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

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.


Results: The radiograph measurements showed very high intra- and interrater agreement. The in- tra-class correlation (ICC) of the intrarater agreement was 0.97 for rater 1 and 0.98 for rater 2. The intrarater 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 inaccurate validity and reliability when compared to 3D-CT scans.
konnte keine Korrelation zwischen beiden Methoden nachgewiesen werden.

**Schlussfolgerung:** Die Vermessung der Schafftversion mit einer kommerziell verfügbaren, digitalen Planungsoftware auf zweidimensionalen Röntgenbildern, nach endoprothetischem Hüftersatz, zeigt trotz guter Reproduzierbarkeit, eine klinisch inakzeptable Ungenauigkeit verglichen mit 3D-CT-Aufnahmen.

**Kernaussagen:**
- Die Vermessung der Schafftversion auf zweidimensionalen Röntgenbildern mit digitaler Planungsoftware ist nicht valide.

**Introduction**

Primary total hip arthroplasty (THA) is one of the most performed orthopedic operations worldwide [1]. However, inaccurate placement of the femoral and acetabular components in THA can lead to dislocation, decreased range of motion (ROM), periprosthetic or bony impingement and component wear [2–6]. If a complication in THA appears, it is necessary to detect the reason and whether wrong component placement could be responsible. Therefore, the surgeon is able to resolve the problem selectively. The easiest way would be to analyze standard radiographs. Previous studies showed that version of acetabular component can be measured on single anteroposterior (AP) radiographs with accurate results considering the measuring method for daily clinical use [7–9]. The stem version can be evaluated using the so-called “Budin method”, a validated protocol for the radiological measurement of stem version [10]. A limitation of this method is the need for a special radiological image, which leads to additional radiation and expense. Another problem regarding exact assessment of the femoral component is its variation in the final position. Sendtner et al. found a range from −19° retroversion to 33° anteverision in cementless THA, which is in accordance with the results of Wines et al., who showed a range from −15° to −52° [11, 12]. Weber et al. developed a new mathematical formula for measuring stem version using the projected prosthetic neck-shaft angle (NSA) on AP radiographs and compared the results with three-dimensional CT scans (3D-CT) [13]. The authors found a high reliability and validity in the evaluation of the stem version in cementless THA, considering the limitation that this method cannot differentiate between anteverision and retroversion.

Today there are several software programs that are used for preoperative planning in THA and total knee arthroplasty (TKA). Some of these programs offer the possibility of postoperative component measuring in AP radiographs. This can be relevant considering claims of recourse of unsatisfied patients. However, to the best of our knowledge, no study has been reported about the validity of these programs with respect to measuring the stem version (SV) after THA.

This retrospective secondary analysis out of a large prospective study 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 when compared to CT scans as the gold standard.

**Patients and Methods**

In the course of a registered, prospective controlled trial (DRKS00000739, German Clinical Trials Register), hip radiographs in two planes (AP and axial) and 3D-CT scans were obtained for patients who underwent minimally invasive THA. This investigation was approved by the local ethics committee (no. 10-121-0263). All procedures were in accordance with the ethical standards of the responsible committee on human experimentation and with the Helsinki Declaration of 1975, as revised in 2000. The current study is a secondary analysis of a larger project [14].

The primary outcome of this larger study was to assess whether the ROM of the prosthetic joint could be improved by computer-assisted functional optimization of the position and containment of the acetabular component.

For this study AP radiographs of 121 patients out of the whole study collective were analyzed. Characteristics of the study group are shown in **Table 1**. The patients were chosen by random. All THAs were performed by four orthopedic surgeons (JG, ES, MWö, TR) from Regensburg University Medical Center. Each surgeon had experience with >200 fluoroscopy and navigation-controlled THAs. Press-fit acetabular components, uncremented hydroxyapatite-coated femoral components (Pinnacle acetabular component, Corail femoral component (both DePuy, Warsaw, Indiana), neutral polyethylene liners and metal heads with a diameter of 32 mm were used in all patients.

All operations were performed in the lateral decubitus position through a minimally invasive, modified Smith-Petersen approach (MicroHip®) [15]. Six week after surgery, full weight-bearing standing radiographs of the whole pelvis AP and the operated hip axial were taken (MULTIX TOP ACS; Siemens, Erlangen, Germany). The radiographers made sure that the pelvis was set parallel to the plane of the film without rotation or flexion of the hip joint and the leg was placed in a neutral position, with the patella pointing forward disregarding the foot progression angle in the event of a tibial version. All radiographs were to be taken under these standardized conditions (focus-film distance 115 cm, 75 kV, automatic exposure). During the same visit a CT scan was obtained from the pelvis down to the femoral condyles (Somatom Sensation 16; Siemens, Erlangen, Germany).

The radiological SV was measured with the help of the “semi-automatic” function of digital planning software (TraumaCad 2.0, BrainLAB Feldkirchen, Germany). For this purpose, an exact circle has to be drawn around the femoral head to assess its center. Then the axes of the component neck and the axis of the stem have to be determined. The angle between these axes is regarded as the neck-shaft angle measured automatically by the software or manually with a 4-point middle line. Both anteverision and retroversion around the axis of the femur cause the projected neck-shaft to appear increased (**Fig. 1**). The true NSA of the stem is known to be 135° and its difference to the vertical axis of the implant (180°) is 45°. This means the higher the version of the stem, the higher the projection-based increase of the NSA. The software promises to recognize the difference between the neck-
The radiograph measurements showed very high intra- and interrater agreement. The ICC of the intrarater agreement was 0.97 for rater 1 and 0.98 for rater 2 (Fig. 2). The intrarrater 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.
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></td>
<td>M1</td>
<td>M2</td>
<td>M1</td>
<td>M2</td>
<td>rater 1</td>
<td>rater 2</td>
<td>rater 1</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>

SD = standard deviation, M = measurement, ICC = intraclass correlation coefficient.
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

14 Renkawitz T, Haimeri M, Dohmen L et al. Minimally invasive computer-navigated total hip arthroplasty, following the concept of femur first and combined anteversion: design of a blinded randomized controlled trial. BMC musculoskeletal disorders 2011; 12: 192
16 Landis JR, Koch GC. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159 – 174

Fig. 3 The Bland-Altman plot shows the mean difference between the average of the measurements of both observers and 3D-CT measurements of stem torsion. The straight line represents the mean value of all differences between the pair of measurements and the dashed lines above and below represent 95% limits of agreement (anteversion (°)).

Abb. 3 Der Bland-Altman-Plot zeigt die mittlere Abweichung zwischen den Durchschnittswerten der Messungen beider Untersucher und der 3D-CT-Messung der Schafttorsion. Die gerade Linie repräsentiert den Mittelwert der Differenzwerte der beiden Messungen und die gestrichelten Linien begrenzen das 95% Übereinstimmungsintervall (Antetorsion(°)).