The handle http://hdl.handle.net/1887/42759 holds various files of this Leiden University dissertation.

**Author:** IJsseldijk, E.A. van  
**Title:** Model-based wear measurements in total knee arthroplasty: development and validation of novel radiographic techniques  
**Issue Date:** 2016-09-01
A model-based approach to measure the minimum joint space width of TKA in standard radiographs
A model-based approach to measure the minimum joint space width of total knee arthroplasty in standard radiographs

E.A. van IJsseldijk\textsuperscript{a}, E.R. Valstar\textsuperscript{a,b}, B.C. Stoel\textsuperscript{c}, R.G.H.H. Nelissen\textsuperscript{a}, B.L. Kaptein\textsuperscript{a}

\textsuperscript{a} Biomechanics and Imaging Group, Department of Orthopaedics, Leiden University Medical Center, PO Box 9600, 2300 RC Leiden, The Netherlands

\textsuperscript{b} Department of Biomechanical Engineering, Faculty of Mechanical, Maritime, and Materials Engineering, Delft University of Technology, Delft, The Netherlands

\textsuperscript{c} Division of Image Processing, Department of Radiology, Leiden University Medical Center, Leiden, The Netherlands

Published as "A model-based approach to measure the minimum joint space width of total knee replacements in standard radiographs", Journal of biomechanics, 2012, 45(12), pages 2171-2175
Abstract

Introduction
Excessive wear is in total knee arthroplasty is detected by measuring the minimum joint space width (mJSW) in anterioposterior radiographs. The accuracy of conventional measurement methods is limited and can be improved using model-based techniques. In this study, the model-based wear measurement (MBWM) is introduced. Its accuracy and reproducibility are assessed and compared to the conventional measurement.

Method
40 anterioposterior radiographs were obtained of a knee prosthesis using a phantom set-up. Both measurement methods were applied and the accuracy and precision were compared. The reproducibility was calculated with an inter- and intra-observer experiment. Three observers measured the mJSW in 30 clinical radiographs with both the conventional measurement and the MBWM and repeated this after 6 weeks. The experiments were conducted with a NexGen mobile bearing and fixed bearing prostheses.

Results
In the phantom experiment, the accuracy (mean of the absolute error) was significantly higher (t-test, $p < 0.01$) for the MBWM as for the conventional measurement (0.15 mm versus 0.43 mm, 0.14 mm versus 0.35 mm for the mobile and fixed bearing respectively). The standard deviation of the measurements is smallest for the MBWM measurement for both prosthesis types (0.16 mm versus 0.47 mm, Levene’s test, $p < 0.01$). In the reproducibility experiment, both the intra- and inter-observer agreements was higher for the MBWM than for the conventional method.

Conclusions
The results show that the MBWM is superior to the conventional measurement in both accuracy and reproducibility. Although the use of a phantom experiment poses some limitations in conveying the findings to clinical practice, this improved mJSW measurement can lead to better wear detection for surgery decisions and research purposes.
3-1 Introduction

Excessive polyethylene wear is an important cause of implant failure in total knee arthroplasty (TKA)\cite{16, 22}. As the incidence of total knee arthroplasty is increasing, the impact of wear problems is expected to increase as well\cite{54}.

In current clinical practice, polyethylene wear is determined in vivo using the minimum joint space width (mJSW), which is assessed in standard radiographs. This diagnostic tool is used to evaluate new prosthesis designs and for decision support for surgical procedures such as isolated polyethylene exchanges\cite{55-57}. The mJSW is obtained in anterioposterior (AP) or mediolateral radiographs\cite{27, 28}. However, the accuracy is limited and measurement errors higher than 1 mm are not exceptional\cite{28}.

The measurement accuracy and precision can be improved by model-based techniques. In our previous work, we described and validated a wear measurement method for model-based roentgen stereophotogrammetric analysis (MBRSA), in which the tibia-femoral distance is obtained based on 3D surface models of the components\cite{58}, using 3D vision techniques\cite{59}.

This approach can be used for standard radiographs as well. The accuracy of the measurement will be lower than the accuracy found in MBRSA, as accurate calibration is not possible and model matching is done with a single X-ray source only. Nonetheless, we hypothesize that the generally applicable, model-based approach will be more accurate and reproducible compared to conventional methods, as more image information is exploited and less dependency is expected to joint space narrowing caused by anterioposterior tilt of the tibial baseplate.

The goal of this study is to evaluate the accuracy and reproducibility of the model-based wear measurements in AP radiographs (MBWM) and compare the results to the conventional measurement method. To determine the accuracy, a phantom set-up was constructed with a known mJSW, in which the measurements were conducted for different positions of the phantom, insert sizes and prosthesis types. The reproducibility was determined using inter- and intra-observer studies with clinical data.
3-2 Materials and Methods

3-2-1 Measurement Methods
We now describe the metal-to-middle (conventional) and model-based measurements. Both methods determine the mJSW as the shortest distance between the tibial tray and the femoral condyles. The first method uses the visible distance in the image itself, whereas the second uses a semi-automatic measurement based on 3-D models that are matched with the image.

Metal-to-middle measurement
The metal-to-middle measurement is the standard method in obtaining the mJSW in the image [27, 28]. A reference line is drawn through the tibial tray at its largest medialateral width. Then, the shortest perpendicular distances are estimated between this reference and the femoral condyles (Figure 3-6 left).

In our experiments, the metal-to-middle method was conducted using a computer software (Digimizer® version 4.0.0.0, MedCalc Software, Mariakerke, Belgium). The image magnification was corrected using the ratio of the known width of the tibial baseplate to the width in the radiograph.

Model-based wear measurement
In the model-based wear measurement (MBWM) method, 2-D/3-D registration is used to match 3-D surface models of the tibial and femur components with the AP radiograph. Then, the minimal medial and lateral distances are automatically measured based on the models (Figure 3-6 right).

The 2-D/3-D registration was conducted in model-based RSA (Version 3.3, Medis Specials, Leiden, The Netherlands). The origin of the laboratory frame was located in the center of the image detector. The x and y coordinates thereby describe the image plane, whereas the z coordinate is the direction perpendicular to the image. It was assumed that the position of the X-ray source was located on the z-axis (e.g. perpendicular to the image center). The DICOM information was used to set the distance between the X-ray source and the detector and the physical pixel spacing of the detector.
The image contours of the tibial and femoral components were selected semi-automatically with canny edge detector [60]. The position and orientation of the models were calculated by minimizing the difference between the image contours and the projected model silhouettes.

![Figure 3-6. Comparison of the measurement methods: (left) the metal-to-middle method, in which the mJSW is obtained in the radiograph, and (right) the MBWM in which the mJSW is obtained semi-automatically based on 3D models](image)

**3-2-2 Experiments**

**Phantom experiment**

The phantom setup consisted of the tibial and femoral components of the knee prostheses, which were inserted into sawbones (Figure 3-7). The setup was placed in an X-ray imaging system (CXDI-series, 169dpi, 12BPP, Canon, New York, USA), according to the anterior-posterior (AP) protocol in standing position as used in our hospital. The X-ray source was positioned 1.2 meters from the detectors and the phantom was positioned approximately 20 cm from the detector.

The actual mJSW was set using radiolucent plates (Plexiglass/PMMA), which had an accurately defined thickness (tolerance 0.05mm). Four different sizes (5, 8, 10
A model-based approach to measure the mJSW of total knee replacements arthroplasty in standard radiographs

and 12 mm) were used and the appropriate plate was placed between the tibial tray and the medial femoral condyle during the acquisition. As contact was possible only for the medial condyle, the lateral mJSW was not measured in the experiment.

For each plate size 10 images were acquired. Among these images, the position (range -10 to 10 cm) and orientation (range -10° to 10°) of the phantom with respect to the image were varied. In addition, the setup was placed in different anterior tilt angles (range 0° to 10°), as illustrated in Figure 3-7.

We repeated the acquisitions for both the fixed bearing and mobile bearing NexGen (Zimmer, Warsaw, IN, USA) total knee prostheses to cover different geometric designs. The size of the fixed and mobile bearings were (5-F) and (4-D), respectively (tibia-femur). Computer aided design (CAD) models were available for all components, except for the tibia component of the fixed prosthesis. For this component a reversed engineered model was created with a 3-D laser scan (Hyscan, Hymarc Tech, Ottawa, Canada), which had a tolerance of 0.020 mm.

In total, 80 images were acquired (10 images x 4 plate sizes x 2 prosthesis types). For all images, the mJSW was calculated with the MBWM and conventional measurement.
method, which was obtained by a clinician. Subsequently, the errors were calculated as the difference with the actual mJSW defined by the plate thickness.

**Statistical analysis**
We calculated the error mean, absolute error mean, standard deviation of the error and error range per prosthesis type and measurement method. The sizes of the errors are tested for statistical significance with unpaired \( t \)-tests. Levene’s tests are used to test for differences in variance between the errors of the measurement methods. Finally, the dependency between the error and actual size was determined using a regression analysis (Pearson’s rho).

**Clinical experiment**
In this experiment, a comparison was made between the inter- and intra-observer variability of the conventional wear measurement and MBWM. Clinical data was used as no ground truth value is required to obtain this measure.

For both the mobile and the fixed bearing prosthesis, 15 bearing AP radiographs were retrieved from the hospital database, in a random order. Both bilateral and unilateral images were included.

Three observers were included in the experiment: a clinician, a researcher and a senior researcher. They were asked to measure the medial and lateral insert thickness in the radiographs using both the conventional and model-based measurement methods. Observers could practice until they felt comfortable with the methods, preventing learning curve effects. To obtain the intra-observer variability, the observers repeated the measurements after a period of at least 6 weeks. In this series, the average measurement duration of the model-based method was also recorded.

**Statistical analysis**
The inter- and intra-observer variability of each measurement method was analyzed with the interclass correlation coefficient (ICC, two-way random, absolute agreement). The difference in spread between the measurements methods was tested with Pitman’s test for correlated measures[61]. Bland-Altman plots were created to detect possible trends in the data[62].
3-3 Results

3-3-1 Phantom Experiment (accuracy)

The measurement errors (i.e. the difference between the measured thickness and the actual thickness of the plate) for each measurement method and prosthesis type are shown in box plots (Figure 3-8). The statistical characteristics of the errors are shown in Table 3-4, split for the measurement method and prosthesis type.

![Figure 3-8. Boxplots showing the measurement errors of the different methods and protheses types](image)

Table 3-4. Statistics over the measurement errors in the phantom experiment.

<table>
<thead>
<tr>
<th></th>
<th>Mobile (N=40)</th>
<th>Fixed (N=40)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>conv</td>
<td>MBWM</td>
</tr>
<tr>
<td>mean (mm)</td>
<td>-0.36*</td>
<td>0.15*</td>
</tr>
<tr>
<td>standard deviation (mm)</td>
<td>0.40</td>
<td>0.06†</td>
</tr>
<tr>
<td>absolute error mean (mm)</td>
<td>0.43</td>
<td>0.15**</td>
</tr>
<tr>
<td>range (mm)</td>
<td>1.90</td>
<td>0.21</td>
</tr>
</tbody>
</table>

conv = conventional measurement; MBWM = model-based wear measurement

* statistically significant difference from 0 mm (t-test, p < 0.05)
† statistically significant difference in variance compared to the conventional measurement (Levene's test, p < 0.01)
** statistically significant difference in mean compared to the conventional measurement (t-test, p < 0.01)
For both prosthesis types, the standard deviation of the measurements is significantly smaller for the model-based measurement than for the conventional measurement (Levene’s test, $p < 0.01$). Also, the model-based measurement had a significantly smaller standard deviation for the mobile prosthesis than for fixed prosthesis (Levene’s test, $p < 0.01$). This is probably due to the implant geometry. The fixed prosthesis type contains thin structures such as the metal rim. These structures produce less pronounced image edges, increasing the localization error.

The average of the measurement error indicates whether a systematic bias (information bias) is present. The data from Table 3-4 shows that only the model-based measurement is unbiased for the fixed prosthesis. For the mobile prosthesis, the model-based measurement shows the smallest bias of the two methods. For both prosthesis types, the absolute error is lower for the MBWM than for the conventional measurement ($p < 0.01$).

No statistically significant correlation was found between the measurement error of the model-based measurement and the true distance for both prosthesis types (Pearson’s rho, $p < 0.05$).

3-3-2 Clinical experiment (reproducibility)

The ICC values found in the inter- and intra-observer study were higher for the model-based method than for the conventional method for any observer and prosthesis type (Table 3-5), indicating a better reproducibility for the first method. On average, the reproducibility was higher for the fixed bearing than for the mobile bearing.

The average measurement durations of the observers in the second measurement series were 1:37, 2:24 and 2:37 (min:sec).

The standard deviations over all measurements with the conventional method and model-based method are 0.37 mm and 0.15 mm respectively. Pitman’s test flagged the difference in spread significant ($p < 0.05$) for two out of three observers.
Table 3-5. Results of the inter and intra-observer variability of the measurement methods for the conventional (conv) and model-based method (mb) in terms of the interclass correlation coefficient.

<table>
<thead>
<tr>
<th></th>
<th>Intra-observer variability</th>
<th>Interobserver variability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observer 1</td>
<td>Observer 2</td>
</tr>
<tr>
<td>ICC values</td>
<td>conv MBWM</td>
<td>conv MBWM</td>
</tr>
<tr>
<td>mobile prosthesis</td>
<td>.945 .963</td>
<td>.926 .983</td>
</tr>
<tr>
<td>fixed prosthesis</td>
<td>.963 .986</td>
<td>.973 .982</td>
</tr>
</tbody>
</table>

The Bland-Altman plot gives the agreement between the measurements, by plotting the difference between the measurement methods (MBWM - conventional) against the mean value (Figure 3-9). Only the first measurement series was used and the mean value over the three observers was used, reducing the data to 60 points. This reduction keeps the plot legible and – more importantly - prevents oversampling, because the observer data contains a high dependency.

![Bland-Altman plot](image)

Figure 3-9. Bland-Altman plot showing the measurement agreement between the two methods.

The limit of agreement between the measurements is 1.27 mm (2 x SD), which is indicated with the broken lines. For the fixed prosthesis, a statistically significant difference of 0.23 mm was found, i.e. the MBWM gives a larger value than the conventional measurement on average (t-test, p = 0.008).
3-4 Discussion

We developed a model based wear measurement (MBWM) that is superior in obtaining the mJSW in comparison with the method currently used in clinical practice. The main advantage is the improvement in precision and reproducibility that was obtained. A higher precision was found in the phantom experiment and a higher inter- and intra-observer reproducibility was found with clinical data. As these experiments were conducted with both a fixed and a mobile prosthesis design, we expect that these findings are generally applicable to other prosthesis designs as well.

Furthermore, we found a lower bias for the MBWM than for the conventional measurement in the phantom experiment. However, the high variability of the conventional measurements makes generalization of this result difficult. Furthermore, bias is of lesser importance than precision, because bias can nullify in relative measurements such as wear-rate measurements.

The average measurement duration of the model-based measurement was approximately 2 minutes, which is adequate for clinical use. We expect that this duration can be decreased by further automation of the contour detection, as the implant shadows are clearly distinguishable in AP radiographs.

The phantom experiment had several limitations. It did not include the soft tissue attenuation that is present in real clinical images. Still, the attenuation is usually limited and the pose estimator remains robust when only 5% of the complete contour is used [63]. Another limitation was that the geometric design of the fixed bearing tibia component in the phantom was different from the design in clinical images, due to this availability. This could have influenced the conventional measurement, because of differences in the metal rim surrounding the tibia plateau. This could explain why the mean difference between the measurements differs for this prosthesis, when the phantom experiment and clinical experiment are compared (-0.23 mm vs. 0.12 mm).

In clinical practice, the outcomes of the mJSW measurements (both conventional and model-based) depend on the articulating points of the femoral component at the moment of X-ray acquisition. Unfortunately, it was not possible to include this
effect in the phantom experiment. Although we expect that congruent liners limit the variability of the articulating point, we hope to eliminate any uncertainty with a retrieval study, in which clinical radiographs are compared directly to retrieved liners. A general limitation of radiograph-based distance measurement is that creep and true wear cannot be distinguished. Instead, it is assumed that creep stabilizes within two months, whereas wear is expected to be a constant process over time[52].

Several other studies describe alternatives to the radiographic wear measurement. Some studies use fluoroscopy to improve the reproducibility, as the alignment between the tibial tray and the radiographic beam can be optimized before the measurement[40, 64]. A standard deviation of 0.15\,\text{mm} was found in this measurement, which is similar to the finding in our work, yet fluoroscopy generally comes with a higher radiation dose for the patient and requires a longer imaging time.

In other researches, a similar model-based wear measurement for calibrated single-source radiographs is described[41, 65]. Although the validation data is limited, this indeed seems to give better results (SD = 0.1\,\text{mm}). However, this method imposes the presence of a calibration object. We think that the applicability to standard radiographs is a considerable advantage of the method we are using.

Based on these results, we conclude that the model-based method is a reliable tool to evaluate the insert thickness in standard radiographs. It can therefore aid in a better timing of insert exchanges, with the aim of decreasing the number of complications. Moreover, the accuracy of the method combined with the advantage that any standard radiograph can be used renders the method interesting for wear studies to compare prosthesis types.

REFERENCES


[27] Miller, T.T., Imaging of knee arthroplasty. European Journal of Radiology,


