Observer variation in ankle fracture classification amongst orthopaedic surgeons in Brunei Darussalam

Mohammad Abdul Hafidz YAKOB [1], SA Jamaludin [2], Ketan PANDE [1]
[1] Department of Orthopaedics, and [2] Department of General Surgery, RIPAS Hospital, Brunei Darussalam

ABSTRACT

Introduction: Classification systems can be used in the treatment of fractures to grade severity of injury, plan treatment, predict and compare results. Different classification systems are available for ankle fractures. These systems are subject to observer variation and have been studied with varying results. The aim of this study was to assess the observer variation of the Lauge-Hansen and AO Weber classification of ankle fractures and to assess the agreement regarding the use of a syndesmotic screw in treatment of ankle fractures. **Materials and Methods:** Antero-posterior and lateral radiographs of patients who have undergone surgery for ankle fractures were collected (n=30). Eight orthopaedic surgeons of varying experience graded the radiographs independently and their inter- and intra-observer agreement was assessed using the kappa statistic. Two assessment sessions were conducted three weeks apart and each session was preceded by a presentation of the fracture classifications studied. **Results:** The results were expressed as mean kappa value. For the Lauge-Hansen classification the inter-observer value was 0.58 and intra-observer value was 0.64 and for AO Weber the values were 0.65 and 0.77 respectively. For the use of syndesmotic screw the values were 0.34 and 0.59 respectively. Level of experience was noted to have no effect on the intra-observer variation. **Conclusion:** AO Weber classification had better observer agreement compared to Lauge-Hansen classification and should be preferred. The decision for using a syndesmotic screw should not be based on radiological findings alone.

Keywords: Observer variation, ankle fracture, AO Weber classification, Lauge-Hansen classification, Kappa statistic

INTRODUCTION

In the orthopaedics specialty, classification systems are used to assess various injuries. These help in grading the severity of an injury, plan appropriate treatment and often can be used to predict outcome. Classification systems can also be used to quantify results, allowing us to compare practise between treatment centres. With ankle fractures, several classification systems are available. The decision for surgery is often based on radiographs of these injuries. With regards to stability of the ankle joint one is also able to assess the integrity of the syndesmosis and

Correspondence author: Mohammad Abdul Hafidz HJ YAKOB
Department of Orthopaedics, RIPAS Hospital,
Bandar Seri Begawan BA 1710, Brunei Darussalam.
Tel: +673 7160177
E mail: fizzy_83@hotmail.com

therefore decide if the use of a syndesmosis screw is warranted.

Interpreting ankle radiographs using classification systems and decision to use a syndesmotic screw is dependent on the observer, and thus subject to observer variability. Measuring the observer variation can help assess the validity of these classification systems. It is important to ensure that a particular classification system is reliable and reproducible justifying its widespread and regular use in clinical practise. Several studies have looked at the inter- and intra-observer variation for ankle fracture classifications which have shown a varying degree of agreeability. [5-7]

This aim of the study was to measure the inter- and intra-observer variation in; (a) two commonly used ankle fracture classifications and, (b) the decision to use a syndesmotic screw amongst Orthopaedic surgeons in the RIPAS Hospital, a tertiary referral centre in Brunei Darussalam.

MATERIALS AND METHODS

Patients who underwent surgical treatment of ankle fractures in RIPAS Hospital were selected (n=30). The preoperative antero-posterior and lateral radiographs of these patients were collected for the study. The two classification systems that were used when interpreting the ankle radiographs were the Lauge-Hansen and the AO Weber classifications.

The Lauge-Hansen classification system (Figure 1) of ankle fractures is based on the position of the foot and the direction of deforming force which is useful in describing the mechanism of injury. There are four types: Supination-adduction, Supination-external rotation, Pronation-abduction and Pronation-external rotation. The AO Weber classification is an anatomical classification which categorises fractures of the distal fibula in relation the distal tibio-fibula joint (syndesmosis): below (Type A), at the level (Type B), and above (Type C) (Figure 2). The syndesmotic screw (Figure 3) is generally used in injuries where the syndesmosis is likely disrupted, which correlates to higher grades of injury within the four types of injuries in the Lauge-Hansen classification and type B and C injuries in the AO Weber classification.

Eight Orthopaedic surgeons of various

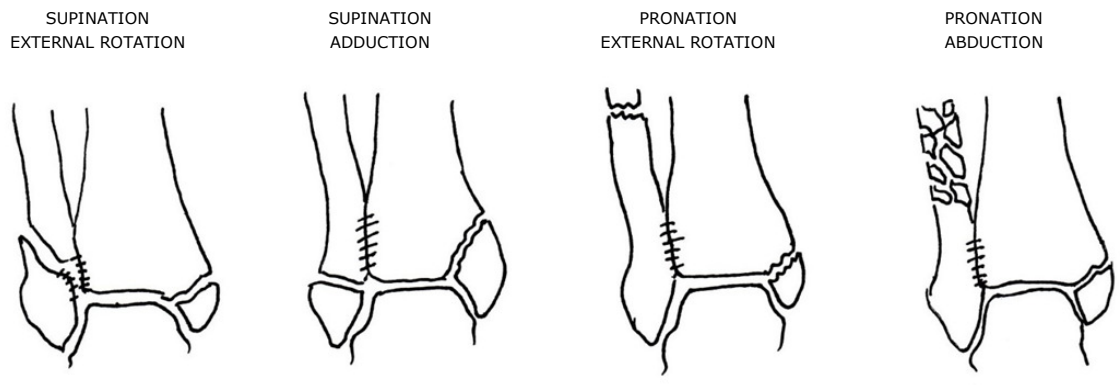

Fig. 1: Lauge-Hansen Classification of ankle fractures.

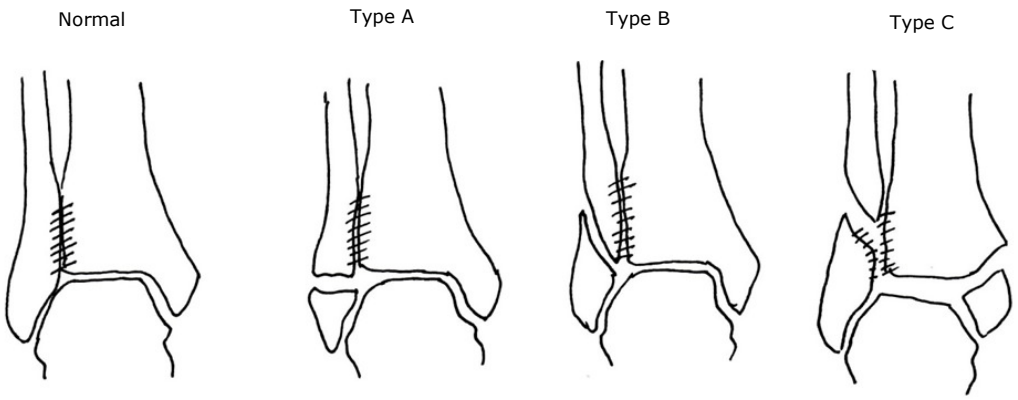

Fig. 2: AO Weber Classification.

Fig. 3: A syndesmotic screw.

degrees of experience participated as observers. The observers reviewed each set of radiographs at two separate occasions three weeks apart and classified each fracture according to Lauge-Hansen and AO Weber classification. The observers were also asked to decide if they would include the use of a syndesmotic screw as part of the operative treatment. The classification of each radiograph was blinded from the observer and the observers classified each radiographs individually without collaborating. Each session was preceded by a presentation on the fracture classifications studied.

For both inter- and intra-observer variation, the mean kappa value was used to represent the overall degree of agreeability within and between observers. Values from the second session were used to study inter-observer variation. As for the intra-observer variation the observations from the first and second sessions were compared for each observer.

To study the effect of experience on observer variation, the observers were divided into two groups, those with more and less than ten years of experience. The intra-observer variation of these groups was then compared. The kappa statistic was used to assess the inter- and intra-observer agreement and was derived using SPSS version 10.0. A kappa value of 1.0 indicates perfect agreement, taking into account that decisions

Table 1: Level of Kappa agreement.

Kappa Agreement	
< 0	Less than chance agreement
0.01–0.20	Slight agreement
0.21–0.40	Fair agreement
0.41–0.60	Moderate agreement
0.61–0.80	Substantial agreement
0.81–0.99	Almost perfect agreement

could have been made by chance, whereas a kappa value of 0.0 would indicate the opposite. This would imply that the closer the kappa value is to 1.0 the better the level of agreement. In interpreting this value, Landis and Kock developed a grading system which has been used to evaluate the level of agreement in this study (Table 1).

RESULT

With the Lauge-Hansen classification the mean kappa value for inter-observer variation was 0.58 with a range of 0.40-0.80. For intra -observer variation the mean kappa value was 0.64 with a range of 0.29-0.94.

The mean kappa value for inter-observer variation with the AO Weber classification was 0.65 with a range of 0.20-0.90. For intra-observer variation the mean kappa value was 0.77 with a range of 0.60-0.90.

For the decision of using a syndesmotic screw the mean kappa values for inter-observer variation and intra-observer variation were 0.34 (0.02-0.86) and 0.59 (0.10-0.90) respectively.

For the Lauge-Hansen classification the kappa values were 0.58 in the experienced group and 0.9 in the non-experienced group for intra-observer variation. For the AO Weber classification it was 0.78 and 0.75 respectively. For the decision to use a syndesmotic screw the kappa values were 0.55 and 0.65 respectively.

DISCUSSION

The results of our study showed that there was a moderate degree of agreement amongst the observers for the Lauge-Hansen classification, however there was substantial agreement when using the AO Weber classification. Intra-observer variation for both classifications was substantial.

With regards to decision for the use of a syndesmotic screw it was seen that inter-observer agreement was fair, however there was a moderate level of agreement for intra-observer variation.

In the study by Thomsen *et al.* four observers classified 94 radiographs of ankle fractures using the Lauge-Hansen and AO Weber classifications. [6] The mean kappa score for inter-observer variation for Lauge-Hansen was 0.60 and for AO Weber it was 0.56. The intra-observer scores were 0.65 and 0.69 respectively. This study showed comparable kappa scores for each classification system, however the authors concluded that the complexity of the subdivisions of the Lauge-Hansen system limited its practical use in the day to day clinical setting.

Malek *et al* included 50 sets of ankle radiographs with five observers to assess agreeability for the AO Weber Classification. [5] Their study concluded that the AO Weber classification had substantial inter (0.61) and intra (0.74) observer agreement.

More recently, Alexandropoulos *et al* had three observers who classified 293 radiographs using the Lauge-Hansen (0.17–0.48) and AO Weber (0.40–0.48) classifications. [7] The study concluded that both the classifications had considerable inter-observer variability which restricts their use in selection of treatment options and prognosis. [4]

Our results were similar to the studies mentioned and we found the AO Weber classification to have a higher level of agreeability. As suggested by Thomsen et al the Lauge -Hansen system has inherent complexity which may be a cause for the lower Kappa scores. [6] Conversely the simpler AO Weber classification allowed for a higher mean kappa score.

Fracture classification systems are subject to a degree of variation as it is based on the subjective observation of the clinician at that time. Other factors that can give rise to variability between observers include; the complexity of the grading system, familiarity with a given system, the level of experience and number of observers. [16]

As already demonstrated the degree of complexity of the Lauge-Hansen classification can give rise to variability. Minute differences between grades within a classification system or a system with numerous grades has been shown to cause discrepancy between observers.

Familiarity with a particular system can also influence the degree of agreement within a group of observers. If instruction in the classification system is given this can increase inter-observer agreement. However this factor was eliminated from our study as instruction on both classification systems where given before each observation session. The Kappa value is also affected by the number of observers in a study group. With increasing number of observers in a group the Kappa value stabilises towards a mean value indicating a decrease in sampling variation and standard error.

With experienced clinicians interpreting radiographs would be more consistent therefore we would expect their intra-observer agreement to be higher. However we found that the experienced group had on average similar kappa scores for the AO Weber classification and use of syndesmotic screw and a lower mean kappa score for the Lauge-Hansen classification.

To our knowledge no study has measured the inter- and intra-observer variation for the decision to use a syndesmotic screw based on ankle radiographs alone however Hermans *et al* measured the correlation of predicted syndesmotic injuries compared to MRI and found that the Lauge-Hansen classification had a sensitivity and specificity of 92% in detecting syndesmotic injuries. However due to the low level of agreement that we found, it is not recommended to base this decision on radiological findings alone.

In conclusion, based on the findings of this study it is recommended that the AO Weber classification be used in day to day practice over the Lauge-Hansen classification because of its higher observer agreement and its simplicity. The decision for using a syndesmotic screw should not be based on radiological findings alone as it only has fair agreement.

REFERENCES

1: Sidor M, Zuckerman JD, Lyon T, et al. The Neer classification system for proximal humeral fractures. An assessment of interobserver reliability and intraobserver reproducibility. J Bone Joint Surg Am. 1993; 75:1745-50.

2: Schneidmüller D, Röder C, Ralf Kraus, et al. Development and validation of a paediatric long-bone fracture classification. A prospective multicentre study in 13 European paediatric trauma centres.

BMC Musculoskeletal Disorders 2011; 12:89.

3: Phillips RS. A plea for simplicity in the classification of ankle fractures. Postgrad Med J. 1968; 44:199-211.

4: Sowman B, Radic R, Kuster M, et al. Distal tibiofibular radiological overlap. Does it always exist? Bone Joint Res. 2012; 1:20–4.

5: Malek IA, Machani B, Mevccha 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.

6: Thomsen NO, Overgaard S, Olsen LH, et al. Observer variation in the radiographic classification of ankle fractures. J Bone Joint Surg BR. 1991; 74:676 -8.

7: Alexandropoulos C, Tsourvakas S, Papachristoc J, et al. Ankle fracture classification: an evaluation of three classification systems : Lauge-Hansen A.O and Broos-Bisschop. Acta Orthop Belg. 2010; 76:521-5.

8: Howard KA, Forrester DM. Classification of Ankle Fractures: An Algorithm. AJR. 1980; 135:1057-63.

9: McRae R. Pocketbook of Orthopaedics and Fractures. Miller A, Conn F, editors. Regional injuries: the ankle. 2nd Ed. Churchill Livingstone; 2006: pg480.

10: Rammelt S, Zwipp H, Grass R. Injuries to the distal tibiofibular syndesmosis: an evidence-based approach to acute and chronic lesions. Foot Ankle Clin. 2008; 13:611-33.

11: Viera AJ, Garrett JM. Understanding Interobserver Agreement: The Kappa Statistic. Fam Med 2005; 37:360-3.

12: Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33:159-74.

13: Wainwright AM, Williams JR, Carr AJ. Interobserver and intraobserver variation in classification systems for fractures of the distal humerus. J Bone Joint Surg. 2000; 82:636-42.

14: Rasmussen S, Madesen PV, Bennicke K. Observer variation in the Lauge-Hansen classification of ankle fractures. Percision improved by instruction. Acta Orthop Scand. 1993; 64:693-4.

15: 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.

16: Pande KC, Babhulkar SS. Observer variation in the use of fracture classification systems. Indian J Orthop. 2002; 36:2.

17: Hermans JJ, Wentink N, Beumer A *et al*. Correlation between radiological assessment of acute ankle fractures and syndesmotic injury on MRI. Skeletal Radiol. 2012; 41:787–801.