Research Article

Reliability of commonly used classification systems of trochanteric fractures of the proximal femur and evaluation of CT in improving the interobserver agreement

Dr Mihir Thanvi[1], Dr Anil Kumar Gupta[2*], Dr Naveen Goyal[2], Dr. Rahul Agarwal[3], Dr. Ritesh Agarwal[3], Dr. Chinmaya Sharma[1]

[1]M.S. Orthopaedics, Senior resident, Department of Orthopaedics, SMS Medical College, Jaipur, Rajasthan, India
[2]M.S. Orthopaedics, Consultant Orthopaedic Surgeon, Jaipur, Rajasthan, India
[3]M.S. Orthopaedics, Senior resident, Department of Orthopaedics, Mahatma Gandhi Medical College, Jaipur, Rajasthan, India

*Corresponding author
Dr Anil Kumar Gupta
Email: drguptaanil2004@yahoo.co.in

Abstract: Objective and Aim of the study was to assess the interobserver reliability of two frequently used classifications for trochanteric femur fractures, the Boyd-Griffin classification and the AO/ASIF classification. Furthermore, the change in the level of agreement after addition of CT to plain radiograph was also evaluated. Fracture patterns of 31 consecutive patients with trochanteric fracture of the proximal femur were analyzed by four different observers (two radiologists and two orthopaedic surgeons) using plain X-Rays (Standard AP & Lateral views) and with combination of CT Scans. The level of agreement was then analyzed. The mean kappa value for radiologists & orthopaedists using BG system was 0.05 and 0.42. The mean kappa value for radiologists & orthopaedists using AO/ASIF without sub-groups was 0.17 and 0.54. The mean kappa value for radiologists & orthopaedists using AO/ASIF with sub-groups was 0.02 and 0.27. The mean kappa value for radiologists & orthopaedists using Boyd-Griffin system was 0.00 and 0.27. The mean kappa value for radiologists & orthopaedists using AO/ASIF excluding sub-groups system was 0.15 and 0.65. The mean kappa value for radiologists & orthopaedists using AO/ASIF with sub-groups was – 0.12 and 0.29. We found a 'slight' reliability for the Boyd-Griffin classification and only a 'fair' reliability for the AO/ASIF classification. Furthermore, our study showed that the reliability of the AO/ASIF classification improved when subgroups of the classification were not provided. Addition of computed tomography to the plain radiographs has improved observer reliability from 'moderate' to 'substantial' levels in comparison to radiographs alone.
Keywords: Trochanteric fractures, Classification, Reliability

INTRODUCTION

An ideal fracture classification system should provide information on fracture geometry and more importantly, it should guide the treatment & prognosis. In order to use it for research purpose the classification should have a high degree of reliability.

Trochanteric femoral fracture treatment is considered to be a common practice in orthopaedic traumatology and the fracture account for about half of all hip fractures.

The reliability of the two frequently used classifications, the Boyd-Griffin classification and the AO/ASIF classification, have been assessed in a limited number of studies. The Boyd-Griffin classification [1] is a very basic classification with only 4 types & describes the location of the fracture line. A more recently developed AO/ASIF classification [2] is designed to provide prognostic information on achieving and maintaining reduction of the fracture in addition to the fracture pattern. In the more complex type of trochanteric fractures adequate radiological

evaluation could be the answer to evaluate an adequate treatment plan and a reliable fracture classification. The value of computed tomography (CT) has been studied for different type of fractures with complicated fracture patterns such as tibial plateau or calcaneal fractures and proved to be superior to plain radiography [3-7]. However, for trochanteric fractures of higher complexity improvement of the reliability of fracture classifications was never assessed with CT in a clinical study. It is possible that better understanding of the fracture type and improved preoperative planning will result in higher agreement and improved clinical outcome [8].

The goal of this study is to assess the interobserver reliability of two frequently used classifications for trochanteric femur fractures, the Boyd-Griffin classification and the AO/ASIF classification. Furthermore, the change in the level of agreement after addition of CT to plain radiograph was also evaluated.

MATERIALS AND METHODS

We enrolled 31 consecutive patients with trochanteric

fracture of the proximal femur. The AP (antero-posterior) view was taken in 15-degree internal rotation to offset the femoral anteversion and a lateral view followed by a CT scan.

The scans were analysed by four different observers – two Radiologists and two Orthopaedic surgeons. They were asked to classify the fracture pattern according to the Boyd-Griffin classification and the AO/ASIF classification. Initially only the X-Ray films were provided and later CT Scans were also provided to mark the change of opinion with addition of CT Scans. The observers were given adequate time for assessment without any time restrictions. The observers were not allowed to discuss their findings with others & were provided with detailed description of both classification systems. Statistical analysis was performed by calculating the Cohen kappa value using SPSS 17.0 statistical software for interobserver reliability. We interpreted the kappa value coefficient according to the guidelines proposed by Landis and Koch: less than 0.00 poor reliability, 0.00– 0.20 slight reliability, 0.21–0.40 fair reliability, 0.41–0.60 moderate reliability, 0.61–0.80 substantial agreement and 0.81–1.00 almost perfect agreement [9].

RESULTS

The mean age of the 31 subjects was 59 years (32-98 years). Fifteen patients were male and 16 female. Table 1 shows the entire kappa values calculated after observers classified the fractures using radiographs alone.

-The mean kappa value for radiologists & orthopaedists using BG system was 0.05 and 0.42.
- The mean kappa value for radiologists & orthopaedists using AO/ASIF without sub-groups was 0.17 and 0.54.
- The mean kappa value for radiologists & orthopaedists using AO/ASIF with sub-groups was 0.02 and 0.27.

Table 2 shows the entire kappa values calculated after observers classified the fractures using radiographs combined with computed tomography.
- The mean kappa value for radiologists & orthopaedists using Boyd-Griffin system was 0.00 and 0.27.
- The mean kappa value for radiologists & orthopaedists using AO/ASIF excluding sub-groups system was 0.15 and 0.65
- The mean kappa value for radiologists & orthopaedists using AO/ASIF with sub-groups was – 0.12 and 0.29.

In this study the reliability of two commonly used classifications for trochanteric femur fractures, the AO/ASIF classification and the Boyd-Griffin classification, was compared. We found a 'slight' reliability for the Boyd-Griffin classification and only a 'fair' reliability for the AO/ASIF classification. Furthermore, our study showed that the reliability of the AO/ASIF classification improved when subgroups of the classification were not provided.

Table 1: Inter-observer agreements using plain radiographs only

Boyd-Griffin classification	Kappa value	SE
Radiologist1 - Radiologist2	0.05	0.06
Orthopaedist1 - Orthopaedist2	0.42	0.15
AO/ASIF (without sub-groups)		
Radiologist 1- Radiologist 2	0.17	0.14
Orthopaedist1-Orthopaedist2	0.54	0.13
AO/ASIF (with sub-group)		
Radiologist 1- Radiologist 2	0.02	0.19
Orthopaedist1-Orthopaedist2	0.25	0.12

Table 2: Inter-observer agreement when combined with CT

Boyd-Griffin classification	Kappa value	SE
Radiologist 1- Radiologist 2	0.0	0.06
Orthopaedist1-Orthopaedist2	0.27	0.16
AO/ASIF (without sub-groups)		
Radiologist 1- Radiologist 2	0.15	0.15
Orthopaedist1-Orthopaedist2	0.65	0.14
AO/ASIF (with sub-group)		
Radiologist 1- Radiologist 2	-0.12	0.13
Orthopaedist1-Orthopaedist2	0.29	0.14

DISCUSSION

Any fracture classification system should be reliable among different observers as well as by the same observer on separate occasions. In addition, the classification should suggest treatment and/or outcome. The kappa statistic provides a method of determining the diagnostic accuracy of a classification system with the calculation adjusting for random matches.

Previous studies have assessed the reliability of either the AO or other commonly used classification systems. Chapman *et al.* [10] to evaluated the reliability of AO/ASIF and Jensen's classification system. They found that the interobserver reliability does not change dramatically when CT is added to plain radiographs. They also found that inter-observer kappa coefficients were also higher for radiologists than orthopaedic surgeons i.e. 0.67 & 0.57 respectively. And concluded that these classification schemes may not be comprehensive in describing fracture pattern and displacement. Finally, both systems failed to provide any prognostic value.

H. Pervez *et al.* [11] took five observers using the Jensen modification of the Evan's classification and the AO classification (with and without subgroups) classified the radiographs of 88 trochanteric hip fractures. For the Jensen classification, the mean kappa value was 0.34. For the AO system with subgroups, the mean kappa value was 0.33. For the AO classification system without subgroups, the mean kappa value was 0.62. Our study also observed similar results.

Van Embden *et al.* [8] in their prospective study compared the reproducibility of two classifications for trochanteric femur fractures: the Jensen classification and the AO/ASIF classification. The inter-observer agreement of the AO/ASIF classification and the Jensen classification was kappa 0.40 and kappa 0.48 respectively. The inter-observer agreement of the AO/ASIF classification without the sub-groups was 0.68 and concluded that both the classifications systems showed poor reliability.

Schipper *et al.* [12] did a prospective study on the reliability of AO/ASIF classification. 20 radiographs were reviewed twice by 15 observers. A substantial improvement in agreement was found when fractures were classified only according to main groups rather including subgroups. The mean kappa value for interobserver reliability was 0.33 and 0.67 without sub-groups. They concluded that the AO/ASIF classification for pertrochanteric fractures is reliable for fracture groups 31A1, A2 or A3. The group classification should be used to compare scientific data and determine the best treatment. Further classification of fracture subgroups leads to poor reproducibility of results.

Wen-Jie Jin *et al.* [13] also studied the reliability of classification systems for intertrochanteric fractures of the proximal femur among experienced orthopaedic surgeons. Five experienced observers using AO, Evans, Kyle and Boyd-Griffins classification systems classified forty trochanteric fractures independently and interobserver variation was assessed. They found that the AO classification system with groups could be used more reliably to measure intertrochanteric fractures of the proximal femur than the other classification systems. However, the reliability of the AO classification with subgroups is not satisfactory. Their kappa values were 0.29 for Boyd-Griffin system and 0.80 for AO system without sub-groups and 0.38 including sub-groups.

We found that AO/ASIF classification system is a better system than Boyd-Griffin system in classifying trochanteric fractures. We also found that the simpler classification of trochanteric fracture into three AO groups (31A1, 31A2, 31A3) is more acceptable. The reliability of both the classification system is affected by the training, knowledge and experience of the observers. We have included CT in addition to plain radiographs which only one study had included [10]. In our study addition of computed tomography to the plain radiographs has improved observer reliability from 'moderate' to 'substantial' levels in comparison to radiographs alone. So CT can be an excellent mode for the diagnosis and classification of trochanteric fractures.

REFERENCES

1. Boyd HB, Griffin LL; Classification and treatment of trochanteric fractures. Arch Surg., 1949; 58(6):853—66.
2. Müller ME, Nazarian S, Koch P, Schatzker J; The comprehensive classification of fractures of the long bones. Springer-Verlag, Berlin/Heidelberg, 1990.
3. Chan PS, Klimkiewicz JJ, Luchetti WT, Esterhai JL, Kneeland JB, Dalinka MK *et al.*; Impact of CT scan on treatment plan and fracture classification of tibial plateau fractures. J Orthop Trauma, 1997; 11(7): 484–489.
4. Doornberg J, Lindenhovius A, Kloen P, van Dijk CN, Zurakowski D, Ring D; Two and three-dimensional com- puted tomography for the classification and management of distal humeral fractures. Evaluation of reliability and diagnostic accuracy. J Bone Joint Surg Am., 2006; 88(8):1795–1801.
5. Haapamaki VV, Kiuru MJ, Koskinen SK; Ankle and foot injuries: analysis of MDCT findings. AJR Am J Roentgenol., 2004;183(3): 615–622.
6. Humphrey CA, Dirschl DR, Ellis TJ; Interobserver reliability of a CT-based fracture classification system. J Orthop Trauma, 2005; 19(9): 616–622.
7. Mustonen AO, Koskinen SK, Kiuru MJ; Acute

knee trauma: analysis of multi- detector computed tomography findings and comparison with conventional radiography. Acta Radiol., 2005; 46(8): 866–874.

8. van Embden D, Rhemrev SJ, Meylaerts SA, Roukema GR; The comparison of two classifications for trochanteric femur fractures: the AO/ASIF classification and the Jensen classification. Injury, 2010; 41(4): 377-381.

9. Landis JR, Koch GG; The measurement of observer agreement for categorical data. Biometrics, 1977; 33(1): 159–174.

10. Chapman CB, Herrera MF, Binenbaum G, Schweppe M, Staron RB, Feldman F *et al.*; Classification of intertrochanteric fractures with computed tomography: a study of intraobserver and interobserver variability and prognostic value. Am J Orthop (Belle Mead NJ), 2003; 32(9): 443-449.

11. Pervez H, Parker MJ, Pryor A, Lutchman L, Chirodian N; Classification of trochanteric fracture of the proximal femur: a study of the reliability of current systems. Injury, Int J Care Injured, 2002; 33(8): 713–715.

12. Schipper IB, Steyerberg EW, Castelein RM, van Vugt AB; Reliability of the AO/ASIF classification for pertrochanteric femoral fractures. Acta Orthop Scand., 2001; 72 (1): 36-41.

13. Jin WJ, Dai LY, Cui YM, Zhou Q, Jiang LS, Lu H; Reliability of classification systems for intertrochanteric fractures of the proximal femur in experienced orthopaedic surgeons. Injury, Int. J. Care Injured, 2005; 36(7): 858-861.