Questionnaire-Reduced Form (acronym AD-R) scores using the opinion of a psychologist after a semi-structured clinical interview as gold standard.

Materials and methods
Patient population and data collection
The present observational study involved consecutively enrolled patients with pulmonary, cardiac, or neurological and neuromuscular disease undergoing in-hospital rehabilitation at the Salvatore Maugeri Foundation, IRCCS, Scientific Institute Division of Respiratory, Cardiac, and Neuromotor Rehabilitation during a period of 6 months in 2010. As a rule, the subjects completed the AD-R within the second to third day from the hospital admission. On the same day, a psychologist independently assessed their anxiety and depression status using a semistructured interview17,18 and decided the appropriate psychological support needed. The psychologist was blinded to the AD-R scores. Exclusion criteria were as follows: the inability to complete questionnaires and a history of a severe psychiatric disease. The protocol was reviewed and approved by an internal review board for ethical protection of subjects (Comitato Tecnico Scientifico), and written informed consent of all the participants was obtained.

Measures
With the aim of making the screening process more rapid and accurate, we developed the ten-item version of the State Anxiety Inventory (STAI-X3)19-21 and the 15-item QD-R both with validated and reliable criteria (concurrent and predictive content).22,23 The reduced form of STAI-X3, consists of 10 items asking the subjects how they feel “right now” that are scored using a 4-point Likert scale (total score 10–40). The QD-R measures depressive symptoms and was originally constructed with reference to Diagnostic and Statistical Manual of Mental Disorders (DSM)-III and meets all of the DSM-IV Revised criteria for major depressive disorder (depressed mood; loss of interest or pleasure; variations in appetite and weight; insomnia/hypersonomnia; psychomotor agitation/slowing; fatigability; self-depreciation; poor concentration; recurrent thoughts of death). For more details on the reduction methods, see Vidotto et al.22 The two questionnaires in the reduced form take ~5 minutes to complete.

It simplifies screening of patients in hospital settings as it is suitable for subjects with mild/moderate or subclinical depression.22,23 The QD-R has 15 items, each consisting of a statement (eg, “The future looks very bleak”) to be answered “yes” or “no” (total score 0–15) and excludes somatic symptoms, thereby avoiding potential confounding by the somatic symptoms in hospitalized patients. The instructions ask that the questions should be answered “thinking about how you feel at this moment”, with the subject being asked to ponder the time span corresponding to that required to complete the survey.

Using Cronbach’s alpha score, the internal consistency of the QD-R is 0.77; any value between 0.7 and 0.8 is considered satisfactory for comparing groups.25 STAI-X3 showed an internal consistency assessed with Cronbach’s alpha of 0.90 in healthy subjects.20

Semistructured clinical interview: gold standard
In order to structure and maintain a single criterion for defining the gold standard, we used a “semistructured clinical interview” form based on and in respect of the DSM-IV anxiety and depression (DSM code 300.4) criteria. Inter-rater agreement with the psychological judgment for anxiety state (Cohen’s K = 3.60; concordance 76%) and for depressive reaction (Cohen’s K = 2.39; concordance 86%) has been found in a previously published study.17 The “semistructured clinical interview” form is divided into three sections: anxiety, depression, and an area in which the diagnostic criteria for such disturbances overlap.

The interview began with a series of unstructured questions with the aim of establishing a cooperative relationship between the patient and the psychologist, and acquiring diagnostically useful information. At the end of the interview, the clinical psychologist had to judge whether the subject showed no anxiety/depression or one or both of these characteristics.
If this was the case, the subject was invited to attend further sessions for clinical psychological support.

Statistical analysis

R software 3.3.0 (R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/; language and environment for statistical computing and graphics) was used to analyze the data sample and analysis of variance to verify the significance of the differences in mean QD-R and STAI-X3 scores between males and females and between the disease groups. The construct validity of the AD-R schedule as a measure of depression and anxiety was assessed by examining the differences in mean value between the clinical groups as classified by the psychologist. Bonferroni’s correction was applied for the type I error inflation due to multiple comparisons.

ROC curves were used to identify the AD-R cutoff points. ROC analysis quantifies the accuracy of diagnostic tests (or further appraisal types) used to discriminate between two states or conditions. The discriminatory accuracy of a diagnostic test is quantified by its ability to suitably classify between subjects with and without disease. A ROC plot displays the performance of a dichotomous classification procedure with continuous or discrete ordinal outcome. In the ROC space, the area under the curve (AUC) measures the performance of a classifying variable and is frequently applied for method comparison. A higher AUC means a better classification. AUCs are computed with trapezoids. In our case, ROC curves were used to identify the AD-R cutoff points. This technique is commonly used in medical decision-making research in order to determine how well a potential classifier discriminates two classes. In the context of this study, the potential classifying variables were the total AD-R scores, and the two classes were the binary classification of the presence/absence of the clinically relevant psychological variables (anxiety and depression).

The confidence intervals (CIs) were computed with bootstrap for AUCs. The 95% CIs of the AD-R cutoff points and the sensitivity and specificity values were computed with bootstrap resampling (stratified manner), and the averaging methods described by Fawcett. In all bootstrap CIs, the subjects were resampled and the modified curve was built before the statistics of interest were computed.

Results

One hundred and seventy-seven subjects (101 males and 76 females) completed the AD-R schedule and the interview with the psychologist at the beginning of their in-hospital rehabilitation period. Tables 1 and 2 show their characteristics. The main pulmonary diseases were asthma, COPD, and respiratory failure; the main cardiac diseases were coronary artery disease (myocardial infarction, angina pectoris), congestive heart failure, and valvular heart disease; and the main neurological or neuromuscular diseases were stroke and myopathy.

Comparing the three disease groups, no significant differences were found either for STAI-X3 scores ($F_{(2,174)}=0.252, P=0.778$) or for QD-R scores ($F_{(2,174)}=0.186, P=0.830$).

Table 3 shows the distribution of the AD-R scores on the basis of the psychologist’s diagnosis of depression and anxiety. Based on the psychologist’s judgment, 53 subjects were “possible case for anxiety” (prevalence =29.9%) and 42 were “possible case for depression” (prevalence =23.7%). There was a significant difference in mean STAI-X3 scores between the subjects with and without clinically relevant anxiety ($t_{(175)}=14.813, P<0.001$), and in mean QD-R scores between the subjects with and without clinically relevant depression ($t_{(175)}=12.864, P<0.001$).

Cutoff scores

The best AUC for STAI-X3 was obtained with a cutoff point of 21.0 for males and 25.0 for females (Figure 1). CI of AUC for STAI-X3 sample of males (equal to 95.5%) was

<table>
<thead>
<tr>
<th>Disease group</th>
<th>Disease</th>
<th>Disease number of subjects</th>
<th>Disease group number of subjects</th>
<th>Disease group age (mean ± SD)</th>
<th>Total number of subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiac</td>
<td>Ischemic heart disease</td>
<td>67</td>
<td>110</td>
<td>61.6±11.1</td>
<td>177 (101 males 76 females)</td>
</tr>
<tr>
<td></td>
<td>Valvular heart surgery</td>
<td>27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Heart failure</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Other cardiac disease</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pulmonary</td>
<td>Asthma</td>
<td>7</td>
<td>47</td>
<td>61.43±13.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chronic obstructive pulmonary disease</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Obstructive sleep apnea syndrome</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Other respiratory disease</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neuromotor</td>
<td>Parkinson</td>
<td>7</td>
<td>20</td>
<td>62.0±12.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stroke</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Osteoarthritis</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2 Characteristics of the sample

<table>
<thead>
<tr>
<th>Employment status</th>
<th>Cardiac disease</th>
<th>Pulmonary disease</th>
<th>Neuromotor disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employed</td>
<td>26</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>Retired/housewife</td>
<td>78</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>Unemployed</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Others</td>
<td>4</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Civil status</th>
<th>Cardiac disease</th>
<th>Pulmonary disease</th>
<th>Neuromotor disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Married</td>
<td>79</td>
<td>30</td>
<td>5</td>
</tr>
<tr>
<td>Widower</td>
<td>17</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>Divorced/single</td>
<td>3</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Others</td>
<td>11</td>
<td>2</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Questionnaires</th>
<th>Mean score ± SD</th>
<th>STAI-X3</th>
<th>QD-R</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>20.33±5.79</td>
<td>21.1±7.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.84±2.92</td>
<td>4.66±3.91</td>
</tr>
</tbody>
</table>

Abbreviations: QD-R, questionnaire depression-reduced; STAI-X3, state anxiety inventory; SD, standard deviation.

92.0%–99.0%, whereas CI of AUC for STAI-X3 sample of females (equal to 92.9%) was 86.5%–99.4%.

The best AUC for QD-R was obtained with a cutoff point of 6.0 for males and 8.0 for females (Figure 2). CI of AUC for QD-R sample of males (equal to 94.9%) was 90.4%–99.4%, whereas CI of AUC for QD-R sample of females (equal to 96.7%) was 92.8%–100.0%.

The difference between sexes in STAI-X3 scores was not statistically significant ($t_{(175)}=-0.952, P=0.342$), whereas female showed higher ($t_{(175)}=-2.415, P=0.017$) QD-R scores (5.51±3.4) than male (4.37±3.2). The bootstrap test for ROCs (2,000 resampling) indicates that the differences between curves for males and females were not significant both for STAI X-3 ($D=0.679, P$-value=0.497), and QD-R ($D=0.356, P$-value=0.721).

Discussion

Granted that assessing depression and anxiety in patients undergoing rehabilitation in a hospital is of major importance, it is necessary to devise an efficient way of completing such assessments. In this study, we searched the cutoff score of the STAI-X3 and QD-R not referring particularly to the specific disease (ie, CHD, COPD), but to the hospitalized condition in general, considering that DSM criteria suggest to pay attention to symptoms that are clearly due to a general medical condition, not to a specific medical condition. When identifying a cutoff score for a routine screening, we also found that the QD-R was sensible and specific for a clinically relevant state of depression worthy of a deeper psychological examination, not to identify a major depressive disorder to be treated with antidepressants. This avoids the risk suggested by some authors that antidepressant medications may be initiated merely based on a positive depression screen.

From a clinical perspective, our findings support the use of AD-R cutoff scores as a means of screening psychological status in rehabilitation and hospital settings. Additionally, there were no differences between the disease groups. This would allow the multidisciplinary team to devise therapeutic interventions designed to improve both physical and psychological symptoms across disease conditions, which may be the best method to optimize functioning.

The AD-R schedule is clearly subdivided in a solid measure of anxiety and another of depression, with different scores and cutoff points. Some questionnaires measuring depression focus narrowly on anhedonia, defined as a reduced ability to experience pleasure; it is too much to expect that ill patients will discriminate the intended meaning from their experience of not wanting to engage in previously pleasurable activities because of pain, fatigue, and other physical impairment.

The use of ROC curves provide information concerning AD-R cutoff values, which allow psychologists to optimize early clinical interventions during rehabilitation or in the provision of secondary prevention by identifying a clinically relevant state of depression and/or anxiety worthy of a deeper examination. In our sample, we found a STAI-X3 cutoff point of 21 for males and 25 for females. This means that a score ≥21 for males and ≥25 for females is indicative of a critical level of anxiety that needs to be evaluated further by a psychologist. For QD-R, we found a cutoff point of 6 for males and 8 for females. This means that a score ≥6 for males and ≥8 for females is indicative of a critical mood.

Table 2 Characteristics of the sample

Table 3 Construct validity of the AD-R schedule

Table 3 Construct validity of the AD-R schedule

<table>
<thead>
<tr>
<th>Clinical judgment</th>
<th>Depression</th>
<th>Anxiety</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of subjects</td>
<td>QD-R score (mean ± SD)</td>
</tr>
<tr>
<td>No</td>
<td>135</td>
<td>3.5±2.4</td>
</tr>
<tr>
<td>Yes</td>
<td>42</td>
<td>8.9±2.4</td>
</tr>
</tbody>
</table>

Abbreviations: AD-R, STAI-X3, and Depression Questionnaire-Reduced Form; QD-R, questionnaire depression-reduced; STAI-X3, state anxiety inventory.
Brief anxiety and Depression: clinical validation and cutoff scores

Figure 1 ROC curves and cut-off scores for STAI-X3.
Notes: (A) ROC curve for STAI-X3, sample of males. (B) ROC curve for STAI-X3, sample of females. Both figures show the value of cutoff point (with the percentages of specificity and sensitivity) and the AUC (with 95% confidence interval).
Abbreviations: AUC, area under the curve; ROC, receiver operating characteristic; STAI-X3, state anxiety inventory.

Figure 2 ROC curves and cut-off scores for QD-R.
Notes: (A) ROC curve for QD-R, sample of males. (B) ROC curve for QD-R, sample of females. Both figures show the value of cutoff point (with the percentages of specificity and sensitivity) and the area under the curve (AUC, with 95% confidence interval).
Abbreviations: AUC, area under the curve; QD-R, questionnaire depression-reduced; ROC, receiver operating characteristic.

Table 4 The CI of cutoff points of STAI-X3 and QD-R on the basis of the ROC method in relation to the clinical judgment expressed by the psychologist after the semistructured interview

<table>
<thead>
<tr>
<th>Questionnaires (x sexes)</th>
<th>Cutoff point</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>PPV (95% CI)</th>
<th>NPV (95% CI)</th>
<th>PLR (95% CI)</th>
<th>NLR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAI-X3 males</td>
<td>21.0</td>
<td>92.90% (82.1%–100%)</td>
<td>86.30% (78.1%–86.3%)</td>
<td>72.2% (54.5%–86%)</td>
<td>96.9% (89.3%–99.6%)</td>
<td>6.78% (3.8%–12.2%)</td>
<td>0.08% (0.02%–0.3%)</td>
</tr>
<tr>
<td>STAI-X3 females</td>
<td>25.0</td>
<td>88.0% (76%–100%)</td>
<td>82.4% (70.6%–92.2%)</td>
<td>70.9% (51.5%–85.9%)</td>
<td>93.4% (81.8%–98.8%)</td>
<td>4.99% (2.7%–9.2%)</td>
<td>0.15% (0.05%–0.4%)</td>
</tr>
<tr>
<td>QD-R males</td>
<td>6.0</td>
<td>86.7% (66.7%–100%)</td>
<td>93.0% (87.2%–97.7%)</td>
<td>68.3% (42.6%–87.8%)</td>
<td>97.6% (91.5%–99.7%)</td>
<td>12.42% (5.6%–27.6%)</td>
<td>0.14% (0.04%–0.5%)</td>
</tr>
<tr>
<td>QD-R females</td>
<td>8.0</td>
<td>88.9% (70.9%–97.6%)</td>
<td>95.92% (86.0%–99.5%)</td>
<td>92.3% (74.9%–99.1%)</td>
<td>94.0% (83.3%–98.8%)</td>
<td>21.78% (18.8%–25.2%)</td>
<td>0.12% (0.02%–0.7%)</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; NLR, negative likelihood ratio; NPV, negative predictive value; PLR, positive likelihood ratio; PPV, positive predictive value; QD-R, questionnaire depression-reduced; ROC, receiver operating characteristic; STAI-X3, state anxiety inventory.
level suggestive of a level of depression that requires a more complete evaluation by a psychologist.

Regarding construct validity, we found higher cutoff scores in females compared to males, as has been reported. These results may be due to sex differences in illness perception: females, compared to males, are more likely to attribute cardiovascular disease (CVD) to causes beyond their control and perceive CVD as a chronic, untreatable condition. Screening, especially for depression, is strongly recommended even in primary care. Furthermore, in our previous paper, QD-R scores significantly correlated with meters walked in the 6-m walking test by 252 patients during cardiovascular rehabilitation, and patients with QD-R scores ranging from 0 to 5 showed a progressive reduction in the total distance walked during the test. In that study, a fall in walking distance corresponded to a value of 6 in the depression score as measured by QD-R. Further research could be performed to observe the trend of functional performance along clinical cutoff points and to evaluate the effectiveness of integrated and multidisciplinary stepped care, and studies with hospitalized subjects.

Limitations
We collected a sample from a single hospital; our results essentially describe what was found in the sample, but the extent to which those results might generalize beyond the center where the study was conducted is unknown. We also studied patients with pulmonary, cardiac, or neurological and neuromuscular diseases with very heterogeneous characteristics. However, this situation reproduces the proportion of patients usually followed by a psychologist during the rehabilitation phase in our institute. Further study with a larger sample and with different diseases would be required to test the validity of the AD-R cutoff scores for the screening of hospitalized patients that need a specific psychological support.

Conclusion
Using these cutoff values, the STAI-X3 and QD-R allow psychologists to optimize early clinical intervention strategies.

Acknowledgments
The authors wish to thank and acknowledge Alessandra Ianni for her help in collecting and organizing the data set. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Disclosure
The authors report no conflicts of interest in this work.

References


