Inter-rater reliability of Böhler and Gissane angles in different calcaneal fracture according to the Essex-Lopresti and Sanders classifications

Rui dos Santos Barroco1, Bruno Rodrigues de Miranda1,2, Herbert Amantéa Fernandes1, Gregory Bittar Pessoa1, Danilo Ryuko Cândido Nishikawa1, Leticia Zaccaria Prates de Oliveira1, Álvaro Diego Pupa de Freitas1, Caio Ivo de Almeida1

1. ABC School of Medicine, Santo André, SP, Brazil.
2. Hospital do Servidor Público Municipal de São Paulo, São Paulo, SP, Brazil.

Abstract

Objective: To evaluate the inter-rater reliability and intra-class correlation coefficients (ICC) of Böhler’s angle and the critical angle of Gissane in calcaneal fractures, stratified by severity and by the Essex-Lopresti and Sanders classifications.

Methods: Retrospective study of radiographs obtained from 97 patients: 67 with calcaneal fractures and 30 with normal lateral radiographs (used as a control group). Böhler’s angle and the angle of Gissane were measured by six raters: two orthopedic surgery residents, two musculoskeletal radiologists, a foot and ankle surgery fellow, and a senior consultant in foot surgery. Statistical analysis of inter-rater reliability was performed for the two angles, in the sample overall and stratified by the different radiographic and CT subtypes of calcaneal fractures.

Results: For the angle of Gissane, the ICC was at best 0.400 (95% CI: 0.250 to 0.581) for normal radiographs, with poor agreement across all classifications and severity stratifications. For Böhler’s angle, the ICC values indicated weak to moderate agreement, with the best reproducibility obtained for the overall sample (0.740; 95% CI: 0.673 to 0.801). In Sanders type 1 fractures, the ICC was 0.704 (95% CI: 0.397 to 0.940), and in Sanders type 2 fractures, 0.762 (95% CI: 0.634 to 0.870).

Conclusion: Böhler’s angle is more reproducible than the critical angle of Gissane, with greater inter-rater reliability among fractures deemed less severe on the Sanders classification, although the overall ICC ranged from weak to moderate at best.

Level of Evidence III; Case Control Study; Diagnostic Studies.

Keywords: Calcaneus; Intra-articular fractures; Radiology; Reproducibility of results.

Introduction

The diagnosis and prognosis of calcaneal fractures are related to their location and to the extent of joint involvement(1). Some classification systems for these fractures use radiographs to determine their patterns, but this approach is limited by the difficulty in clearly defining the fragments(2).

For the radiographic evaluation of calcaneal fractures, Böhler’s angle (BA) is frequently used to guide treatment, and is mentioned in several studies as a determinant of prognosis(3). The crucial (or critical) angle of Gissane (AG) is also classically cited as a useful measure in the diagnosis of calcaneal fractures, but its reproducibility is lower compared to BA(4).

With the advent of computed tomography (CT), the determination of fracture patterns has evolved, and new classification systems have emerged. Among these, the Sanders classification is most popular; it uses an oblique coronal CT view to classify the fracture according to the degree of joint involvement of the posterior subtalar facet. Prognosis can also be
estimated from the amount of fragments and the location of the fracture lines\(^{(5)}\). However, CT has higher cost and limited availability compared to plain radiographs. It may not be available at all emergency departments, which makes radiography still essential in establishing diagnosis and identifying factors implicated in fracture severity\(^{(6)}\). However, questions remain regarding the inter-rater reliability of these measurements\(^{(4)}\).

Within this context, the present study aimed to evaluate inter-rater reliability in measurement of BA and AG in calcaneal fractures with different severity levels and typologies on the Essex-Lopresti and Sanders classifications. Our hypothesis is that more severe fractures would hinder measurement of angular radiographic parameters, thus limiting their reliability.

**Methods**

This study was approved by the relevant institutional Research Ethics Committee and registered on Plataforma Brasil (CAAE: 97546918.5.0000.0082). A retrospective investigation was conducted on radiographic images obtained from 97 patients seen at a tertiary orthopedic surgery service. Of these, 67 radiographs represented calcaneal fractures and 30 were normal; the latter were used as a control group (or comparators) for measurement of inter-rater reliability. The inclusion criteria were research patients with calcaneal fractures who were treated at our service between the years 2004 and 2018 and for whom plain radiographs and CT scans were available on the medical record. The exclusion criteria were patients who refused to participate in the study, who did not have the necessary imaging tests on file, or whose images were of insufficient quality for evaluation.

At the study service, lateral radiographs of the ankle are obtained using a standardized technique, with the patient in lateral recumbent position on the imaged side with the lower limb fully extended and in contact with the table. The beam was directed perpendicular to the film, centered on the calcaneus (approximately 2.5cm distal to the medial malleolus), with a source-to-image distance of 100cm.

All fractures were analyzed by an independent senior consultant in foot and ankle surgery and divided into groups, who classified them according to the Essex-Lopresti method; 24 fractures were of the tongue type and 43 of the joint depression type. The Sanders CT classification was also applied by the same senior consultant, who identified six type 1 fractures, 14 type 2 fractures, 23 type 3 fractures, and 14 type 4 fractures.

Six examiners took part in measurements of inter-rater reliability: two orthopedics residents, one first-year and one third-year; two radiologists with a special interest in musculoskeletal imaging; a foot and ankle surgery fellow; and a second senior consultant in foot and ankle surgery. All received printouts of the radiographs, previously organized by the independent consultant (who did not participate in the measurements) and sequentially numbered at random, without nominal identification of the patients, so that evaluation was blinded in relation to patient identity and among assessors. AG and BA measurements were performed using a goniometer. None of the examiners involved in performing the angular measurements had prior access to the radiographs or to the results of the other examiners.

AG is measured at the intersection of two straight lines, one drawn along the posterior facet and the other tangent to the anterior beak. Its normal range is from 105° to 135°\(^{(7)}\). BA is formed by a line drawn from the highest point of the anterior tuberosity to the highest point of the posterior facet and another line drawn from the highest point of the posterior facet tangent to the superior extremity of the calcaneal tuberosity. Its normal range is from 20° to 40°\(^{(8)}\).

Data were collected using a standardized form and entered into an Excel spreadsheet. This was then exported into SPSS Version 10.1 (SPSS Inc., Chicago, IL) for statistical analysis.

A significance level of 0.05 was adopted. Intraclass correlation coefficients (ICC) were calculated for the assessment of agreement of the AG and BA measurements obtained by between the six observers, both in the sample overall and within separate subgroups defined by the Essex-Lopresti and Sanders tomographic classifications. Means were calculated with 95% confidence intervals. Levels of agreement were interpreted as follows: 0 to 0.2, weak; 0.21 to 0.4, fair; 0.41 to 0.6, moderate; 0.61 to 0.8, strong; and 0.81 to 1, almost perfect agreement.

**Results**

The sample consisted of 40 (41.2%) women and 57 (58.8%) men. The mean age was 44.9 years (range, 17 to 73 years) (Table 1). Of the radiographic images obtained, 52 (53.6%) were of the right foot and 45 (46.4%), of the left foot. Regarding the Sanders CT classification, there were six (9.0%) type 1 fractures, 24 (35.8%) type 2 fractures, 23 (34.3%) type 3 fractures, and 14 (20.9%) type 4 fractures.

According to the Essex-Lopresti radiographic classification, 43 (44.3%) fractures were of the joint depression type and 24 (24.7%) of the tongue type; 30 radiographs (30.9%) were normal (Table 1). Observed ICC values for the AG were, at best, 0.400 (95% CI: 0.250 to 0.581) for normal radiographs, indicating poor inter-rater agreement in this stratum of radiographs. ICC values were even lower in the fracture group, across all fracture classifications and severity stratifications (Table 2). For BA, the ICC values indicated of weak to moderate agreement between raters, with the best reproducibility obtained for the overall sample (0.740; 95% CI: 0.673 to 0.801), as well as in Sanders type 1 fractures (0.704; 95% CI: 0.397 to 0.940), and in Sanders type 2 fractures, 0.762 (95% CI: 0.634 to 0.870) (Table 2).

Separate assessment of only those scans classified as Sanders type 1 showed no evidence of difference between raters in measurement of the angle of Gissane (p=0.101), but there was a significant difference in measurements of Böhler’s angle (p=0.014). Among those fractures classified as Sanders type 2, there was evidence of significant differences between...
In the most severe fracture patterns according to the Sanders classification, the measurements again show major differences between raters. Assessment of only those scans classified as Sanders type 3 showed evidence of significant differences between raters in measurements of AG (p<0.001), but no significant difference in measurements of BA (p=0.464). Finally, analysis of only those fractures classified as Sanders type 4 again showed evidence of significant differences between raters for both measures, AG and BA (p<0.001), demonstrating low reproducibility.

When analyzing the sample stratified by Essex-Lopresti classification, there were differences in reliability and reproducibility for both angles (BA and AG) across different scenarios.

In the sample with joint depression-type calcaneal fracture according to the Essex-Lopresti classification (Table 3, Figures 1 and 2), there was evidence of significant differences between observers in measurements of AG (p<0.001) and BA (p<0.001).

Assessment of only those radiographs classified as showing a tongue-type calcaneal fracture on the Essex-Lopresti scheme (Table 4, Figures 3 and 4) showed evidence of significant differences between raters in measurements of AG (p<0.001), but no significant difference in measurements of BA (p=0.081).

Even when analyzing the sample of normal radiographs (i.e., those showing no evidence of fractures), there were significant differences between raters in measurements of both angles (AG, p<0.001; BA, p=0.004; Figures 5 and 6), demonstrating, once again, the limitation of angular measurements for the reproducible assessment of the shape of the calcaneus.

### Discussion

The crucial angle of Gissane and Böhler’s angle are classically used in the diagnosis of calcaneal fractures. Böhler described his eponymous angle in 1931 as a diagnostic tool, establishing normal values between 30° and 35°. More recent studies have demonstrated a wider range of variation, with between 20 and 45° being considered normal. Despite its widespread use in the literature, the original description of the technique for measuring the angle is confusing. In fact, the original article does not specifically describe obtaining...
Table 3. Estimated means (and 95% CI) for measurements of the critical angle of Gissane and Böhler’s angle obtained by different raters in a sample of joint depression-type calcaneal fractures (Essex-Lopresti classification) (n=43)

<table>
<thead>
<tr>
<th>Observer</th>
<th>Angle</th>
<th>Gissane</th>
<th>Böhler</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resident 1</td>
<td></td>
<td>111.4 (106.5; 116.3)</td>
<td>11.9 (6.3; 17.5)</td>
</tr>
<tr>
<td>Resident 2</td>
<td></td>
<td>126.0 (122.4; 129.7)</td>
<td>12.1 (7.1; 17.2)</td>
</tr>
<tr>
<td>Foot and ankle fellow</td>
<td></td>
<td>110.0 (104.2; 115.8)</td>
<td>9.3 (4.9; 13.6)</td>
</tr>
<tr>
<td>Radiologist 1</td>
<td></td>
<td>97.5 (91.4; 103.5)</td>
<td>11.4 (5.7; 17.0)</td>
</tr>
<tr>
<td>Radiologist 2</td>
<td></td>
<td>107.7 (101.1; 114.3)</td>
<td>6.6 (1.7; 11.6)</td>
</tr>
<tr>
<td>Senior consultant</td>
<td></td>
<td>95.1 (87.3; 102.9)</td>
<td>14.5 (9.7; 19.3)</td>
</tr>
</tbody>
</table>

p-value

<table>
<thead>
<tr>
<th>Multiple comparisons</th>
<th>p-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resident 1 vs. Resident 2</td>
<td>&lt;0.001</td>
<td>&gt;0.999</td>
</tr>
<tr>
<td>Resident 1 vs. Foot and ankle fellow</td>
<td>&gt;0.999</td>
<td>&gt;0.999</td>
</tr>
<tr>
<td>Resident 1 vs. Radiologist 1</td>
<td>0.002</td>
<td>&gt;0.999</td>
</tr>
<tr>
<td>Resident 1 vs. Radiologist 2</td>
<td>&gt;0.999</td>
<td>0.298</td>
</tr>
<tr>
<td>Resident 1 x Senior consultant</td>
<td>0.002</td>
<td>&gt;0.999</td>
</tr>
<tr>
<td>Resident 2 vs. Foot and ankle fellow</td>
<td>&lt;0.001</td>
<td>0.597</td>
</tr>
<tr>
<td>Resident 2 vs. Radiologist 1</td>
<td>&lt;0.001</td>
<td>&gt;0.999</td>
</tr>
<tr>
<td>Resident 2 vs. Radiologist 2</td>
<td>&lt;0.001</td>
<td>0.066</td>
</tr>
<tr>
<td>Resident 2 vs. Senior consultant</td>
<td>&lt;0.001</td>
<td>&gt;0.999</td>
</tr>
<tr>
<td>Foot and ankle fellow vs. Radiologist 1</td>
<td>0.013</td>
<td>&gt;0.999</td>
</tr>
<tr>
<td>Foot and ankle fellow vs. Radiologist 2</td>
<td>&gt;0.999</td>
<td>0.666</td>
</tr>
<tr>
<td>Foot and ankle fellow vs. Senior consultant</td>
<td>0.013</td>
<td>0.001</td>
</tr>
<tr>
<td>Radiologist 1 vs. Radiologist 2</td>
<td>0.093</td>
<td>0.500</td>
</tr>
<tr>
<td>Radiologist 1 x Senior consultant</td>
<td>&gt;0.999</td>
<td>&gt;0.999</td>
</tr>
<tr>
<td>Radiologist 2 x Senior consultant</td>
<td>0.071</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Data expressed as estimated means with 95% confidence intervals; p-values corrected by the sequential Bonferroni procedure.

Figure 1. Estimated means (and 95% CI) for measurements of the critical angle of Gissane obtained by different raters in a sample of joint depression-type calcaneal fractures (Essex-Lopresti classification) (n=43).

Figure 2. Estimated means (and 95% CI) for measurements of Böhler’s angle obtained by different raters in a sample of joint depression-type calcaneal fractures (Essex-Lopresti classification) (n=43).
Table 4. Estimated means (and 95%CI) for measurements of the critical angle of Gissane and Böhler’s angle obtained by different raters in a sample of tongue-type calcaneal fractures (Essex-Lopresti classification) (n=24)

<table>
<thead>
<tr>
<th>Observer</th>
<th>Gissane</th>
<th>Böhler</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resident 1</td>
<td>116.8 (112.1; 121.4)</td>
<td>14.9 (11.1; 18.8)</td>
</tr>
<tr>
<td>Resident 2</td>
<td>123.4 (117.6; 129.2)</td>
<td>13.5 (10.1; 16.9)</td>
</tr>
<tr>
<td>Foot and ankle fellow</td>
<td>112.0 (107.3; 116.7)</td>
<td>9.3 (4.6; 14.0)</td>
</tr>
<tr>
<td>Radiologist 1</td>
<td>110.8 (105.9; 115.7)</td>
<td>11.1 (6.4; 15.8)</td>
</tr>
<tr>
<td>Radiologist 2</td>
<td>109.7 (105.8; 113.5)</td>
<td>8.0 (2.8; 13.2)</td>
</tr>
<tr>
<td>Senior consultant</td>
<td>108.2 (102.7; 113.8)</td>
<td>13.2 (9.6; 16.8)</td>
</tr>
</tbody>
</table>

p-value

Multiple comparisons

- Resident 1 vs. Resident 2: 0.254 >0.999
- Resident 1 vs. Foot and ankle fellow: 0.456 0.431
- Resident 1 vs. Radiologist 1: 0.230 >0.999
- Resident 1 vs. Radiologist 2: 0.013 0.214
- Resident 1 x Senior consultant: 0.043 >0.999
- Resident 2 vs. Foot and ankle fellow: 0.004 0.895
- Resident 2 vs. Radiologist 1: 0.002 >0.999
- Resident 2 vs. Radiologist 2: <0.001 0.485
- Resident 2 vs. Senior consultant: <0.001 >0.999
- Foot and ankle fellow vs. Radiologist 1: >0.999 >0.999
- Foot and ankle fellow vs. Radiologist 2: >0.999 >0.999
- Foot and ankle fellow vs. Senior consultant: >0.999 >0.999
- Radiologist 1 vs. Radiologist 2: >0.999 >0.999
- Radiologist 1 x Senior consultant: >0.999 >0.999
- Radiologist 2 x Senior consultant: >0.999 0.673

Data expressed as estimated means with 95% confidence intervals; p-values corrected by the sequential Bonferroni procedure.

Figure 3. Estimated means (and 95%CI) for measurements of the critical angle of Gissane obtained by different raters in a sample of tongue-type calcaneal fractures (Essex-Lopresti classification) (n=24).

Figure 4. Estimated means (and 95%CI) for measurements of Böhler’s angle obtained by different raters in a sample of tongue-type calcaneal fractures (Essex-Lopresti classification) (n=24).
Barroco et al. Inter-rater reliability of Böhler and Gissane angles in different calcaneal fracture according to the Essex-Lopresti and Sanders classifications

Several studies have questioned the use of both angles due to low inter-rater agreement and poor reproducibility. In current clinical practice, the angle of Gissane is rarely used in the diagnosis and management of patients with calcaneal joint fractures, but Böhler’s angle is still considered a useful tool in determining fracture severity, and has been cited by some authors as a parameter of restoration of normal anatomy after osteosynthesis. Su et al. reported preoperative use of Böhler’s angle as a determining tool in surgical decision-making, demonstrating a relationship between a decreased angle and the need for surgical treatment, and establishing a relationship between BA values in the postoperative period and functional prognosis. Paley and Hall also reported a prognostic correlation between lower BA values and worse postoperative functional outcomes. Barroco et al. found a trend toward correlation between more severe calcaneal fractures according to the Sanders CT classification and lower BA values.

In the present study, we observed an ICC of 0.4 at most for the angle of Gissane, indicative of weak agreement between observers, across all types of joint fractures as well as in healthy, non-fractured calcaneus films—a fact also observed in the aforementioned study by Knight, in which this angle presented low reproducibility, sensitivity, and specificity even for fracture diagnosis. We observed a mean ICC of 0.305 for AG in tongue-type fractures, and 0.134 in joint depression-type fractures. Measurement of Böhler’s angle showed poor to moderate agreement, with mean ICCs of 0.621 for joint depression-type fractures and 0.444 for tongue-type fractures. The highest reliability was obtained in joint fractures of Sanders types 1 and 2. The ICCs found in this study are in agreement with the current literature. When analyzing tongue-type fractures, differences in AG values between observers were significant. In joint depression-type fractures, there were significant differences in agreement in the values of both angles.

We used the Sanders CT classification to categorize patients with joint fractures into subgroups, with the aim of evaluating inter-rater agreement in determining radiographic angles and establishing parameters with fracture patterns and degrees of joint involvement. One potential limitation of our study is that even CT classification systems are associated with inconsistent agreement between raters. Furey showed moderate agreement with use of the Sanders classification. Bhatia also demonstrated high degrees of variability and inconsistency in interpretation, resulting in weak to moderate reliability. In our study, we chose to have an independent senior orthopedist classify the fractures in an attempt to reduce this possible assessment bias.

The literature also describes some factors that may be determinants of this low reproducibility of radiographic angles of the calcaneus. Obliquity when taking lateral radiographs is one of these factors. In most cases, patients who present to the radiology service with suspected calcaneal fractures are immobilized and in pain, factors that impair proper alignment.

**Figure 5.** Estimated means (and 95%CI) for measurements of the critical angle of Gissane obtained by different raters in a sample of normal radiographs (Essex-Lopresti classification) (n=30).

**Figure 6.** Estimated means (and 95%CI) for measurements of Böhler’s angle obtained by different raters in a sample of normal radiographs (Essex-Lopresti classification) (n=30).
positioning of the foot during radiographic examination\(^{(8)}\). Gonzalez et al.\(^{(8)}\) concluded that oblique beam incidence makes it difficult to measure Böhler’s angle, with anterior and caudal inclination leading to a reduction in the angle, while posterior and cephalad inclination increase its value. The level of training of the rater performing the measurement is also important. In the same study, Gonzalez concluded that more experienced orthopedic surgeons showed greater accuracy in determining Böhler’s angle\(^{(8)}\).

**Conclusion**

In our sample, Böhler’s angle was generally more reliable than the angle of Gissane for the radiographic evaluation of calcaneal fractures. Even so, inter-rater agreement in measurement of Böhler’s angle was generally weak to moderate, with better reliability only among those fractures deemed less severe according to the Sanders CT classification (types 1 and 2). In more severe fractures as classified by the Sanders method, both angles proved to be unreliable.

**Authors’ contributions:** Each author contributed individually and significantly to the development of this article: RSB *(https://orcid.org/0000-0002-2870-2261) Conceived and planned the activities that led to the study, wrote the article, participated in the review process, approved the final version; BRM *(https://orcid.org/0000-0002-5306-2972) Conceived and planned the activities that led to the study, wrote the article, participated in the review process, approved the final version; HAF *(https://orcid.org/0000-0001-9920-5636) Wrote the article, participated in the review process, approved the final version; GBP *(https://orcid.org/0000-0003-4632-9672) Participated in the review process, approved the final version; DRNC *(https://orcid.org/0000-0003-0227-2440) Participated in the review process, approved the final version; LZPO *(https://orcid.org/0000-0001-5849-5841); ADPF *(http://orcid.org/0000-0002-5306-2972) Conceived and planned the activities that led to the study, wrote the article, participated in the review process, approved the final version; GBP *(https://orcid.org/0000-0003-4632-9672) Participated in the review process, approved the final version; CIA *(http://orcid.org/0000-0002-8951-1450). All authors read and approved the final manuscript. *ORCID (Open Researcher and Contributor ID) [10].

**References**