JNS

CLINICAL ARTICLE

J Neurosurg 127:32–35, 2017

Reliability assessment of the Biffl Scale for blunt traumatic cerebrovascular injury as detected on computer tomography angiography

Paul M. Foreman, MD,[1] Christoph J. Griessenauer, MD,[2] Kimberly P. Kicielinski, MD, MSPH,[1] Philip G. R. Schmalz, MD,[1] Brandon G. Rocque, MD, MS,[1] Matthew R. Fusco, MD,[3] Joseph C. Sullivan III, MD,[4] John P. Deveikis, MD,[1] and Mark R. Harrigan, MD[1]

Departments of [1]Neurosurgery and [4]Radiology, University of Alabama at Birmingham, Alabama; [2]Neurosurgical Service, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts; and [3]Department of Neurosurgery, Vanderbilt University, Nashville, Tennessee

OBJECTIVE Blunt traumatic cerebrovascular injury (TCVI) represents structural injury to a vessel due to high-energy trauma. The Biffl Scale is a widely accepted grading scheme for these injuries that was developed using digital subtraction angiography. In recent years, screening CT angiography (CTA) has been used to identify patients with TCVI. The reliability of this scale, with injuries assessed using CTA, has not yet been determined.

METHODS Seven independent raters, including 2 neurosurgeons, 2 neuroradiologists, 2 neurosurgical residents, and 1 neurosurgical vascular fellow, independently reviewed each presenting CTA of the neck performed in 40 patients with confirmed TCVI and assigned a Biffl grade. Ten images were repeated to assess intrarater reliability, for a total of 50 CTAs. Fleiss' multirater kappa (κ) and interclass correlation were calculated as a measure of interrater reliability. Weighted Cohen's κ was used to assess intrarater reliability.

RESULTS Fleiss' multirater κ was 0.65 (95% CI 0.61–0.69), indicating substantial agreement as to the Biffl grade assignment among the 7 raters. Interclass correlation was 0.82, demonstrating excellent agreement among the raters. Intrarater reliability was perfect (weighted Cohen's κ = 1) in 2 raters, and near perfect (weighted Cohen's κ > 0.8) in the remaining 5 raters.

CONCLUSIONS Grading of TCVI with CTA using the Biffl Scale is reliable.

https://thejns.org/doi/abs/10.3171/2016.7.JNS16849

KEY WORDS traumatic cerebrovascular injury; trauma; Biffl Scale; Biffl grade; dissection; reliability; interventional neurosurgery

Blunt traumatic cerebrovascular injury (TCVI) represents a structural defect in a vessel wall that is directly attributable to high-energy, nonpenetrating trauma.[3] Overall incidence of TCVI among patient admissions for blunt trauma is estimated at 1%.[4,5,11] Mechanisms of acute cerebral ischemia include thromboembolism and hemodynamic failure, with contemporary studies reporting overall ischemic stroke rates of 9%–12%, with rates as high as 26% in untreated patients.[4,11]

In 1999 the Denver group developed what came to be known as the Biffl Scale for the grading of TCVI.[1] This scale was not only intended to provide prognostic and therapeutic information, but to allow for systematic investigation of these injuries. Despite the widespread acceptance of the Biffl Scale, its reliability has not been formally evaluated. We tested the inter- and intrarater reliability of the Biffl Scale across a spectrum of clinicians by using widely available CT angiography (CTA).

Methods

A prospective study of TCVI was done at a single center in patients treated between January 2007 and December 2011. During this time, all patients admitted after blunt

ABBREVIATIONS CA = carotid artery; CTA = CT angiography; DSA = digital subtraction angiography; ICC = interclass correlation; TCVI = traumatic cerebrovascular injury; VA = vertebral artery.

SUBMITTED April 4, 2016. **ACCEPTED** July 26, 2016.

INCLUDE WHEN CITING Published online October 21, 2016; DOI: 10.3171/2016.7.JNS16849.

trauma with evidence of extracranial TCVI on screening neck CTA underwent digital subtraction angiography (DSA). The database that was maintained for this study was reviewed to identify a total of 40 cases in which TCVI was identified by screening CTA and then confirmed by follow-up DSA. This series of cases included 20 carotid artery (CA) injuries and 20 vertebral artery (VA) injuries. This study was performed with approval from the institutional review board.

Seven raters, including 2 neurosurgeons, 2 neuroradiologists, 2 neurosurgical residents, and 1 neurosurgical vascular fellow, independently reviewed each CTA and assigned a Biffl grade (Table 1); examples of Biffl Grade I–IV injuries are provided in Fig. 1. Per interpretation of CTA and DSA by the senior author (M.R.H.), the following distribution of Biffl-graded TCVIs was studied. The CA injuries included the following: 3 Grade I injuries, 9 Grade II injuries, 6 Grade III injuries, and 2 Grade IV injuries. The VA injuries included the following: 3 Grade I injuries, 6 Grade II injuries, 2 Grade III injuries, and 9 Grade IV injuries.

The distribution of selected images represented the distribution of incidence of presenting injuries at our institution; no Grade V injuries were available. Reviewers were blind to previous image interpretation and to all clinical information not contained within the single available CTA. All CTAs were acquired on a 40-section multidetector scanner. The images included axial, coronal, and sagittal slices of 6 mm, 3 mm, and 3 mm, respectively. Ten cases were repeated to assess intrarater reliability, for a total of 50 CTAs. Repeated images included: 1 Grade I CA injury, 2 Grade II CA injuries, 1 Grade III CA injury, 1 Grade IV CA injury, 1 Grade I VA injury, 1 Grade II VA injury, 1 Grade III VA injury, and 2 Grade IV VA injuries. Repeat images were randomly inserted into the image set.

Sample size was calculated using a method designed by Walter et al.,[13] which demonstrated that increasing the number of raters will decrease the number of observations required to achieve an adequate sample size. Using 0.58 as the minimum acceptable level of interrater reliability and 0.80 as the desired level of interrater reliability, based on $p = 0.05$ and 80% power, the sample size required for 7 raters is at least 20 images (22% error margin).

Fleiss' multirater kappa (κ) and interclass correlation (ICC) was calculated as a measure of overall agreement of Biffl grade assignment among the 7 raters. For intrarater reliability analysis, weighted Cohen's κ was used to assess repeat measurement agreement for each rater. Agreement measured by κ was interpreted as almost perfect with κ values between 0.81 and 1.00, substantial with κ values between 0.61 and 0.80, moderate with κ values between 0.41 and 0.60, fair with κ values between 0.21 and 0.40, and poor with κ values between 0 and 0.20.[6] An ICC > 0.75 was considered to have excellent agreement, with an ICC between 0.40 and 0.75 classified as fair to good, and < 0.40 considered poor agreement. All statistical analysis was performed using online programs (http://www.statstodo.com/CohenKappa_Pgm.php; https://department.obg.cuhk.edu.hk/researchsupport/IntraClass_correlation.asp) and SPSS version 21.0 software (IBM Corp.).

TABLE 1. Biffl Scale for traumatic cerebrovascular injury

Injury Grade	Description
I	Luminal irregularity or dissection w/ <25% luminal narrowing
II	Dissection or intramural hematoma w/ ≥25% luminal narrowing
III	Pseudoaneurysm
IV	Occlusion
V	Transection w/ free extravasation

Modified with permission from Biffl et al: Blunt carotid arterial injuries: implications of a new grading scale. *J Trauma* 47:845–853, 1999.

Results

Interrater Reliability

Fleiss' multirater κ was 0.65 (95% CI 0.61–0.69), indicating substantial agreement as to the Biffl grade assignment among the 7 raters. The ICC was 0.82, indicating excellent agreement among the raters (Table 2).

Intrarater Reliability

Intrarater reliability was perfect (weighted Cohen's $\kappa = 1$) in 2 raters and near perfect (weighted Cohen's $\kappa > 0.8$) in the remaining 5 raters (Table 2).

Overall Correlation Between DSA and CTA Grading

Of a total of 280 TCVI grades assigned by 7 independent reviewers based on CTA (10 CTAs repeated for intrarater reliability were not included; with 7 reviewers this totaled 70 TCVI grades), 211 (75.4%) grades matched the DSA grade assigned by the senior author at the time of angiography.

Discussion

Biomedical grading scales allow for the characterization of pathology, thus facilitating decision making, communication between physician and patient, communication among physicians, and systematic investigation. For a grading scale to be robust, it must be both valid and reliable. In the present study we evaluated the reliability—a test of consistency and reproducibility—of the Biffl Scale using CTA for the evaluation of TCVI, and found a substantial to excellent agreement among raters (interrater reliability) and a near-perfect agreement within a single rater (intrarater reliability).

The 5-tier Biffl Scale was originally published in 1999 in an effort to create a grading scale with prognostic and therapeutic implications that would also serve as a common language for future research.[1] The original description was derived from DSA and was applied to only the CA; subsequently, the scale was expanded to include VA injury as well. This scale is now widely accepted as a common language, enabling interphysician communication and systematic research. Moreover, TCVI subtypes, as described by the Biffl Scale, correlate with prognosis; higher-grade CA injuries carry a significantly higher risk of ischemic stroke as compared with other subtypes.[1] Although the reported stroke rate associated with a particular injury grade has varied among publications, the origi-

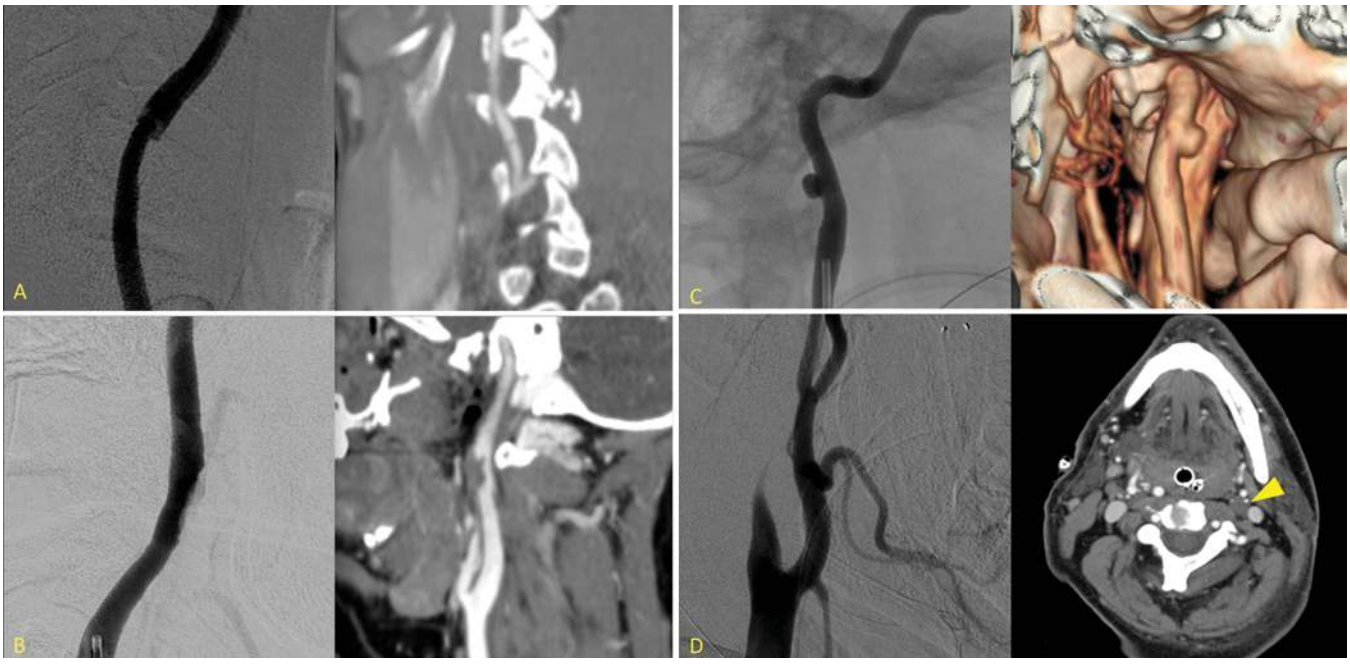

FIG. 1. Examples of Biffl Grade I–IV injuries; DSA (**left**) and CTA (**right**). Type I VA injury (**A**), Type II internal carotid artery (ICA) injury (**B**), Type III ICA injury (**C**), and Type IV ICA injury (**D**). Figure is available in color online only.

nal description attributed stroke rates of 3%, 11%, 33%, 44%, and 100% for carotid injury Grades I–V, respectively.[1] With the exception of the very rare Grade V injury, we treat all TCVIs (Grades I–IV) with 325 mg daily aspirin as first-line therapy. Despite its prevalent use, the reliability of this scale had not been formally tested.

Raters were selected from a spectrum of physicians involved in the care of patients with TCVI. Despite the improved sensitivity of CTA interpretations performed by neuroradiologists,[9] a formal neuroradiology interpretation may not be available due to the temporal nature of traumatic injury. Additionally, circumstances may dictate prompt clinical decision making, necessitating the initial review of the CTA by a non-neuroradiologist, or even a non-radiologist physician. Thus, it is useful to assess the reliability of the Biffl Scale as interpreted by both radiologist and non-radiologist physicians involved in the care of patients with TCVI.

Despite all patients in this study having a TCVI confirmed on CTA and DSA, only the CTA was used for inter- and intrarater assessment. Currently CTA is the diagnostic modality of choice for screening traumatically injured patients at risk for TCVI, with DSA reserved for select cases (i.e., symptomatic despite medical management, and high pretest probability with negative noninvasive imaging) or in patients in whom endovascular treatment is anticipated. Prospective studies assessing the accuracy of 16-section multidetector CTA compared with DSA in trauma patients at risk for TCVI found sensitivity, specificity, positive predictive values, and negative predictive values of 74%–97.7%, 86%–100%, 65%–99.3%, and 90%–99.3%, respectively.[2,7]

However, a 2013 systematic review, which included the above-mentioned studies among others, concluded that accuracy of CTA varied considerably across centers and suggested that CTA had a high specificity but low sen-

TABLE 2. Reliability of the Biffl Scale using CTA

Rater	Value (95% CI)		
	Fleiss' Multirater κ	ICC	Weighted Cohen's κ
Interrater reliability	0.65 (0.61–0.69)	0.82 (0.75–0.89)	
Intrarater reliability			
Neurosurgeon 1			1
Neurosurgeon 2			0.82 (0.59–1.04)
Neuroradiologist 1			1
Neuroradiologist 2			0.89 (0.7–1.09)
Resident neurosurgeon 1			0.91 (0.75–1.07)
Resident neurosurgeon 2			0.91 (0.75–1.07)
Vascular fellow			0.91 (0.75–1.07)

sitivity.[9] Variability was believed to be due to diagnostic threshold, number of available CT slices, and training, with increased sensitivity thought to be the result of an increased number of slices and neuroradiology training. This finding highlights the benefit of modern CT scanners and formally trained radiologists to both improve patient care and allow for rigorous scientific inquiry. As CT scanners with larger numbers of detectors become more widely used, the accuracy of TCVI diagnosis, and the reliability of the imaging for practitioners using it to distinguish TCVI grades, is likely to improve.

The current study identified a correlation of just 75% between the CTA and DSA grades. Although the dynamic nature of these injuries could play a role, this is most likely the result of injuries falling in a gray area among Grades I, II, and III; Grade IV injuries are readily apparent. It is conceivable that formal training could provide a standardized method of grading TCVI in this gray area, thus improving accuracy. However, given that all injuries are treated with aspirin as first-line therapy, it would be reasonable to combine Grades I–III in the context of a multicenter trial to improve diagnostic accuracy among participating institutions.

Interrater reliability of the Biffl Scale was substantial ($\kappa = 0.65$) to excellent (ICC = 0.82). This degree of reliability is similar to accepted techniques to measure atherosclerotic CA stenosis, which have been used in major clinical trials.[8,10,12] Given the unpredictable spectrum of traumatic pathology and its frequent association with artifact and concomitant injuries, the reliability was believed to be robust and capable of supporting future large-scale clinical studies.

Limitations of the Study

This study has several limitations that merit discussion. Images used in the study were obtained using a 40-section multidetector CT; more contemporary scanners, with more detectors, are more accurate.[9] There was a relative paucity of some injury grades, including no Grade V injuries, and an abundance of others; this was the result of varied incidences of different grades of injury affecting the CA and VA. The selected images were representative of the incidences of identified lesions at our institution. It is also worth noting that the included reviewers did not undergo formal training in the assignment of Biffl grades; this was part of an effort to improve the generalizability of the results.

Conclusions

Grading of TCVI imaged with CTA and categorized using the Biffl Scale is reliable. This finding affirms the scale's use in clinical practice as a means of reliable communication among physicians, and authenticates its use in clinical studies.

References

1. Biffl WL, Moore EE, Offner PJ, Brega KE, Franciose RJ, Burch JM: Blunt carotid arterial injuries: implications of a new grading scale. **J Trauma 47:**845–853, 1999
2. Eastman AL, Chason DP, Perez CL, McAnulty AL, Minei JP: Computed tomographic angiography for the diagnosis of blunt cervical vascular injury: is it ready for primetime? **J Trauma 60:**925–929, 2006
3. Fusco MR, Harrigan MR: Cerebrovascular dissections: a review. Part II: blunt cerebrovascular injury. **Neurosurgery 68:**517–530, 2011
4. Griessenauer CJ, Fleming JB, Richards BF, Cava LP, Curé JK, Younan DS, et al: Timing and mechanism of ischemic stroke due to extracranial blunt traumatic cerebrovascular injury. **J Neurosurg 118:**397–404, 2013
5. Kerwin AJ, Bynoe RP, Murray J, Hudson ER, Close TP, Gifford RR, et al: Liberalized screening for blunt carotid and vertebral artery injuries is justified. **J Trauma 51:**308–314, 2001
6. Landis JR, Koch GG: The measurement of observer agreement for categorical data. **Biometrics 33:**159–174, 1977
7. Malhotra AK, Camacho M, Ivatury RR, Davis IC, Komorowski DJ, Leung DA, et al: Computed tomographic angiography for the diagnosis of blunt carotid/vertebral artery injury: a note of caution. **Ann Surg 246:**632–643, 2007
8. North American Symptomatic Carotid Endarterectomy Trial Collaborators: Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis. **N Engl J Med 325:**445–453, 1991
9. Roberts DJ, Chaubey VP, Zygun DA, Lorenzetti D, Faris PD, Ball CG, et al: Diagnostic accuracy of computed tomographic angiography for blunt cerebrovascular injury detection in trauma patients: a systematic review and meta-analysis. **Ann Surg 257:**621–632, 2013
10. Stapf C, Hofmeister C, Hartmann A, Seyfert S, Koch HC, Mohr JP, et al: Interrater agreement for high grade carotid artery stenosis measurement and treatment decision. **Eur J Med Res 5:**26–31, 2000
11. Stein DM, Boswell S, Sliker CW, Lui FY, Scalea TM: Blunt cerebrovascular injuries: does treatment always matter? **J Trauma 66:**132–144, 2009
12. Walker MD, Marler JR, Goldstein M, Grady PA, Toole JF, Baker WH, et al: Endarterectomy for asymptomatic carotid artery stenosis. Executive Committee for the Asymptomatic Carotid Atherosclerosis Study. **JAMA 273:**1421–1428, 1995
13. Walter SD, Eliasziw M, Donner A: Sample size and optimal designs for reliability studies. **Stat Med 17:**101–110, 1998

Disclosures

The authors report no conflict of interest concerning the materials or methods used in this study or the findings specified in this paper.

Author Contributions

Conception and design: Foreman, Harrigan. Acquisition of data: Foreman, Griessenauer, Kicielinski, Schmalz, Rocque, Fusco, Sullivan, Deveikis. Analysis and interpretation of data: Foreman, Griessenauer, Kicielinski, Rocque, Sullivan, Deveikis, Harrigan. Drafting the article: Foreman. Critically revising the article: all authors. Reviewed submitted version of manuscript: all authors. Approved the final version of the manuscript on behalf of all authors: Foreman. Statistical analysis: Griessenauer, Kicielinski. Study supervision: Harrigan.

Correspondence

Paul M. Foreman, Department of Neurosurgery, University of Alabama at Birmingham, Faculty Office Tower 1005, 510 20th St. South, Birmingham, AL 35294. email: pforeman@uabmc.edu.