



<http://www.diva-portal.org>

This is the published version of a paper published in *Injury*.

Citation for the original published paper (version of record):

Juto, H., Moller, M., Wennergren, D., Edin, K., Apelqvist, I. et al. (2016)

Substantial accuracy of fracture classification in the Swedish Fracture Register: Evaluation of AO/OTA-classification in 152 ankle fractures.

*Injury*, 47(11): 2579-2583

<https://doi.org/10.1016/j.injury.2016.05.028>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-128964>



## Substantial accuracy of fracture classification in the Swedish Fracture Register: Evaluation of AO/OTA-classification in 152 ankle fractures



Hans Juto<sup>a,\*</sup>, Michael Möller<sup>b</sup>, David Wennergren<sup>b</sup>, Klas Edin<sup>a</sup>, Ida Apelqvist<sup>a</sup>, Per Morberg<sup>a</sup>

<sup>a</sup> Department of Surgical and Perioperative Science (Orthopedics), Sunderby Research Unit, Umeå University, Umeå, Sweden

<sup>b</sup> Department of Orthopedics and Trauma, Sahlgrenska University Hospital, Gothenburg/Mölndal, Sweden

### ARTICLE INFO

#### Article history:

Received 14 March 2016

Accepted 19 May 2016

#### Keywords:

Ankle fractures  
Classification  
Agreement  
Accuracy  
Register

### ABSTRACT

**Introduction:** Since 2011 the Swedish Fracture Register (SFR) has collected information on fracture epidemiology, treatment and outcome. More than 112 000 fractures, including 11 600 ankle fractures, have been registered so far. The accuracy of the classification is crucial when using the register in future research and quality improvement of care. In this study we examine the degree of accuracy of the AO/OTA-classification in which ankle fractures are recorded in the SFR.

**Methods:** 152 randomly selected ankle fractures registered in the SFR were classified independently by a reference group of three orthopedic surgeons on two occasions. The agreed AO/OTA-classification of the reference group was regarded as the gold standard classification for each case. The originally recorded classification in the SFR was subsequently compared with the classification of the reference group and accuracy calculated.

**Results:** The agreement between the classification in the SFR and of the reference group was 74%, corresponding to kappa 0.67, 95% CI (0.58–0.76) for AO/OTA group level. An agreement of 88% equivalent to kappa 0.77, 95% CI (0.67–0.87) was noted for AO/OTA type level. The agreement corresponds to substantial according to Landis and Koch. The kappa value of inter- and intraobserver agreement ranged from 0.67 to 0.93.

**Conclusions:** The study results show substantial agreement between classifications made in the SFR and gold standard classification. The finding is equivalent to or higher than in previous studies. Consequently, we conclude that classifications of ankle fractures in the SFR are accurate and valid.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

### Introduction

The Swedish Fracture Register (SFR) collects information on fracture epidemiology, given treatments and outcome after treatment. Data on all extremity, pelvic and spine fractures in adults and all long bone fractures in children are included. Together with the ICD-10 code for diagnosis, a fracture classification is performed by the attending physician upon registration in the register. The entire process of register creation and implementation was recently described [1]. In December 2015, the SFR contained information about more than 112 000 fractures,

including 11 600 ankle fractures, recorded at 26 different departments.

The SFR uses the AO/OTA-classification for ankle fractures and the reliability of the classification system has been evaluated in several previous studies. Craig and Dirschl demonstrated, in a group of six observers with different experience, an interobserver agreement with a mean kappa value of 0.77 on AO/OTA type level and 0.61 on AO/OTA group level [2]. In the study by Thomson, Olsen and Nielsen, with as many as 74 observers a mean kappa of 0.57 was reached for interobserver agreement when using the Weber classification system (which corresponds to type level of the AO/OTA-classification) [3]. Four other studies show similar results [4–7]. The range of mean kappa values in earlier studies was at type level 0.57–0.77 and at group level 0.45–0.61.

To facilitate fracture classification in the SFR, the process online uses pictograms with the AO/OTA-classification drawings. In addition typical X-rays on each fracture group are available online as well as text explanations on the characteristics of each fracture

**Abbreviations:** SFR, Swedish Fracture Register; AO, Arbeitsgemeinschaft für osteosynthesefragen; OTA, Orthopaedic Trauma Association; ICD-10, International Classification of Diseases Tenth Revision.

\* Corresponding author at: Department of Orthopedics, Sunderby Hospital, SE-971 80 Luleå, Sweden.

E-mail address: [hans.juto@nll.se](mailto:hans.juto@nll.se) (H. Juto).

<http://dx.doi.org/10.1016/j.injury.2016.05.028>

0020-1383/© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

group. The AO/OTA-classification of ankle fractures (segment 44) contains type (3 options), group (9 options) and subgroup (27 options). The group level is used in the SFR except for A2 and A3 and it is possible to select twelve different options when recording ankle fractures: A1, A2.1/2, A2.3, A3.1/2, A3.3, B1, B2, B3, C1, C2, C3 and “not able to classify”.

Entry of data into the SFR is performed by the attending physicians who have a wide variation of orthopedic experience. Registration, including classification, is mainly done at the Accident & Emergency departments by residents with limited experience of fracture classification. The accuracy of the classification is of utmost importance when using the register for research purposes and interpretation of the data. The aim of our study was to examine the accuracy of the classification of ankle fractures in the SFR. A secondary aim was to evaluate inter- and intra-observer reliability in classification of ankle fractures.

## Methods

### Material

In 2013, 3040 cases of ankle fractures were registered in the SFR at 17 different orthopedic departments. The most frequently reporting department registered 657 ankle fractures and the least reporting department registered only two. For this study, 180 patients with ankle fractures (AO/OTA 44) injured in 2013, were randomly selected from the SFR. The extracted cases of ankle fractures had been treated at 16 different units. The departments were contacted and asked to deliver all digitized radiographs for their cases, which they all agreed to.

Radiographs from one department (22 cases) arrived after the due date and were therefore excluded. Information and classification on one case had been removed from the register at the

patient's own request. Images of five cases from different departments could not be retrieved and were not classified (Fig. 1).

Subsequently radiographs for 152 fractures (84%) from 15 departments were included in the study. In two of the cases a CT-scan had been conducted but the classification of the reference group was for logistical reasons based only on plain X-rays. The radiographs for all except two cases, included at least three standardized views (anteroposterior, mortise and lateral). For several cases there were additional images (range 2–11) of the ankle and most commonly we received four images (71% of cases). In 26 cases (17%) we also received images of proximal fibula. In 25 cases (16%) there were only X-rays with casts taken. Mean age among the 152 patients included in the study was 56 years (range 13–101) and 92 of them were females (ratio 1.5–1).

### Classification

Three orthopedic surgeons with different level of experience formed a reference group. The observers (HJ, KE, PM) were a senior registrar, a consultant and a senior consultant respectively. All were well familiar with classifying and treating ankle fractures. The reference group were presented with the same radiographic images that had been available to the attending physician at the time of the original classification upon registration in the SFR. The three observers independently, in their own pace and without any clinical information, classified the fractures. To get a classification as correct as possible, the descriptions of the fracture classes on the SFR website as well as the AO online surgery reference was available to the reference group at all times [8,9]. The intention of the study methodology was also to use the criteria proposed by Audigé et al. [10].

The results of the first classification was sealed. One month later the same three observers repeated the classification on the same

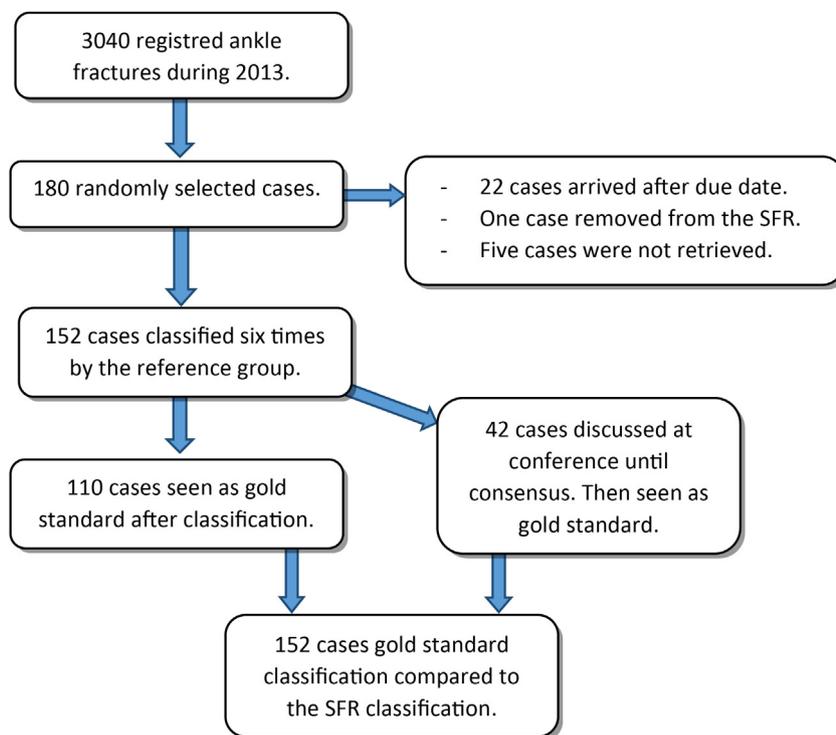


Fig. 1. Flow chart of ankle fractures included in the study.

fractures. If all six observations (three observers on two occasions) or at least five out of six observations corresponded, this classification was regarded as the final result i.e. the gold standard classification for that case. If less than five observations corresponded, the classification of the fracture was discussed among the three raters in a conference until consensus was reached. In this way a final classification, considered the gold standard classification, was obtained for each fracture. This result was then, in the final step of the validation process, compared with the classification recorded in the SFR at the various participating departments.

**Statistics**

The statistical analysis was done using the IBM SPSS statistical program, version 23 and inter and intra-observer reliability determined using the Cohen’s kappa statistic. 95% confidence intervals was calculated from kappa value  $\pm 1.96 \cdot SE$ . For interpretation of the results, the agreement of the kappa value was ranked as follows: 0–0.2 slight, 0.2–0.4 fair, 0.4–0.6 moderate, 0.6–0.8 substantial and 0.8–1.0 excellent or almost perfect, according to the guidelines of Landis and Koch [11].

A sample size analysis was made based on a power of 0.9 and a significance level of 0.05. The calculation showed that 144 cases were needed. We estimated a possible loss of 20% and therefore 180 cases were initially selected from the SFR.

**Ethics**

Ethical approval for the study was obtained from the Regional Ethics Committee at Umeå University (Dnr: 2014/143-31).

**Results**

*Classification of the reference group*

In 88 cases (58%), the reference group classified the fractures identically in six out of six observations and in 22 cases (14%), five out of six observations. In 42 cases (28%) the classification in the reference group corresponded in four or less out of six observations. These cases were discussed at a conference until consensus of the classification was reached. This classification was then considered as the gold standard classification for each case.

*Inter- and intraobserver agreement*

The interobserver agreement amongst the raters in the reference group was 88–93% (kappa 0.80–0.88) for AO/OTA type and 73–78% (kappa 0.67–0.73) for AO/OTA group (Table 1). The agreement between the different raters of the reference group and gold standard classification was 90–95% for AO/OTA type and 75–86% for AO/OTA group corresponding to mean kappa values of 0.83–0.90 and 0.69–0.83 respectively (Table 2). The intraobserver

**Table 1**  
Interobserver agreement for raters in the reference group on AO/OTA type and group level at first and second observation.

Rater	Observation	AO/OTA type			AO/OTA group		
		PA	Kappa	95% CI	PA	Kappa	95% CI
1 vs 2	1	89%	0.81	0.72–0.89	78%	0.73	0.65–0.81
	2	89%	0.81	0.72–0.89	78%	0.73	0.65–0.81
1 vs 3	1	88%	0.80	0.71–0.88	78%	0.73	0.65–0.81
	2	89%	0.81	0.73–0.90	76%	0.70	0.62–0.78
2 vs 3	1	93%	0.88	0.81–0.95	78%	0.73	0.65–0.81
	2	89%	0.81	0.73–0.90	73%	0.67	0.58–0.76

PA: Percentage of agreement.

**Table 2**  
Agreement on AO/OTA type and group level between each rater in the reference group and gold standard classification.

Rater	Observation	AO/OTA type			AO/OTA group		
		PA	Kappa	95% CI	PA	Kappa	95% CI
1	1	91%	0.85	0.77–0.93	86%	0.83	0.76–0.90
	2	93%	0.87	0.80–0.94	86%	0.82	0.75–0.89
2	1	94%	0.89	0.82–0.96	86%	0.82	0.75–0.89
	2	93%	0.88	0.81–0.95	86%	0.82	0.75–0.89
3	1	95%	0.90	0.84–0.97	84%	0.80	0.73–0.87
	2	90%	0.83	0.75–0.91	75%	0.69	0.61–0.78

PA: Percentage of agreement.

**Table 3**  
Intraobserver agreement of the raters in the reference group on AO/OTA type and group level.

Rater	AO/OTA type			AO/OTA group		
	PA	Kappa	95% CI	PA	Kappa	95% CI
1	88%	0.80	0.71–0.88	79%	0.74	0.67–0.82
2	96%	0.93	0.88–0.98	88%	0.86	0.79–0.92
3	94%	0.89	0.83–0.96	85%	0.81	0.74–0.88

PA: Percentage of agreement.

agreement in the reference group varied from 88 to 96% (kappa 0.80–0.93) on AO/OTA type level and 79–88% (kappa 0.74–0.86) on AO/OTA group level (Table 3). All kappa values for inter- and intraobserver agreement correspond to “substantial” or “almost perfect” according to Landis and Koch [11]. None of the results differed significantly from each other.

*Accuracy in the SFR*

Accuracy, defined as agreement between the classification in the SFR and the reference group classification was 74% corresponding to kappa 0.67, 95% CI (0.58–0.76) for AO/OTA group. This corresponds to “substantial agreement” according to Landis and Koch [11]. Compared to AO/OTA type level, the overall accuracy rose to 88%, equivalent to kappa of 0.77, 95% CI (0.67–0.87). When comparing the classification done at the different participating departments with the reference group classification for each case, the kappa values ranged from 0.31 to 0.83. A small number of cases per department is the cause of the large confidence intervals (Table 4).

*Agreement according to AO/OTA group*

The distribution between the classification in the SFR and the gold standard classification according to different fracture classes is demonstrated in Table 5. A variation in agreement between the different AO/OTA groups can be seen. In the A1 and B1-group there is 91% and 85% agreement respectively between the SFR and the gold standard classification whereas in the B2 and C2 group an agreement of 46% and 25% is seen. Table 5 also shows misclassification of each case. For example two cases defined as A1 fractures in the gold standard classification were wrongly classified as B2 fractures in the SFR. Similarly, seven cases of B1 fractures were wrongly classified as A1 fractures in the SFR. In a similar way six cases of B2 fractures and five cases of B3 fractures were wrongly classified as B1 and B2 fractures.

**Discussion**

The aim of this study was to use a method that gave a classification as correct as possible in a randomly selected sample of ankle fractures from the SFR. This was achieved by strictly using

**Table 4**  
Agreement between the original classification in the SFR for each department and the gold standard classification.

Department	Number of cases	AO/OTA ankle fracture type				AO/OTA ankle fracture group			
		Number of agreed	PA	Kappa	95% CI	Number of agreed	PA	Kappa	95% CI
1	5	4	80	A	A	4	80	0.74	0.29–1.00
2	13	8	62	A	A	6	46	0.31	0.04–0.57
3	8	6	75	A	A	4	50	0.36	–0.04–0.76
4	10	9	90	0.80	0.44–1.00	8	80	0.7	0.36–1.00
5	37	33	89	0.80	0.62–0.98	27	73	0.67	0.50–0.84
6	5	5	100	1.00	A	3	60	0.55	0.15–0.94
7	1	1	100	A	A	1	100	A	A
8	9	9	100	1.00	A	8	89	0.83	0.54–1.00
9	3	2	67	A	A	1	33	A	A
10	5	4	80	0.55	A	3	60	0.52	0.10–0.95
11	11	9	82	0.61	0.22–1.00	9	82	0.76	0.47–1.00
12	30	28	93	0.87	0.70–1.00	25	83	0.79	0.63–0.96
13	6	6	100	1.00	A	5	83	0.79	0.46–1.00
14	1	1	100	A	A	1	100	A	A
15	8	8	100	1.00	A	7	88	0.77	0.44–1.00
Total	152	133	88	0.77	0.67–0.87	112	74	0.67	0.58–0.76

A: Could not be calculated due to small numbers or full agreement.

PA: Percentage of agreement.

**Table 5**  
Crosstable of distribution between the original classification in the SFR and the gold standard classification.

		Gold standard classification													Total SFR
		A1	A2.1/2	A2.3	A3.1/2	A3.3	B1	B2	B3	C1	C2	C3	NF	NA	
SFR classification	A1	21	1	0	0	0	7	2	0	0	0	0	0	0	31
	A2.1/2	0	9	0	0	0	0	0	0	0	0	0	1	0	10
	A2.3	0	0	1	0	0	0	1	0	0	0	0	0	0	2
	A3.1/2	0	0	0	0	0	0	0	0	0	0	0	0	1	1
	A3.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	B1	2	0	0	0	0	46	6	0	0	1	0	0	1	56
	B2	0	0	0	0	0	1	11	5	1	1	0	0	0	19
	B3	0	0	0	0	0	0	3	13	0	0	0	0	0	16
	C1	0	0	0	0	0	0	1	0	5	0	1	0	0	7
	C2	0	0	0	0	0	0	0	0	2	1	0	0	0	3
	C3	0	0	0	0	0	0	0	0	1	1	5	0	0	7
	NA	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Total GS	23	10	1	0	0	54	24	18	9	4	6	1	2	152
	Agreement		91%	90%	100%	–	–	85%	46%	72%	56%	25%	83%	–	–

NA: Not able to classify.

NF: No fracture.

GS: Gold standard.

the guidelines of the AO/OTA-classification and performing classifications independently six times on each fracture and if necessary discussing until consensus was reached. A similar method was recently used when validating tibia fracture classification in the SFR. [12]. To our knowledge this is the best way to define a gold standard of fracture classification with which the original classification in the SFR can be compared and accuracy evaluated.

We observed a general agreement of kappa 0.77 between classification of ankle fractures in the SFR and the gold standard classification of AO/OTA type level and kappa 0.67 of AO/OTA group level. This corresponds to a substantial agreement according to Landis and Koch [11]. Despite the fact that many different physicians with a wide range of experience performed the actual classification recorded in the SFR, the agreement is as high as, or even higher than, the interobserver agreement in the previous studies (Table 6). We believe that the main reason for this is that the guidelines for the AO/OTA-classification are available online at the SFR-website [8] when the physicians classify fractures during registration in the SFR.

The study by Wennergren et al. showed moderate to substantial agreement (kappa 0.75 and 0.56 for AO/OTA type and group respectively) between gold standard and classification in the SFR [12]. They studied fractures in all three segments of the tibia which gives 27 fracture groups and reported some disagreement in segment assignment. This did not occur in the current study on ankle fractures with 12 fracture groups within one segment. This is

**Table 6**  
Mean kappa value of interobserver agreement in previous studies on AO/OTA-classification of ankle fractures. Interobserver agreement and agreement between the SFR and gold standard in the current study are included.

Study	AO/OTA type	AO/OTA group
Thomsen et al. 1991	0.57	–
Craig and Dirschl 1998	0.77	0.61
Brage et al. 1998	0.70	–
Thomsen et al. 2002	0.57	–
Malek et al. 2006	0.61	–
Alexandropoulos et al. 2010	–	0.45
Juto et al. 2016	0.82	0.71
SFR vs gold standard classification	0.77	0.67

probably the main explanation for the higher kappa values in the current study. In the current study we extended the methodology even further by randomly selecting fractures from all participating departments in the SFR. This further strengthens the conclusion that the classification of fractures in the SFR is accurate.

When classifying ankle fractures there are mainly three questions to answer. The level of the fibula fracture, the number of fractured malleoli and the fragmentation of the fibula fracture. The answers might result in 27 different subgroups in the AO/OTA-classification. However the subgroup classification is seldom used. The SFR uses the group level with 9 options with the exception for A2 and A3 which are split into subgroups for clarification. Together with the option “not able to classify”, the SFR ends up with 12 possible options.

In the different fracture groups there seems to be different difficulties in classifying correctly. A certain pattern of the misclassifications can be found. The A1-group seems to be difficult to differentiate from the B1-group rather than from the rest of type A-fractures. In contrast, the B2 and B3 fractures seems to be more closely “related” to each other and less likely to be misclassified as type A or C.

Consequently we detected three areas of difficulties when classifying ankle fractures according to the AO/OTA-classification system. First, the level of the fibula fracture that differentiates between type A and type B fractures. In the current study 2 cases of A1 fractures had been misclassified as B1 fractures and 7 cases of B1 fractures had been misclassified as A1 fractures (Table 5). It might be helpful to consider the trauma mechanism as described by Lauge-Hansen [13] and this is also included in the AO/OTA-classification system [9]. Accordingly, not only the level but also the pattern of the fibula fracture differentiates between a type A and a type B fracture. Type A (i.e. A1.3) is a transverse fracture as a result of an avulsion to the fibula and type B is an oblique fracture as a result of an external rotation or twisting force to the distal fibula.

Secondly, a difficulty arises when the number of fractured malleoli in type B fractures are to be assessed. In both medial and posterior malleoli fractures it is difficult to judge when to name a small fragment a malleolus fracture or a ligament rupture with a small bony avulsion. This is probably the reason for misclassification of several B2 fractures as B3 fractures and vice versa (Table 5).

The third difficulty is to differentiate between C1 and C2-fractures. The division is between a simple and a multi fragmented fibula fracture. In our experience this is often hard to perceive on plain radiographs. However type C fractures constitute only approximately 10% of all ankle fractures. No obvious pattern of misclassification could therefore be detected.

A possible limitation to the study is that all observers work in the same department leading to a potential bias. The high interobserver agreement could be affected by this but it is somewhat contradicted by the high intraobserver agreement.

## Conclusions

The study results show substantial agreement between classifications made in the SFR and the gold standard classification. The finding is equivalent to or higher than in previous studies. Consequently, we conclude that classifications of ankle fractures in the SFR are accurate and valid.

## Conflict of interest

There are no conflicts of interest reported for any of the authors.

## Contributions

The original design of the study was made by MM and DW with adjustments made by HJ and PM. IA collected and presented radiographs and HJ, KE and PM classified the fractures. HJ made the statistical analysis and compiled the results. HJ wrote the first version of the manuscript but all authors contributed with revisions and further elaboration.

## Acknowledgement

We thank the Department of Orthopedics and Trauma at Stockholm South General Hospital, Östersund Hospital, Karlstad Hospital, Sahlgrenska University Hospital, Kalmar Hospital, Skövde Hospital, Gävle Hospital, Sunderby Hospital, Eskilstuna Hospital, Västerås Hospital, Örebro Hospital, Lidköping Hospital, Hudiksvall Hospital, Alingsås Hospital, Uddevalla Hospital and Borås Hospital for kindly helping us to get access to the radiographs needed. Also thanks to Magnus Hultin, Associate Professor, Department of Surgical and Perioperative Science, Umeå University for help with statistics and Linda Akrami, Centre of Registers, Västra Götaland for the help of getting access to data from the SFR. The study has been funded by the Department of Research and Education, Norrbotten County Council, Sweden.

## References

- [1] Wennergren D, Ekholm C, Sandelin A, Möller M. The Swedish fracture register: 103,000 fractures registered. *BMC Musculoskelet Disord* 2015;16:338.
- [2] Craig 3rd WL, Dirschl DR. Effects of binary decision making on the classification of fractures of the ankle. *J Orthop Trauma* 1998;12:280–3.
- [3] Thomsen NO, Olsen LH, Nielsen ST. Kappa statistics in the assessment of observer variation: the significance of multiple observers classifying ankle fractures. *J Orthop Sci* 2002;7:163–6.
- [4] Alexandropoulos C, Tsourvakas S, Papachristos J, Tselios A, Soukoulis P. Ankle fracture classification: an evaluation of three classification systems: Lauge-Hansen, A: O. and Broos-Bisschop. *Acta Orthop Belg* 2010;76:521–5.
- [5] Brage ME, Rockett M, Vraney R, Anderson R, Toledano A. Ankle fracture classification: a comparison of reliability of three X-ray views versus two. *Foot Ankle Int* 1998;19:555–62.
- [6] Malek IA, Machani B, Mevcha AM, Hyder NH. Inter-observer reliability and intra-observer reproducibility of the Weber classification of ankle fractures. *J Bone Joint Surg Br* 2006;88:1204–6.
- [7] Thomsen NO, Overgaard S, Olsen LH, Hansen H, Nielsen ST. Observer variation in the radiographic classification of ankle fractures. *J Bone Joint Surg Br* 1991;73:676–8.
- [8] Svenska Frakturregistret (Swedish Fracture Register web site). Available at: <http://www.frakturregistret.se> (accessed 15.12.15).
- [9] Barbosa P, Bonnaire F, Kojima K, AO Surgery Reference-Malleoli. Available at: <https://www2.aofoundation.org/wps/portal/surgery?showPage=diagnosis&bone=Tibia&segment=Malleoli> (accessed 15.12.15).
- [10] Audigé L, Bhandari M, Kellam J. How reliable are reliability studies of fracture classifications? A systematic review of their methodologies. *Acta Orthop Scand* 2004;75:184–94.
- [11] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
- [12] 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.
- [13] Lauge-Hansen N. Fractures of the ankle. II. Combined experimental-surgical and experimental-roentgenologic investigations. *Arch Surg* 1950;60:957–85.