


Inter- and Intra-Observer Reliability of the New Comprehensive Classification of Intertrochanteric Fracture of the Femur

Xiaopei Xu, Yang Liu, Shuo Diao, Junlin Zhou 

Department of Orthopedic Surgery, Beijing Chaoyang Hospital, Capital Medical University, Beijing, 100020, People's Republic of China

Correspondence: Junlin Zhou, Department of Orthopedic Surgery, Beijing Chaoyang Hospital, Capital Medical University, 8 Gongren Tiyuchang Nanlu, Chaoyang District, Beijing, 100020, People's Republic of China, Tel/Fax +86-01085231227, Email junlinzhou_article@outlook.com

Objective: To evaluate the reliability of the new Comprehensive Classification in clinical practice by comparing it with Evans-Jensen (1975), AO/OTA (2007), and AO/OTA (2018).

Methods: A total of 192 patients with intertrochanteric fractures were included in the study. The fractures were classified using the Comprehensive Classification system and the three traditional methods by eight observers, consisting of four high-experienced and four less-experienced surgeons. Interobserver and intraobserver agreements were assessed using kappa statistics and compared across the different classification systems.

Results: The study included 192 patients. The average inter-observer Kappa values using CT scan for Evans-Jensen (1975), AO/OTA (2007), AO/OTA (2018), and Comprehensive Classification were 0.361 ± 0.06 , 0.348 ± 0.05 , 0.408 ± 0.06 , and 0.492 ± 0.02 , respectively. The average inter-observer Kappa values based on 3DR were 0.411 ± 0.04 , 0.396 ± 0.03 , 0.442 ± 0.03 , and 0.524 ± 0.03 , respectively. The average intra-observer Kappa values using CT scan were 0.467 ± 0.05 , 0.453 ± 0.06 , 0.542 ± 0.04 , and 0.631 ± 0.02 , respectively. The average intra-observer Kappa values using 3DR were 0.503 ± 0.05 , 0.482 ± 0.03 , 0.561 ± 0.03 , and 0.672 ± 0.02 , respectively. The mean kappa value of Inter- and intra-observer for the Comprehensive Classification were higher than Evans-Jensen (1975) and AO/OTA (2007/2018) using CT scans and 3DR ($P < 0.05$). Additionally, the high-experienced group using 3DR had the highest agreement, at approximately 75.86%, when comparing the implant type corresponding to the fracture classification evaluated by the observer with the implant type selected by the observer.

Conclusion: The Comprehensive Classification system emerges as a reliable and superior tool for the classification of intertrochanteric fractures, offering improved agreement rates and practical treatment insights compared to traditional methods. This classification has guiding significance for treatment.

Keywords: intertrochanteric fractures, hip fractures, femur, fracture classification, reliability

Introduction

Femoral intertrochanteric fractures is a common type of fracture in clinical practice, with significant differences in fracture incidence rates in different regions and countries.¹ North America and Asia are considered as moderate-risk areas. Considering population density and aging trends, by 2050, more than half of hip fractures globally will occur in Asia.¹ Due to its complex fracture pattern and poor patient condition, it poses significant challenges and economic burdens to clinical work.² It is well known that in trauma assessment, the main purpose of fracture classification systems is to unify the language used to describe the injury pattern. A good system should cover a wide range of fracture types, guide treatment, and be repeatable by different practitioners.³

Currently, the classic classification systems commonly used in clinical practice include Evans-Jensen (1975)⁴ and AO/OTA (2007/2018).^{5,6} Both Evans-Jensen and AO/OTA are based on X-ray examination, and classified based on the shape of the fracture line, the number of “intermediate bone blocks”, and the stability of the lateral wall. They also guide the selection of internal fixation devices based on the concept of the lateral wall, but lack classification of posterior

coronal bone blocks and small trochanter displacement.⁷ With the continuous development and maturity of computed tomography (CT) scans, CT-based classification systems, such as the Shoda classification⁸ and Kijima classification,⁹ continue to emerge. These systems have significantly improved the accuracy of diagnosis, helped clinical physicians to obtain a better understanding of the fracture pattern, and provided more Comprehensive Classification methods, thereby improving the value of treatment decision-making.^{10,11}

Recently, Chang et al⁷ proposed a new Comprehensive Classification of femoral intertrochanteric fractures based on 3D-CT reconstruction (3DR). It consists of 4 types (A1 - A4), each with 4 subtypes. A1 (20%) is simple two - part fractures like isolated single trochanter fractures, etc., treatable by DHS/DHB. A2 (62.5%) features lesser trochanter and posterior coronal fractures. Short cephalomedullary nails are usually used, and A2.4 may need long nails or cerclage. A3 (15.5%) has a ruptured lateral wall, needing long nails and specific distal locking. A4 (2%) is medial comminution, demanding combined treatments such as intramedullary and extramedullary systems or prosthesis replacement. See Table 1 and Figure 1 for details. According to the new Comprehensive Classification method, we found that this classification is more intuitive and complete, and pays more attention to the anatomical structure and treatment options of the fractures. At the same time, it fills the gap in the A2.1 subgroup of AO/OTA (2018) and improves rare subtypes. Currently, many studies have analyzed the reliability of traditional classification systems under X-ray and found that their reliability is not ideal and cannot be completely relied upon.¹²⁻¹⁴ Therefore, with the development and widespread use of CT scans and 3DR, whether the reliability of these traditional classification systems will be significantly improved remains to be seen. Additionally, the reliability of the new Comprehensive Classification method in clinical use is still unclear. Therefore, the purpose of this study was to compare the intra-observer and inter-observer reliability of three widely used classification systems (Evans-Jensen, AO/OTA (2007), AO/OTA (2018)), and the new Comprehensive Classification method in CT scan and 3DR, to further evaluate whether these classifications are suitable for widespread use in clinical practice and further guide the treatment of fractures.

Table 1 A Sophisticated Fracture Classification of Proximal Femur Trochanteric Region AO/OTA-31A

Types	Subtypes			
A1	A1.1	A1.2	A1.3	A1.4
Simple two-part fractures	Isolated single trochanteric fracture. (1) GT; (2) LT	Non-displaced fracture (1) Occult and incomplete (2) Complete with no displacement	Two-part displaced fracture. (1) LT with the shaft; (2) LT with head-neck; (3) LT bisected	Two-part displaced fracture with additional posterior coronal fragment (1) Posterior part of GT; (2) +Posterior crest
A2	A2. 1	A2.2	A2.3	A2.4
With LT fragment, or plus partial LW fracture, posterior coronal fragment	=1 Isolated LT	=1 Posterior banana-like fragment (GT + LT)	≥2 Posterior intermediate coronal fragments	Large posterior coronal fragments, with. (1) Distal cortex extension >2 cm, (2) Anterior expansion to entry portal on lateral cortex
A3	A3.1	A3.2	A3.3	A3.4
With complete LW broken, transverse anterior cortex fracture, proximal and distal part	Reverse oblique fracture. LW fracture line is above/at/ below the entry portal of the head-neck implant	Transverse intertrochanteric fracture	Comminuted intertrochanteric fracture	Pantrochanteric fracture (five-part). The fracture line at the AMC inferior corner is simple and can be reduced to direct contact
A4	A4. 1	A4.2	A4.3	A4.4
With further AMC inferior corner comminution, lack of medial compression force transmission structure	Comminution of the AMC at extracapsular	Comminution of the AMC at intracapsular; the clinical calcar	Comminution of the AMC at both extra and intracapsular	Segmental fracture Concomitant ipsilateral neck and trochanter fractures

Abbreviations: GT, greater trochanter; LT, lesser trochanter; LW, lateral wall; AMC, anteromedial cortex.

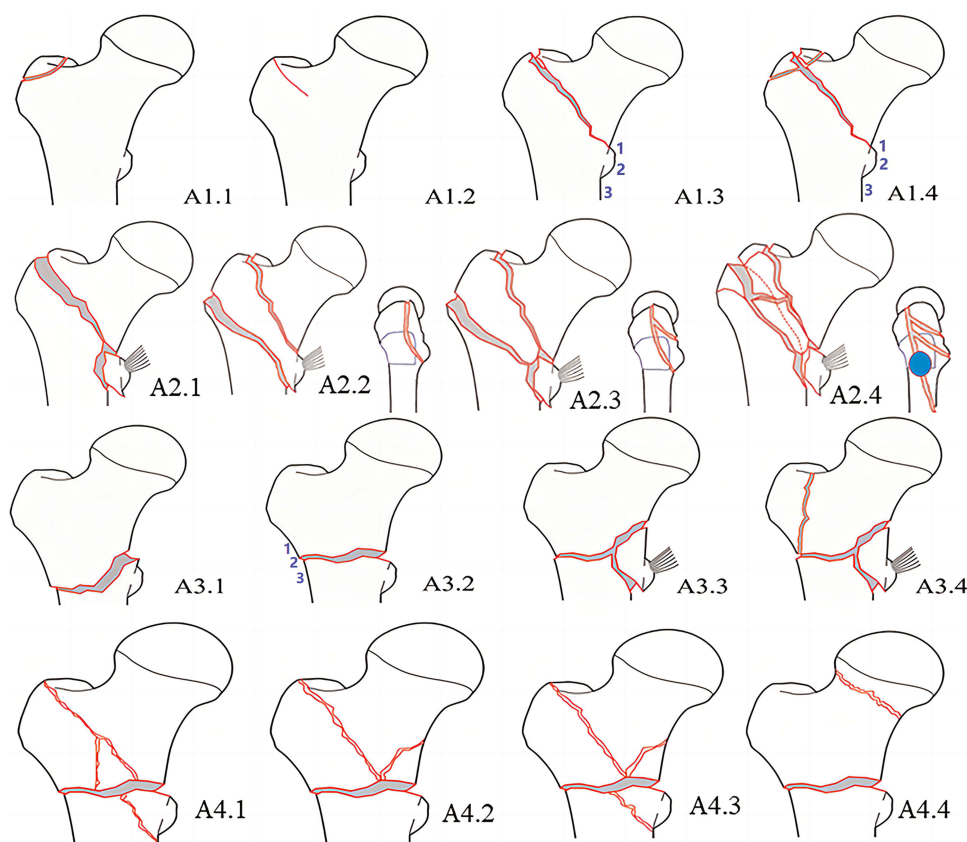


Figure 1 A sophisticated fracture classification of proximal femur trochanteric region AO/OTA-31A.

Material and Methods

Study Design, Setting, and Participants

We conducted a multicenter observational study involving orthopedic surgeons from three tertiary hospitals to determine the intra- and inter-observer reliability of the new four-by-four Comprehensive Classification system. The study was approved by the institutional review board, waiving the requirement for formal written informed consent due to the retrospective nature of the study.

A total of 192 patients with intertrochanteric fractures treated at the Orthopaedic Department of Beijing Chaoyang Hospital from January 2019 to January 2022 were collected. We excluded the following patients based on exclusion criteria: (i) exclude all patients with previous injuries to the proximal femur and hip diseases; (ii) old fractures (time interval between trauma and imaging > 3 weeks); (iii) patients with bone defects; (iv) patients with pathological fractures associated with primary bone tumors and bone metastases; (v) patients with incomplete, inaccurate and unclear imaging data.

Surgeons Grouped by Expertise

The cases were evaluated by eight evaluators representing two different levels of training for hip fractures from three tertiary hospitals. The high-experience group consisted of four senior orthopedic surgeons with at least 10 years of experience, and the low-experience group consisted of four orthopedic trauma residents with less than 5 years of experience. At least one senior orthopedist and one resident at each hospital participated in the assessment.

Image Data Processing and Evaluation

All imaging data of the included patients were saved in DICOM (Digital Imaging and Communications in Medicine) format, and the basic patient information, including name and ID number, was hidden through image processing

software. Two professionally trained staff with no involvement in the study except for 3D reconstruction performed 3D reconstruction using MicroDicom (version 2.92, MicroDicom Ltd, Sofia, Bulgaria). Fractures were graded according to Evans-Jensen, AO/OTA (2007), AO/OTA (2018), and Comprehensive Classification.⁴⁻⁷ Subgroup analysis was performed by all observers who reviewed CT scans and 3DR of 192 patients. A schematic diagram of the four fracture classifications and a corresponding handout, supplemented with text, were provided to the observer during the review. To prevent communication between observers during the study, observers were given image folders with different codes.

All observers were asked to assess whether fractures were stable or unstable according to the dichotomous classification of the different fracture types. Using the Evans-Jensen system, the stable patterns were I and IIA, and all the other fractures were considered unstable Using the AO/OTA (2007) and AO/OTA (2018) systems, the stable patterns were A1 and A2.1; A1, respectively; all the other fractures were considered unstable Using the comprehensive system, the stable patterns were A1, and all the other fractures were considered unstable Additionally, observers in both groups selected surgical treatment and implant type based on CT scans and 3DR of the fractures.

After four weeks, the original image folders were re-numbered and re-classified by the same method mentioned above. Only the first author of the study was informed and recorded the patient data, imaging data, and coding data in detail but did not participate in the review.

Statistical Analysis

SPSS 26.0 (IBM, USA) was used for data derivation and statistical analysis. Chi-square test and independent sample *t*-test were used to compare the demographic characteristics of included and excluded data. Intra- and inter-observer reliability was calculated by Cohen’s kappa statistic. Landis and Koch’s criteria are used to explaining the intensity of Kappa values (less than 0.00, 0.00–0.20, 0.21 –0.40, 0.41–0.60, 0.61 –0.80 and greater than 0.80 indicate poor, slight, fair, moderate, substantial, almost perfect agreement, respectively).¹⁵ In all classification systems, the Kappa value of Intra- and inter-observer reliability is expressed as $\bar{x} \pm s$. The overall comparison was analyzed by One-way analysis of variance analysis (ANOVA), and the multiple comparisons was analyzed by LSD-*t* test. Comparisons between the high-experience and low-experience groups were performed using an independent sample *t*-test.

Results

The study included 192 patients with a mean age of 77 years (range 67–98), comprising 129 females and 63 males, with 113 having left intertrochanteric fractures. A total of 3072 evaluations were conducted, and the results of the mean kappa of inter-observer reliability in CT scan and 3DR for Evans-Jensen, AO/OTA (2007), AO/OTA (2018), and Comprehensive Classification were presented in Table 2. The mean kappa of the intra-observer reliability of the classification systems was reported in Table 3.

Inter-Observer Agreement

When both CT scan and 3DR were utilized, it was revealed that there were significant differences ($P < 0.05$) in the mean Kappa values of the four classification types among all observers, within the high/low experience groups, and during the

Table 2 Comparison of Inter-Observer Reliability of Different Classifications When Using CT Scan and 3DR

	CT Scan				3DR			
	ALL	High-Experience	Low-Experience	P Value	ALL	High-Experience	Low-Experience	P Value
Evans-Jensen	0.361±0.06 ^{α,β}	0.377±0.05 ^{α,β}	0.357±0.05 ^α	0.592	0.411±0.04 ^{α,β}	0.423±0.04 ^{α,β}	0.401±0.05 ^{α,β}	0.466
AO/OTA(2007)	0.348±0.05 ^{α,β}	0.352±0.04 ^{α,β}	0.343±0.04 ^α	0.761	0.396±0.03 ^{α,β}	0.403±0.02 ^{α,β}	0.386±0.03 ^{α,β}	0.382
AO/OTA(2018)	0.408±0.06 ^α	0.436±0.02 ^α	0.379±0.04 ^α	0.044	0.442±0.03 ^α	0.464±0.01 ^{α,†}	0.405±0.04 ^{α,†}	0.029
Comprehensive	0.492±0.02	0.513±0.01	0.496±0.02	0.179	0.524±0.03	0.547±0.01	0.526±0.02	0.109

Notes: α: compared with Comprehensive classification $p < 0.05$. β: compared with AO/OTA(2018) classification $p < 0.05$.

Table 3 Comparison of Intra-Observer Reliability of Different Classifications When Using CT Scan and 3DR

Classification	CT Scan				3DR			
	ALL	High-Experience	Low-Experience	P Value	ALL	High-Experience	Low-Experience	P Value
Evans-Jensen	0.467±0.05 ^{α,β}	0.486±0.03 ^{α,β}	0.452±0.04 ^{α,β}	0.223	0.503±0.05 ^{α,β}	0.517±0.02 ^{α,β}	0.491±0.03 ^α	0.199
AO/OTA(2007)	0.453±0.06 ^{α,β}	0.463±0.04 ^{α,β}	0.447±0.05 ^{α,β}	0.635	0.482±0.03 ^{α,β}	0.512±0.03 ^{α,β}	0.465±0.04 ^{α,β}	0.109
AO/OTA(2018)	0.542±0.04 ^α	0.564±0.02 ^{α,†}	0.515±0.03 ^{α,†}	0.035	0.561±0.03 ^α	0.582±0.01 ^α	0.531±0.03 ^α	0.018
Comprehensive	0.631±0.02	0.637±0.01	0.627±0.02	0.406	0.672±0.02	0.682±0.01	0.675±0.02	0.554

Notes: α: compared with Comprehensive classification p < 0.05. β: compared with AO/OTA (2018) classification p<0.05.

stability assessment. Using CT scans, the Comprehensive Classification had the highest mean Kappa value of 0.492 ± 0.02 among all observers, showing moderate agreement. Among them, the high - experience group had a mean Kappa value of 0.513 ± 0.01, also at a moderate agreement level. In contrast, the mean Kappa values of the other three classifications, Evans - Jensen, AO/OTA (2007), and AO/OTA (2018), were 0.361, 0.348, and 0.408 respectively, within the range of fair agreement. Further comparison revealed that the mean Kappa values of AO/OTA (2018) and the Comprehensive Classification were significantly different from the other three classification types (P < 0.05). Using 3DR, the mean Kappa value of the Comprehensive Classification reached 0.524 ± 0.03, indicating substantial agreement; the mean Kappa value of AO/OTA (2018) was 0.442 ± 0.03, showing moderate agreement. When assessing the Comprehensive Classification and AO/OTA (2018), a statistically significant difference was observed between the high - experience group and the low - experience group (P < 0.05). Overall, the Comprehensive Classification performed excellently in terms of inter - observer consistency in 3DR technology. See Table 2 for details.

The Comprehensive Classification had the highest mean Kappa value in the inter - observer assessment of stability, reaching 0.695 ± 0.02 in CT scan, showing substantial agreement; and reaching 0.755 ± 0.01 in 3DR, approaching almost perfect agreement. See Table 4 for details.

Intra-Observer Agreement

Analogous to the inter-observer agreement scenario, when using both CT scan and 3DR, substantial variations in the mean Kappa values of the four classification types were detected among all observers, high/low experience groups, and during stability assessment (P < 0.05). The Comprehensive Classification had the highest mean Kappa values of 0.631 ± 0.02 in CT scan and 0.672 ± 0.02 in 3DR. In contrast, the mean Kappa values of the other three classifications, Evans - Jensen, AO/OTA (2007), and AO/OTA (2018), were 0.467±0.05, 0.453±0.06, and 0.542±0.04 in CT scan. In both CT scan and 3DR imaging techniques, the Comprehensive Classification exhibited showed substantial agreement among both the high - experience and low - experience groups. Meanwhile, the mean Kappa values of AO/OTA (2018) and the Comprehensive Classification were significantly different from the other three (P < 0.05). See Table 3 for details. In the intra - observer assessment of stability, the mean Kappa value of the Comprehensive Classification was as high as 0.826 ± 0.01 in 3DR, reaching almost perfect agreement, and as 0.772±0.01 in CT scan had the substantial agreement, See

Table 4 In Assessment of Stability, Comparison of Inter-Observer and Intra-Observer Reliability of Different Classifications When Using CT Scan and 3DR

	Inter-Observer		Intra-Observer	
	CT Scan	3DR	CT Scan	3DR
Evans-Jensen	0.448±0.03 ^{α,β}	0.527±0.04 ^{α,β}	0.524±0.02 ^{α,β}	0.573±0.03 ^{α,β}
AO/OTA(2007)	0.415±0.05 ^{α,β}	0.479±0.03 ^{α,β}	0.476±0.04 ^{α,β}	0.552±0.03 ^{α,β}
AO/OTA(2018)	0.512±0.04 ^α	0.593±0.04 ^α	0.593±0.05 ^α	0.675±0.02 ^α
Comprehensive	0.695±0.02	0.755±0.01	0.772±0.01	0.826±0.01

Notes: α: compared with Comprehensive classification p < 0.05. β: compared with AO/OTA(2018) classification p<0.05.

Table 5 The Degree of Agreement Between the Type of Implant and the Observer's Selection in the Comprehensive Classification

	CT Scan (%)	3DR (%)
ALL	62.42±14.1	66.61±13.2
High-experience	68.73±12.8	75.86±8.1
Low -experience	56.11±13.0	57.36 ±10.5

Table 4 for details. In CT scan, the agreement rates of different experience groups varied. The high - experience group had a relatively higher rate than the low - experience group. In 3DR, the high - experience group also demonstrated a notably higher agreement rate than the low - experience group.

In the Comprehensive Classification assessment, the internal fixation type selected by the observer was compared with the internal fixation type corresponding to the fracture classification assessed by the observer. The high - experience group based on 3DR had the highest degree of agreement, approximately 75.86%. See Table 5 for details.

Discussion

The pivotal discovery of our study is the outstanding reliability of the newly proposed Comprehensive Classification for intertrochanteric femur fractures, which is crucial for its adoption in broad clinical settings.^{14,16} Our evaluation, grounded in CT scans and 3DR, revealed that the Comprehensive Classification holds a significant edge over the three other classifications with respect to overall performance between high/low experience groups, and stability. In CT scans, the mean Kappa value for inter-observer agreement is highest for the Comprehensive Classification, indicating “moderate agreement”. Moreover, higher-experience groups demonstrate greater agreement compared to low-experience groups. Meanwhile, the mean Kappa value for intra-observer agreement remains highest, indicating “substantial agreement”, with no significant difference between high and low-experience groups. In 3DR, the new Comprehensive Classification demonstrates the best inter-observer agreement, with both high- and low-experience groups displaying “substantial” agreement. Furthermore, the mean Kappa values of AO/OTA (2018) and Comprehensive Classification were found to have statistical differences compared to other classifications. In studies of fracture stability, the Comprehensive Classification showed the highest mean Kappa values among both inter- and intra-observer consistency analyses. Additionally, we noted that in the intra-observer assessment, the mean Kappa value for the Comprehensive Classification reached almost perfect levels when using 3DR.

At present, there are numerous studies analyzing the reliability of traditional classification systems under X-ray. For example, Parker et al¹² conducted a statistical analysis of the reliability of the Evans-Jensen classification and AO/OTA (2007) classification. The results showed that the Evans-Jensen and complete grouping of the AO/OTA classification cannot be fully relied upon. Furthermore, the subgroups of AO/OTA were found to be confusing and incomplete, and it was acceptable to simplify the classification into three AO groups (31A1, 31A2, 31A3) for femoral shaft fractures. Dai et al¹³ also reached similar conclusions in their subsequent studies. Chan et al¹⁴ analyzed the revised AO/OTA (2018) standard and found it to be unreliable at both the group and subgroup levels. Additionally, it cannot provide consistent guidance for critical treatment decisions, such as determining stability or instability based on grouping. These studies all indicate that the reliability of these traditional classification systems under X-ray is not ideal and cannot be fully relied upon.

Fracture classification reliability is influenced by several factors, and understanding these is crucial for improving accuracy and consistency. Variations in fracture line number, orientation, irregularity, as well as differences in fragment size, shape, distribution, and degree of comminution, all contribute to the complexity and uncertainty of fracture classification. For instance, when fractures involve multiple irregular branching lines and severe comminution of bone fragments, there is an increased likelihood of inter-observer discrepancies in classification, which consequently reduces reliability. Observer experience also plays a significant role. Analysis revealed notable differences in classification accuracy among observers with varying experience levels. Experienced clinicians were better able to handle complex

fractures, achieving higher inter-observer reliability. To improve accuracy, ongoing training focused on fracture identification and classification could enhance the skills of less experienced clinicians, improving overall system reliability. Additionally, imaging quality is crucial. The radiological parameters of CT scans and 3D reconstructions directly affect fracture visibility. To eliminate variability, we used the same imaging equipment throughout the study, ensuring consistent imaging conditions and minimizing device-related discrepancies.

In this study, when using CT scans, the Evans-Jensen and AO/OTA (2007/2018) classifications showed moderate and fair reliability, respectively, which is an improvement from under X-ray but still not ideal. In comparison, the new Comprehensive Classification showed moderate and substantial agreement, respectively. We believe that the limitations of the traditional fracture classification are due to their reliance on conventional film systems, which are not optimal for describing fracture patterns. This could also explain why the use of 3DR did not result in a substantial improvement in the reliability of Evans-Jensen and AO/OTA classifications.

The latest revision of the AO/OTA classification includes a classification criterion for the thickness of the lateral wall, as many researchers have pointed out that the thickness and status of the lateral wall are crucial determinants of implant positioning and fracture stability.^{17,18} Lateral wall stability is a reliable indicator for patients to choose a dynamic hip screw (DHS), and lateral wall thickness is considered a predictive factor for secondary lateral wall fractures after DHS implantation.¹⁷ When choosing a more simple, skilled, and economical DHS has always been a controversial issue in clinical practice. From a comprehensive point of view, the stability of the fracture pattern is the key point.

A good classification system should demonstrate good inter-observer and intra-observer reliability, help translate clinical research into practice, and promote evidence-based interventions, especially in cases of ongoing debate.¹⁴ An excellent classification system should accurately describe fracture characteristics, reflect the degree of injury, guide treatment choices, help predict prognosis, and prevent complications.⁷ Currently, two types of internal fixation devices are commonly used in clinical practice, including the extramedullary device represented by DHS and the intramedullary device represented by Proximal Femoral Nail Anti-rotation (PFNA).¹⁹ Our criteria for selection are primarily the stability of the fracture pattern and the likelihood of a secondary fractures of the lateral cortex. The widely accepted stability criteria, namely the typical fracture line direction from the greater tuberosity (proximal and lateral) to the lesser tuberosity (distal and medial), is considered unstable if there is additional comminution extending to the medial femoral condyle, and the fracture line is referred to as a “reverse oblique fractures” when it runs transversely, which is also unstable.²⁰ In addition, when treating with intramedullary nailing, the distal extension of the lesser tuberosity fragment and the anterior extension of the posterior coronal fracture line to the entry point of the implant in the head-neck region are considered predictive factors for instability after short cephalomedullary nailing.^{21–23} Using the criterion of measuring the thickness of the lateral cortex below the trochanteric ridge, a thickness greater than 20.5 mm is associated with a significant decrease in secondary lateral cortex fractures after DHS implantation.¹⁷

The new Comprehensive Classification proposed by Chang et al⁷ builds upon the advantages of both the old and new versions of AO/OTA classification while retaining the principles of classification based on fracture lines. Within this research team, Zhang et al²⁴ produced fracture maps, heat maps, and 3D measurements of fractures of the lateral wall, and concluded that there is a distribution pattern of coronal fracture lines causing damage to the lateral wall. The coronal fracture lines in the lateral wall are concentrated in the posterior superior aspect, with more dense fracture lines in the posterior aspect. Furthermore, partial disruption of the lateral cortical bone extending to the coronal posterior aspect of the femur was also observed.²⁵ The new Comprehensive Classification summarized the above studies and proposed different treatment recommendations based on factors such as the fracture line, degree of fracture comminution, the nail entry channel in the lateral cortex, displacement of the lesser tuberosity, and support of the medial femoral condyle. In this study, we found that the Comprehensive Classification had better inter- and intra-observer reliability among surgeons with different levels of experience when using 3DR. Even in terms of stability, the intra-observer agreement reached the nearly perfect consistency. This may be due to the more clearly defined classification criteria, the omission of lateral wall thickness measurement, and the more comprehensive classification that reduced the impact of posterior coronal plane structural fractures.

Two other 3D-CT-based classification systems have also demonstrated significant promise in offering enhanced accuracy in fracture assessment, thereby assisting in the formulation of optimal treatment strategies, especially for elderly

patients with osteoporosis and complex fractures. One system proposed a comprehensive approach with three main categories and six subgroups, demonstrating high intra- and inter-observer reliability, and effectively identifying fractures missed on plain radiographs.²⁶ Another six-part classification, incorporating fracture stability analysis, offers enhanced prognostic value and better distinguishes stable from unstable fractures, thereby improving treatment decisions and patient outcomes.²⁷ Future research should focus on validating these systems in multicenter studies and assessing their integration with advanced surgical techniques to further improve fracture management.

We believe that a more explicit stability assessment can quickly formulate treatment plans that meet the requirements of Enhanced Recovery After Surgery (ERAS),²⁸ and the richer description of the lateral wall fracture line and medial support expands the indications for the combined use of auxiliary devices and intramedullary and extramedullary fixation. In our study, when selecting internal fixation based on the Comprehensive Classification, high-experienced surgeons achieved a consistency rate of approximately 76.32%, which stems from its effective clinical treatment guidelines. We believe that using the Comprehensive Classification can better utilize its value in guiding surgical strategies.

In addition, there are some limitations in this study. First, all included data are from patients with acute intertrochanteric fractures, without clinical and imaging follow-up. Therefore, the impact of fracture classification on patient prognosis cannot be thoroughly evaluated. Similarly, the accuracy of classification has not been verified by surgery and other means. Secondly, we did not analyze the subgroups and subgroups of different classification systems separately, which may greatly increase our workload. Furthermore, only the reliability of different classification systems was studied, and the time required for each observer to read and understand different classification systems when applying fracture classification was not measured, which lacked analysis of routine clinical practice. Another point is that we did not directly compare with the review under X-ray, and could not determine whether there was a significant improvement under CT scan and 3DR. The advantage of this study is that we included a large amount of patient data, and orthopedic doctors with different levels of experience from three tertiary hospitals were chosen as observers, which minimizes bias as much as possible. In addition, by evaluating the consistency of the stability of the Comprehensive Classification and analyzing the selection of internal fixation, the value of this new classification in the clinical application was fully evaluated.

Conclusion

In conclusion, our study demonstrates that the Comprehensive Classification is a reliable tool for classifying intertrochanteric fractures, with better inter- and intra-observer agreement compared to the other three traditional fracture classifications. Additionally, the sub-classification of the Comprehensive Classification is reasonable, compensating for missing and completing rare subtypes, and providing guidance for treatment. Therefore, we recommend the widespread use of Comprehensive Classification in clinical practice.

Data Sharing Statement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics Approval and Consent to Participate

This study was approved by the Ethics Committee of Beijing Chaoyang Hospital. Ethics Committee of Beijing Chaoyang Hospital waived the requirement for written informed consent because the study was retrospective, it did not have any adverse effect on patients' health, and it reported anonymized patient data. The authors announce that all methods were performed in accordance with the relevant guidelines and regulations. Strict confidentiality of patient data and adherence to the Declaration of Helsinki.

Acknowledgments

We appreciate all co-authors for their contributions to this study and the writing of this manuscript.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

The correspondence author Junlin Zhou and Xiaopei Xu discloses receipt of the following financial support for the research, authorship, and publication of this article: National Natural Science Foundation of China (82272469) and Beijing Key Clinical Specialty Project and Gold Seed Research Fund (CYJZ202423).

Disclosure

The authors declare that they have no competing interests.

References

- Kanis JA, Odén A, McCloskey EV, Johansson H, Wahl DA, Cooper C. A systematic review of Hip fracture incidence and probability of fracture worldwide. *Osteoporos Int*. 2012;23(9):2239–2256. doi:10.1007/s00198-012-1964-3
- Chang SM, Hou ZY, Hu SJ, Du SC. Intertrochanteric Femur Fracture Treatment in Asia: what We Know and What the World Can Learn. *Orthop Clin North Am*. 2020;51(2):189–205. doi:10.1016/j.ocl.2019.11.011
- Bhandari M, Swiontkowski M. Management of Acute Hip Fracture. *N Engl J Med*. 2017;377(21):2053–2062. doi:10.1056/NEJMcp1611090
- Jensen JS, Michaelsen M. Trochanteric femoral fractures treated with McLaughlin osteosynthesis. *Acta Orthop Scand*. 1975;46(5):795–803. doi:10.3109/17453677508989266
- Marsh JL, Slongo TF, Agel J, et al. Fracture and dislocation classification compendium - 2007: orthopaedic Trauma Association classification, database and outcomes committee. *J Orthop Trauma*. 2007;21(10 Suppl):S1–133. doi:10.1097/00005131-200711101-00001
- Meinberg EG, Agel J, Roberts CS, Karam MD, Kellam JF. Fracture and Dislocation Classification Compendium-2018. *J Orthop Trauma*. 2018;32(1):S1–S170. doi:10.1097/BOT.0000000000001063
- Chang SM, Wang ZH, Tian KW, Sun GX, Wang X, Rui YF. A sophisticated fracture classification system of the proximal femur trochanteric region (AO/OTA-31A) based on 3D-CT images. *Front Surg*. 2022;9:919225. doi:10.3389/fsurg.2022.919225
- Shoda E, Kitada S, Sasaki Y, et al. Proposal of new classification of femoral trochanteric fracture by three-dimensional computed tomography and relationship to usual plain X-ray classification. *J Orthop Surg (Hong Kong)*. 2017;25(1):2309499017692700. doi:10.1177/2309499017692700
- Kijima H, Yamada S, Konishi N, et al. The reliability of classifications of proximal femoral fractures with 3-dimensional computed tomography: the new concept of comprehensive classification. *Adv Orthop*. 2014;2014:359689. doi:10.1155/2014/359689
- Reid JJ, Kunkle BF, Greene AT, Eichinger JK, Friedman RJ. Variability and reliability of 2-dimensional vs. 3-dimensional glenoid version measurements with 3-dimensional preoperative planning software. *J Shoulder Elbow Surg*. 2022;31(2):302–309. doi:10.1016/j.jse.2021.07.011
- Tatro JM, Gilbertson JA, Schroder LK, Cole PA. Five to Ten-Year Outcomes of Operatively Treated Scapular Fractures. *J Bone Joint Surg Am*. 2018;100(10):871–878. doi:10.2106/JBJS.17.00673
- Pervez H, Parker MJ, Pryor GA, Lutchman L, Chirodian N. Classification of trochanteric fracture of the proximal femur: a study of the reliability of current systems. *Injury*. 2002;33(8):713–715. doi:10.1016/S0020-1383(02)00089-X
- 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*. 2005;36(7):858–861. doi:10.1016/j.injury.2005.02.005
- Chan G, Hughes K, Barakat A, et al. Inter- and intra-observer reliability of the new AO/OTA classification of proximal femur fractures. *Injury*. 2021;52(6):1434–1437. doi:10.1016/j.injury.2020.10.067
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174. doi:10.2307/2529310
- Klaber I, Besa P, Sandoval F, et al. The new AO classification system for intertrochanteric fractures allows better agreement than the original AO classification. An inter- and intra-observer agreement evaluation. *Injury*. 2021;52(1):102–105. doi:10.1016/j.injury.2020.07.020
- Hsu CE, Shih CM, Wang CC, Huang KC. Lateral femoral wall thickness. A reliable predictor of post-operative lateral wall fracture in intertrochanteric fractures. *Bone Joint J*. 2013;95B(8):1134–1138. doi:10.1302/0301-620X.95B8.31495
- Gao Z, Lv Y, Zhou F, et al. Risk factors for implant failure after fixation of proximal femoral fractures with fracture of the lateral femoral wall. *Injury*. 2018;49(2):315–322. doi:10.1016/j.injury.2017.11.011
- Kaplan K, Miyamoto R, Levine BR, Egol KA, Zuckerman JD. Surgical management of Hip fractures: an evidence-based review of the literature. II: intertrochanteric fractures. *J Am Acad Orthop Surg*. 2008;16(11):665–673. doi:10.5435/00124635-200811000-00007
- Ahn J, Bernstein J. Fractures in brief: intertrochanteric Hip fractures. *Clin Orthop Relat Res*. 2010;468(5):1450–1452. doi:10.1007/s11999-010-1263-2
- Rehme J, Woltmann A, Brand A, von Rüden C. Does auxiliary cerclage wiring provide intrinsic stability in cephalomedullary nailing of trochanteric and subtrochanteric fractures. *Int Orthop*. 2021;45(5):1329–1336. doi:10.1007/s00264-020-04795-4
- Schopper C, Faschingbauer M, Moeller RT, Gebhard F, Duerselen L, Seitz A. Modified Candy-Package technique vs Cerclage technique for refixation of the lesser trochanteric fragment in pertrochanteric femoral fractures. A biomechanical comparison of 10 specimens. *Injury*. 2020;51(8):1763–1768. doi:10.1016/j.injury.2020.06.008
- Lee WC, Chou SM, Tan CW, Chng LS, Yam G, Chua T. Intertrochanteric fracture with distal extension: when is the short proximal femoral nail antitraction too short. *Injury*. 2021;52(4):926–932. doi:10.1016/j.injury.2020.10.059

24. Zhang Y, Sun Y, Liao S, Chang S, Huang T. Three-Dimensional Mapping of Medial Wall in Unstable Pertrochanteric Fractures. *Biomed Res Int.* 2020;2020(1):8428407. doi:10.1155/2020/8428407
25. Ma Z, Yao XZ, Chang SM. The classification of intertrochanteric fractures based on the integrity of lateral femoral wall: letter to the editor, Fracture morphology of AO/OTA 31-A trochanteric fractures: a 3D CT study with an emphasis on coronal fragments. *Injury.* 2017;48(10):2367–2368. doi:10.1016/j.injury.2017.07.030
26. Kalia RB, Arora SS, Sarkar B, Paul S, Singh S. A comprehensive 3D CT based classification of intertrochanteric fracture. *J Clin Orthop Trauma.* 2022;30:101912. doi:10.1016/j.jcot.2022.101912
27. Wang Z, Li K, Peng C. Six-Part Classification of Femoral Intertrochanteric Fractures: a Classification Method to Improve the Diagnosis Rate of Unstable Fractures. *Orthop Surg.* 2024;16(3):637–653. doi:10.1111/os.13998
28. Yin M, Yan Y, Fan Z, et al. The efficacy of Enhanced Recovery after Surgery (ERAS) for elderly patients with intertrochanteric fractures who received surgery: study protocol for a randomized, blinded, controlled trial. *J Orthop Surg Res.* 2020;15(1):91. doi:10.1186/s13018-020-01586-w

International Journal of General Medicine

Publish your work in this journal

The International Journal of General Medicine is an international, peer-reviewed open-access journal that focuses on general and internal medicine, pathogenesis, epidemiology, diagnosis, monitoring and treatment protocols. The journal is characterized by the rapid reporting of reviews, original research and clinical studies across all disease areas. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/international-journal-of-general-medicine-journal>

Dovepress

Taylor & Francis Group