Evaluating the Reliability and Reproducibility of the AO and Lauge-Hansen Classification Systems for Ankle Injuries

MENG-CHEN YIN, MD; XUE-FEI YUAN, MD; JUN-MING MA, MD; YE XIA, MD; TAO WANG, MD; XIAO-LI XU, MD; YIN-JIE YAN, MD; JIN-HAI XU, MD; JIE YE, MD; ZHENG-YI TONG, MD; YAN-QI FENG, MD; HONG-BO WANG, MD; XUE-QUN WU, MD; WEN MO, MD

abstract

Ankle injuries are responsible for more than 5 million emergency department visits each year. The AO and Lauge-Hansen classification systems are widely used in the clinical diagnosis of ankle injuries. This study aimed to analyze the intraobserver reliability and interobserver reproducibility of the AO and Lauge-Hansen classification systems. In addition, the authors explored the differences among physicians’ classification responses and evaluated the clinical value for diagnosis. Fifty-six patients with an ankle injury with complete clinical and radiologic data were enrolled. The definition of injury type, the index score typing methods, and the specific study criteria were explained in detail. Five observers, who were orthopedic surgeons, determined the classifications according to both the AO and Lauge-Hansen systems. The classification was repeated 1 month later. Cronbach’s alpha and Cohen’s kappa test were used to determine interobserver reliability and intraobserver reproducibility. The physicians conducted 560 classifications (56 cases × 5 physicians × 2 times per patient). Average inter- and intraobserver kappa values for the AO system were 0.708 and 0.608, respectively. Average inter- and intraobserver kappa values for the Lauge-Hansen system were 0.402 and 0.398, respectively. Cronbach’s alpha coefficient was 96.7% for the AO system and 76.0% for the Lauge-Hansen system. The Lauge-Hansen classification system is a comprehensive yet cumbersome system. Comparatively, the AO classification system is easier to understand. This study shows that the AO classification system has more reliability and reproducibility, and thus has more value in clinical practice, than the Lauge-Hansen classification system. [Orthopedics. 2015; 38(7):e626-e630.]

The authors are from the Department of Orthopaedics (MY, JM, YX, TW, XX, YY, JX, YT, YF, HW, XW, WM), Longhua Hospital; and the Department of Gynaecology (XY), Yueyang Hospital of Integrated Traditional Chinese and Western Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, China.

Drs Yin, Yuan, and Ma contributed equally to this work and should be considered as equal first authors. The authors have no relevant financial relationships to disclose.

This study was supported by The Project of Shanghai City Health System Excellent Discipline Leader Training (grant XBR2013104) and the construction of the national 5 years’ major drug discovery projects—Chinese medicine (malignant tumor disease) clinical evaluation (grant 2011ZX09302-006-04).

Correspondence should be addressed to: Wen Mo, MD, Department of Orthopaedics, Longhua Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai 200032, China (mw2218@hotmail.com).

Received: June 30, 2014; Accepted: September 2, 2014.
doi: 10.3928/01477447-20150701-62
Ankle injuries are responsible for more than 5 million emergency department visits per year. Ankle sprains comprise 85% of these injuries, and ankle fractures account for the remaining 15%. Ankle fractures are intra-articular fractures, accounting for 3.9% of fractures in adults, and they are the most common type of lower limb fracture. Intra-articular fractures often occur during physical exercise, strenuous labor, and other activities. The most common causes of ankle fractures are falls (37.5%), followed by inversion injuries (31.5%) and sports-related injuries (10.2%).

The mechanism of an ankle fracture is complex, and its precise definition is often difficult. In addition to damaging the bone itself, the ligament can be damaged. Therefore, a better understanding of the mechanism of injury may help surgeons choose the proper treatment and may provide a basis for prognosis.

Currently, there are different views and understandings of the mechanisms in clinical practice. The AO and Lauge-Hansen classification systems are widely used in the clinical diagnosis of ankle injuries. The AO classification system is simple to understand, and it stresses the coordinating role of the fibula and syndesmosis of the ankle joint. The Lauge-Hansen classification system emphasizes the different stages of pathological damage in addition to the fracture pattern, and it stresses the understanding of damage to the ankle ligament system. It can provide a comprehensive assessment of ankle injuries.

This study aimed to analyze the intra-observer reliability and interobserver reproducibility of the AO and Lauge-Hansen classification systems. In addition, the authors explored the differences among physicians’ classification responses and evaluated the clinical value for diagnosis.

**MATERIALS AND METHODS**

**Clinical Data**

All patients in the authors’ orthopedic department who were surgically treated for ankle injuries between October 2008 and October 2013 were retrospectively analyzed. Patients older than 18 years with an ankle injury diagnosis who had complete clinical and radiographic case data were included in this study. Patients who underwent surgery were excluded if they had epiphyseal fractures, tumors, or tuberculosis. A total of 56 patients (32 men and 24 women) with a mean age of 63.3±4.3 years were enrolled. There were 25 left ankle fractures and 31 right ankle fractures.

A specialist collected and numbered the examination results and imaging data (W.M.). Imaging data and the patients’ medical history and symptoms were not marked with any patient information. The study was conducted in accordance with the principles of the Declaration of Helsinki, and the study protocol was approved by the ethics committee of Longhua Hospital, Shanghai University of Traditional Chinese Medicine. Written informed consent for surgery was obtained from all patients. Because of the retrospective nature of the study, patient consent for inclusion was waived.

**Research Methods**

Five orthopedic surgeons (M.Y., Y.X., T.W., Y.Y., X.X.), with clinical experience and training in classifying ankle injuries using the Lauge-Hansen and AO classification systems, were recruited for study participation, and they were issued original patient documentation. The definition of injury type, index score typing methods, and specific study criteria were explained in detail.

After training was complete, the authors selected 6 ankle fractures in patients with complete radiographical data and medical history and tested the physicians’ mastery of these 2 classification systems. After passing the test, the 5 physicians classified the ankle injuries of 56 patients who had a history of symptoms and radiographical data, and the original literature on the Lauge-Hansen and AO classification systems was used as a reference. One month later, the radiographical data were classified and scored again. The 2 sets of classification responses were recorded by a physician who was not involved in the study, and then the reliability and reproducibility of both classification systems were analyzed. If there was a parting difference among the 5 physicians, the authors identified it as inconsistent and investigated the reasons for the inconsistency.

**Statistical Methods**

Reliability and reproducibility were quantified by using the Cronbach’s alpha coefficient, and the Cohen’s kappa value was calculated using SPSS for Windows version 20.0 (IBM, Armonk, New York). The authors analyzed the results of the 5 physicians and the Cronbach’s alpha coefficient and the Cohen’s kappa value. The inter- and intraobserver results were evaluated for the reliability and reproducibility of the AO and Lauge-Hansen classification systems.

Based on the agreement of 2 reviewers (J.M., X.W.), kappa coefficient values ranging from -1 to 1 were used; the better the kappa value, the greater the consistency, and the smaller the kappa value, the worse the consistency. A-1 indicated complete disagreement, 0 was random agreement, and 1 was complete agreement. The standard by Landis and Koch was used, which is a widely used criteria for interpreting the agreement of a statistic score; values of 0.81 to 1 represented excellent consistency, 0.61 to 0.8 was substantial agreement, 0.41 to 0.6 was moderate agreement, 0.21 to 0.4 was fair agreement, and 0 to 0.2 was slight agreement.

**RESULTS**

The physicians conducted 560 classifications (56 cases × 5 physicians × 2 per patient). Reliability analysis of the AO classification system for the 5 physicians resulted in kappa values of 0.763, 0.675, 0.751, 0.804, and 0.620, respectively, for the first round of classification, and 0.720, 0.838, 0.553, 0.597, and 0.758, respective-
ly, for the second round of classification (1 month later). The average kappa value was 0.708, and the Cronbach’s alpha coefficient was 96.7%. The results indicated that the reliability of the AO classification system was substantially consistent. Reliability analysis of the Lauge-Hansen classification system for the 5 physicians resulted in kappa values of 0.282, 0.413, 0.516, 0.437, and 0.233, respectively, for the first round of classification, and 0.404, 0.329, 0.426, 0.436, and 0.545, respectively, for the second round of classification (1 month later). Average kappa value was 0.402, and Cronbach’s alpha coefficient was 76.0%. The results indicated that the reliability of the Lauge-Hansen classification system was substantially consistent (Table 1).

Reproducibility analysis of the AO classification system for the 5 physicians resulted in kappa values of 0.753, 0.624, 0.724, 0.751, and 0.687, respectively. Average kappa value was 0.698, which indicated that the reproducibility of the AO classification system was substantial. Reproducibility analysis of the Lauge-Hansen classification system for the 5 physicians resulted in kappa values of 0.535, 0.436, 0.242, 0.339, and 0.438, respectively. Average kappa value was 0.398, indicating that the reproducibility of the Lauge-Hansen classification system was fair (Table 2).

### Table 1

<table>
<thead>
<tr>
<th>Physician</th>
<th>Kappa Value of AO System</th>
<th>Kappa Value of L-H System</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-B</td>
<td>0.763</td>
<td>0.282</td>
</tr>
<tr>
<td>A-C</td>
<td>0.675</td>
<td>0.413</td>
</tr>
<tr>
<td>A-D</td>
<td>0.751</td>
<td>0.516</td>
</tr>
<tr>
<td>A-E</td>
<td>0.804</td>
<td>0.437</td>
</tr>
<tr>
<td>B-C</td>
<td>0.620</td>
<td>0.233</td>
</tr>
<tr>
<td>B-D</td>
<td>0.720</td>
<td>0.404</td>
</tr>
<tr>
<td>B-E</td>
<td>0.838</td>
<td>0.329</td>
</tr>
<tr>
<td>C-D</td>
<td>0.553</td>
<td>0.426</td>
</tr>
<tr>
<td>C-E</td>
<td>0.597</td>
<td>0.436</td>
</tr>
<tr>
<td>D-E</td>
<td>0.758</td>
<td>0.545</td>
</tr>
<tr>
<td>Average</td>
<td>0.708</td>
<td>0.402</td>
</tr>
</tbody>
</table>


### Table 2

<table>
<thead>
<tr>
<th>Physician</th>
<th>Kappa Value of AO System</th>
<th>Kappa Value of L-H System</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.753</td>
<td>0.535</td>
</tr>
<tr>
<td>B</td>
<td>0.624</td>
<td>0.436</td>
</tr>
<tr>
<td>C</td>
<td>0.724</td>
<td>0.242</td>
</tr>
<tr>
<td>D</td>
<td>0.751</td>
<td>0.339</td>
</tr>
<tr>
<td>E</td>
<td>0.637</td>
<td>0.438</td>
</tr>
<tr>
<td>Average</td>
<td>0.698</td>
<td>0.398</td>
</tr>
</tbody>
</table>


### Discussion

An ideal classification system should be comprehensive, simple, highly reliable, and highly reproducible, and it should provide clinical treatment and prognosis guidance. In addition, it should be a necessary part of reporting on clinical series for peer review and subsequent widespread dissemination. The Lauge-Hansen and AO classification systems are the most commonly used classification systems for ankle injuries in clinical practice and, to the authors’ knowledge, this study is the first to evaluate their reliability and reproducibility.

In 1942, Dane Lauge-Hansen first developed a classification system that was based on autopsy examinations combined with clinical observations and radiographic images. This system was published in several articles, and it had a significant influence on fracture research. Because this classification system based the observed injury on a presumed mechanism of injury, it was a historical advancement. Modern textbooks still regard it as an important tool for the treatment of ankle fractures. Lauge-Hansen performed sufficient research on the trauma pathology and pathogenesis for ankle injuries. The classification system was based on foot position and the direction of the external injuring force. Lauge-Hansen’s classic study is the current basis for understanding the mechanisms that cause ankle fractures. Supination was composed of the components of inward rotation (adduction and inversion), and pronation was composed of outward rotation (abduction and eversion). As a result, the following 4 classifications were developed: supination-adduction, supination-eversion, pronation-abduction, and pronation-eversion.

In 1972, Weber developed the AO classification system, which was a purely radiographical classification system based on the Danis classification. It was designed to facilitate communication between providers and to assist with providing consistent documentation for clinical and research purposes. According to the position of the fibula fracture line and tibiofibular syndesmosis, ankle fractures were categorized into easy to understand types: infrasyndesmotic, with or without a medial lesion (type A); transsyndesmotic, with or without a medial or posterior lesion (type B); and suprasyndesmotic, with or without a medial or posterior lesion (type C). The AO Group maintains the view that the integrity of the ankle mortise is dependent on the following: a normal length of the fibula and precise location of the tibia fibula notch; and the integrity of 3 parts: the lower front tibiofibular liga-

---

Table 1: Reliability of the AO and Lauge-Hansen Classification System

<table>
<thead>
<tr>
<th>Physician</th>
<th>Kappa Value of AO System</th>
<th>Kappa Value of L-H System</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-B</td>
<td>0.763</td>
<td>0.282</td>
</tr>
<tr>
<td>A-C</td>
<td>0.675</td>
<td>0.413</td>
</tr>
<tr>
<td>A-D</td>
<td>0.751</td>
<td>0.516</td>
</tr>
<tr>
<td>A-E</td>
<td>0.804</td>
<td>0.437</td>
</tr>
<tr>
<td>B-C</td>
<td>0.620</td>
<td>0.233</td>
</tr>
<tr>
<td>B-D</td>
<td>0.720</td>
<td>0.404</td>
</tr>
<tr>
<td>B-E</td>
<td>0.838</td>
<td>0.329</td>
</tr>
<tr>
<td>C-D</td>
<td>0.553</td>
<td>0.426</td>
</tr>
<tr>
<td>C-E</td>
<td>0.597</td>
<td>0.436</td>
</tr>
<tr>
<td>D-E</td>
<td>0.758</td>
<td>0.545</td>
</tr>
<tr>
<td>Average</td>
<td>0.708</td>
<td>0.402</td>
</tr>
</tbody>
</table>


Table 2: Reproducibility of the AO and Lauge-Hansen Classification System

<table>
<thead>
<tr>
<th>Physician</th>
<th>Kappa Value of AO System</th>
<th>Kappa Value of L-H System</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.753</td>
<td>0.535</td>
</tr>
<tr>
<td>B</td>
<td>0.624</td>
<td>0.436</td>
</tr>
<tr>
<td>C</td>
<td>0.724</td>
<td>0.242</td>
</tr>
<tr>
<td>D</td>
<td>0.751</td>
<td>0.339</td>
</tr>
<tr>
<td>E</td>
<td>0.637</td>
<td>0.438</td>
</tr>
<tr>
<td>Average</td>
<td>0.698</td>
<td>0.398</td>
</tr>
</tbody>
</table>

membrane.\textsuperscript{9} A high level of the proposed fibula fracture and the inferior tibiofibular means that the ligament injury is more serious and the risk of ankle mortise failure is greater. This classification is used for guidance in ankle fracture surgery to provide more attention to the relationship and stability between the fibula fracture line and tibiofibular syndesmosis.

In the clinical setting, there are different understandings of these classification systems and the mechanisms of injury among many physicians. Therefore, the purposes of the current study were to analyze the reliability and reproducibility of the 2 classification systems, to explore the reasons for differences in classification responses from physicians, and to evaluate the clinical value for diagnosis.

The Lauge-Hansen classification system puts more emphasis on the pathological process, so it is more complex and more difficult to understand than the AO classification system. Some scholars have performed intra- and interobservations. Thomsen et al\textsuperscript{20} reported kappa values of 0.55 for interobserver variation and 0.65 for intraobserver variation. Nielsen et al\textsuperscript{21} demonstrated interobserver variation of 68\%, whereas the intraobserver variation ranged between 81\% and 95\% for the fractures classified. The figures in their study show that the reliability and reproducibility are moderate. However, subsequent studies\textsuperscript{22} have demonstrated poor reliability and challenged the validity of the work by Nielsen et al. The results of the current study demonstrated kappa values of 0.402 for interobserver variation and 0.398 for intraobserver variation, which are lower than the kappa values for the AO classification system. All of the studies suggest that the reliability of the Lauge-Hansen classification system is acceptable. However, observer variation in the current study was only for the class of injury, and the authors did not explore the reliability of the classification system for the stage of injury.

In the course of the study, the current authors found that the Lauge-Hansen classification system was inadequate and may not be entirely reliable or reproducible. It is necessary to obtain more information on the circumstances of how the patient was injured, such as the position of his or her foot and the direction of the external force. It is also worth noting that the pronation-abduction class showed particularly poor radiographical definition. Atypical fracture lines make classification challenging. If a fracture is medial malleolus and lateral malleolus in nature, is it considered a supination or pronation injury? Furthermore, the Lauge-Hansen classification system puts more emphasis on the lateral ligament injury, but this is difficult to detect on a radiograph. For example, in the first stage of pronation-abduction and pronation-eversion classification types, damage is often seen first to the medial deltoid ligament, but it cannot be diagnosed unless magnetic resonance imaging is obtained or pain and swelling are reported during examination.

Furthermore, this study found some commonalities between the 2 classification systems. According to the AO classification system, type A fractures of the lateral malleolus are usually stable and seldom displaced. These fractures correspond with supination-adduction injuries according to the Lauge-Hansen classification system. Type B distal fibular fracture lines are usually oblique or spiral oblique in nature. These injuries often present with shortening of the distal fibular fracture fragment or lateral displacement. According to the Lauge-Hansen classification system, type B fractures are considered supination-eversion and pronation-abduction injuries. The mechanism of type C fractures can cause high supination-eversion, pronation-abduction, or Maisonneuve fractures along with isolated posterior malleolar fractures. The type of injury may make physicians suspicious of high fibular fracture. However, the current authors found that the Lauge-Hansen classification system cannot categorize a simple fracture of the medial malleolus into the AO classification system, such as the first stage of the pronation-abduction type and the pronation-eversion type of the Lauge-Hansen classification system.

The AO classification system is purely radiographical, and it is based on the relationship between the position of the fibula fracture line and tibiofibular syndesmosis. It is simpler and easier for physicians to understand. The current study also confirms that the AO classification system has a high consistency and is much higher in reliability (AO: 0.708 vs Lauge-Hansen: 0.402) and reproducibility (AO: 0.608 vs Lauge-Hansen: 0.398) than is the Lauge-Hansen classification system.

The Lauge-Hansen and AO classification systems for the diagnosis of ankle injuries have been widely used in clinical applications, and their clinical value has been recognized in the literature. This study focused on these 2 classification systems and confirmed that their reliability and reproducibility are acceptable.

However, the current study has limitations. First, the authors only included 56 patients, which is a relatively small sample size, and some of the patients lacked complete clinical data, which may cause bias on the mechanism of injury. In addition, the seniority and experience of physicians who participated in the study was an important factor affecting the accuracy of classification. The seniority of the 5 physicians was relatively low, and although they were orthopedic surgeons, they did not specialize in ankles. These 2 points may have caused inconsistency in the understanding of the classification systems, which may have caused a deviation in the results.

Therefore, in future clinical work, high-quality, large sample, and multicenter studies should be conducted to provide orthopedic surgeons with the best evidence-based information. To improve the trial design quality and the level of performance, future trials should follow the CONSORT statement as a guideline for reporting clinical trials.\textsuperscript{23,24}
CONCLUSION

The Lauge-Hansen classification system is a comprehensive yet cumbersome system. Comparatively, the AO classification system is easier to understand. This study shows that the AO classification system has more reliability and reproducibility, and thus has more value in clinical practice. Despite these challenges and shortcomings, the Lauge-Hansen classification system stands as the basis for understanding the mechanics of ankle fractures. Therefore, its value should not be neglected because it elucidates the importance of recognizing ankle ligament injuries that occur with fractures seen on radiographs. Ankle ligament injuries must be carefully considered when treating ankle fractures.

REFERENCES