Injury, Int. J. Care Injured 45 (2014) 146–150

Contents lists available at SciVerse ScienceDirect

Injury

journal homepage: www.elsevier.com/locate/injury

Reliability of injury grading systems for patients with blunt splenic trauma

D.C. Olthof [a], C.H. van der Vlies [b], M.J. Scheerder [c], R.J. de Haan [d], L.F.M. Beenen [c], J.C. Goslings [a,*], O.M. van Delden [c]

[a] Trauma Unit Department of Surgery, Academic Medical Center, Amsterdam, The Netherlands
[b] Department of Surgery, Maasstad Ziekenhuis, Rotterdam, The Netherlands
[c] Department of Radiology, Academic Medical Center, Amsterdam, The Netherlands
[d] Clinical Research Unit, Academic Medical Center, Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 1 May 2012
Accepted 31 July 2012

Keywords:
Blunt abdominal trauma
Splenic injury
Interobserver reliability
Intraobserver reliability
Injury grading system(s)

ABSTRACT

Objectives: The most widely used grading system for blunt splenic injury is the American Association for the Surgery of Trauma (AAST) organ injury scale. In 2007 a new grading system was developed. This 'Baltimore CT grading system' is superior to the AAST classification system in predicting the need for angiography and embolization or surgery. The objective of this study was to assess inter- and intraobserver reliability between radiologists in classifying splenic injury according to both grading systems.

Methods: CT scans of 83 patients with blunt splenic injury admitted between 1998 and 2008 to an academic Level 1 trauma centre were retrospectively reviewed. Inter and intrarater reliability were expressed in Cohen's or weighted Kappa values.

Results: Overall weighted interobserver Kappa coefficients for the AAST and 'Baltimore CT grading system' were respectively substantial (kappa = 0.80) and almost perfect (kappa = 0.85). Average weighted intraobserver Kappa's values were in the 'almost perfect' range (AAST: kappa = 0.91, 'Baltimore CT grading system': kappa = 0.81).

Conclusion: The present study shows that overall the inter- and intraobserver reliability for grading splenic injury according to the AAST grading system and 'Baltimore CT grading system' are equally high. Because of the integration of vascular injury, the 'Baltimore CT grading system' supports clinical decision making. We therefore recommend use of this system in the classification of splenic injury.

Introduction

The spleen is the most commonly injured solid organ in blunt abdominal trauma.[1,2] Contrast-enhanced multidetector computed tomography (MDCT) is the gold standard diagnostic examination for splenic trauma because of its speed, widespread availability, diagnostic accuracy, and relatively non invasive nature. The most widely used grading system for blunt splenic injuries is the American Association for Surgery of Trauma (AAST) organ injury scale.[3,4] The AAST grading system, first introduced in 1989, is based on anatomic disruption of the spleen, as found during laparotomy (Appendix A). CT-based injury grading systems, derived from the AAST scale, also exist.[5]

Prior studies showed that grade of injury on the CT scan alone is a poor predictor for successful outcome of nonoperative management (NOM).[6–8] Recent evidence suggests that vascular injuries including active splenic haemorrhage (the so-called 'contrast blush'), pseudoaneurysms and post-traumatic arteriovenous fistulas are associated with an increased failure rate of NOM.[9,10] Furthermore, it was shown that the higher the grade of splenic injury according to the AAST, the greater the risk of vascular injury.[11] To date, vascular injuries are not integrated in the AAST grading system. Therefore, in 2007 Marmery and colleague radiologists validated a new grading system for the classification of splenic injury[12] (Appendix B). The presence of a contrast blush is a key factor in this grading system (further referred to as 'Baltimore CT grading system'). The 'Baltimore CT grading system' is superior to the AAST system in predicting the need for angiography and embolization or splenic surgery in patients sustaining blunt splenic injury and therefore its use may be preferred over the AAST grading system.[5]

The objective of the present study was to assess inter- and intraobserver reliability between radiologists in classifying blunt splenic injury according to the AAST and 'Baltimore CT grading system'.

* Corresponding author at: Trauma Unit Department of Surgery, Academic Medical Center, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands. Tel.: +31 20 5666019.
E-mail address: j.c.goslings@amc.nl (J.C. Goslings).

0020–1383/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.injury.2012.08.013

Patients and methods

All patients with blunt splenic injuries admitted between 1998 and 2008 to the level 1 trauma centre of the Academic Medical Centre, the Netherlands, were identified from the hospital's trauma registry. The presence and quality of CT images, made during initial trauma screening in the trauma room, were verified by a senior radiologist (LB).

Contrast-enhanced CT scans with 8 mm slices or thinner collimation and with images obtained during the portal-venous phase were selected.

Computed tomography scanning protocol

CT scans were obtained on a 4 slice scanner (Sensation 4, Siemens Medical Solutions, Forcheim, Germany) before 2008 and a 64 slice scanner (Siemens Sensation 64) after 2008. Images were acquired 70 s after intravenous administration of 100 ml contrast material (mainly Ultravist 300). The CT scans were independently scored by two senior radiologists (>10 years of experience, observer 1 and 2) and one radiology resident (4 years of experience, observer 3) on a PACS system (Impax 4,5, AGFA Gevaert, Belgium). The CT scans were scored twice with a time interval of ≥2 months. The scans were presented to the observers in a random order. Since observers could have been involved in initial trauma screening in the trauma room of a number of the clinical cases included in the study, they were blinded for patients' name, identity number and his or her clinical course.

Investigated parameters

Splenic Injury was scored according to the ordinal grading systems of the AAST (Appendix A) and the 'Baltimore CT grading system' (Appendix B). AAST grades I–III and grades IV and V were additionally dichotomized into low versus high grade splenic injury, respectively. This distinction is commonly applied in literature and has therapeutic implications. Since low reliability for scoring the presence (or absence) of a contrast blush can negatively influence the reliability of the 'Baltimore CT grading system', we additionally assessed the reliability of this parameter. The presence of a contrast blush was documented according to the following nominal categories: intraparenchymal (in the splenic parenchyma or subcapsular space) or intraperitoneal (into the peritoneum). A contrast blush was defined as a well-circumscribed, peri-splenic or intraparenchymal contrast collection that was hyperdense with respect to the rest of the splenic parenchyma.[13] Lastly, we assessed if in our study population a relation was observed between the grade of splenic injury (scored according to the AAST grading system) and the presence of vascular injury. In addition to the presence and type of contrast blush, this includes the presence of pseudoaneurysms and post traumatic arteriovenous fistulas.

Statistical analysis

Relevant patient characteristics (sex, age and Injury Severity Score) and the relation between vascular injury and splenic injury grade were summarized using descriptive statistics. For reporting vascular injury in relation to splenic injury grade, the mean value of the three observers (round 2) was calculated. Inter- and intraobserver reliability were expressed in Kappa coefficients. The Kappa statistic estimates the proportion of agreement among or within observers after chance agreement has been removed. Binary and nominal data were expressed in Cohen's Kappa values, whereas ordinal data were expressed in weighted Kappa values. Average (weighted) Kappa values of pair of observations were considered as an overall index for concordance among or within observers.[14]

Kappa values were arbitrarily classified according to Landis and Koch[15] with values <0 indicating no agreement, 0–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1 as almost perfect agreement. Statistical uncertainty of kappa of pair of observations was expressed in a 95% confidence interval.

Results

CT scans of 88 patients were analysed. Five patients were excluded due to incomplete data. The study population consisted of 83 patients. 88% of the patients was male. Median age was 29 years (range 17–86). Median Injury Severity Score was 24 (range 4–66).

Interobserver reliability

The interobserver reliability values of all parameters are shown in Table 1. The overall Kappa coefficients for the AAST grading system and 'Baltimore CT grading system' were substantial (kappa = 0.80) and almost perfect (kappa = 0.85), respectively. The average Cohen's Kappa for the presence or absence of a contrast blush was substantial (kappa = 0.76). Average Cohen's Kappa's for the subtypes of blushes (intraperitoneal and intraparenchymal) were 'substantial' (kappa = 0.68; data not presented) and 'fair' (kappa = 0.41; data not presented) respectively.

In general, the point estimates of the Kappa values of the most experienced observers (observer 1 and 2) were higher, indicating better interobserver agreement. The point estimates of the Kappa values of observer 2 against observer 3 were lower. Appendix C shows where disagreement was predominantly situated. 23 out of 34 CT scans graded as AAST splenic injury grade 3 by observer 2 were graded differently by observer 3. 9 out of the 14 CT scans graded as AAST splenic injury grade 4 by observer 2 were graded as AAST splenic injury grade 5 by observer 3. CT scans that were graded as splenic injury grade 4A by observer 2 according the 'Baltimore grading system', were graded as grade 2 ($n = 3$), grade 3 ($n = 4$) or grade 4B ($n = 1$) by observer 3. Interobserver Kappa values after a time interval of ≥2 months showed the same patterns (data not presented).

Table 1
Interobserver reliability: (weighted) Kappa values with 95% confidence intervals ($n = 83$ CT scans).

	(weighted) Kappa of pair of observations (95% confidence interval)	Overall Kappa[a]
AAST[b] grading system	1,2: 0.83 (0.77–0.90) 2,3: 0.75 (0.65–0.84) 1,3: 0.81 (0.70–0.91)	0.80
AAST low vs. high grade splenic injury[c]	1,2: 0.74 (0.60–0.89) 2,3: 0.72 (0.57–0.87) 1,3: 0.78 (0.64–0.91)	0.75
'Baltimore CT grading system'[b]	1,2: 0.91 (0.87–0.95) 2,3: 0.79 (0.71–0.88) 1,3: 0.84 (0.77–0.91)	0.85
Contrast blush[c]	1,2: 0.83 (0.70–0.96) 2,3: 0.69 (0.52–0.86) 1,3: 0.75 (0.59–0.90)	0.76

1: observer 1; 2: observer 2; 3: observer 3.
[a] Overall Kappa indicates the average (weighted) Kappa values of pair of observations.
[b] Ordinal data expressed in weighted Kappa values.
[c] Binary and nominal data expressed in Cohen's Kappa values.