Evaluation of the Traumatic Coma Data Bank Computed Tomography Classification for Severe Head Injury

PIETER E. VOS,1 ANCO C. VAN VOSKUILEN,1 TJEMME BEEMS,2 PAUL F. M. KRABBE,3 and OSCAR J. M. VOGELS1

ABSTRACT

This study determines the interrater and intrarater reliability of the Traumatic Coma Data Bank (TCDB) computed tomography (CT) scan classification for severe head injury. This classification grades the severity of the injury as follows: I = normal, II = diffuse injury, III = diffuse injury with swelling, IV = diffuse injury with shift, V = mass lesion surgically evacuated, or VI = mass lesion not operated. Patients with severe closed head injury were included. Outcome was assessed using the Glasgow Outcome Score (GOS) at 3 and 6 months. Four observers, two of them classifying the scans twice, independently evaluated CT scans. Of the initial CT scans of 63 patients (36 males, 27 females; age, 34 ± 24 years), 6.3% were class I, 26.9% class II, 28.6% class III, 6.3% class IV, 22.2% were class V, and 9.6% class VI. The overall interrater and intrarater reliability was 0.80 and 0.85, respectively. Separate analyses resulted in higher inter- and intrarater reliabilities for the mass lesion categories (V and VI), 0.94 and 0.91, respectively, than the diffuse categories (I–IV) 0.71 and 0.67. Merging category III with IV, and V with VI resulted in inter- and intrarater reliabilities of 0.93 and 0.78, respectively. Glasgow outcome scores after 6 months were as follows: 19 dead (30%), one vegetative (2%), five severely disabled (8%), 17 moderately disabled (27%), and 21 good recovery (33%). Association measures (Sommers' D) between CT and GOS scores were statistically significant for all observers. This study shows a high intra- and interobserver agreement in the assessment of CT scan abnormalities and confirms the predictive power on outcome when the TCDB classification is used.

Key words: computed tomography scan; Glasgow Outcome Scale; outcome; prognostic factor; severe head injury; traumatic brain injury; Traumatic Coma Data Bank classification

INTRODUCTION

Accurate prediction of long-term outcome after emergency admission to the hospital is useful as an aid to clinical decision making in patients with (severe) head injury (Signorini et al., 1999). Age, Glasgow Coma Score (GCS), pupil reactivity, and Injury Severity Score have been shown to be independent predictors of outcome. The practical reliability of the GCS and pupil reactivity is high and is accepted overall (Braakman et al., 1977; Teasdale et al., 1978).

The treatment of patients with head injury has been...
improved considerably by the use of computed tomography (CT), especially with respect to quick assessment of the presence of surgical mass lesions. In 1991, Marshall et al. proposed a CT scan classification of head injury that added to the category of diffuse injury the presence of surgical mass lesions (Marshall et al., 1991). This Trauma Coma Databank (TCDB) classification utilized the status of the mesencephalic cisterns, the degree of midline shift in millimeters, and the presence or absence of intracerebral or extracerebral hematoma to define six categories (Table 1). Four categories indicate diffuse injury of increasing severity, whereas two categories indicate the presence of a mass lesion (Marshall et al., 1991). Although not proven on an individual basis, the TCDB classification can be used to predict the development of intracranial hypertension or a possible fatal outcome, and it is currently widely used (Lobato et al., 1997; Marshall et al., 1991; Stocchetti et al., 2000). If, for example, the CT classification is used as an aid in randomization procedures in treatment trials of patients with severe head injury, the power may be increased with a reduction in sample size (Machado et al., 1999).

Although the TCDB classification is an adjunct to clinical parameters and is easy to use, it does not take into account all the possible prognostic factors visible on CT and has clearly some limitations. An important limitation is that the classification is partially based upon arbitrary assessments and the accuracy of measured volumes of focal mass lesions. Furthermore, TCDB categories V and VI are, in part, retrospective in nature because of the surgical decision to operate any (mass) lesion or not. Besides, this CT classification method has never been validated for its interrater and intrarater reliability. The purpose of this study was to investigate the overall intrarater and interrater variability of the TCDB classification when used by nonradiologists. Differences among different raters in the accuracy of assessing volumes of mass lesions and the diffuse injury categories were also determined separately. The correlation between the TCDB classification and outcome as measured with the Glasgow Outcome Scale at 3 and 6 months was also measured in this study.

**MATERIALS AND METHODS**

Patients with severe closed head injury admitted to the University Medical Centre Nijmegen between November 1996 and September 1998 were included. Severe head injury was defined as a GCS of ≤8 following resuscitation (Teasdale and Jennett, 1974). Patients had an immediate CT scan taken after resuscitation. Exclusion criteria were as follows: no CT scan taken or a CT scan taken after 12 h. Data collected from medical records included age, sex, and GCS. Outcome was assessed using the Glasgow Outcome Score (GOS; Jennett and Bond, 1975) at 3 and 6 months, by one rater (P.E.V.) during the regular patient visits to the clinic or from the charts of the regular visits (in a few patients). All initial CT scans were independently evaluated by four observers, two of them classifying the scans twice, with a time interval of at least 2 weeks. The CT scans of all patients were classified according to the TCDB criteria (Marshall et al., 1991; Table 1). All scans were covered with a cardboard form to blind the raters for date, name, age, and sex. Individual standardized forms were used to enter coded CT scan data. Predefined criteria were used to ensure that all raters used the same set of criteria: the status of the mesencephalic cisterns was considered present, absent, or compressed; the degree of midline shift (taken as the displacement in millimeters of the septum pellucidum) was measured in millimeters. The volume of low/mixed or high density lesions, either intracerebral or extracerebral, was estimated according to the ellipsoidal method (Pasqualin et al., 1991). This volume is \( \frac{4}{3} \pi \times 0.5 A \times 0.5 B \times 0.5 C \). A is the largest diameter, B is the diameter perpendicular to A, and C is the vertical diameter estimated by the number of slices in which the lesion is present.

**Table 1. Traumatic Coma Data Bank Computed Tomography Categories for Head Injury**

<table>
<thead>
<tr>
<th>Category</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diffuse injury I</td>
<td>No visible intracranial pathology</td>
</tr>
<tr>
<td>Diffuse injury II</td>
<td>Cisterns are present with midline shift of 0–5 mm and/or lesion densities present, no high or mixed density lesions at ≥25 mL</td>
</tr>
<tr>
<td>Diffuse injury III</td>
<td>Cisterns compressed or absent, with midline shift of 0–5 mm, no high or mixed density lesions of ≥25 mL</td>
</tr>
<tr>
<td>Diffuse injury IV</td>
<td>Midline shift of &gt;5 mm, no high or mixed density lesion of ≥25 mL</td>
</tr>
<tr>
<td>Evacuated mass lesion (V)</td>
<td>Any lesion surgically evacuated</td>
</tr>
<tr>
<td>Nonevacuated mass lesion (VI)</td>
<td>High or mixed density lesion of ≥25 mL, not surgically evacuated</td>
</tr>
</tbody>
</table>

multiplied with slice thickness. The presence of a mass lesion with a volume greater than 25 mL that was not surgically evacuated denotes a TCDB score of V. Any patient that is operated upon for an intracranial mass lesion irrespective of size falls into category V (Table 1).

**Statistical Analyses**

With the “Variance Components” module of SPSS 9.0, a specific application of analysis of variance, the proportion of the total variance uniquely attributable to the object of measurement (TCDB classification), and the other sources of variance (raters and measurement error) were estimated. Subsequently, the interobserver agreement and the intraobserver agreement were determined based on the estimated proportions of variance by computing their associated intraclass coefficients (Shavelson et al., 1989). Analysis was performed for the standard TCDB, and four additional analyses were performed on recoded TCDB scores. Because a score of V is not based on a subjective evaluation, its assignment fully depends on the fact whether a surgical treatment has been performed or not. The first transformation we performed was to merge the scores V and VI together (A). To define a CT scan as TCDB category V, information was gathered from the patient charts. To evaluate the contribution of the diffuse categories and the mass lesion categories, two separate analyses were done on category I–IV (B) and category V and VI (C). A fourth analysis (D) was done by recoding the original TCDB scores in three diffuse categories and one mass volume category defined as follows: normal = I, diffuse injury = II, diffuse injury with swelling or shift = III (merging category III and IV), and mass lesion = IV (merging category V and VI).

To investigate the relation of the TCDB classification with outcome (GOS at 3 and 6 months), the original TCDB categories (I–VI) were included in the analysis. Association measures (Sommers’ D) were computed for each of the four assessors between their raw CT scores and the GOS scores at 3 and 6 months (Siegel and Castellan, 1988). Sommers’ D is a measure of association between two ordinal variables that ranges from −1 to 1. This statistic is particularly appropriate when two obser-

<table>
<thead>
<tr>
<th>TCDB category</th>
<th>Interrater ICC</th>
<th>Intrarater ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>I–VI (overall)</td>
<td>0.80</td>
<td>0.85</td>
</tr>
<tr>
<td>A: I–VI (recoding V = VI)</td>
<td>0.83</td>
<td>0.87</td>
</tr>
<tr>
<td>B: I–IV (separately)</td>
<td>0.71</td>
<td>0.67</td>
</tr>
<tr>
<td>C: V–VI (separately)</td>
<td>0.94</td>
<td>0.91</td>
</tr>
<tr>
<td>*D: I–IV (IV = III, V = IV, VI = IV)</td>
<td>0.93</td>
<td>0.78</td>
</tr>
</tbody>
</table>

TCDB, Trauma Coma Data Bank; ICC, intraclass coefficient; *D, the original TCDB classification recoded as I = normal, II = diffuse injury, III = diffuse injury with swelling or shift, and IV = mass lesion.
vations of two ordered variables (TCDB classification and GOS) have many ties (same score for both variables). In addition, Sommers’ D offers a measure of association that is suited to deal with a situation where one of the two variables is of particular importance or when one is a dependent variable and the other is an independent variable (Siegel and Castellan, 1988).

To assess discrimination, the receiver operating characteristic (ROC) curve was calculated to estimate the severity and specificity for the different scores of the classification in relation to the clinical outcome (measured by GOS scores at 3 and 6 months). For this purpose, the GOS was dichotomized in unfavorable (GOS 1–3) versus favorable (GOS 4–5) outcome. In addition, the area under the ROC curve was estimated as a measure of predictive discrimination.

RESULTS

Sixty-three patients (36 males, 27 females; age range, 4–82 years; mean age [±SD], 34 [±24] years) had a Glasgow Coma Score of ≤8 after resuscitation. The classification results of the principal rater (P.E.V.) were 6.3% class I, 26.9% class II, 28.6% class III, 6.3% class IV, 22.2% class V, and 9.6% for class VI (Table 2). Statistical analysis revealed an overall interrater reliability of 0.80 and an overall intrarater reliability of 0.85 for the four raters (Table 3). Recoding scores V and VI as a V resulted in an interrater reliability of 0.83 and an intrarater reliability of 0.87 (Table 3A). When category I–IV was analyzed separately, interrater reliability and intrarater reliability were 0.71 and 0.67, respectively (Table 3B). Separate analysis of category V and VI resulted in an interrater reliability of 0.82 and an intrarater reliability of 0.83 (Table 3C). Merging category III with IV and V with VI resulted in an overall interrater and intrarater reliability of 0.93 and 0.78, respectively (Table 3D).

**TABLE 4. ASSOCIATION MEASURES (SOMMER’S D) BETWEEN THE COMPUTED TOMOGRAPHY CORES AND THE GOS CORES AT 3 AND 6 MONTHS FOR EACH OF THE FOUR ASSESSORS**

<table>
<thead>
<tr>
<th>Assessor</th>
<th>GOS 3</th>
<th>GOS 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>−0.53</td>
<td>−0.45</td>
</tr>
<tr>
<td>2</td>
<td>−0.40</td>
<td>−0.33</td>
</tr>
<tr>
<td>3</td>
<td>−0.47</td>
<td>−0.42</td>
</tr>
<tr>
<td>4</td>
<td>−0.45</td>
<td>−0.38</td>
</tr>
</tbody>
</table>

All association measures are statistically significant (p < 0.01).

GOS, Glasgow Outcome Score.

The average reading time to obtain a TCDB classification of the initial CT scan of patients with severe head injury was approximately 6 min.

Glasgow outcome scores of the 63 patients after 6 months were as follows: 19 dead (30%), one vegetative (2%), five severely disabled (8%), 17 moderately disabled (27%), and 21 good recovery (33%) (Table 2). When the GOS was dichotomized in poor outcome versus good outcome, the association measures (Sommers’ D) between the CT scores and the GOS scores at 3 and 6 months were statistically significant for all the four raters (Table 4). The area under the curve for the ROC was 0.79 (SD = 0.03) at month 3 and 0.74 (SD = 0.03) at month 6 (Fig. 1). Not one of the six CT scores showed a relatively high sensitivity related with a low specificity or vice versa.

DISCUSSION

In this study, four clinical observers without special training in neuroradiology made use of a predefined set of criteria as defined by the TCDB Study to classify the initial CT scan of patients with severe head injury (Marshall et al., 1991). In this study, we used the intraclass
correlation coefficient (ICC) to determine intra- and interobserver agreement. The computations performed by the ICC are identical to the standard kappa (Streiner and Norman, 1995). The main reason for applying the ICC method is that this approach is far more flexible. Based on an underlying theory, all the sources of variability in a single study can be systematically combined, utilising all the data to estimate the variance between subjects and the various components of error (Shavelson et al., 1989). Therefore, in one single analysis, more than two raters, repeated measurements, and even evaluations of different diagnostic tools may be included. The overall agreement between the four raters (interobserver reliability) when all TCDB scores were considered proved to be high (0.80), whereas the intraobserver agreement was even higher (0.85). The volumes of low/mixed or high density lesions either intracerebral or extracerebral were assessed with the ellipsoid method (Pasqualin et al., 1991). Separation analysis of the diffuse categories (I–IV) and the mass lesion categories (V–VI) showed lower ICC scores for the diffuse categories underscoring the subjectivity of these categories (Table 3B,C). In this respect it is suggested that the TCDB classification may show a better interobserver agreement when diffuse injury III and IV are taken together and category V and VI are taken together. Recording the TCDB scores in normal (I), diffuse injury (II), diffuse injury with swelling or shift (III), and mass lesion (IV) did indeed show an higher interrater reliability coefficient of 0.93. But this procedure assumes that determination of a mass volume greater than 25 mL has 100% agreement among the different raters. Little disagreement occurred among the different raters in measuring the exact volume as shown by an interrater variability coefficient of 0.94 when this key aspect of category V and VI was assessed separately. This is in line with the study of Stocchetti, where the gold standard for measuring the volume of CT mass lesions was a computer based tracing by a neuroradiologist (Stocchetti et al., 2000). Out of 19 lesions with a volume of >25 mL, 17 could be identified by three intensive care unit staff members with a special training in the identification and measurement of areas and volumes on the CT slices when using the ellipsoid method.

The average reading time using different methods for the assessment of volumes of mass lesions on CT scans of patients with severe head injury has been shown to lie between 4 min and 27 sec, and 7 min and 44 sec (Stocchetti et al., 2000). In this study, we found the average reading time for the TCDB classification to be 6 min.

For this study, we agreed on stringent TCDB criteria for making the classification. This means that the results of this study are the potential agreement, which might be achieved based on the TCDB classification, rather than what is achieved in daily practice.

The main goal of this study was to verify interrater and intrarater reliability for the initial CT scan of patients with severe head injury. Up to 51% of patients with severe head injury may develop significant changes on repeat CT scans within 48 h (Lobato et al., 1997). The time window for inclusion in this study was set at 12 h. This was based on the finding that the majority of new mass lesions not present at the admission CT scan develop within 12 h after injury and the fact that treatment trials often use the CT scan taken within a 12-h window after injury to randomize patients (Harders et al., 1996; Marmarou et al., 1999; Smith and Miller, 1991; Yamaki et al., 1990).

The high interobserver agreement found in this study is comparable with the reliability of clinical parameters such as age, the GCS, and pupillary reactions used in the assessment of the severity of head injury (Braakman et al., 1977; Signorini et al., 1999; Teasdale et al., 1978).

Sommers’ D analysis confirmed the relationship between the CT scan diagnosis based on the TCDB classification and outcome after severe head injury for all assessors (Table 4). In the original paper by Marshall outcome was determined in 746 TCDB patients at discharge from the hospital. The CT scan diagnosis was a highly significant independent predictor of mortality when age and the Glasgow Coma motor score were included in a logistic regression model (Marshall et al., 1991). Even with the relatively small number of patients included in our study, a strong association between any TCDB diagnosis and outcome at 3 and 6 months after injury was found.

Although the TCDB classification seems widely accepted and shows high reliability there are limitations that hamper its applicability. The lack of perfect intraobserver and interobserver agreement must be attributed to differences between observers in assessing volumes of mass lesions and boundary lines of anatomical or pathological brain structures or misclassifying the basal cisterns as normal, compressed or absent. In this respect, the distinction between category III (diffuse injury with swelling) and IV (diffuse injury with shift) and II and III (diffuse injury without swelling) is important. It is known that absent or compressed basal cisterns on the first CT scan is an ominous predictor of outcome in severe head injury (Toutant et al., 1984). However, the classification of diffuse injury with or without swelling and with or without midline shift is arbitrary.

The major distinction between category V and VI is whether the patient is operated or not. This means that the differentiation between category V and VI is made by treatment, not by assessing the CT characteristics. Be-
sides category V patients may have a mass lesion greater than 25 mL (category VI) but also smaller volumes are categorized as V if they are operated upon. Redefining category V and VI as any lesion with a volume greater than 25 mL will probably enhance the predictive power of the TCDB classification.

Several other CT classifications exist (Teasdale et al., 1992) but not any that has been better evaluated than the TCDB classification. Based on a subgroup of head-injured patients with subarachnoid haemorrhage, Greene et al. proposed a simple CT grading scale with four categories that incorporated thickness of subarachnoid blood as well as the presence of mass lesions (Greene et al., 1995). The study demonstrated a significant association between CT diagnosis and outcome at the time of discharge from acute hospitalization. Although the information needed for completion of the CT forms was extracted and documented independently by a neurosurgeon and a neuroradiologist no intra- and interrater variability was determined (Greene et al., 1995). The TCDB classification does not take into account the presence of traumatic subarachnoid haemorrhage. The incidence of CT documented traumatic subarachnoid haemorrhage in patients with severe traumatic brain injury is 23–63% (Kakarieka, 1997). It is most common in patients with subdural hematoma or hemorrhagic contusion and associated with a worse prognosis (Greene et al., 1995; Kakarieka, 1997). Two large randomized controlled trials showed a promising modest increase in the proportion of favorable outcome in patients with severe head injury treated with nimodipine (Bailey et al., 1991; Harders et al., 1994). But after reanalysis it was stated that the overall benefit of treating unselected head injured patients with nimodipine is unlikely to be clinically relevant and that more detailed studies at this topic are necessary (Murray et al., 1996). Adding subarachnoid haemorrhage in any CT classification system may enhance the predictive power on outcome and have consequences for treatment and outcome issues and in clinical trials (Greene et al., 1995; Kakarieka, 1997). It should be noted however that adding subarachnoid hemorrhage as another CT parameter open to subjective error may decrease the intra- and interobserver agreement of any CT classification.

This study also showed predictive power of the TCDB scores on the GOS at 3 and 6 months. However, no scores were detected for which the sensitivity/specificity combination may be considered as sufficiently adequate for clinical practice. Hence, combinations of other clinical parameters with the CT classification may still result in more accurate prediction of outcome and seems required (Machado et al., 1999).

In conclusion, the TCDB CT classification system for severely head injured patients has a high interobserver- and intraobserver reliability when used by clinicians without special training in neuroradiology. The agreement among different raters in the assessment of mass lesions is higher than for the diffuse categories. This study also confirmed the strong relationship between TCDB CT category and outcome. The appropriateness of the TCDB classification can be improved if the retrospective category V is omitted or if category V and VI are combined in one category that assesses the volume of any mass lesion. Probably also combining the diffuse categories III and IV may further enhance the predictive power. These issues as well as including subarachnoid hemorrhage as a CT parameter have to be the subject of a larger prospective study.

ACKNOWLEDGMENTS

We thank Dr. Hans Merx, neuroradiologist, for assistance with the TCDB criteria.

REFERENCES


Address reprint requests to:
Pieter E. Vos, M.D., Ph.D.
University Medical Centre Nijmegen
P.O. Box 9101
6500 HB, Nijmegen, The Netherlands

E-mail: p.vos@czzoneu.azn.nl