Original Article

Validity and screening properties of three depression rating scales in a prospective sample of patients with severe traumatic brain injury

Marcelo L. Schwarzbold, Alexandre P. Diaz, Jean C. Nunes, Daniel S. Sousa, Alexandre Hohl, Ricardo Guarnieri, Marcelo N. Linhares, Roger Waltz

1 Center of Applied Neurosciences (CeNAp), Hospital Universitário da Universidade Federal de Santa Catarina (HU-UFSC), Florianópolis, SC, Brazil. 2 Neurosurgery Service, Hospital Governador Celso Ramos, Florianópolis, SC, Brazil. 3 Department of Clinical Medicine, HU-UFSC, Florianópolis, SC, Brazil.

Objective: To evaluate the validity and utility of the Hamilton Rating Scale for Depression (HAM-D), Beck Depression Inventory (BDI), and Hospital Anxiety and Depression Scale (HADS) as screening tools for depression after severe traumatic brain injury (TBI).

Methods: Forty-six consecutive survivors of severe TBI were evaluated at a median of 15 months after injury. Receiver operating characteristic (ROC) analysis was performed using HAM-D, BDI, and HADS as predictors, and the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) as gold standard.

Results: The area under the curve (AUC) for HAM-D was 0.89, and the optimal cutoff point was 7 (sensitivity 92.9%, specificity 78.1%); for the BDI, the AUC was 0.946 and the optimal cutoff point was 14 (sensitivity 92.3%, specificity 96.7%); for the HADS, the AUC was 0.947 and the optimal cutoff point was 9 (sensitivity 100%, specificity 80.7%); and for the HADS depression subscale, the AUC was 0.937 and the optimal cutoff point was 6 (sensitivity 92.9%, specificity 83.9%). There were no statistically significant differences among the AUCs.

Conclusion: Our findings support a high validity and utility for the HAM-D, BDI, and HADS as screening tools for depression in patients with severe TBI, without major changes in standard cutoff points.

Keywords: Traumatic brain injury; depression; ROC curve; screening

Introduction

Patients who sustained traumatic brain injury (TBI) often exhibit neurobehavioral symptoms, especially depression. Rates of depression vary widely among studies and are largely dependent on the sampling method, but usually exceed those found in the general population, ranging from 18 to 61%. In addition, the occurrence of a first major depressive episode is disproportionately high after head injury. Depression can be then considered a sequela of TBI, although its psychosocial determinants and biological mechanisms remain largely unknown. Depression after TBI has been associated with poor functional, social, and occupational outcomes as well as decreased health-related quality of life, even when compared with other neurobehavioral sequelae. Improving diagnosis and treatment of depression has the potential to mitigate the general burden of TBI.

In primary care and other non-psychiatric settings, depression is frequently misdiagnosed. Studies have shown that less than one-half of patients with major depression receive a formal diagnosis and appropriate treatment. On the other hand, overdiagnosis and overtreatment have also been described in community settings. To overcome these issues, the use of depression rating scales has been suggested as a valid and practical strategy for screening for depression in patients with general medical conditions. Screening tools cannot be used to diagnose depression, but they can identify cases requiring a more detailed assessment, reducing the chances of non-recognition of the disorder or misuse of antidepressants. Most depression rating scales were originally designed to measure the severity of depressive symptoms, but subsequent studies indicated their utility as screening tools by investigating and defining appropriate cutoff points for specific populations.

A particularly powerful method for defining diagnostic cutoff points is the receiver operating characteristic (ROC) analysis. Thus far, few studies have performed ROC analysis in the context of screening for depression following TBI. Most authors found that depression rating scales are highly valid, as indicated by high area under the curve (AUC) values, but others reported poor...
These studies assessed patients with heterogeneous severity of injury, most commonly recruited from rehabilitation centers. Addressing more homogeneous and well-defined samples will add to current knowledge and clarify the usefulness of particular instruments in specific populations.

In the present study, patients consecutively admitted to intensive care due to severe TBI were evaluated in the chronic phase by using a well-established structured psychiatric interview to diagnose depression and three well-known rating scales for depression. ROC analysis was then performed to determine the validity and optimal diagnostic cutoff points of the depression rating scales.

Methods

Subjects

The protocols were approved by the Human Research Ethics Committee of Universidade Federal de Santa Catarina (UFSC) and written informed consent was given by patients and relatives. The study was conducted on patients consecutively admitted from 2006 to 2011 to Hospital Governador Celso Ramos, a Level I Trauma Center in the metropolitan area of Florianópolis, state of Santa Catarina, in southern Brazil. Patients were included if they sustained a severe TBI as defined by a Glasgow Coma Scale (GCS) score of 8 or lower on admission; if they were 16 years of age or older at the time of injury; and if they lived in the Florianópolis metropolitan area. Cases of gunshot injury were not included. Demographic, clinical, and psychiatric data were obtained prospectively on the basis of direct evaluation of patients or by interviewing family members, according to our research protocol.13-16

Psychiatric assessment

Psychiatric interviews were conducted by two experienced psychiatrists (MLS and APD) in outpatient hospital facilities or at the patient’s home. Most patients were accompanied by family members (usually a spouse or parents) who were able to provide additional information if necessary. At first, the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) was applied to determine psychiatric diagnoses (Table 1).17 The diagnosis of personality changes (not covered by the SCID-I) was defined based on the DSM-IV-TR criteria, which include symptoms of apathy, disinhibition, affective lability, suspiciousness, aggression, or combined types.18 The SCID-I has been translated and validated for the Brazilian population.19 Depression was defined by the presence of a major depressive episode in major depressive disorder at the time of the interview (that is, patients with previous depressive episodes that did not meet criteria when the interview was conducted were not considered cases of depression). The dichotomous variable anxiety disorders and obsessive-compulsive disorder (OCD) (Table 1) covered the diagnosis of panic disorder, generalized anxiety disorder, posttraumatic stress disorder, and OCD. The dichotomous variable substance abuse or dependence included problems with any substances addressed by the SCID-I.

After SCID-I administration, the same psychiatrist applied the following depression rating scales: the 17-item Hamilton Rating Scale for Depression (HAM-D),20 the Beck Depression Inventory (BDI),21 and the Hospital Anxiety and Depression Scale (HADS).22 These instruments have been translated and validated for use in Brazil.23-25 Since a considerable proportion of patients had a low educational level (Table 1) and the first evaluated patients demonstrated difficulty in reading the self-report scales (BDI and HADS), the interviewers took the approach of neutrally reading items for all patients, leaving them free to choose the option they considered the most appropriate.

Statistical analysis

Statistical analyses were carried out using SPSS version 17.0 for Windows (SPSS Inc., Chicago, IL, USA) and MedCalc 12.7 for Windows (MedCalc Software, Ostend, Belgium). ROC curves were plotted using GraphPad Prism 6 for Windows (GraphPad Software Inc., La Jolla, CA, USA). The Kolmogorov-Smirnov goodness-of-fit test was used to determine whether a numerical variable could be assumed to be normally distributed. If this was the case, an unpaired t test was performed to compare patients with and without depression; otherwise, a Mann-Whitney U test was applied. The associations between depression and independent categorical variables were analyzed using logistic regression and expressed as odds ratios (OR) with a 95% confidence interval (95%CI).

Empirical ROC curves were generated using SCID-I diagnosis of depression as the gold standard and HAM-D, BDI, or HADS scores as predictors. In the absence of gold-standard tests in psychiatry, structured interviews are generally accepted as such for practical and definite purposes.26 A separate analysis was performed for the depression subscale of the HADS. Sensitivity, specificity, and positive and negative predictive values at possible cutoff points (coordinates of the curve) were calculated with 95% exact binomial confidence intervals. The AUC values were calculated by trapezoidal summation with 95% exact binomial confidence intervals and compared between curves and with the null hypothesis (AUC value of 0.5).27 The coordinate nearest to the upper left corner and the Youden index for each curve were defined to illustrate the selection of optimal cutoff points.28

Results

Figure 1 shows the sampling flow. Of 99 survivors of severe TBI who were consecutively admitted to an intensive care unit (ICU), 68 fulfilled the inclusion criteria and 46 (67.6%) completed the study. Depression affected 30.4% of patients, which was also the frequency of substance abuse or dependence and personality changes. The most common type of personality changes was the apathetic type (42.8%), followed by the aggressive (21.4%) and disinhibited (21.4%) types; two patients
(14.4%) had a combined type with apathy and aggression. Anxiety disorders were diagnosed in 19.4% of the sample; only one patient had OCD, with onset prior to TBI and comorbidity with generalized anxiety disorder.

Table 1 describes the demographic, clinical, and psychiatric characteristics of the sample, as well as comparisons between patients with and without depression. Patients underwent psychiatric evaluation at a median of 15 months after hospital admission (range, 9-35 months). They were predominantly male (82.6%), young (mean ± standard deviation [SD] of age at injury, 31.1 ± 11.5 years; range, 16-68 years), and most (63.1%) had not completed secondary education (mean ± SD of years of education, 10.2 ± 4.9; range, 0-23 years). Traffic incidents were the main cause of TBI (77.3%). The mean ± SD of the GCS score at admission was 5.4 ± 1.9;
the median (interquartile range) was 6 (3-7). Univariate analysis revealed an inverse association between depression and duration of mechanical ventilation (p = 0.03), as well as a trend toward association between depression and length of ICU stay (p = 0.08). Depression was not associated with personality changes or substance abuse or dependence, but there was a trend toward association with anxiety disorders (p = 0.08). As expected, patients with depression exhibited higher HAM-D, BDI, and HADS scores (p < 0.001).

Figure 2 displays ROC curves generated using depression diagnosed by the SCID-I as gold standard and HAM-D, BDI, HADS, and HADS depression subscale scores as predictors. Confidence intervals of coordinates were also plotted. The optimal cutoff point for each curve was highlighted, as determined as the point closest to the upper left corner and the Youden index. For each curve both criteria indicated the same coordinates: for the HAM-D, a score of 7 or more led to a sensitivity of 92.9% and specificity of 78.1%; for the BDI, a score of 14 or more led to a sensitivity of 92.3% and specificity of 96.7%; for the HADS, a score of 9 or more led to a sensitivity of 100% and specificity of 80.7%; and for the HADS depression subscale, a score of 6 or more led to a sensitivity of 92.9% and specificity of 83.9%.

Table 2 lists AUC values as well as selected coordinates and their respective sensitivity, specificity, and predictive values, with confidence intervals. For the HAM-D, the AUC was 0.89; for the BDI, the AUC was 0.946; for the HADS, the AUC was 0.947; and for the HADS depression subscale, the AUC was 0.937. All curves had an AUC significantly greater than 0.5 (p < 0.0001 for all comparisons), but did not differ between each other (p > 0.05 for all comparisons).

Figure 1 Recruitment of patients for psychiatric evaluation in chronic phase of severe traumatic brain injury. Adult patients were consecutively admitted to intensive care due to non-missile trauma. Survivors who were not in a persistent vegetative state and lived in the metropolitan region of Florianópolis, southern Brazil, were eligible for the study.

Discussion

Improvements in the prehospital and hospital care of patients with TBI have contributed to a reduction in TBI mortality in the last decade. This is certainly positive news, but it may lead to a greater demand for health services for survivors. Improvements in diagnosis and treatment of depression may reduce the overall burden of TBI, because depression is frequent in these patients and strongly affects health-related quality of life.

The frequencies of psychiatric disorders in the present study were similar to those reported in the literature, with depression and personality changes as the most frequent psychiatric disorders following TBI. A trend toward association between depression and anxiety disorders was found, reflecting a high comorbidity which has been already reported in patients with TBI and is well known in the general population. Depression was associated with duration of mechanical ventilation and ICU stay, but in a direction opposite to what might be expected, considering the high prevalence of emotional symptoms in general ICU survivors. The basis of these associations remains elusive. Most studies on the topic did not find a relationship between depression and indicators of TBI severity, as occurs with posttraumatic stress disorder, for example. Thus far, few studies have employed ROC analysis to assess the validity of depression rating scales in patients with TBI. In ROC analysis, an AUC greater than 0.9 is considered indicative of high overall validity for a test, while 0.7 to 0.9 indicates moderate validity, and 0.5 to 0.7, low validity. One previous study assessed patients with TBI of different severities (predominantly mild cases) and found an AUC of 0.97 using the Patient Health...
Questionnaire-9. Another study\textsuperscript{10} reported an AUC of 0.88 for the Beck Depression Inventory-II in a sample consisting of military veterans with mainly mild and moderate TBI; this sample also had other distinctive characteristics, such as a high incidence of posttraumatic stress disorder and older age. In a community sample\textsuperscript{9} of patients with mild to severe TBI (45% of severe cases), an AUC of 0.82 was found for the HADS depression subscale; in this study, the definition of depression was broader, with the inclusion of dysthymia. More recently, AUCs of 0.87 for the HADS and 0.9 for the Depression Anxiety Stress Scales were reported in a sample with heterogeneous severity of TBI (44% of severe cases) recruited from a rehabilitation center.\textsuperscript{11} In contrast, the HADS performed virtually at chance level in an Omani sample\textsuperscript{12} of outpatients, with an AUC of 0.53; as discussed by the authors, cultural and language-related factors may underlie this exception. In line with most previous investigations, the present study found high validity and discriminatory ability for all the evaluated depression rating scales.

A cutoff point of 7 led to the optimal trade-off between sensitivity and specificity for the HAM-D. This value is quite compatible with cutoff points for the general population (in which a value of 7 or less is generally considered normal and is also a criterion of remission) and for patients with stroke (a disease sharing several clinical and pathophysiological similarities with TBI).\textsuperscript{31,32} It is of practical importance for the clinician that the cutoff point of 7 also had a high negative predictive value (NPV), but higher cutoff points (such as 11 or 15) would be necessary for a high positive predictive value (PPV). In the case of the BDI, the cutoff point of 14 led to elevated sensitivity, specificity, and predictive values. This score is in the range originally indicated to characterize mild to moderate depression,\textsuperscript{21} but it is up to 4 points above the cutoff for patients with stroke considering similar sensitivity and specificity.\textsuperscript{33,34} In the case of the HADS, one review\textsuperscript{35} of studies conducted in patients with general medical conditions described optimal cutoff points ranging from 3 to 11 for the depression subscale, with 8 as

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{ROC_curves.png}
\caption{Receiver operating characteristic (ROC) curves generated using depression diagnosed by the Structured Clinical Interview for DSM-IV Axis I Disorders as gold standard and scores on the 17-item Hamilton Rating Scale for Depression (HAM-D), Beck Depression Inventory (BDI), and Hospital Anxiety and Depression Scale (HADS) as predictors in patients with severe traumatic brain injury. Curves and coordinates are represented by solid lines with circles. The white circle in each curve corresponds to the optimal cutoff point as defined by the Youden index and the criterion of the point closest to the upper left corner. Dashed lines represent the confidence interval of coordinates.}
\end{figure}
the cutoff most commonly found; this is also the score originally suggested by the authors of the scale.\textsuperscript{22} With regard to TBI, one previous study\textsuperscript{3} found a sensitivity of 62% and specificity of 92% for a score of 8 or higher on the HADS, whereas another\textsuperscript{11} found a sensitivity of 80% and specificity of 73% for this same cutoff point. In the present study, the optimal cutoff point for the HADS depression subscale was 6, but a score of 8 or more determined additionally high specificity and predictive values while retaining acceptable sensitivity.

Although the present results are consistent with previous research, with the advantage of having been obtained from a homogeneous and well-defined sample, some limitations must be discussed. Firstly, the sample size is relatively small, as reflected by the large confidence intervals in many coordinates of the curves. Even so, to our knowledge, only one previous study\textsuperscript{11} performed a similar investigation of a larger number of severe cases of TBI (55 patients), which were recruited from a rehabilitation center and pooled for analysis with mild and moderate cases. Secondly, almost one-third of eligible patients did not complete the study. Loss to follow-up is a common problem in studies addressing long-term outcomes of TBI, with sample sizes typically declining by half within 2 years of injury.\textsuperscript{36,37} This appears to be an issue inherent to TBI populations.\textsuperscript{36} However, a previous investigation\textsuperscript{7} with the same cohort of patients of the present study found no difference between subjects who were and were not assessed in the chronic phase regarding demographic and early-phase clinical variables such as sex, age, GCS score, Marshall computed tomography classification, associated trauma, length of hospital stay, and basic laboratory parameters. Although this observation does not rule out a selection bias in the chronic phase, it indicates a lack of effect of the available data for this cohort. Finally, one may argue that reading self-report scales to patients or applying them in the same occasion of SCID-I administration may have overestimated reporting of depressive symptoms. Again, this cannot be completely ruled out, but it seems unlikely, as discussed by other authors,\textsuperscript{10} because determining the diagnostic properties of the scales was not an objective at the time of the interviews. Moreover, in a previously mentioned study,\textsuperscript{12} the interviewers read the HADS to all subjects and the scale performed poorly. In fact, the approach of reading self-report scales as neutrally as possible may be a valuable strategy for subjects with low educational attainment, a common situation in developing countries.

In conclusion, the present results are indicative of high validity for the HAM-D, BDI, and HADS as instruments for assessment of depression in patients with severe TBI. They are also generally consistent with standard recommended diagnostic cutoff points, suggesting that these scales are useful as screening tools for depression in this population without major adjustments. However, due to the current paucity of data in the literature, further studies with larger samples are necessary to confirm these recommendations, particularly in the setting of severe TBI. Addressing the utility of psychometric tools for the screening of other neurobehavioral problems following TBI (such as personality changes or anxiety disorders) and investigating the role of biomarkers not only in mortality\textsuperscript{14,15} but also in long-term sequelae\textsuperscript{16} are important directions for future research.

### Table 2

Statistics and selected coordinates of the ROC curves generated using HAM-D, BDI, and HADS scores as predictors and SCID-I diagnosis of depression as gold standard in patients with severe traumatic brain injury

<table>
<thead>
<tr>
<th>Scale and cutoff points</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HAM-D</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 7*</td>
<td>92.9 (66.1-99.8)</td>
<td>78.1 (60.0-90.7)</td>
<td>65.0 (40.8-84.6)</td>
<td>96.2 (80.4-99.9)</td>
<td>0.890 (0.762-0.963)</td>
</tr>
<tr>
<td>≥ 8</td>
<td>85.7 (57.2-98.2)</td>
<td>81.3 (63.6-92.8)</td>
<td>66.7 (41.0-86.7)</td>
<td>92.9 (76.1-99.2)</td>
<td>0.947 (0.836-0.992)</td>
</tr>
<tr>
<td>≥ 11</td>
<td>78.6 (49.2-95.3)</td>
<td>90.6 (75.0-98.0)</td>
<td>78.6 (49.2-95.3)</td>
<td>90.6 (75.0-98.0)</td>
<td>0.946 (0.831-0.992)</td>
</tr>
<tr>
<td>≥ 15</td>
<td>57.1 (28.9-82.3)</td>
<td>100 (89.1-100)</td>
<td>100 (63.1-100)</td>
<td>84.2 (68.7-94.0)</td>
<td>0.890 (0.762-0.963)</td>
</tr>
<tr>
<td>BDI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 3</td>
<td>100 (75.3-100)</td>
<td>23.3 (9.9-42.3)</td>
<td>36.1 (20.8-53.8)</td>
<td>100 (59.0-100)</td>
<td>0.946 (0.831-0.992)</td>
</tr>
<tr>
<td>≥ 14*</td>
<td>92.3 (64.0-99.8)</td>
<td>96.7 (82.8-99.9)</td>
<td>92.3 (64.0-99.8)</td>
<td>96.7 (82.8-99.9)</td>
<td>0.947 (0.836-0.992)</td>
</tr>
<tr>
<td>≥ 17</td>
<td>84.6 (54.6-98.1)</td>
<td>100 (88.4-100)</td>
<td>100 (69.2-100)</td>
<td>93.7 (79.2-99.2)</td>
<td>0.946 (0.836-0.992)</td>
</tr>
<tr>
<td><strong>HADS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 9*</td>
<td>100 (76.8-100)</td>
<td>80.7 (62.5-92.5)</td>
<td>70.0 (45.7-88.1)</td>
<td>100 (85.8-100)</td>
<td>0.947 (0.836-0.992)</td>
</tr>
<tr>
<td>≥ 10</td>
<td>85.7 (57.2-98.2)</td>
<td>83.9 (66.3-94.5)</td>
<td>70.6 (44.0-89.7)</td>
<td>92.9 (76.5-99.1)</td>
<td>0.947 (0.836-0.992)</td>
</tr>
<tr>
<td>≥ 17</td>
<td>71.4 (41.9-91.6)</td>
<td>96.8 (83.3-99.9)</td>
<td>90.9 (58.7-99.8)</td>
<td>88.2 (72.5-96.7)</td>
<td>0.947 (0.836-0.992)</td>
</tr>
<tr>
<td>≥ 21</td>
<td>50.0 (23.0-77.0)</td>
<td>100 (88.8-100)</td>
<td>100 (59.0-100)</td>
<td>81.6 (65.7-92.3)</td>
<td>0.947 (0.836-0.992)</td>
</tr>
<tr>
<td><strong>HADS depression subscale</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 2</td>
<td>100 (76.8-100)</td>
<td>41.9 (24.5-60.9)</td>
<td>43.7 (26.4-62.3)</td>
<td>100 (75.3-100)</td>
<td>0.937 (0.822-0.987)</td>
</tr>
<tr>
<td>≥ 6*</td>
<td>92.9 (66.1-99.8)</td>
<td>83.9 (66.3-94.5)</td>
<td>72.2 (46.5-90.3)</td>
<td>96.3 (61.0-99.9)</td>
<td>0.937 (0.822-0.987)</td>
</tr>
<tr>
<td>≥ 8</td>
<td>78.6 (49.2-95.3)</td>
<td>96.8 (83.3-99.9)</td>
<td>91.7 (61.5-99.8)</td>
<td>90.9 (75.7-98.1)</td>
<td>0.937 (0.822-0.987)</td>
</tr>
<tr>
<td>≥ 11</td>
<td>57.1 (28.9-82.3)</td>
<td>100 (88.8-100)</td>
<td>100 (63.1-100)</td>
<td>83.8 (68.0-93.8)</td>
<td>0.937 (0.822-0.987)</td>
</tr>
</tbody>
</table>

AUC = area under the curve; BDI = Beck Depression Inventory; HADS = Hospital Anxiety and Depression Scale; HAM-D = 17-item Hamilton Rating Scale for Depression; PPV = positive predictive value; NPV = negative predictive value; ROC = receiver operating characteristic; SCID-I = Structured Clinical Interview for DSM-IV Axis I Disorders.

Values expressed as percentage (95% confidence interval).

* Cutoff point indicated by the Youden index and corresponding to the coordinates closest to the upper left corner.
Acknowledgements
The authors thank Prof. Emil Kupek for statistical assistance. This work was supported by PRONEX Program (NENASC Project) and PPSUS Program of Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina (FAPESC) and the National Science and Technology Institute for Translational Medicine (INCT-TM).

Disclosure
The authors report no conflicts of interest.

References