Published in final edited form as:

J Neurosurg. 2017 July ; 127(1): 32–35. doi:10.3171/2016.7.JNS16849.

Reliability Assessment of the Biffl Scale for Blunt Traumatic Cerebrovascular Injury as Detected on Computer Tomography Angiography

Paul M. Foreman, MD¹, Christoph J. Griessenauer, MD², Kimberly P. Kicielski, MD¹, Philip G. R. Schmalz, MD¹, Brandon G. Rocque, MD, MS¹, Matthew R. Fusco, MD⁴, Joseph C. Sullivan III, MD⁵, John P. Deveikis, MD¹, and Mark R. Harrigan, MD¹

¹Department of Neurosurgery, University of Alabama at Birmingham, Birmingham, AL

²Neurosurgical Service, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA

⁴Department of Neurosurgery, Vanderbilt University, Nashville, TN

⁵Department of Radiology, University of Alabama at Birmingham, Birmingham, AL

Abstract

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 computed tomography angiography (CTA) has been used to identify patients with TCVI. The reliability of this scale, using CTA, has not yet been determined.

Methods—Seven independent raters, including two neurosurgeons, two neuroradiologists, two neurosurgical residents, and one neurosurgical vascular fellow independently reviewed a presenting CTA of the neck performed on 40 patients with confirmed TCVI and assigned a Biffl grade. Ten images were repeated to assess intra-rater reliability, for a total of 50 CTAs. Fleiss' multi-rater kappa (κ) and interclass correlation (ICC) were calculated as a measure of inter-rater reliability. Weighted Cohen's kappa (κ) was used to assess intra-rater reliability.

Results—Fleiss' multi-rater kappa was 0.65 (95% CI 0.61 – 0.69), indicating substantial agreement as to the Biffl grade assignment among the seven raters. Interclass correlation was 0.82 demonstrating excellent agreement among the raters. Intra-rater reliability was perfect (weighted Cohen's kappa = 1) in two raters and near perfect (weighted Cohen's kappa > 0.8) in the remaining 5 raters.

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

Corresponding author: Paul M. Foreman MD, Faculty Office Tower 1005, 510 20th Street South, Birmingham, AL 35294, Phone: (205) 934-7100, pforeman@uabmc.edu.

Conflict of Interest: None

Funding: None

Keywords

traumatic cerebrovascular injury; trauma; Biffl Scale; Biffl grade; dissection; reliability

Introduction

Blunt traumatic cerebrovascular injury (TCVI) represents a structural defect in a vessel wall that is directly attributable to high-energy, non-penetrating trauma.³ Overall incidence of TCVI among blunt trauma admission 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 seek to test the inter- and intra-rater reliability of the Biffl Scale across a spectrum of clinicians using widely available computed tomography angiography (CTA).

Methods

A prospective study of TCVI was done at a single center from January 2007 to December 2011. During this time, all patients admitted after blunt trauma with screening neck CTA evidence of extracranial TCVI 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 a follow-up DSA. This series of cases included 20 carotid artery injuries and 20 vertebral artery injuries. This study was performed with approval from the Institutional Review Board.

Seven raters, including two neurosurgeons, two neuroradiologists, two neurosurgical residents, and one neurosurgical vascular fellow independently reviewed each CTA and assigned a Biffl grade (Table 1); examples of Biffl grade I–IV injuries are provided in Figure 1. Per interpretation of CTA and DSA by the senior author (MRH), the following distribution of Biffl graded TCVIs were studied. Carotid artery (CA) injuries included: 3 grade I injuries, 9 grade II injuries, 6 grade III injuries, and 2 grade IV injuries. Vertebral artery (VA) injuries included: 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 intra-rater 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 inter-rater reliability and 0.80 as the desired level of inter-rater reliability, based on $p = 0.05$ and 80% power, the sample size required for 7 raters is at least 20 images (22% error margin).¹³ With seven raters, reviewing five duplicated images, the intra-rater reliability should have a 95% confidence interval of ± 0.135 .

Fleiss' multi-rater kappa (κ) and interclass correlation (ICC) was calculated as a measure of overall agreement of Biffl grade assignment among the seven raters. For intra-rater reliability analysis, weighted Cohen's kappa (κ) was used to assess repeat measurement agreement for each rater. Agreement measured by kappa (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 Interclass correlation greater than 0.75 was considered to have excellent agreement, with an ICC between 0.40 and 0.75 classified as fair to good, and less than 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 21.0 (IBM Corp. Armonk, NY).

Results

Inter-rater reliability

Fleiss' multi-rater kappa (κ) was 0.65 (95% CI 0.61 – 0.69), indicating substantial agreement as to the Biffl grade assignment among the seven raters. Interclass correlation was 0.82 indicating excellent agreement among the raters (Table 2).

Intra-rater reliability

Intra-rater reliability was perfect (weighted Cohen's kappa = 1) in two 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 intra-rater 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, 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. The present study 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 (inter-rater reliability) and a near perfect agreement within a single rater (intra-rater reliability).

The five-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 carotid artery; subsequently the scale was expanded to include vertebral artery injury as well. This scale is now widely accepted as common language, enabling inter-physician communication and systematic research. Moreover, TCVI subtypes, as described by the Biffl Scale, correlate with prognosis with higher grade carotid artery injuries carrying a significantly higher risk of ischemic stroke as compared to other subtypes.¹ While the reported stroke rate associated with a particular injury grade has varied among publications, the original description attributed stroke rates of 3%, 11%, 33%, 44%, and 100% for carotid injuries 1–5, respectively.¹ With the exception of the very rare grade V injury, the authors of the current manuscript treat all TCVI (grade I–IV) with aspirin 325 mg daily 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 that the initial review of the CTA by a non-neuroradiologist, and even 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 CTA and DSA confirmed TCVI, only the CTA was used for inter- and intra-rater 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, high pretest probability with negative non-invasive imaging) or in patients where 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, 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 sensitivity.⁹ Variability was felt to be due to diagnostic threshold, number of available CT slices, and training, with increased sensitivity with increased number of slices and neuroradiology training.⁹ 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 to distinguish TCVI grades, is likely to improve. The current study identified a correlation of just 75% between the CTA and DSA grades. While 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, as 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 1–3 in the context of a multi-center trial to improve diagnostic accuracy among participating institutions.

Inter-rater 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 carotid artery stenosis, which have been employed in major clinical trials.^{8,10,12} Given the unpredictable spectrum of traumatic pathology and its frequent association artifact and concomitant injuries, the reliability was felt to be robust and capable of supporting future large-scale clinical studies.

Limitations

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 Vs, and 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 done in an effort to improve the generalizability of the results.

Conclusion

Grading of TCVI imaged with CTA 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*. 1999; 47:845–853. [PubMed: 10568710]
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*. 2006; 60:925–929. discussion 929. [PubMed: 16688051]
3. Fusco MR, Harrigan MR. Cerebrovascular dissections: a review. Part II: blunt cerebrovascular injury. *Neurosurgery*. 2011; 68:517–530. discussion 530. [PubMed: 21135751]
4. Griessenauer CJ, Fleming JB, Richards BF, Cava LP, Cure JK, Younan DS, et al. Timing and mechanism of ischemic stroke due to extracranial blunt traumatic cerebrovascular injury. *J Neurosurg*. 2013; 118:397–404. [PubMed: 23216467]
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*. 2001; 51:308–314. [PubMed: 11493789]
6. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33:159–174. [PubMed: 843571]

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.* 2007; 246:632–642. discussion 642–633. [PubMed: 17893500]
8. North American Symptomatic Carotid Endarterectomy Trial C. Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis. *N Engl J Med.* 1991; 325:445–453. [PubMed: 1852179]
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.* 2013; 257:621–632. [PubMed: 23470509]
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.* 2000; 5:26–31. [PubMed: 10657286]
11. Stein DM, Boswell S, Sliker CW, Lui FY, Scalea TM. Blunt cerebrovascular injuries: does treatment always matter? *J Trauma.* 2009; 66:132–143. discussion 143–134. [PubMed: 19131816]
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.* 1995; 273:1421–1428. [PubMed: 7723155]
13. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med.* 1998; 17:101–110. [PubMed: 9463853]

Figure 1. DSA (left) and CTA (right) examples of Biffi grade I–IV injuries. (A) Type I vertebral artery injury, (B) Type II internal carotid artery injury, (C) Type III internal carotid artery injury, (D) Type IV internal carotid artery injury.

Table 1Biffi Scale for traumatic cerebrovascular injury²

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

Table 2

Reliability of the Biffi Scale using CTA

	Fleiss's multi-rater kappa (95% CI)	Interclass correlation (95% CI)	Weighted Cohen's kappa
Inter-rater reliability	0.65 (95% CI 0.61–0.69)	0.82 (95% CI 0.75–0.89)	
Intra-rater reliability			
- Neurosurgeon			1
- Neurosurgeon			0.82 (95% CI 0.59–1.04)
- Neuroradiologist			1
- Neuroradiologist			0.89 (95% CI 0.7–1.09)
- Resident neurosurgeon			0.91 (95% CI 0.75–1.07)
- Resident neurosurgeon			0.91 (95% CI 0.75–1.08)
- Vascular fellow			0.91 (95% CI 0.75–1.07)