

Interobserver Variability in the Assessment of CT Imaging Features of Traumatic Brain Injury

Kimberly A. Chun,¹ Geoffrey T. Manley,^{2,3} Shirley I. Stiver,^{2,3} Ashley H. Aiken,¹ Nicholas Phan,^{2,3}
Vincent Wang,^{2,3} Michele Meeker,^{2,3} Su-Chun Cheng,⁴ A.D. Gean,^{1,3} and Max Wintermark¹

Abstract

The goal of our study was to determine the interobserver variability between observers with different backgrounds and experience when interpreting computed tomography (CT) imaging features of traumatic brain injury (TBI). We retrospectively identified a consecutive series of 50 adult patients admitted at our institution with a suspicion of TBI, and displaying a Glasgow Coma Scale score ≤ 12 . Noncontrast CT (NCT) studies were anonymized and sent to five reviewers with different backgrounds and levels of experience, who independently reviewed each NCT scan. Each reviewer assessed multiple CT imaging features of TBI and assigned every NCT scan a Marshall and a Rotterdam grading score. The interobserver agreement and coefficient of variation were calculated for individual CT imaging features of TBI as well as for the two scores. Our results indicated that the imaging review by both neuroradiologists and neurosurgeons were consistent with each other. The kappa coefficient of agreement for all CT characteristics showed no significant difference in interpretation between the neurosurgeons and neuroradiologists. The average Bland and Altman coefficients of variation for the Marshall and Rotterdam classification systems were 12.7% and 21.9%, respectively, which indicates acceptable agreement among all five reviewers. In conclusion, there is good interobserver reproducibility between neuroradiologists and neurosurgeons in the interpretation of CT imaging features of TBI and calculation of Marshall and Rotterdam scores.

Key words: accuracy; computed tomography; imaging scores; interobserver agreement; traumatic brain injury

Introduction

EVERY YEAR, TRAUMATIC BRAIN INJURY (TBI) leaves more than 5 million people with significant disabilities that require lifelong medical care at an annual cost of more than \$56 billion for the United States alone. As a consequence of the wars in Iraq and Afghanistan, about 20% of U.S. soldiers deployed to the Middle East are suffering from TBI, many of whom will require long-term medical care.

The identification and classification of TBI patients most likely to benefit from different types of medical and surgical treatment is a difficult challenge (Doppenberg et al., 1997; Marshall et al., 1992; Narayan et al., 2002). Currently, the Glasgow Coma Scale (GCS), a clinical scale that reflects level of consciousness, is the main tool for classification of patients in TBI clinical trials (Saatman et al., 2008). Although GCS has been found to be extremely useful in the clinical management

and prognosis of TBI, it grades only clinical symptom severity, but it does not distinguish between different types of injury. For instance, a GCS score of 8 or less is typically classified as severe TBI, but different lesions on CT may be responsible for the patient's GCS score. It is difficult to effectively treat multiple types of injury based only on the clinical severity or GCS score (Saatman et al., 2008). Another limitation to the GCS score is that it is difficult to measure GCS on admission because TBI patients are frequently intubated and/or sedated. Moreover, the GCS score has been shown to reliably predict mortality at 24 h after injury, but has only limited ability to predict longer-term outcome (Steyerberg et al., 2008).

Imaging characteristics from noncontrast head computed tomography (NCT) scans in TBI patients have become pivotal in the acute evaluation of TBI, and NCT scanning can identify the presence and extent of structural damage in these patients (Maas et al., 2007). CT imaging features of TBI have been

¹Department of Radiology, Neuroradiology Section, ³Brain and Spinal Injury Center, and ⁴Department of Epidemiology and Biostatistics, University of California–San Francisco, San Francisco, California.

²Department of Neurological Surgery, University of California–San Francisco, and San Francisco General Hospital, San Francisco, California.

TABLE 1. MARSHALL COMPUTED TOMOGRAPHY (CT) CLASSIFICATION

Category	Definition
Diffuse injury I (no visible pathology)	No visible intracranial pathology seen on CT scan
Diffuse injury II	Cisterns are present with midline shift 0–5 mm and/or lesion densities present; no high or mixed density lesion >25 cc may include bone fragments and foreign bodies
Diffuse injury III (swelling)	Cisterns compressed or absent with midline shift 0–5 mm; no high or mixed density lesion >25 mm
Diffuse injury IV (shift)	Midline shift >5 mm; no high or mixed density lesion >25 cc
Evacuated mass lesion	Any lesion surgically evacuated
Nonevacuated mass lesion	High or mixed density lesion >25 cc; not surgically evacuated

Reproduced from Maas et al., 2007.

combined into scores such as the Marshall (Marshall et al., 1991) and the Rotterdam scores (Maas et al., 2005). Marshall and Rotterdam scores determined at the time of admission to the hospital have been shown to correlate with long-term clinical outcome following TBI. By stratifying TBI patients into categories of varying severity and types of injuries, NCT imaging might allow improved triaging to appropriate treatments that more effectively target a patient's specific injuries (Borg et al., 2004).

An important issue that has been raised in the use of CT characteristics in the prediction of clinical outcome is how reliably different observers can interpret CT imaging features of TBI, when considered individually or combined into the Marshall and Rotterdam scores. In October 2007, the National Institute of Neurological Disorders and Stroke, together with the Brain Injury Association of America, the Defense and Veterans Brain Injury Center, and the National Institute of Disability and Rehabilitation Research, put forth a consensus opinion that the Marshall and Rotterdam classification systems needed additional validation of their reliability and standardization (Saatman et al., 2008).

The purpose of our study was to determine the interobserver variability between observers with different backgrounds and levels of experience when interpreting CT characteristics of TBI.

Methods

Study design

Imaging data obtained as part of standard clinical care at our institution were retrospectively reviewed with the approval of the institutional review board. At our institution, an NCT scan of the brain is a routine component of care for all patients with suspected TBI.

From the San Francisco General Hospital Trauma Registry, which identifies all trauma patients admitted to our hospital, we identified 50 consecutive patients with TBI and an admission GCS score ≤ 12 admitted between November 2004 and June 2005. Victims of penetrating trauma were excluded from this study. Patient charts were reviewed and demographic data were recorded.

Imaging reviews

The NCT studies were anonymized and sent to five readers. Five reviewers with different backgrounds and

levels of experience independently reviewed 50 NCT scans. Two of the readers were neuroradiologists (A.A. and M.W.). A.A. has 7 years of experience and M.W. has 11 years of experience. The remaining three readers were neurosurgeons (S.S., N.P., and V.W.). S.S. has 10 years of experience, N.P. has 12 years of experience, and V.W. has 4 years of experience.

The different reviewers each evaluated the NCT exams for standard imaging features of TBI, including: the presence or absence of skull fracture, epidural hematoma (EDH), subdural hematoma (SDH), subarachnoid hemorrhage (SAH), intraventricular hemorrhage (IVH), contusion, brain parenchymal hematoma, traumatic axonal injury (TAI), and mass effect manifested as a midline shift >5 mm or effacement/absence of the basal cisterns. The reviewers each assigned two numerical scores to every NCT based on the guidelines set by the Marshall and Rotterdam classification scales, as shown in Tables 1 and 2 (Maas et al., 2005, 2007). The midline shift was measured quantitatively. The intracranial hematoma volume was calculated by adding the area of hemorrhage for each slice and multiplying each area by the slice thickness.

TABLE 2. ROTTERDAM COMPUTED TOMOGRAPHY CLASSIFICATION

Predictor value	Score
Basal cisterns	
Normal	0
Compressed	1
Absent	2
Midline shift	
No shift or shift ≤ 5 mm	0
Shift >5 mm	1
Epidural mass lesion	
Present	0
Absent	1
Intraventricular blood or subarachnoid hemorrhage	
Absent	0
Present	1
Sum score	+1

In the Rotterdam scoring system, one point is added as a sum score to make the Rotterdam grade numerically total 6 points, consistent with the motor score of the Glasgow Coma Scale and with the Marshall classification. Reproduced from Maas et al., 2005.

TABLE 3. DIAGNOSTIC PERFORMANCE OF EACH INDIVIDUAL REVIEWER FOR DETECTION OF ANY ABNORMALITY ON THE CT SCANS OF 50 TBI PATIENTS

	<i>Sensitivity</i>	<i>Specificity</i>	<i>NPV</i>	<i>PPV</i>	<i>Accuracy</i>
Neuroradiologist #1	95.0%	80.0%	80.0%	95.0%	93.0%
Neuroradiologist #2	95.0%	90.0%	82.0%	98.0%	94.0%
Neurosurgeon #1	91.0%	60.0%	60.0%	91.0%	86.0%
Neurosurgeon #2	100.0%	90.0%	100.0%	98.0%	98.0%
Neurosurgeon #3	98.0%	60.0%	86.0%	91.0%	91.0%
Average	95.8%	76.0%	81.6%	94.6%	92.4%

NPV, negative predictive value; PPV, positive predictive value; TBI, traumatic brain injury; CT, computed tomography.

Statistical analysis

Sensitivity, specificity, positive and negative predictive values, and accuracy were calculated to assess the diagnostic performance of each reviewer and for each of the CT features of TBI. The standard against which the accuracy of the reviewer's interpretation was gauged was determined by comparison to the consensus opinion among the different reviewers after they completed their individual reviews. The prevalence of each individual CT feature of TBI was calculated and compared against the consensus interpretation.

The interobserver agreement was determined using Cohen's kappa statistics for qualitative features, and Bland and Altman's coefficient of variation for quantitative and ordinal features. To evaluate the overall agreement between reviewers in their interpretation of individual CT imaging features, the kappa coefficients for all pairs of readers were averaged and graded as follows: values below 0.2 indicate poor agreement; between 0.21 and 0.4, fair agreement; between 0.41 and 0.6, moderate agreement; between 0.61 and 0.8, good agreement; and >0.8, near perfect agreement (Altman, 1991).

Bland and Altman's coefficient of variation for quantitative and ordinal measurements was calculated as a ratio of the absolute difference between the measurements performed by the two reviewers over the mean value of the observers (Rosner, 1995). A coefficient of variation that is less than 30% indicates acceptable reliability, and one that is greater than 30% indicates unacceptable variability (Rosner, 1995).

Results

Study patients

Fifty consecutive patients were retrospectively identified who met our inclusion criteria. The median age of the patients was 40 years, the interquartile range (IQR) was 32–51 years, and the range was 18–65 years. Thirty-two patients were males and 18 were females. The median admission GCS score was 8, the IQR was 5–10, and the range was 3–12.

By consensus, 23 patients (46.0%) were found to have a skull fracture, 2 patients (4.0%) had EDH, 23 patients (46.0%) had an SDH, 30 patients (60.0%) had SAH, 7 patients (14.0%) had IVH, 17 (33.3%) patients had contusions, 11 patients (22.0%) had TAI, 2 patients (3.9%) presented with parenchymal hematomas, and 21 patients (42.0%) displayed a midline shift >5 mm or had effacement of the basal cisterns.

Accuracy of reader interpretations of CT imaging features of TBI

The accuracy, sensitivity, specificity, and positive and negative predictive values for each reader in their ability to differentiate an abnormal from a normal CT in TBI patients are shown in Table 3. Table 4 further describes accuracy, sensitivity, specificity, positive and negative predictive values, averaged over all readers, for detection of any CT abnormality, skull fracture, EDH, SDH, SAH, IVH, contusions, parenchymal hematomas, midline shift >5 mm, and patent basal cisterns. Results from this diagnostic performance

TABLE 4. AVERAGE DIAGNOSTIC PERFORMANCE OF ALL FIVE REVIEWERS WHEN EVALUATING FOR INDIVIDUAL CT IMAGING FEATURES OF TBI PATIENTS

	<i>Frequency (n = 50)</i>	<i>Sensitivity</i>	<i>Specificity</i>	<i>NPV</i>	<i>PPV</i>	<i>Accuracy</i>
Any abnormality	41 (82.0%)	95.8%	76.0%	81.6%	94.6%	92.4%
Skull fracture	23 (46.0%)	90.8%	84.0%	91.0%	84.2%	87.2%
EDH	2 (4.0%)	90.0%	97.6%	99.6%	79.2%	97.4%
SDH	23 (46.0%)	78.0%	94.6%	85.2%	92.2%	87.0%
SAH	30 (60.0%)	94.4%	84.6%	92.0%	90.4%	90.6%
IVH	7 (14.0%)	94.6%	97.2%	96.8%	83.4%	94.6%
Contusions	17 (34.0%)	65.6%	90.6%	85.6%	77.0%	82.6%
Parenchymal hematomas	2 (4.0%)	73.4%	98.0%	98.4%	71.8%	96.4%
TAI	11 (22.0%)	63.4%	94.2%	91.6%	73.4%	87.8%
Midline shift >5 mm	18 (36.0%)	90.0%	89.4%	94.0%	84.6%	89.8%
Effaced/absent basal cisterns	21 (42.0%)	92.6%	90.2%	93.8%	89.4%	91.2%

EDH, epidural hemorrhage; SDH, subdural hemorrhage; SAH, subarachnoid hemorrhage; IVH, intraventricular hemorrhage; TAI, traumatic axonal injury; NPV, negative predictive value; PPV, positive predictive value; TBI, traumatic brain injury; CT, computed tomography.

TABLE 5. INTEROBSERVER AGREEMENT (COHEN'S KAPPA COEFFICIENTS) BETWEEN PAIRS OF READERS WHEN EVALUATING ANY ABNORMAL CT FEATURE OF TBI

	<i>Any CT feature of TBI</i>
Neuroradiologist #1 + Neuroradiologist #2	0.585
Neuroradiologist #1 + Neurosurgeon #1	0.509
Neuroradiologist #1 + Neurosurgeon #2	0.681
Neuroradiologist #1 + Neurosurgeon #3	0.411
Neuroradiologist #2 + Neurosurgeon #1	0.466
Neuroradiologist #2 + Neurosurgeon #2	0.754
Neuroradiologist #2 + Neurosurgeon #3	0.599
Neurosurgeon #1 + Neurosurgeon #2	0.555
Neurosurgeon #1 + Neurosurgeon #3	0.413
Neurosurgeon #2 + Neurosurgeon #3	0.706
Average	0.568

TBI, traumatic brain injury; CT, computed tomography.

analysis indicate that the imaging interpretation by both neuroradiologists and neurosurgeons were consistent with each other and accurate. All the reviewers had comparable performances, some more sensitive, others more specific. Sensitivity was highest for detection of the presence of any CT abnormality, IVH, and SAH. Contusions, parenchymal hematomas, and TAI were identified with the lowest rates of sensitivity. The readers demonstrated highest specificity in their evaluation of EDH, IVH, and parenchymal hematomas. These same CT features (EDH, IVH, and parenchymal hematomas) were also associated with the highest accuracy among the readers. There was no significant difference in terms of accuracy of the readers depending on the GCS (mean accuracy for detection of any CT abnormality: 92.3% in the group with GCS scores 0–8, and 91.2% in the group with GCS scores 9–12; $p = 0.998$).

Interobserver agreement

The average interobserver agreement (Cohen's kappa coefficient) between all readers for detection of any abnormality in the CT imaging of TBI was 0.568, indicating moderate

agreement between all readers (Table 5). There was no significant difference in terms of kappa coefficient depending on the GCS (mean accuracy for any abnormality: 0.547 in the group with GCS scores 0–8, 0.567 in the group with GCS scores 9–12; $p = 0.392$).

The kappa coefficients between different pairs of reviewers for each CT feature of TBI are displayed in Table 6. The kappa coefficients ranged from 0.411 (Neuroradiologist #1 and Neurosurgeon #3) to 0.754 (Neurosurgeon #2 and Neuroradiologist #2). The kappa coefficient of agreement for detection of any CT abnormality demonstrated fair to good agreement between reviewers, with no significant difference between the neurosurgeons and neuroradiologists. Interobserver agreement was fair for detection of contusions and parenchymal hematomas; moderate for skull fractures, EDH, SDH, and TAI; and good agreement between reviewers was achieved for identification of SAH, IVH, midline shift >5 mm, and patent basal cisterns.

The average Bland and Altman coefficients of variation between all reviewers in their assignments of Marshall and Rotterdam classification scores were 12.7% and 21.9%, respectively ($p < 0.001$) (Table 7). This indicates acceptable agreement among all five reviewers for both the Marshall and Rotterdam scores.

The average Bland and Altman coefficients of variation between reviewers' calculations and measurement of EDH, SDH, parenchymal hematoma, and contusion volumes, as well as midline shift, are shown in Table 8. Acceptable agreement between all five readers was demonstrated for EDH volume, parenchymal volume, and midline shift measurements, with average Bland and Altman coefficients of 7.2%, 9.6%, and 28.2%, respectively, for these features. The average Bland and Altman coefficients of variation for SDH and contusion volumes were 51.0% and 51.4%, respectively, indicating poor agreement between the five readers for these measurements.

Discussion

In 1991, Marshall and associates proposed a CT classification system that identifies six different groups of patients with TBI, based on the type and severity of abnormalities seen on

TABLE 6. AVERAGE INTEROBSERVER AGREEMENT (COHEN'S KAPPA COEFFICIENTS) BETWEEN ALL READERS IN COMPARISON TO THE INTEROBSERVER AGREEMENT BETWEEN ONLY NEURORADIOLOGISTS, ONLY NEUROSURGEONS, AND NEURORADIOLOGISTS AND NEUROSURGEONS WHEN EVALUATING FOR EACH CT FEATURE OF TBI

	<i>Frequency (n = 50)</i>	<i>All reviewers</i>	<i>Neuroradiology versus neuroradiology</i>	<i>Neurosurgery versus Neurosurgery</i>	<i>Neuroradiology versus Neurosurgery</i>
Any abnormality	41 (82.0%)	0.568	0.585	0.558	0.570
Skull fracture	23 (46.0%)	0.590	0.547	0.630	0.577
EDH	2 (4.0%)	0.577	0.791	0.431	0.615
SDH	23 (46.0%)	0.531	0.733	0.419	0.554
SAH	30 (60.0%)	0.669	0.496	0.712	0.676
IVH	7 (14.0%)	0.683	0.571	0.623	0.731
Contusions	17 (34.0%)	0.332	0.628	0.214	0.342
Parenchymal hematomas	2 (4.0%)	0.400	0.298	0.470	0.383
TAI	11 (22.0%)	0.423	0.418	0.302	0.484
Midline shift >5 mm	18 (36.0%)	0.652	0.762	0.639	0.641
Effaced/absent basal cisterns	21 (42.0%)	0.691	0.842	0.630	0.697

EDH, epidural hemorrhage; SDH, subdural hemorrhage; SAH, subarachnoid hemorrhage; IVH, intraventricular hemorrhage; TAI, traumatic axonal injury; TBI, traumatic brain injury; CT, computed tomography.

TABLE 7. BLAND AND ALTMAN COEFFICIENT OF VARIATION BETWEEN PAIRS OF REVIEWERS WHEN ASSESSING THE MARSHALL AND ROTTERDAM SCORES

	Marshall score	Rotterdam score
Neuroradiologist #1 + Neuroradiologist #2	9.2%	30.9%
Neuroradiologist #1 + Neurosurgeon #1	15.1%	31.6%
Neuroradiologist #1 + Neurosurgeon #2	11.4%	31.5%
Neuroradiologist #1 + Neurosurgeon #3	10.5%	33.0%
Neuroradiologist #2 + Neurosurgeon #1	16.0%	33.7%
Neuroradiologist #2 + Neurosurgeon #2	13.6%	28.9%
Neuroradiologist #2 + Neurosurgeon #3	9.2%	36.7%
Neurosurgeon #1 + Neurosurgeon #2	13.0%	26.4%
Neurosurgeon #1 + Neurosurgeon #3	15.3%	21.0%
Neurosurgeon #2 + Neurosurgeon #3	13.2%	30.3%
Average	12.7%	21.9%

CT scan (Marshall et al., 1991). To improve outcome prediction, the Rotterdam CT scoring system was developed in 2005 by Maas and colleagues, and is based on adding individual CT characteristics to the Marshall classification. These individual CT characteristics include: IVH and traumatic SAH (tSAH), and a detailed differentiation of mass lesions and basal cisterns (Maas et al., 2005). The ability to stratify TBI patients is improved by adding tSAH, because this characteristic has been shown to be a strong independent predictor of outcome and mortality (Eisenberg et al., 1990; Greene et al., 1996, 1995; Kakarieka et al., 1994; Ono et al., 2001; Selladurai et al., 1992; Servadei et al., 2002). Furthermore, based on studies showing that patients with an epidural hematoma have a much better clinical prognosis than those with a subdural or intracerebral hematoma, the Rotterdam CT classification system distinguishes the type of mass lesion to provide additional discriminative value (Chestnut et al., 2000; Genarelli et al., 1982). Both the Marshall and Rotterdam classification systems have been found to be predictive of clinical outcome, and are included in the international guidelines for clinical management and prognosis of TBI (Chestnut et al., 2000).

The primary purpose of this study was to determine the agreement between independent observers of varying backgrounds when evaluating CT imaging features of TBI, in-

cluding those used to categorize patients according to the Marshall and Rotterdam CT scoring systems. Our data show that there is no significant difference between multiple readers from varying backgrounds when assessing typical CT imaging features of TBI. Regardless of background, the reviewers scored an average kappa coefficient corresponding to an agreement of fair or better for all CT characteristics. Furthermore, the diagnostic accuracy achieved by all five readers was consistently high and did not significantly differ across the various CT imaging features.

The Marshall score showed better reproducibility than the Rotterdam score. This is likely because the Marshall score is more of a summary score. Indeed, features in the Marshall system are not independent of one another; rather, they are connected. On the contrary, in the Rotterdam scoring system, each feature stands alone and gets counted individually. This makes the Rotterdam score a more reliable score (Maas et al., 2005), but also makes it more prone to variability.

We acknowledge several limitations to our paper. A limitation of kappa statistics is that they are influenced by prevalence (Viera and Garrett, 2005). The calculation of the kappa coefficient is based on the difference between how much an agreement is actually present compared to how much an agreement is due to chance alone. If the prevalence of a certain feature is low, it is likely that different readers agree on the absence of that feature just by chance, resulting in a lower kappa coefficient. Also, our study sample was limited in size.

This was a single-center study, and patterns of CT interpretation may be influenced by fact that the reviewers have worked together. Further validation across multiple institutions would be important for future studies.

In conclusion, interobserver reproducibility is very good between multiple readers with varying backgrounds when interpreting CT imaging features of TBI, and in calculating the Marshall and Rotterdam scores. Thus NCT can be considered a reliable tool to stratify TBI patients into categories of varying severity and types of injuries, and might be used to improve triaging to appropriate treatments that more effectively target a patient's specific injuries.

Acknowledgments

Max Wintermark receives funding from the National Center for Research Resources (grant KL2 RR024130). The content of the article is solely the responsibility of the authors

TABLE 8. BLAND AND ALTMAN COEFFICIENTS OF VARIATION AVERAGES ACROSS ALL REVIEWERS FOR MARSHALL AND ROTTERDAM SCORES, AS WELL AS FOR EDH VOLUME, SDH VOLUME, PARENCHYMAL HEMATOMA VOLUME, CONTUSION VOLUME, AND MIDLINE SHIFT MEASUREMENT

	Median	IQR	Standard deviation	Overall	Neuroradiology versus neuroradiology	Neurosurgery versus neurosurgery	Neuroradiology versus neurosurgery
Marshall score	13.1	4.6	2.4	12.7%	9.2%	13.8%	12.6%
Rotterdam score	21.6	4.4	2.6	21.9%	18.7%	22.4%	22.3%
EDH volume	5.1	9.0	5.0	7.2%	4.1%	9.4%	8.4%
SDH volume	46.5	9.7	10.0	51.0%	39.1%	45.6%	54.4%
Parenchymal hematoma volume	9.7	3.0	2.3	9.6%	13.3%	8.2%	9.6%
Contusion volume	48	17.5	11.4	51.4%	39.0%	55.2%	51.6%
Midline shift measurement	25.1	9.9	7.1	28.2%	24.1%	30.6%	27.7%

EDH, epidural hemorrhage; SDH, subdural hemorrhage; IQR, interquartile range.

and does not necessarily represent the official views of the National Institute of Neurological Disorders and Stroke, the National Center for Research Resources, the National Institutes of Health, or the other sponsors.

Author Disclosure Statement

No competing financial interests exist.

References

- Altman, D.G. (1991). *Practical Statistics for Medical Research*. Chapman and Hall: London, pps. 404–408.
- Borg, J., Holm, L., Cassidy, J.D., Peloso, P.M., Carroll, L.J., von Holst, H., and Ericson, K. (2004). Diagnostic procedures in mild traumatic brain injury: results of the WHO Collaborating Centre Task Force on Mild Traumatic Brain Injury. *J. Rehabil. Med.* (43 Suppl.), 61–75.
- Chestnut, R., Ghajar, J., and Maas, A.I. (2000). Management and prognosis of severe traumatic brain injury. Part 2: Early indicators of prognosis in severe traumatic brain injury. *J. Neurotrauma* 17, 557–627.
- Doppenberg, E.M., Choi, S.C., and Bullock, R. (1997). Clinical trials in traumatic brain injury. What can we learn from previous studies? *Ann. N.Y. Acad. Sci.* 825, 305–322.
- Eisenberg, H.M., Gary, H.E., Jr., Aldrich, E.F., Saydjari, C., Turner, B., Foulkes, M.A., Jane, J.A., Marmarou, A., Marshall, L.F., and Young, H.F. (1990). Initial CT findings in 753 patients with severe head injury. A report from the NIH Traumatic Coma Data Bank. *J. Neurosurg.* 73, 688–698.
- Gennarelli, T.A., Spielman, G.M., Langfitt, T.W., Gildenberg, P.L., Harrington, T., Jane, J.A., Marshall, L.F., Miller, J.D., and Pitts, L.H. (1982). Influence of the type of intracranial lesion on outcome from severe head injury. *J. Neurosurg.* 56, 26–32.
- Greene, K.A., Jacobowitz, R., Marciano, F.F., Johnson, B.A., Spetzler, R.F., and Harrington, T.R. (1996). Impact of traumatic subarachnoid hemorrhage on outcome in nonpenetrating head injury. Part II: Relationship to clinical course and outcome variables during acute hospitalization. *J. Trauma* 41, 964–971.
- Greene, K.A., Marciano, F.F., Johnson, B.A., Jacobowitz, R., Spetzler, R.F., and Harrington, T.R. (1995). Impact of traumatic subarachnoid hemorrhage on outcome in nonpenetrating head injury. Part I: A proposed computerized tomography grading scale. *J. Neurosurg.* 83, 445–452.
- Kakarika, A., Braakman, R., and Schakel, E.H. (1994). Clinical significance of the finding of subarachnoid blood on CT scan after head injury. *Acta Neurochir. (Wien.)* 129, 1–5.
- Maas, A.I., Hukkelhoven, C.W., Marshall, L.F., and Steyerberg, E.W. (2005). Prediction of outcome in traumatic brain injury with computed tomographic characteristics: a comparison between the computed tomographic classification and combinations of computed tomographic predictors. *Neurosurgery* 57, 1173–1182; discussion 1173–1182.
- Maas, A.I., Steyerberg, E.W., Butcher, I., Dammers, R., Lu, J., Marmarou, A., Mushkudiani, N.A., McHugh, G.S., and Murray, G.D. (2007). Prognostic value of computerized tomography scan characteristics in traumatic brain injury: results from the IMPACT study. *J. Neurotrauma* 24, 303–314.
- Marshall, L.F., Eisenberg, H., Jane, J.A., Marshall, S.B., and Klauber, M.R. (1991). A new classification of head injury based on computerized tomography. *J. Neurosurg.* 75, S14–S20.
- Marshall, L.F., Marshall, S.B., Klauber, M.R., Van Berkum Clark, M., Eisenberg, H., Jane, J.A., Luerksen, T.G., Marmarou, A., and Foulkes, M.A. (1992). The diagnosis of head injury requires a classification based on computed axial tomography. *J. Neurotrauma* 9(Suppl. 1), S287–S292.
- Narayan, R.K., Michel, M.E., Ansell, B., Baethmann, A., Biegon, A., Bracken, M.B., Bullock, M.R., Choi, S.C., Clifton, G.L., Contant, C.F., Coplin, W.M., Dietrich, W.D., Ghajar, J., Grady, S.M., Grossman, R.G., Hall, E.D., Heetderks, W., Hovda, D.A., Jallo, J., Katz, R.L., Knoller, N., Kochanek, P.M., Maas, A.I., Majde, J., Marion, D.W., Marmarou, A., Marshall, L.F., McIntosh, T.K., Miller, E., Mohberg, N., Muizelaar, J.P., Pitts, L.H., Quinn, P., Riesenfeld, G., Robertson, C.S., Strauss, K.I., Teasdale, G., Temkin, N., Tuma, R., Wade, C., Walker, M.D., Weinrich, M., Whyte, J., Wilberger, J., Young, A.B., and Yurkewicz, L. (2002). Clinical trials in head injury. *J. Neurotrauma* 19, 503–557.
- Ono, J., Yamaura, A., Kubota, M., Okimura, Y., and Isobe, K. (2001). Outcome prediction in severe head injury: analyses of clinical prognostic factors. *J. Clin. Neurosci.* 8, 120–123.
- Rosner, B. (1995). *Fundamentals of Biostatistics*. Duxbury Press: Belmont, CA, p. 327.
- Saatman, K.E., Duhaim, A.C., Bullock, R., Maas, A.I., Valadka, A., and Manley, G.T. (2008). Classification of traumatic brain injury for targeted therapies. *J. Neurotrauma* 25, 719–738.
- Selladurai, B.M., Jayakumar, R., Tan, Y.Y., and Low, H.C. (1992). Outcome prediction in early management of severe head injury: an experience in Malaysia. *Br. J. Neurosurg.* 6, 549–557.
- Servadei, F., Murray, G.D., Teasdale, G.M., Dearden, M., Iannotti, F., Lapiere, F., Maas, A.J., Karimi, A., Ohman, J., Persson, L., Stocchetti, N., Trojanowski, T., and Unterberg, A. (2002). Traumatic subarachnoid hemorrhage: demographic and clinical study of 750 patients from the European brain injury consortium survey of head injuries. *Neurosurgery* 50, 261–267; discussion 267–269.
- Steyerberg, E.W., Mushkudiani, N., Perel, P., Butcher, I., Lu, J., McHugh, G.S., Murray, G.D., Marmarou, A., Roberts, I., Habbema, J.D., and Maas, A.I. (2008). Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. *PLoS Med.* 5, e165; discussion e165.
- Viera, A.J., and Garrett, J.M. (2005). Understanding interobserver agreement: the kappa statistic. *Fam. Med.* 37, 360–363.

Address correspondence to:

Max Wintermark, M.D., M.A.S.

University of California—San Francisco

Department of Radiology, Neuroradiology Section

505 Parnassus Avenue, Box 0628

San Francisco, CA 94143-0628

E-mail: Max.Wintermark@radiology.ucsf.edu