Comparison of 5 Different Methods for Measuring Stress Radiographs to Improve Reproducibility During the Evaluation of Knee Instability

Yong Seuk Lee, Seung Hwan Han, Joon Jo, Kyu-sung Kwak, Kyung Wook Nha and Jun Ho Kim

DOI: 10.1177/0363546510396182

The online version of this article can be found at:
http://ajs.sagepub.com/content/39/6/1275

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
American Orthopaedic Society for Sports Medicine

Additional services and information for _The American Journal of Sports Medicine_ can be found at:

Email Alerts: http://ajs.sagepub.com/cgi/alerts

Subscriptions: http://ajs.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

>> Version of Record - Jun 2, 2011

OnlineFirst Version of Record - Feb 24, 2011

What is This?
Comparison of 5 Different Methods for Measuring Stress Radiographs to Improve Reproducibility During the Evaluation of Knee Instability

Yong Seuk Lee,*† MD, Seung Hwan Han,‡ MD, Joon Jo,‡ MD, Kyu-sung Kwak,§ MD, PhD, Kyung Wook Nha,‖ MD, and Jun Ho Kim,‡ MD

Investigation performed at Ajou University School of Medicine, Suwon, Korea

Background: Although stress radiography is frequently used to assess abnormal knee instability, the reliability and reproducibility for an evaluation of anterior-posterior instability of the knee may be affected by a variety of factors.

Hypothesis: Different measurement methods result in different levels of reliability and reproducibility for instability; there may be a novel method that is more reliable and relatively unaffected by slight changes in flexion and rotation.

Study Design: Cohort study (diagnosis); Level of evidence, 2.

Methods: Stress radiographs of 40 patients with anterior cruciate ligament injury or reconstruction and 40 patients with posterior cruciate ligament injury or reconstruction were taken using the Telos device. The values of 4 conventional methods (medial-medial, lateral-lateral, middle-middle, and peripheral-central) and 1 newly devised method (Blumensaat line–anterior tibia) were compared. Intraclass correlation coefficients were examined to assess intraobserver and interobserver reliability of the measurements. For an evaluation of the reproducibility of each method, stress radiographs were taken twice (before and after the examination at the outpatient clinic) on the same day, and the values from the first and second stress radiographs were compared.

Results: In the anterior drawer test, as to measurement reliability, the ranges of intraclass correlation coefficients were 0.713 to 0.889 for medial-medial, 0.624 to 0.812 for lateral-lateral, 0.834 to 0.932 for middle-middle, 0.722 to 0.892 for peripheral-central, and 0.891 to 0.963 for Blumensaat line–anterior tibia. As to test-retest reproducibility, the mean differences (SD) of displacement between the first and second radiographs were 1.0 (0.8) mm for medial-medial, 2.4 (2.3) mm for lateral-lateral, 1.7 (1.6) mm for middle-middle, 1.2 (0.6) mm for peripheral-central, and 0.5 (0.7) mm for Blumensaat line–anterior tibia. In the posterior drawer test, as to measurement reliability, the ranges of intraclass correlation coefficients were 0.859 to 0.958 for medial-medial, 0.773 to 0.915 for lateral-lateral, 0.859 to 0.951 for middle-middle, 0.852 to 0.958 for peripheral-central, and 0.893 to 0.961 for Blumensaat line–anterior tibia. As to test-retest reproducibility, the mean differences (SD) of displacement between the first and second radiographs were 1.6 (1.3) mm for medial-medial, 1.8 (1.7) mm for lateral-lateral, 1.7 (1.5) mm for middle-middle, 1.4 (1.1) mm for peripheral-central, and 1.1 (1.2) mm for Blumensaat line–anterior tibia.

Conclusion: Different methods of measuring stress radiographs resulted in different levels of reliability and reproducibility. In the anterior drawer test, the Blumensaat line–anterior tibia method showed the best measurement reliability and test-retest reproducibility. In the posterior drawer test, the Blumensaat line–anterior tibia method showed favorable measurement reliability and reproducibility, but the superiority could not be demonstrated.

Keywords: knee; instability; stress radiograph; measurement method; reliability; reproducibility

Stress radiographs have been introduced to measure and document sagittal knee instability, and some authors have emphasized their superiority over both the arthrometer and clinical examination, particularly in the posterior cruciate ligament (PCL) deficient knee.5,8,11,13,15 However, although stress radiography is commonly used to make a better quantification of abnormal knee instability, the position of the knee landmarks is affected by even slight changes in flexion and rotation that occur as the joint is stressed.16

Some studies have examined the reliability of stress radiographs and affecting variables, particularly flexion and rotation.9,12,14,16 Stress radiography is generally accepted to be a reliable method, but further studies are recommended to improve the reproducibility and identify a better tool.4,14 The important problems are that peripheral landmarks are inaccurate if the rotation is changed, and central methods show the most changes if flexion is altered.14,16 Wirz et al16 recommended a central peripheral measuring method to determine the translation of the tibia after comparing 4 measurements methods.
The central method is generally little influenced with rotation, and more posterior points move less as the knee is flexed. Based on the above mentioned concept, it is believed that peripheral landmarks that are positioned at the central portion of the knee joint would show the most constant results regardless of the affecting variables. Therefore, 2 landmarks (Blumensaat line as a femoral landmark and anterior cortical line of proximal tibia as a tibial landmark) were chosen because they can be easily checked and are located at the central portion of the knee joint, and the posterior end of the Blumensaat line is located more posterior than is the femoral shaft axis.

The purposes of this study were (1) to evaluate the efficacy of the various methods of stress radiographs by means of intraobserver and interobserver reliability and test-retest reproducibility and (2) to prove the effectiveness of our novel stress radiograph, namely Blumensaat line–anterior tibia (BAT) method, in practical use.

**MATERIALS AND METHODS**

Stress radiographs of 40 patients with anterior cruciate ligament (ACL) injuries and 40 patients with PCL injuries were taken using the Telos device (Austin & Associates, Fallston, Maryland). Nonsurgically and surgically treated patients were enrolled. There were 4 ACL-deficient knees (nonsurgically managed) and 36 reconstructed knees in the ACL patients. There were 12 PCL-deficient and 28 reconstructed knees in the PCL patients. We performed stress radiographs after 6 months from initial trauma or reconstruction. We also included referred patients from other hospitals for the evaluation of reconstructed ligaments. All patients provided informed consent. We excluded patients if they refused to have the stress radiographs. This study was approved by the institutional review board. There were 32 men and 8 women with a mean age of 28 years (range, 19-48 years) in the ACL group and 34 men and 6 women with a mean age of 32 years (range, 22-54 years) in the PCL group. Stress radiographs were taken using the Telos device (Austin & Associates, Fallston, Maryland). Nonsurgically and surgically treated patients were enrolled. There were 4 ACL-deficient knees (nonsurgically managed) and 36 reconstructed knees in the ACL patients. There were 12 PCL-deficient and 28 reconstructed knees in the PCL patients. We performed stress radiographs after 6 months from initial trauma or reconstruction. We also included referred patients from other hospitals for the evaluation of reconstructed ligaments. All patients provided informed consent. We excluded patients if they refused to have the stress radiographs. This study was approved by the institutional review board. There were 32 men and 8 women with a mean age of 28 years (range, 19-48 years) in the ACL group and 34 men and 6 women with a mean age of 32 years (range, 22-54 years) in the PCL group.

Stress radiographs after 6 months from initial trauma or reconstruction. We also included referred patients from other hospitals for the evaluation of reconstructed ligaments. All patients provided informed consent. We excluded patients if they refused to have the stress radiographs. This study was approved by the institutional review board. There were 32 men and 8 women with a mean age of 28 years (range, 19-48 years) in the ACL group and 34 men and 6 women with a mean age of 32 years (range, 22-54 years) in the PCL group.

Statistical Analysis

The statistical analysis to prove the reliability was performed in 2 phases, separately in anterior drawer and
posterior drawer tests. The reliability with regard to measurement was checked by the intraobserver and interobserver agreement of the measurement of the identical radiograph, and the reproducibility with regard to taking the radiograph was evaluated by test-retest reproducibility. The SAS version 9.13 (SAS Institute Inc, Cary, North Carolina) was used, and the statistical significance was set at $P < .05$.

First, the intraobserver and interobserver reliability was assessed by measuring the displacement on the same radiograph by 2 separate raters (2 of the authors) for the 5 different methods. The same measurement was repeated after 3 weeks. The reliability was evaluated by intraclass correlation coefficient (ICC), which is defined as the ratio of the variance between patients to the total variance. The ICC values can theoretically range from 0 to 1, with a higher value indicating that less variance is due to other factors such as differences between observers. The ICC was calculated from a 2-way random effects model, for absolute agreement, namely ICC (type 2,1). The ranges of ICCs were from the minimum to the maximum values of intrarater and interrater reliability for each method. The reference values were 0.00 to 0.40, poor; 0.40 to 0.75, fair to good; 0.75 to 1.00: excellent, albeit arbitrary.\textsuperscript{1,2}

Second, the test-retest reproducibility of taking the repeated radiographs was assessed by measuring the difference between the displacements on the 2 separate radiographs. The measurement reliability was confirmed by the aforementioned analysis, and the measurement of a single rater was used in the following analysis of test-retest reproducibility. Individual means (SD) and their 95% confidence intervals (CI) were calculated. The mean difference in the displacements in repeated radiographs (second radiograph – first radiograph) and SD were analyzed. The stress radiography with smaller difference in displacements and SD was regarded to be better than those with larger differences.

RESULTS

In the anterior drawer test, the BAT method showed the best measurement reliability and test-retest reproducibility.
As to measurement reliability, the ICCs and their 95% CIs were as follows: MM: ICCs, 0.713 to 0.889 (95% CIs, 0.775 to 1.245); LL: ICCs, 0.624 to 0.812 (95% CIs, 1.696 to 3.104); Mid-Mid: ICCs, 0.834 to 0.932 (95% CIs, 1.21 to 2.19); PC: ICCs, 0.722 to 0.892 (95% CIs, 1.016 to 1.384); and BAT: ICCs, 0.891 to 0.963 (95% CIs, 0.286 to 0.714) (Table 1). Generally, the BAT method showed the best ICCs followed by Mid-Mid, PC, MM, and LL methods. As to test-retest reproducibility, the mean differences (SD) of displacement between the first and second radiograph was 1.0 (0.8) for MM, 2.4 (2.3) for LL, 1.7 (1.6) for Mid-Mid, 1.2 (0.6) for PC, and 0.5 (0.7) for BAT (Table 2). Generally, the BAT method showed the smallest mean difference and SD followed by MM, PC, Mid-Mid, and LL methods.

In the posterior drawer test, the BAT method showed favorable measurement reliability and test-retest reproducibility, but the superiority could not be demonstrated. As to measurement reliability, the ICCs and their 95% CIs were as follows: MM: ICCs, 0.859 to 0.951 (95% CIs, 1.241 to 2.159); LL: ICCs, 0.852 to 0.958 (95% CIs, 1.063 to 1.737); and BAT: ICCs, 0.893 to 0.961 (95% CIs, 0.732 to 1.467) (Table 3). As to test-retest reproducibility, the mean differences (SD) of displacement between the first and second radiographs were 1.6 (1.3) for MM, 1.8 (1.7) for LL, 1.7 (1.5) for Mid-Mid, 1.4 (1.1) for PC, and 1.1 (1.2) for BAT (Table 4).

**DISCUSSION**

The principal findings of this study were that the BAT method showed the best measurement reliability and test-retest reproducibility in the anterior stress radiographs. However, the superiority could not be demonstrated in the posterior stress radiographs, although they showed favorable measurement reliability and test-retest reproducibility.

A precise evaluation of knee instability is important because the surgical indication and clinical results are

![Figure 2. Methods for measuring the medial-medial (A), lateral-lateral (B), middle-middle (C), peripheral-central (D), and Blumen-saat line–anterior tibia (E) in the posterior stress radiographs. The distances between arrows were measured.](image-url)
determined by this parameter. Stress radiographs have some advantages over other tools, and an evaluation of an ACL-injured knee is relatively easier than that of a PCL-injured knee. Although stress radiographs are often replaced by more or less precise tools, the posterior stress test is generally accepted for quantifying posterior laxity.\textsuperscript{3,6,10,15} The advantages of stress radiographs are they are a noninvasive, objective, reproducible method for recording the amount of sagittal translation and skeletal displacement, which eliminates errors from compliance of the soft tissues that occur with arthrometers.\textsuperscript{5} However, the accuracy of stress radiographs is affected by a range of factors, such as the lack of patient relaxation due to pain, rotation of the limb, or rotation of the x-ray beam and different flexion of the knee joint.\textsuperscript{9} Some studies evaluating the reliability of stress radiographs reported relatively satisfactory results.\textsuperscript{4,6,7,14,16} It is believed that reproducibility is also important because it allowed for consistent planning. To the best of our knowledge, there are few reports on reproducibility, and more studies are recommended.\textsuperscript{14}

### TABLE 1

<table>
<thead>
<tr>
<th>Instrarater</th>
<th>Interrater</th>
<th>1</th>
<th>2</th>
<th>Lower Limit</th>
<th>Upper Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>MM</td>
<td></td>
<td>0.713</td>
<td>0.758</td>
<td>0.713</td>
<td>0.889</td>
</tr>
<tr>
<td>LL</td>
<td></td>
<td>0.745</td>
<td>0.624</td>
<td>0.889</td>
<td>0.853</td>
</tr>
<tr>
<td>Mid-Mid</td>
<td></td>
<td>0.932</td>
<td>0.834</td>
<td>0.934</td>
<td>0.853</td>
</tr>
<tr>
<td>PC</td>
<td></td>
<td>0.722</td>
<td>0.786</td>
<td>0.892</td>
<td>0.722</td>
</tr>
<tr>
<td>BAT</td>
<td></td>
<td>0.897</td>
<td>0.891</td>
<td>0.963</td>
<td>0.891</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Individual intrarater and interrater reliability and the range (minimum, maximum) were given. Intrarater, between the measurements by a rater with 3-week interval; interrater, between the raters in each time of measurement. BAT, Blumensaat line–anterior tibia; LL, lateral-lateral; Mid-Mid, middle-middle; MM, medial-medial; PC, peripheral-central.

### TABLE 2

<table>
<thead>
<tr>
<th>First Radiograph</th>
<th>Second Radiograph</th>
<th>Difference Between First and Second Radiographs</th>
<th>95% CI of Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>MM</td>
<td>3.4</td>
<td>4.4</td>
<td>1</td>
</tr>
<tr>
<td>LL</td>
<td>7.1</td>
<td>9.5</td>
<td>2.4</td>
</tr>
<tr>
<td>Mid-Mid</td>
<td>6.2</td>
<td>7.9</td>
<td>1.7</td>
</tr>
<tr>
<td>PC</td>
<td>48.7</td>
<td>49.9</td>
<td>1.2</td>
</tr>
<tr>
<td>BAT</td>
<td>48.9</td>
<td>49.4</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\textsuperscript{a} BAT, Blumensaat line–anterior tibia; CI, confidence interval; LL, lateral-lateral; Mid-Mid, middle-middle; MM, medial-medial; PC, peripheral-central.

### TABLE 3

<table>
<thead>
<tr>
<th>Instrarater</th>
<th>Interrater</th>
<th>1</th>
<th>2</th>
<th>Lower Limit</th>
<th>Upper Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>MM</td>
<td></td>
<td>0.859</td>
<td>0.958</td>
<td>0.859</td>
<td>0.958</td>
</tr>
<tr>
<td>LL</td>
<td></td>
<td>0.894</td>
<td>0.814</td>
<td>0.773</td>
<td>0.915</td>
</tr>
<tr>
<td>Mid-Mid</td>
<td></td>
<td>0.914</td>
<td>0.951</td>
<td>0.859</td>
<td>0.951</td>
</tr>
<tr>
<td>PC</td>
<td></td>
<td>0.852</td>
<td>0.872</td>
<td>0.858</td>
<td>0.958</td>
</tr>
<tr>
<td>BAT</td>
<td></td>
<td>0.893</td>
<td>0.894</td>
<td>0.893</td>
<td>0.961</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Individual intrarater and interrater reliability and the range (minimum, maximum) were given. Intrarater, between the measurements by a rater with 3-week interval; interrater, between the raters in each time of measurement. BAT, Blumensaat line–anterior tibia; LL, lateral-lateral; Mid-Mid, middle-middle; MM, medial-medial; PC, peripheral-central.
encountered in the same patients although the technicians were well trained and followed the manual. This was possibly due to multiple affecting variables as mentioned previously. In this situation, the treatment plan could not be determined precisely, and different interpretations could occur during the follow-up period. Therefore, there is a need for a new measurement method that is less sensitive to positioning errors.

To quantify the translation of the tibia, landmarks must be determined and easily identified with some rotation and difference in flexion. In addition, they must be close to the joint line. Landmarks farther from the center of the knee tend to show more distortion, particularly with rotation. Therefore, the central landmark is most constant through the rotation arc. More posterior points on the femoral condyles move less than does the femoral shaft axis as the knee is flexed.

TABLE 4
Displacement (in millimeters) of the First and Second Radiographs, Mean Difference, and 95% CI of the Mean Difference for Each Method in Posterior Drawer Stress Radiographs

<table>
<thead>
<tr>
<th>Method</th>
<th>First Radiograph Mean (mm)</th>
<th>First Radiograph SD (mm)</th>
<th>Second Radiograph Mean (mm)</th>
<th>Second Radiograph SD (mm)</th>
<th>Difference Between First and Second Radiographs Mean (mm)</th>
<th>Difference Between First and Second Radiographs SD (mm)</th>
<th>95% CI of Mean Difference Lower</th>
<th>95% CI of Mean Difference Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>MM</td>
<td>8.7</td>
<td>5.3</td>
<td>10.3</td>
<td>5.4</td>
<td>1.6</td>
<td>1.3</td>
<td>1.202</td>
<td>1.998</td>
</tr>
<tr>
<td>LL</td>
<td>6.5</td>
<td>5</td>
<td>8.3</td>
<td>5.4</td>
<td>1.8</td>
<td>1.7</td>
<td>1.279</td>
<td>2.321</td>
</tr>
<tr>
<td>Mid-Mid</td>
<td>7.5</td>
<td>4.5</td>
<td>9.2</td>
<td>4.9</td>
<td>1.7</td>
<td>1.5</td>
<td>1.241</td>
<td>2.159</td>
</tr>
<tr>
<td>PC</td>
<td>43.1</td>
<td>5.9</td>
<td>44.5</td>
<td>6.1</td>
<td>1.4</td>
<td>1.1</td>
<td>1.063</td>
<td>1.737</td>
</tr>
<tr>
<td>BAT</td>
<td>43.2</td>
<td>5.9</td>
<td>44.3</td>
<td>6.1</td>
<td>1.1</td>
<td>1.2</td>
<td>0.732</td>
<td>1.467</td>
</tr>
</tbody>
</table>

*BAT, Blumensaat line–anterior tibia; CI, confidence interval; LL, lateral-lateral; Mid-Mid, middle-middle; MM, medial-medial; PC, peripheral-central.

Figure 3. Different flexion of the checked radiographs (A, anterior; B, posterior stress radiographs) using the same manuals in the same patients.

Figure 4. Different rotation of the checked radiographs (A, anterior; B, posterior stress radiographs) using the same manuals in the same patients.
This phenomenon is also similar in the tibial side. However, we could not identify a constant posterior and central position landmark. For these reasons, the posterior end of the Blumensaat line was chosen as a femoral landmark, and the connecting point between the anterior cortex of the proximalibia and tibial plateau was selected as a tibial landmark.

This BAT method showed superior reproducibility in ACL-injured patients. On the other hand, although the BAT method showed the best results, all 5 methods are suitable for PCL-injured patients. This suggests that slight changes in rotation or flexion are permitted with this BAT method, although the problem of relaxation and large changes in rotation or flexion were not solved.

We checked both sides, but only the injured side was used for the analysis for a number of reasons. First, measurement differences are made at both sides, and the rotation and flexion could be different between sides. Second, the difference itself is smaller than is absolute measurement in case of mild instability, and the results could be magnified with small measurement error.

It is very important to define how much time occurred before and after examination for these radiographs. We repeated the same measurement after 3 weeks for the evaluation of ICCs. However, we checked stress radiographs twice (before and after the examination at the outpatient clinic) on the same day for the evaluation of test-retest reproducibility. It is possible to have some inherent bias in this type of radiograph, as the patient would have memory within 1 to 2 hours of their experiences in this circumstance. However, we intended this effect. In other words, this can minimize patient factor among variables because the patients would have memory of their experiences. Furthermore, there is little possibility of fundamental change of ligament characteristics.

CONCLUSION

Different methods for measuring stress radiographs resulted in different levels of reproducibility for both anterior and posterior instability using the Telos device. In the anterior drawer test, the BAT method showed the best measurement reliability and test-retest reproducibility. In the posterior drawer test, the BAT method showed favorable measurement reliability and reproducibility, but the superiority could not be demonstrated.

REFERENCES