A Computerized Analysis-by-Synthesis Algorithm Improves Precision of Linear Wear Measurements in Total Hip Replacements

Claudio Dora\textsuperscript{a} (MD), Kathrin Burckhardt\textsuperscript{b} (PhD), Gábor Székely\textsuperscript{b} (PhD), Valentin Rousson\textsuperscript{d} (PhD), Juerg Hodler\textsuperscript{c} (MD), Christian Gerber\textsuperscript{a} (MD)

\textsuperscript{a}Department of Orthopedics, University of Zurich, Balgrist, Forchstrasse 340, CH-8008 Zurich, Switzerland

\textsuperscript{b}Computer Vision Laboratory, Swiss Federal Institute of Technology, ETH-Zentrum, Sternwartstrasse 7, CH-8092 Zurich, Switzerland

\textsuperscript{c}Department of Radiology, University of Zurich, Balgrist, Forchstrasse 340, CH-8008 Zurich, Switzerland

\textsuperscript{d}Biostatistics Unit, University of Zurich, 8001 Zurich, Switzerland

Correspondence:
Dora Claudio
Department of Orthopedics, University of Zurich, Balgrist
CH-8008 Zurich, Switzerland

claudio.dora@balgrist.ch

Phone: 0041 44 386 12 91

Fax: 0041 44 386 12 90

Running Title: High precision wear measurement
Summary

Precision is important if small polyethylene wear rates are to be detected early and in small sample sizes. Using an automatic, computerized algorithm relying on a synthetic X-ray generated from the CAD model of the implant may significantly improve precision of linear wear measurements.

Purpose: To test the hypothesis that the proposed method significantly improves the in vivo precision when compared to the nowadays widely used Hip Analysis Suite™ software (version 8.0.1.7).

Methods: Two anterior-posterior pelvic X-rays were taken at the same day for a total of 18 total hip implants and wear was measured by three observers using the two methods under study. Expected real wear was zero and assumed as reference value. Two sources of variability were estimated, one due to the instrument and one due to different operators, and a Wilcoxon Signed Rank Test was used for statistical comparisons.

Results: The overall precision (instrument variability and multi-operator variability) given a 95% precision limit was significantly higher using the proposed method (±0.15 mm) when compared to the Hip Analysis Suite (±0.84 mm, p<0.001).

Conclusions: Due to the availability of the full CAD data of the implant and to further automation of the measuring process, the newly proposed method allowed to improve precision of PE wear measurement by a factor of 5.6.
Introduction

Measurement of femoral head penetration into the polyethylene (PE) liner from sequential radiographs has become a standard of assessing in vivo linear wear. Roentgen Stereophotogrammetric Analysis (RSA) (1) is considered the most accurate method of determining the magnitude of relative displacements of components on radiographic images. It requires the inclusion of specific markers during surgery and is therefore only suitable for prospective investigations. Additionally, the acquisitions of a calibration object, the need for two simultaneously acquired x-ray images from different directions as well as complex and expensive technical requirements are disadvantages of RSA as a routine clinical tool.

Earlier manual techniques such as those of Charnley (2) and Livermore (3) are associated with high variability among different users and lack of the precision to yield useful information in shorter in-vivo time periods or in cases with relatively small amounts of wear (4).

In the past decade computer-assisted methods using image analysis techniques have been developed in an attempt to improve the precision of radiographic measurements in the absence of tantalum markers (5, 6). These methods have in common that only particular geometric features, such as single image points, the contours of the implants or the coordinates of external markers are used to locate the head and the cup. This limits the precision of the methods as only a part of the available information is used. Moreover, additional error is introduced by the features which have to be extracted from the image, i.e. by their identification and the determination of their position. This error is particularly important in manual methods such as the ones of Charnley (2) and Livermore (3) but may also play a role if this geometric information is gained by a computerized procedure such as in RSA(1) or in the methods of Devane (5) and Martell (6). In the latter cases the magnitude of the error depends on the suitability of the image analysis algorithm, on the shape of the individual feature, and on its contrast.

Indeed, precision of the measuring method is of major importance to enable differences of wear rates between implants to be detected early and in small sample sizes. Therefore, we proposed a new method based on an Analysis-by-Synthesis algorithm, that allows evaluating the full intensity information without the need of prior extraction of geometric data. This is made possible by localizing the implants using intensity based registration (7). The implant
position is determined in the image by minimizing the difference between its synthetic x-ray image generated using a CAD model of the prosthetic component and its original radiographic projection (7). An in-vitro validation of this method under simulation of wear revealed an accuracy in terms of the closeness of agreement of 10 µm (7) between the test result and the reference value. The precision was studied in vivo and resulted to be 50 µm (7). This computer-assisted method is well suitable for radiographic routine examination. It can be performed on a standard personal computer and relies on standardized conventional radiographs and on the CAD data of the implants.

The aim of the present study was to compare the overall in-vivo precision of this approach with the currently widely used Hip Analysis Suite™ (University of Chicago, Chicago, IL) software developed by Martell et al (6). The study was performed with the hypothesis that the Analysis-by-Synthesis method is more precise.
Materials and Methods

Patients and Image Material

In the context of a regular follow-up visit after total hip replacement, where an anterior-posterior pelvic x-ray was taken on a routine base, a second anterior-posterior pelvic radiograph was taken for 17 patients who had given informed consent. Approval was obtained from our Institutional Review Board prior to performing this study. One of the participating patients had a bilateral THR resulting in a total number of 18 total hip implants to be measured.

In order to simulate two different x-ray examinations with all positional and expositional variability, which all may result in errors when locating the head and cup in the image, the second pelvic x-rays of the volunteers were taken after the orthopedic consultation by a radiological technician about 1 to 2 hours later. In the meantime, x-ray images of other patients not participating in the study were acquired. The conditions should hence be similar as in two radiographs of a THR follow-up study taken at a greater time interval. As the two pelvic x-rays were acquired within the same follow-up examination, the expected real wear between these radiographs was zero. Theoretical positional changes of the head within the cup between the two radiographs due to incomplete contact of the femoral bearing on the polyethylene liner are assumed to contribute as random error to the measured wear. All patients had a Fitmore™ or a Fitek™ cup (Zimmer, inc, Warsaw ind. (formerly Centerpulse®, Winterthur). The outside diameter of the cups ranged from 46mm to 58mm. The diameter of the (metallic) femoral head was always 28mm.

The radiographs were measured by a postdoc in computer vision (observer I), a clinician (observer II), and a software engineer (observer III) using the two methods under study.

Techniques of measuring wear

Wear measurements using the Hip Analysis Suite (John M. Martell, MD, University of Chicago Medical Center, 5841 S Maryland Ave, MC 3079, Chicago, IL 60637-1463) were performed according to the user manual as well as to the description of the inventor of the technique (6). Usually, the latest follow-up radiograph and the 6 week post-operative image are loaded onto the computer system. For this study, the first and second same day radiographs were used.
As in our study only anterior-posterior radiographs were available, the software was run in a 2D mode.

The lowest points of the ischial tuberosities, three points on the contour of the head, and three points of the cup's elliptical opening face were manually marked by mouse clicks in the radiograph. They defined the orientation of the pelvis, the acetabular inclination and anteversion, and an initial estimate of the position of the implant components. The software then computed automatically the final position of the femoral head relative to the cup by using an edge detection algorithm and by matching circles to the implants' contours (see Fig.1). The surface of the cup under study is flattened at the pole, which made a manual marking with three additional points on its outer contour necessary. The fact that the cup did not have an exactly hemispherical shape possibly increased the error in its localization, because its oblate contour was approximated by a circle, and the manual mode led to an increased influence of the operator.

The position of the head is represented in the coordinate system shown in Fig.2. This system has its origin at the cup center. Its horizontal axis is parallel to the line connecting the tuberosity points, its vertical axis parallel to the perpendicular to this line. The amount of wear is given by the length of the wear vector. This is the vector connecting the center of the head in the earlier image \((c^A_{x,y})\) with the one in the later image \((c^B_{x,y})\). It is termed positive if the head has moved in a superior, and negative if it has moved in an inferior direction. The direction of wear is defined by the (smaller) angle of the wear vector with the vertical coordinate axis. It is positive or negative in case of a medial or a lateral displacement, respectively.

Wear measurements using the Analysis-by-Synthesis algorithm are based on the CAD models of the prosthetic components. Using a computer vision algorithm, a synthetic x-ray projection of the model is matched to the image area of the head or the cup in the radiograph. The user loads an early and a later radiograph of the THR follow-up study as well as the CAD data of head and cup. He initiates the determination of the position of the implants by fitting the synthetic contour of each component to its silhouette in the radiographic image using the mouse (Fig.3). The definitive and precise position is then found by a 3D-2D registration algorithm running on the computer system as follows: The CAD model is virtually moved
within the x-ray unit. At each new position, the synthetic x-ray projection of the component is computed and compared with the original image. The final position is achieved when the difference between the synthetic and the original image is minimal.

The head's position is represented in a coordinate system which has its origin also at the cup center (see Fig.4). The directions of the axes of the coordinate system are, however, not defined by bony points but by the inclination angle of the cup. This angle indicated with $\psi$ is automatically determined by the registration algorithm. One coordinate axis runs along the major axis of the elliptical opening face of the cup, the other one along the minor axis. Wear is represented by the wear vector, i.e. the difference of the head's coordinates between the earlier radiograph ($v_{u,v}^A$) and the later one ($v_{u,v}^B$, see Fig.4). The two components of this vector are positive if the head has moved i) in a superior and ii) in a medial direction.

**Methods**

A conversion of the measurement results was necessary, because different coordinate systems are used in the two techniques. Compared to the coordinate system used in the Hip Analysis Suite, the coordinate system used in the Analysis-by-Synthesis method is rotated by the inclination angle of the cup $\psi$ (see the Figures 2 and 4). Hence, a rotation by $-\psi$ was performed, which means that the following transformation was applied to the vectors describing the position of the head:

$$
\begin{align*}
\hat{c}_x &= (\cos \psi, -\sin \psi) \cdot (-c_x, c_y), \\
\hat{c}_y &= (\sin \psi, \cos \psi) \cdot (-c_x, c_y)
\end{align*}
$$

The change of the sign of $c_u$ was necessary to invert the direction of the $u$-axis. The wear vector and the amount of wear in each image pair were then calculated like in the Hip Analysis Suite. For cross-validation, the inverse transformation was applied to the positional vectors in the Hip Analysis Suite.

To reduce the influence of the operator, the average of the three measurements performed by observer I, II, and III was calculated. The mean value ($\bar{M}$) of the 18 averaged wear measurements was taken as a measure of the accuracy by detecting a possible bias in the measurement, which should be zero according to the reference value (no wear).
The overall precision (in contrast to accuracy) of a measurement process is affected by the variability due to the instrument and by the multi-operator variability. The standard deviation (SD) of the 18 averaged wear measurements was taken as an estimate for the variability due to the instrument, mainly excluding the multi-operator variability by averaging over three measurements. A Wilcoxon Signed Rank Test was used to test if the erroneously measured (averaged) amounts of wear significantly differed between the two instruments.

To get insight into multi-operator variability, the variances of the three measurements performed by operator I, II, and III were calculated. The root of the mean of the variances gave a measure for the operator variability. A statistical comparison of this quantity in the two methods was achieved by inserting the just-mentioned variances in a second Wilcoxon Signed Rank Test.

The overall precision of the methods (instrument variability plus multi-operator variability) was estimated by summing up the squares of instrument variability and multi-operator variability. One third of the squared multi-operator variability was subtracted from the sum. This subtraction was necessary to eliminate the part of the operator error still contained in the estimated instrument precision, despite the averaging over the three measurements. The root of the resulting number multiplied by factor 1.96 gave the 95% precision limit of the two methods.
Results

The results are shown in Tab.1 and Tab.2. The components of the wear vector are indicated by $\Delta c_{x,y} = c_{x,y}^B - c_{x,y}^A$ or by $\Delta c_{u,v} = c_{u,v}^B - c_{u,v}^A$ if the coordinate system of the Hip Analysis Suite or the one of the Analysis-by-Synthesis method was used, respectively. The symbol $|\Delta c|$ describes the amount of wear according to the representation in the Hip Analysis Suite.

The bars (e.g.) indicate averaged measurements.

The mean values of the averaged wear measured using both methods are small compared to the standard deviations. This indicates a sufficient accuracy or “trueness” and means that a systematic bias can be excluded in both of the measuring techniques.

The standard deviation of the measured amount of wear using the Analysis-by-Synthesis algorithm was four times smaller when compared to the Hip Analysis Suite indicating a 4-fold improvement of the instrument precision. The Wilcoxon Signed Rank Test showed also that the erroneously measured amount of wear was significantly smaller using the Analysis-by-Synthesis algorithm ($p<0.001$).

The difference between the calculated standard deviations for the $\Delta c_u$- and the $\Delta c_v$-vector in the results of the Hip Analysis Suite has to be noted: The standard deviation in the $v$-direction was 3 times the standard deviation in the $u$-direction. The reason for this was probably that the cups did not have an exactly hemispherical shape. As the fitted circles did not match the oblate cup contour at the poles, their position in the direction of the cups’ rotational axis, which is in the direction of $v$, was not well defined. In contrast, in the Analysis-by-Synthesis method the standard deviations were isotropic as the CAD data matched the real shape of the cup.

In Tab. 2, the estimates of the operator variability of each technique are shown. Regarding the measured amount of wear, $|\Delta c|$, the operator variability using the Analysis-by-Synthesis method was by a factor 15 smaller than using the Hip Analysis Suite. The Wilcoxon Signed Rank Test confirmed the significance of the difference between the two methods ($p<0.001$).

The overall precision of the measuring process (instrument and multi-operator variability) assuming a 95% precision limit resulted $\pm 0.84$ mm using the Hip Analysis Suite and $\pm 0.15$
mm using the Analysis-by-Synthesis algorithm.
Discussion

No systematic bias was observed in of the two wear measuring methods which both provide wear estimates with sufficient accuracy when compared to the reference value. This is in line with the published in-vitro evaluations of both methods using a phantom hip wear model showing an accuracy of 10µm for the Analysis-by-Synthesis algorithm (7) and of 60µm for the Hip-Analysis Suite (8) used in the 2D mode. A validation of radiographic wear measuring techniques with direct measurements obtained from retrieved specimens exists for the Hip Analysis Suite only and showed excellent correlation and good agreement (9).

The overall precision of a measuring process is affected by instrument variability and multi-operator variability. According to the calculated standard deviations of the averaged measurements for both methods, measuring wear using the Analysis-by-Synthesis method resulted in a 4 times smaller instrument variability when compared to the Hip Analysis Suite. In our opinion, the superiority of the Analysis-by-Synthesis algorithm in terms of instrument precision was due to the fact that, using intensity based registration for the localization of the implant, all available geometric information was taken into account. The Hip Analysis Suite, however, relied only on the implants’ contours, which had to be extracted from the image. This limited the precision of the method, as only a part of the available information was used and errors occurred at the localization of the geometric features. Moreover, additional error was introduced, because the acetabular contour was fitted by a circle although the cups did not have an exactly hemispherical shape, which is, however, also the case in many other types of cups.

According to the mean variance of three measurements performed by 3 different operators, the Analysis-by-Synthesis algorithm resulted in a 15-times reduced multi-operator variability when compared to the Hip Analysis Suite. This was most likely mainly due to the higher level of automation in the Analysis by-Synthesis method. Using the Hip Analysis Suite, for example, the coordinates of the tuberosity points defined manually by the user entered directly the computation of the position of the femoral head. In the Analysis-by-Synthesis method, however, all interactively gained information served only as initial estimate and was further refined by the registration algorithm, reducing the dependency on the individual decision of the user. The more measurements have to be made manually the more the multi-
operator variability will increase.

We were aware of the possibility that an investigator may obtain better precision using a program which he has helped to develop than using a program he is unfamiliar with. We have minimized this effect by having one observer (observer II) who was not involved in the development of the program and not familiar with its use. Both, the clinician and the software engineer were instructed and trained on both programs by the first observer before starting the present study. In order to limit the error while using the Hip Analysis Suite, the instructions in the manual were followed exactly and the manual mode of initializing the localization of the cup was chosen to account for the non-spherical shape of the acetabular component. The small multi-operator variability of the Analysis-by-Synthesis method suggests that the investigator has a negligible influence on the result. On the other hand, the relatively great multi-operator variability in the Hip Analysis Suite, suggests that the overall precision could possibly be higher if the user was more familiar with this method.

The overall precision of the measuring process (instrument variability and multi-operator variability) can be expressed as a 95% precision limit of ± 0.84 mm using the Hip Analysis Suite and of ±0.15 mm using the Analysis-by-Synthesis algorithm. Therefore, real wear must have exceeded approximately 0.84 mm and 0.15 mm before it can be reliably detected using the two measuring processes, respectively. However, wear over years is usually calculated by applying linear regression to the data of wear versus duration of use and as averages on a group of patients rather than on a single individual. Therefore, as long as the error due to the overall precision of the method is random, i.e., as likely to be positive as negative, the lack of precision of a single measurement can be overcome by an adequate number of patients included in a retrospective study and wear can be measured with higher precision. The more precise is the measuring instrument used, the smaller needs to be the group of patients to draw conclusions. Thus, the main interest to have a more precise measuring process is that it would allow comparing wear rates of different materials or of different conditions between smaller groups of patients. For example, one can assume a difference in the wear rate of 0.05 mm/year between two implants and a standard variation of the wear rate within each group of ± 0.025 mm. Then, the expected difference (ED) in wear after 3 and 5 years would be 0.15
mm and 0.25mm, and the standard deviations (SD) of the wear rates within the groups would amount to 0.04mm and 0.06mm, respectively (obtained as 0.025·sqrt (3) and 0.025·sqrt(5)). Including the error bounds of the measuring process, these standard deviations become sqrt(0.04^2+(0.004/2)^2)mm = 0.420 mm and sqrt(0.06^2+(0.004/2)^2)mm = 0.421 mm using the Hip Analysis Suite. Using the Analysis-by-Synthesis algorithm, they become sqrt(0.04^2+(0.15/2)^2)mm = 0.086 mm and sqrt(0.06^2+(0.15/2)^2)mm = 0.093 mm. An approximate formula for the number of patients needed to be enrolled in each group to get a probability (i.e. a power) of 90% to detect a significant difference between both groups at the 5% level is given by 21 ·(SD/ED)^2. After three and five years, one can thus calculate that this number should be approximately 165 and 60 using the Hip Analysis Suite, and 7 and 3 using the Analysis-by-Synthesis algorithm, respectively.

Several limitations of the Analysis-by-Synthesis algorithm must be considered. First, a possible source of error may be given by the machining tolerance of the implants which can reach up to a deviation of ±0.5mm between the CAD data and the real prosthetic shape. This deviation is suspected to result in a systematic error in the range of 1µm (12). It is, however, eliminated by the calculation of the wear, i.e. the difference of the head coordinates in two images. Hence, no significant impact could be observed, neither in vitro (12) nor in this study. Second, the fact that linear PE wear is measured from relative movements of the femoral head against the metal back of the cup is a limitation of the presented method. Any further seating of the PE inlay within the taper of the metal back are thus misinterpreted as PE wear. This is a limitation of all in vivo wear measurements made on pelvic x-rays with the exception of RSA, measuring the displacement of the femoral head relative to a group of tantalum beads embedded in the rim of the polyethylene insert. A comparison of femoral head penetration using RSA and the Hip Analysis Suite by Bragdon et al.(10) showed that the Hip Analysis Suite overestimated the total penetration by about 0.05mm due to this settling of the insert into the shell. However, when this bedding-in period was excluded by measuring the median steady state wear between the 2 and 5 year x-rays no difference in the true wear rate between the two methods was identified. Third, theoretical positional changes of the ball within the cup between the two x-rays acquired the same day, while undetected by the
Analysis-by-Synthesis algorithm, could have been misinterpreted as error when detected by
the Hip Analysis Suite. However, such a misinterpretation is very improbable considering the
higher accuracy and smaller instrument variability of the Analysis-by-Synthesis algorithm.
Fourth, the Analysis-by-Synthesis algorithm was performed in the frontal plane only. Head
displacement out of the plane of the radiograph was therefore not taken into account. As both
methods were run in a 2D mode in the study, their performance was limited in the same way.
Theoretically, the Analysis-by-Synthesis method can be arranged for 3D measurement as it is
already the case for the Hip Analysis Suite. We assume that the technical improvements
leading to a higher precision in the frontal plane have the same effects in the axial plane.

Despite these limitations we conclude that due to the availability of the full CAD data of the
implants and to further automation of the measuring process by intensity based registration,
the Analysis-by-Synthesis algorithm allows PE wear to be measured with a significantly
higher precision. This enables differences in wear rates to be reliably detected in a clinical
setting after a shorter time period and with much smaller patient collectives.
Acknowledgment

The authors thank the company Zimmer® for providing the CAD data of the hip implant components used in this study.
References


7) Burckhardt K, Dora C, Gerber C, Hodler J, Székely G. Measuring orthopaedic implant wear on standard radiographs with a precision in the \(10\ \mu\text{m}\)-range. Medical Image Analysis 2006; 10(4): 520-529.


Tables

Table 1: The instrument error of the two techniques, i.e. the mean (M) and the standard deviation (SD) of the averaged wear measurements using Hip Analysis Suite and Analysis-by-Synthesis.

<table>
<thead>
<tr>
<th></th>
<th>Hip Analysis Suite</th>
<th>Analysis-by-Synthesis method</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M(\Delta c_x, \Delta c_y)$ [mm]</td>
<td>-0.089, -0.094</td>
<td>-0.013, -0.005</td>
</tr>
<tr>
<td>$M(\Delta c)$ [mm]</td>
<td>-0.143</td>
<td>-0.011</td>
</tr>
<tr>
<td>$M(\Delta c_u, \Delta c_v)$ [mm]</td>
<td>-0.006, -0.126</td>
<td>0.001, -0.011</td>
</tr>
<tr>
<td>$SD(\Delta c_x, \Delta c_y)$ [mm]</td>
<td>0.213, 0.204</td>
<td>0.043, 0.060</td>
</tr>
<tr>
<td>$SD(\Delta c)$ [mm]</td>
<td>0.286</td>
<td>0.073</td>
</tr>
<tr>
<td>$SD(\Delta c_u, \Delta c_v)$ [mm]</td>
<td>0.090, 0.284</td>
<td>0.049, 0.051</td>
</tr>
</tbody>
</table>

Table 2: The inter-observer variability of the two techniques, i.e. the roots of the mean variance (VAR) of the three measurements per image pair.

<table>
<thead>
<tr>
<th></th>
<th>Hip Analysis Suite</th>
<th>Analysis-by-Synthesis method</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sqrt{M(\text{VAR}(\Delta c_x^{I,III}, \Delta c_y^{I,III})))}$ [mm]</td>
<td>0.203, 0.268</td>
<td>0.012, 0.015</td>
</tr>
<tr>
<td>$\sqrt{M(\text{VAR}(\Delta c^{I,III})))}$ [mm]</td>
<td>0.339</td>
<td>0.023</td>
</tr>
<tr>
<td>$\sqrt{M(\text{VAR}(\Delta c_u^{I,III}, \Delta c_v^{I,III})))}$ [mm]</td>
<td>0.087, 0.324</td>
<td>0.011, 0.016</td>
</tr>
</tbody>
</table>
Figure legends

Fig.1:
User interface in the Hip Analysis Suite. The contour lines show the final positions of the head and the cup. The vector shows the direction of wear.

Fig.2:
The coordinate system used in Hip Analysis Suite.

Fig.3:
User interface in the Analysis-by-Synthesis method. On the right, the contours of head and cup as fitted by the user are shown.

Fig.4:
The coordinate system used in the Analysis-by-Synthesis method.
Fig. 1: The user interface of Hip Analysis Suite. The contour lines show the final position of head and cup, and the vector the direction of the wear.
Fig. 2: The coordinate system used in Hip Analysis Suite.
Fig. 3: The user interface in the Analysis-by-Synthesis method. On the right, the contours of head and cup fitted by the user are shown.
Fig. 4: The coordinate system used in the Analysis-by-Synthesis method.