Validation of the classification of surgically treated acetabular fractures in the Swedish Fracture Register

Madelene Albrektsson [a,b,*], Olof Wolf [c], Anders Enocson [d], Mikael Sundfeldt [a,b]

[a] Department of Orthopaedics, Sahlgrenska University Hospital, Gothenburg/Mölndal, Sweden
[b] Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
[c] Department of Orthopaedics, Institute of Surgical Sciences, Uppsala University Hospital, Uppsala, Sweden
[d] Department of Molecular Medicine and Surgery, Karolinska University Hospital, Karolinska Institute, Stockholm, Sweden

ARTICLE INFO

ABSTRACT

Objectives: To validate the classification of surgically treated acetabular fractures in the Swedish Fracture Register (SFR) and to investigate the intra- and interrater reliability of the Judet-Letournel / AO/OTA classification systems.

Methods: Surgically treated acetabular fractures were randomly selected from the SFR ($n = 132$) and 124 fractures were classified independently by three experienced orthopedic pelvic surgeons at two different occasions. A gold standard classification was established for each case after these two sessions or, if necessary, after a discussion session. The gold standard classification was compared to the registered SFR classification to assess the validity of SFR data. Accuracy and intra- and interrater agreement were evaluated using Cohen's kappa with interpretation according to Landis and Koch.

Results: There was moderate agreement between the established gold standard classification and the SFR (kappa 0.43). The level of agreement differed between classification groups. The intrarater agreement was substantial to almost perfect and interrater agreement was moderate to substantial.

Conclusions: The accuracy of acetabular fracture classifications in the SFR was moderate and comparable to previous validation studies from the SFR on other fracture types. As the accuracy differed between fracture groups, care should be taken when analyzing data from the SFR on specific acetabular fracture groups.

Introduction

Out of 100 000 persons, three to eight, sustain an acetabular fracture each year and the incidence is increasing among the elderly population and decreasing in the younger population [1–5]. The etiology of an acetabular fracture is mainly either low energy trauma in the elderly population or high energy trauma in younger patients. During the last decades there has been a comprehensive development in the treatment of these fractures [4,5].

The classification of acetabular fractures is recognized as difficult. Although the Judet-Letournel classification is the most commonly used classification worldwide, it has been criticized for both being too complex especially among less experienced users [6], but also for being oversimplified and incomplete [7–9]. Alternative classification systems have been presented, some of which are based on computed tomography (CT) instead of plain radiographs

[8–10], but to date there is no superior alternative to the Judet-Letournel classification.

The Swedish Fracture Register (SFR) is a national quality register that was established in 2011. It has rapidly grown and now includes all fractures of the extremities, pelvis and spine, both surgically and non-surgically treated, making the register unique [11,12]. In spring 2021, the national coverage was 100% [13]. The register collects information on patient age, sex, fracture classification and the chosen treatment including any reoperations. Validation studies from the SFR have been performed for ankle, tibia, proximal humerus, femur and distal radial fractures and shown moderate to substantial accuracy [14–18].

Fractures of the acetabulum have been included in the register since 2012, and in September 2021 more than 3000 patients treated for an acetabular fracture had been registered (www.frakturregistret.se).

The aim of the current study was to assess the accuracy of the classification of surgically treated acetabular fractures in the SFR, thereby validating data reported to the SFR. Intra- and interrater

---

* Corresponding author.
E-mail address: madelene.albrektsson@vgregion.se (M. Albrektsson).

reliability of the Judet-Letournel / AO/OTA classification systems was investigated as a secondary aim.

Methods

In total, 132 acetabular fractures in patients aged 16 years or higher, treated operatively between January 1st 2014 and October 15th 2020, were randomly selected from the SFR in November 2020. These were initially treated at 24 different hospitals. The radiology departments of the treating hospitals were contacted by letter and asked to send the preoperative computed tomography (CT) of their respective patients. A total of 124 fractures (123 patients) were included in the study and eight fractures were excluded due to either missing or poor-quality preoperative imaging, penetrating injury (gunshot), or the presence of a periprosthetic fracture in which case a different classification system is warranted (Fig. 1).

The SFR uses the AO/OTA classification of acetabular fractures from 1996 [19] which is a modification of the Judet-Letournel classification [20,21] adapted to the AO/OTA system for fracture classification but consisting of the same fracture groups as described by Judet-Letournel (Table 1).

Three experienced orthopedic pelvic surgeons (AE, OW and MS) from three different university hospitals in Sweden were chosen as expert raters to classify the fractures at two different occasions, five weeks apart. The first author (MA) demonstrated the CT images, including 3D reconstructions, for the three expert raters using the Agfa HealthCare® imaging system (Belgium) on an online video conference platform (Microsoft Skype®). The raters made their own individual analysis and were not allowed to discuss the classification with each other. The images and explanatory text available to the users of the SFR for guidance in fracture classification were available to the raters during the classification sessions.

Each fracture was classified a total of six times (two times by each rater) and an identical classification in at least five out of six times was accepted as the gold standard (GS). GS was established for 67 out of the 124 cases. One case, however, was commented on as perhaps not being a pure acetabular fracture but a pelvic ring fracture. For that reason, that case was not immediately accepted as the gold standard, but instead was included for discussion at a final third session.

The third and final session was held for the raters to discuss the remaining fractures in which GS had not been established. The images were demonstrated once again during this session and all the three raters shared their opinion and openly discussed their reasoning. Consensus was reached in all cases regardless of previous classifications, and the final gold standard was established.
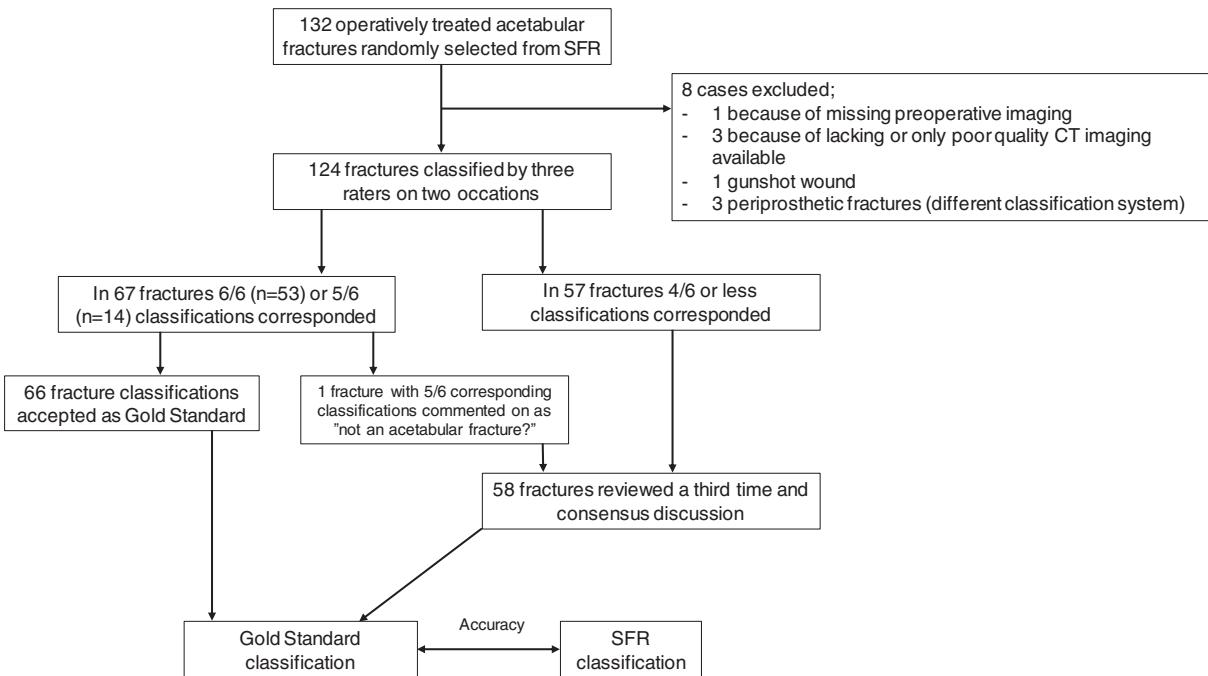

Fig. 1. Flowchart of how the study was conducted.

Table 1
AO/OTA fracture classification groups used in SFR and corresponding group according to the Judet-Letournel classification system.

AO/OTA group as shown in SFR	Judet-Letournel classification	Abbreviations
62-A1	Posterior wall	PW
62-A2.1/2	Posterior column	PC
62-A2.3	Associated posterior column and posterior wall	PC+PW
62-A3.1	Anterior wall	AW
62-A3.2/3	Anterior column	AC
62-B1.1/2	Pure transverse	TRANS
62-B1.3	Associated transverse and posterior wall	TRANS+PW
62-B2	T-shaped	T
62-B3	Associated anterior column and posterior hemitransverse	AC+PHT
62-C	Associated both column	ABC

Table 2

Agreement measures for categorical data according to Landis and Koch.

Kappa Statistic	Strength of Agreement
<0.00	Poor
0.00–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1.00	Almost Perfect

Statistics

Sample size calculations to determine the number of fractures needed for this study were conducted. An estimate of the distribution of fracture groups was made using data from the SFR prior to the data extraction for this study. These numbers were used to calculate the probability of agreement "by chance". Percent agreement of 0.6 (60%) was assumed. To achieve a confidence interval of 95% that did not span more than one category on the kappa coefficient scale presented by Landis and Koch, a relative error of 20% (i.e. kappa +/- 0.1) was accepted [22]. Sample size was calculated to 111 fractures. A loss of approximately 10% was estimated and therefore 132 fractures were selected from the SFR. Cohen's kappa was used for analyzing agreement for accuracy and inter- and intraobserver agreement and the strength of agreement was interpreted as categorized by Landis and Koch (Table 2). Statistical analyses were conducted using SPSS Statistics (version 25, IBM corporation, USA).

Results

Demographics and fracture classification

Of the 124 acetabular fractures, 93 (75%) had occurred in men. Age at the time of injury ranged from 18 to 95 years, with a mean age of 59 for men and 67 for women (Table 3).

According to the established gold standard classification, the group C fractures (both columns) were found to be most common (26%, $n = 32$), followed by the anterior column fractures (21%, $n = 26$) and posterior wall fractures (19%, $n = 23$). The distribution

Table 3

Age and sex distribution.

	Men (n = 93, 75%)	Women (n = 31, 25%)	Total (n = 124)
Mean age (range)	59 (18–95)	67 (25–93)	61 (18–95)
Median age (range)	63 (18–95)	73 (25–93)	65 (18–95)

Table 5

Distribution of acetabular fractures according to how the classification of the three raters on two occasions corresponded.

Number of corresponding classifications	Number of fractures
6/6	53
5/6	14
4/6	30
3/6	22
2/6	5
0/6	0

Table 6

Accuracy SFR vs Gold Standard.

	PA (percent agreement)	Kappa (95% CI)
AO/OTA group	51%	0.43 (0.34–0.52)
AO/OTA type (A/B/C)	69%	0.5 (0.37–0.62)

of fracture groups differed somewhat between the established GS and the registrations made in the SFR (Table 4).

The first two classification sessions resulted in a total of 67 out of the 124 fractures with corresponding classifications in five or six out of six classifications (Table 5).

Accuracy SFR vs gold standard

The SFR fracture classification corresponded to the GS in 63 out of the 124 cases (51%). Cohen's kappa coefficient was 0.43 (0.34–0.52 95% CI), corresponding to moderate agreement according to Landis and Koch [22, Table 2]. When joining the classification groups into; type A, B and C, the percent agreement was 69% ($n = 85/124$) and the kappa value was 0.5 (0.37–0.62 95% CI) (Table 6).

Not counting the anterior wall fracture (A3.1), as there was only one (100%), the highest percent agreement (91%) was obtained for posterior wall fractures (A1) where 21 out of 23 were correctly classified. The PW fractures were followed by fractures of both columns (C) at 63% (20/32) and the combined transverse and posterior wall fractures (B1.3) at 57% (4/7) (Table 7). The other fracture groups had a percent agreement of 50% or less.

Reliability

Intrarater agreement

The intrarater agreement for each individual expert rater between the first two sessions for the full classification code was

Table 4

Age distribution according to fracture group (GS) and number of fractures for each fracture group according to the gold standard classification and according to the SFR registration. (GS=Gold Standard, SFR=Swedish Fracture Register).

Fracture group	Mean age	Gold Standard		SFR	
		Number of fractures	%	Number of fractures	%
A1 (PW)	51	23	19	26	21
A2.1/2 (PC)	65	2	2	7	6
A2.3 (PC+PW)	54	7	6	4	3
A3.1 (AW)	79	1	1	11	9
A3.2/3 (AC)	70	26	21	13	10
B1.1/2 (TRANS)	59	7	6	7	6
B1.3 (TRANS+PW)	45	7	6	9	7
B2 (T)	69	9	7	10	8
B3 (AC+PHT)	67	7	6	10	8
C (ABC)	62	32	26	26	21
Not able to classify	64	3	2	1	1

Table 7
Accuracy Swedish Fracture Register (SFR) vs Gold Standard (GS) for each fracture group.

Fracture group	Number of correctly classified fractures in SFR (total number according to GS)	PA (percent agreement)
A1 (PW)	21 (23)	91%
A2.1/2 (PC)	1 (2)	50%
A2.3 (PC+PW)	1 (7)	14%
A3.1 (AW)	1 (1)	100%
A3.2/3 (AC)	10 (26)	38%
B1.1/2 (TRANS)	1 (7)	14%
B1.3 (TRANS+PW)	4 (7)	57%
B2 (T)	2 (9)	22%
B3 (AC+PHT)	1 (7)	14%
C (ABC)	20 (32)	63%
Not able to classify	1 (3)	33%

substantial for raters one and three and almost perfect for rater two according to the Landis and Koch interpretation of kappa values (Table 8). The mean intrarater kappa value was 0.74.

Interrater agreement

The interrater agreement between the three expert raters for the full classification code ranged from 0.5 to 0.68 which corresponds to moderate to substantial agreement (Table 9). The mean kappa value for interrater agreement was 0.59. For the AO/OTA type (A, B or C) the kappa values ranged from 0.6 to 0.77 corresponding to substantial agreement.

Discussion

The agreement of acetabular fracture classification in SFR compared to the established gold standard in this study was found to be moderate with a kappa value of 0.43. For other fracture locations, validation studies on SFR data have reported on either moderate or substantial agreement [14–18]. A lower agreement on acetabular fracture classification was expected since the classification system for acetabular fractures is known to be complex [6–10]. However, the analysis of intra- and interrater reliability showed that the classification is reliable once the observer was well familiar with it. The substantial intrarater agreement and the moderate interrater agreement of the present study is very similar to a previous study by Zhang et al [10].

The sex distribution with a male predominance of 75% in this study population was comparable to other studies [2,4,23]. Furthermore, the distribution of fracture groups was, to a great extent, comparable to previous studies [2,4,21,23]. With the exception of anterior and posterior column fractures, the distribution of fracture groups in the current study greatly corresponded to the study

conducted by Ochs et al. in which 1266 patients with unilateral acetabular fractures were reviewed [4]. Anterior column fractures were more common in our material (21% compared to 12%) and posterior column fractures were less common (1.6% compared to 7.5%). The differences may be explained by the age difference of the two study cohorts, where the mean age of our study group was 61 years and 47 years in the study Ochs et al. conducted. Ferguson et al. reported on a difference in the distribution of fracture groups between patients aged under and over 60 years [2]. In patients younger than 60 years, 7.2% sustained an anterior column fracture as opposed to 19% in the group older than 60 years. The younger group had a higher prevalence of posterior column fractures than the older group (2.4% compared to 0.4%).

The Judet-Letournel classification has been criticized for not including all fracture patterns and that as much as 20% of acetabular fractures do not fit into any of the ten fracture categories [9,10]. In our study, 2.4% of the acetabular fractures were judged as impossible to classify and in the SFR the corresponding number was 0.8%. The proportion of non-classifiable fractures in the SFR agrees with the results reported by Ochs et al. (0.8%) [4]. These results support the completeness of the Judet-Letournel classification system.

Cohen's kappa coefficient was developed to eliminate the chance that raters were in agreement by making random guesses when unsure of how to score, or in this case, classify a fracture [24]. The interpretation of Cohen's kappa value as suggested by Landis and Koch is widely used [22]. However, the interpretation has been a subject for discussion. McHugh reported that a low agreement among raters can nonetheless be described as substantial and suggests a different interpretation of Cohen's kappa [25]. Comparing GS classification to the SFR, the kappa values of 0.43 and 0.5 correspond to weak agreement instead of moderate when interpreted according to McHugh.

Regardless of which interpretation of Cohen's kappa that is used, we found that the level of accuracy for the SFR established in this study needs to be interpreted with some caution. It is important to be aware of the present study results when further analyzing data on acetabular fractures from the SFR. Some fracture groups, such as posterior wall and both column fractures, had a better agreement than others, ensuring valid data when these fracture groups are studied individually. Other fracture groups seem much more difficult to classify correctly and care should be taken before making assumptions when analyzing data on these groups. Data from the SFR on acetabular fractures as a cohort can be studied and analyzed with good validity.

Strengths and limitations

For this study, only patients with surgically treated acetabular fractures were selected and the validation therefore applies to

Table 8
Intrarater agreement, comparing the classifications of each rater between two seminars. (PA=percent agreement).

	Rater 1		Rater 2		Rater 3	
	PA	Kappa (95% CI)	PA	Kappa (95% CI)	PA	Kappa (95% CI)
AO/OTA group	74%	0.7 (0.61–0.78)	86%	0.84 (0.77–0.91)	73%	0.67 (0.58–0.76)
AO/OTA type (A/B/C)	89%	0.82 (0.74–0.91)	94%	0.9 (0.83–0.97)	81%	0.72 (0.62–0.82)

Table 9
Interrater kappa values with 95% CI.

	Rater 1 vs Rater 2		Rater 1 vs Rater 3		Rater 2 vs Rater 3	
	Seminar 1	Seminar 2	Seminar 1	Seminar 2	Seminar 1	Seminar 2
AO/OTA group	0.6(0.5–0.69)	0.6(0.5–0.69)	0.63(0.54–0.72)	0.68(0.6–0.77)	0.5(0.4–0.59)	0.54(0.45–0.63)
AO/OTA type (A/B/C)	0.75(0.65–0.85)	0.72(0.61–0.82)	0.69(0.59–0.8)	0.77(0.68–0.87)	0.6(0.48–0.71)	0.66(0.55–0.77)

that group of patients. However, most available studies are only performed on surgically treated patients. Since surgically treated patients constitute approximately 25% (www.frakturregistret.se) of the patients in the SFR, it could be argued that we only validated a small selection of the patients. The generalizability of our results could have been greater if we had included both surgically and non-surgically treated patients.

Sample size calculations were conducted to ensure that all fracture groups were included. Some fracture groups, however, were only represented with a very small amount of fractures. This makes it difficult to draw any certain conclusions regarding those specific fractures.

Conclusions

The accuracy of acetabular fracture classifications in the SFR was moderate and comparable to previous validation studies from the SFR on other fracture types. As the accuracy differed between fracture groups, care should be taken when analyzing data from the SFR on specific acetabular fracture groups.

Ethical approval

Ethical approval for this study was obtained from the Swedish Ethical Review Agency (ID 2020–03775).

Declarations of interest

None.

References

[1] Laird A, Keating JF. Acetabular fractures: a 16-year prospective epidemiological study. J Bone Joint Surg British Vol 2005;87:969–73.
[2] Ferguson TA, Patel R, Bhandari M, Matta JM. Fractures of the acetabulum in patients aged 60 years and older: an epidemiological and radiological study. J Bone Joint Surg Br 2010;92:250–7.
[3] Rinne PP, Laitinen MK, Huttunen T, Kannus P, Mattila VM. The incidence and trauma mechanisms of acetabular fractures: a nationwide study in Finland between 1997 and 2014. Injury 2017;48:2157–61.
[4] Ochs BG, Marintschev I, Hoyer H, Rolauffs B, Culemann U, Pohlemann T, et al. Changes in the treatment of acetabular fractures over 15 years: analysis of 1266 cases treated by the German Pelvic Multicentre Study Group (DAO/DGU). Injury 2010;41:839–51.
[5] Lundin N, Huttunen TT, Berg HE, Marcano A, Felländer-Tsai L, Enocson A. Increasing incidence of pelvic and acetabular fractures. A nationwide study of 87,308 fractures over a 16-year period in Sweden. Injury 2021;52:1410–17.
[6] Beaulé PE, Dorey FJ, Matta JM. Letournel classification for acetabular fractures. Assessment of interobserver and intraobserver reliability. J Bone Joint Surg Am 2003;85:1704–9.
[7] Harris JH Jr, Lee JS, Coupe KJ, Trotscher T. Acetabular fractures revisited: part 1, redefinition of the Letournel anterior column. AJR Am J Roentgenol 2004;182:1363–6.
[8] Harris JH Jr, Coupe KJ, Lee JS, Trotscher T. Acetabular fractures revisited: part 2, a new CT-based classification. AJR Am J Roentgenol 2004;182:1367–75.
[9] Herman A, Tenenbaum S, Ougortsin V, Shazar N. There Is No Column: a New Classification for Acetabular Fractures. J Bone Joint Surg Am 2018;100:e8.
[10] Zhang R, Yin Y, Li A, Wang Z, Hou Z, Zhuang Y, et al. Three-column classification for acetabular fractures: introduction and reproducibility assessment. J Bone Joint Surg Am 2019;101:2015–25.
[11] Wennergren D, Ekholm C, Sandelin A, Möller M. The Swedish fracture register: 103,000 fractures registered. BMC Musculoskelet Disord 2015;16:338.
[12] Wennergren D, Möller M. Implementation of the Swedish Fracture Register. Unfallchirurg 2018;121:949–55.
[13] Möller M. The Swedish Fracture Register. Annual Report 2020;2021:63.
[14] Juto H, Möller M, Wennergren D, Edin K, Apelqvist I, Morberg P. Substantial accuracy of fracture classification in the Swedish Fracture Register: evaluation of AO/OTA-classification in 152 ankle fractures. Injury 2016;47:2579–83.
[15] Wennergren D, Ekholm C, Sundfeldt M, Karlsson J, Bhandari M, Möller M. High reliability in classification of tibia fractures in the Swedish Fracture Register. Injury 2016;47:478–82.
[16] Wennergren D, Stjernstrom S, Moller M, Sundfeldt M, Ekholm C. Validity of humerus fracture classification in the Swedish fracture register. BMC Musculoskelet Disord 2017;18:251.
[17] Knutsson SB, Wennergren D, Bojan A, Ekelund J, Moller M. Femoral fracture classification in the Swedish Fracture Register - a validity study. BMC Musculoskelet Disord 2019;20:197.
[18] Bergvall M, Bergdahl C, Ekholm C, Wennergren D. Validity of classification of distal radial fractures in the Swedish fracture register. BMC Musculoskelet Disord 2021;22:587.
[19] Fracture and dislocation compendiumOrthopaedic Trauma Association Committee for Coding and Classification. J Orthop Trauma 1996;10(Suppl 1):1–154 v–ix.
[20] Judet R, Judet J, Letournel E. Fractures of the acetabulum: classification and surgical approaches for open reduction. Preliminary report. J Bone Joint Surg Am. 1964;46:1615–46.
[21] Letournel E. Acetabulum fractures: classification and management. Clin Orthop Relat Res 1980:81–106.
[22] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
[23] Matta JM. Fractures of the acetabulum: accuracy of reduction and clinical results in patients managed operatively within three weeks after the injury. J Bone Joint Surg Am 1996;78:1632–45.
[24] Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960:37–46.
[25] McHugh ML. Interrater reliability: the kappa statistic. Biochem Med 2012;22:276–82.