Digital Planning Software Fails to Reflect Stem Torsion on Plain Radiographs after Total Hip Arthroplasty

Digitale Planungssoftware ist nicht geeignet die Torsion des Implantatschaftes nach endoprothetischem Hüftersatz zu bestimmen

Key words
- hip arthroplasty
- stem version
- anterior posterior radiographs
- planning software
- CT

Zusammenfassung

Zielsetzung: Ziel dieser Studie war es die Validität einer kommerziell verfügbaren Planungssoftware zur Vermessung von zweidimensionalen Röntgenbildern, im Vergleich zu CT-Aufnahmen zu bestimmen.


Abstract

Purpose: The purpose of this study was to evaluate the validity of commercially available planning software on plain radiographs after THA compared to CT scans as the gold standard.

Patients and Methods: In a prospective clinical study, anteroposterior (AP) radiographs and three-dimensional CT scans (3D-CT) were obtained for 121 patients, who underwent minimally invasive, cementless THA with a straight tapered stem, in a lateral decubitus position. For measuring SV, we used digital planning software (TraumaCad 2.0, BrainLAB Feldkirchen, Germany). Two independent raters repeated the analysis after a six-week interval. Radiological measurements were compared with 3D-CT measurements by an independent, blinded external institute. This investigation was approved by the local ethics commission (no. 10-121-0263) and is a secondary analysis of a larger project (DRKS00 000 739, German Clinical Trials Register May-02 – 2011).

Results: The radiograph measurements showed very high intra- and interrater agreement. The intra-class correlation (ICC) of the interrater agreement was 0.97 for rater 1 and 0.98 for rater 2. The interrater reliability was 0.99 using the mean values of both rater measurements. The mean difference between the average radiograph measurement and the 3D-CT-based measurement was 0.41° (SD 11.24°) (range: -33.85°–22.50°; 95% limits of agreement: -21.63 – 22.45), but there was no correlation found between both methods.

Conclusion: Measuring stem version with the help of commercially available digital planning software on plain radiographs after THA has high intra- and interrater reliability but clinically unacceptable validity and reliability when compared to 3D-CT scans.
Measuring stem torsion after THA on plain radiographs with digital planning software is not valid.

Citation Format:
Fig. 1 Radiographic assessment of stem version (26°) calculated automatically by the software, the 45° are the difference between the true neck-shaft angle of 135° of the implant and the vertical stem axis (180°).

Abb. 1 Vermessung der Schaft torsion (26°), automatisch errechnet durch die Software, die 45° sind die Differenz zwischen dem wahren Hals-Schaftwinkel von 135° und der vertikalen Schaftachse (180°).

Table 1 Characteristics of the study group.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>n = 121</th>
</tr>
</thead>
<tbody>
<tr>
<td>gender (female) (%)</td>
<td>66 (55)</td>
</tr>
<tr>
<td>age (yrs)</td>
<td>62.7 (SD 0.6)</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>27.0 (SD 4.1)</td>
</tr>
<tr>
<td>treatment side (right) (%)</td>
<td>67 (55)</td>
</tr>
<tr>
<td>femoral component size (IQR)</td>
<td>11 (2)</td>
</tr>
<tr>
<td>femoral component geometry (%)</td>
<td>Std 60 (49.5); HO 61 (51.5)</td>
</tr>
<tr>
<td>OP time (min)</td>
<td>71.4 (12.5)</td>
</tr>
<tr>
<td>kellgren (IQR)</td>
<td>8 (1)</td>
</tr>
<tr>
<td>length of incision (cm) (SD)</td>
<td>10.4 (1.3)</td>
</tr>
</tbody>
</table>

BMI: body mass index; HO: high-offset stem; Std: standard stem; IQR: interquartile range.

Statistical methods
Statistical analyses were performed with IBM SPSS Statistics® 23.0 (SPSS Inc., Chicago, IL, USA) and R version 3.2.1. Data are presented as mean, standard deviation and range. The accuracy of the radiographs was assessed using Bland-Altman plots and clinical evaluation. Bland-Altman plots illustrate the accuracy of the radiograph measurements compared to the 3D-CT-based measurement (gold standard) by plotting the gold standard on the x-axis and the difference of both measurements on the y-axis. The dashed lines in the graph represent the 95% limits of agreement (mean ± 1.96 SD). Intra- and inter-rater agreement (precision) was assessed by the intraclass correlation coefficient (ICC) and graphically by scatter plots and standard Bland-Altman plots (i.e., the x-axis shows the average of both measurements).

Results ▼

Precision
The radiograph measurements showed very high intra- and inter-rater agreement. The ICC of the intrarater agreement was 0.97 for rater 1 and 0.98 for rater 2 (Fig. 2). The interrater reliability was 0.99 using the mean values of both rater measurements. The 95% limits of agreement range between −7.4 and 6.6 for the intrarater agreement and between −6.9 and 6.2 for the interrater agreement. The mean difference is close to null in both cases (Fig. 2).

Accuracy ▼
Due to the excellent intra- and interrater agreement, we used the mean value of the four measurements for the Bland-Altman plot. The mean difference between the average radiograph measurement and the 3D-CT-based measurement was 0.41° (SD 11.24°) (range: −33.85°–22.50°; 95% limits of agreement: −21.63–22.45) (Fig. 3). In all, 43/121 (36%) of the radiological measurements of prosthetic SV were within a tolerance limit of 5° compared with 3D-CT. The Bland-Altman plot shows that there was no systematic error of the radiograph measurements. Table 2 summarizes the measurements on plain radiographs performed by the two raters and by 3D-CT.

Discussion ▼

Malpositioning of components in THA leads to pain, reduced range of motion and early instability [2–5, 18]. So far, CT has been the gold standard for postoperative assessment of THA components because of its high accuracy and reliability [10, 19]. Today several software programs promise the ability to measure SV on standard AP radiographs. We aimed to investigate the objectivity, reliability and validity of measuring stem version with the help of commercially available planning software on plain radiographs after THA compared to CT scans as the gold standard. We found excellent intra- and interrater reliability of the software. The raters even had the same outliers in both their measurements compared to 3D-CT. In regard to the software’s reliability and validity, the mean prosthetic version of the stem measured by both raters was close to the mean version measured by 3D-CT but without correlation between the two techniques. Reasons for inaccuracy could be differences in picking landmarks, ignorance of the femoral tilt, the angular difference between the long axis of the femoral stem

shaft angle and the projection-based angle and converts it into the degree of version of the stem. The software is not able to differentiate between anteverversion and retroversion. For that reason the second clinical standard plane (axial) has to be consulted. All radiological measurements were performed by two independent examiners (MW, MS), who repeated the measurements after a six-week interval. The raters were blinded to the 3D-CT values as well as to each other’s results. In addition, 3D-CT assessment of prosthetic stem version was obtained by an independent, blinded external institute (MeVis Medical Solutions, Bremen, Germany), as described by Sendtnet et al. [11]. Correlation was characterized as poor (0.00–0.20); fair (0.21–0.40); moderate (0.41–0.60); good (0.61–0.80) or excellent (0.81–1.00) [16]. As the generally accepted range of stem anteverversion is between 10° and 20°, we defined a tolerance limit of 5° compared with 3D-CT as clinically acceptable [17].
and the mechanical axis on a sagittal radiograph. That difference is due to the fact that the stem of the prosthesis follows the natural anterior bow of the proximal femur [20]. In summary, we found high reliability but no validity for the use of digital planning software for measuring SV after THA. Nevertheless, we think the software can be a useful tool for a first approximate determination of major rotational errors of the femoral component in a painful hip after THA.

There are several limitations when measuring SV with the help of digital planning software on plain radiographs. First, software is not able to differentiate between anteversion and retroversion. Therefore, a second axial radiograph is needed to distinguish between the two. Second, we found the handling of the software itself challenging. Nevertheless, we found excellent intra- and inter-rater reliability. The exact determination of the axis of the neck and the stem is prone to error, because it must be done by hand so landmark selection is inaccurate. Even minimal changes of the position of one axis lead to a high change of the value of SV. A more accurate way for determination of the axes could be to draw concentric circles into the neck and the stem and use their midpoints as orientation for the axes as described by Weber et al. [21]. Third, the software bases its measurements on the known NSA. A factor that has a high influence on the NSA is the position of the patient. This means any internal or external rotation of the leg and any extension or flexion of the hip can lead to a misinterpretation of the degree of the stem version. To avoid that impact, a rigorously

### Table 2

<table>
<thead>
<tr>
<th>rater 1</th>
<th>rater 1</th>
<th>rater 2</th>
<th>rater 2</th>
<th>average</th>
<th>average</th>
<th>average</th>
<th>3D-CT</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>M2</td>
<td>M1</td>
<td>M2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>7.1</td>
<td>7.4</td>
<td>7.7</td>
<td>7.7</td>
<td>7.2</td>
<td>7.6</td>
<td>7.4</td>
</tr>
<tr>
<td>SD</td>
<td>14.5</td>
<td>14.8</td>
<td>15.2</td>
<td>15.3</td>
<td>14.5</td>
<td>15.3</td>
<td>14.8</td>
</tr>
<tr>
<td>minimum</td>
<td>–25.0</td>
<td>–26.0</td>
<td>–26.0</td>
<td>–26.0</td>
<td>–25.0</td>
<td>–26.0</td>
<td>–25.5</td>
</tr>
<tr>
<td>maximum</td>
<td>48.0</td>
<td>46.0</td>
<td>45.0</td>
<td>45.0</td>
<td>47.0</td>
<td>45.0</td>
<td>46.0</td>
</tr>
<tr>
<td>ICC</td>
<td>0.97</td>
<td>0.98</td>
<td>0.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
standardized radiological technique as used by us in the study protocol is essential to ensure exact positioning of the patient and to minimize projection errors [22]. Furthermore, the quality of the radiograph is very important. Although our measurements were obtained under optimized conditions following a strict protocol, the radiological image in two cases was inadequate. So there might be another limitation for the method in clinical practice. A different option for measuring SV after THA on plain radiographs is the Budin method. In short, this method uses a posteroanterior radiograph of the hip in 90° flexion and 30° abduction and with 90° flexion of the knee [10]. Lee et al. found a high correlation between the radiological and CT measurements (r = 0.88, p < 0.001) with excellent intra- (0.94) and interrater reliability (0.93) [10]. Another accurate method for the assessment of SV is to use a mathematical formula, which puts the projected NSA of the stem on plain AP radiographs in relation to the true NSA of the implant. This method was described by Weber et al. and was valid compared to 3D-CT (r = 0.88, p < 0.001) [13].

Conclusion

In conclusion, measuring stem version with the help of commercially available digital planning software on plain radiographs after THA has high intra- and interrater reliability but clinically inacceptable validity and reliability when compared to 3D-CT scans.

References