Journal of Biomechanics 46 (2013) 1387–1391

Contents lists available at SciVerse ScienceDirect

Journal of Biomechanics

journal homepage: www.elsevier.com/locate/jbiomech
www.JBiomech.com

Short communication

Validation of the in vivo volumetric wear measurement for total knee prostheses in model-based RSA

E.A. van IJsseldijk [a,*], B. Lebel [b], B.C. Stoel [c], E.R. Valstar [a,d], S. Gouzy [b], C. Vielpeau [b], B.L. Kaptein [a]

[a] Biomechanics and Imaging Group, Department of Orthopaedics, Leiden University Medical Center, The Netherlands
[b] Department of Trauma and Orthopaedic Surgery, Centrale Hospital Universitaire Caen, France
[c] Department of Radiology, Leiden University Medical Center, The Netherlands
[d] Department of Biomechanical Engineering, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Delft, The Netherlands

ARTICLE INFO

Article history:
Accepted 26 February 2013

Keywords:
Total knee arthroplasty
Polyethylene wear
Image-based measurements
Model-based RSA

ABSTRACT

Implant failure related to polyethylene wear remains an important issue in total knee arthroplasty. Polyethylene wear is usually assessed in vivo by measuring the remaining insert thickness on X-ray images of the knee. To reflect the amount of wear debris more accurately, a 3-dimensional overlap measurement has been suggested, which is based on implant component models which are matched on calibrated stereo X-ray images using model-based roentgen stereophotogrammatic analysis. The goal of this study was to determine the influence of pose estimation, insert thickness deviation and variation in the femoral-tibial contact location on the accuracy and precision of the measurement using simulations and a phantom experiment.

We found that the pose estimation was the largest source of variation. The 95% prediction interval varied between 111 and 283 mm^3, which is approximately 100–200% of the detected volumetric wear. Insert thickness variation resulted in prediction intervals of 74–174 mm^3. Variation of the femoral-tibial contact location in the phantom experiment gave a prediction interval of 40 mm^3. Large differences in the detected wear volume were found for different flexion angles. At most 56% of the true wear volume was detected (129 of 230 mm^3, 30° of flexion).

In summary, both the accuracy and precision of the volumetric wear measurement were low. The prediction interval of the volumetric wear measurement is at least as large as the measurement outcome itself. This is an important limitation to the applicability of the volumetric wear measurement in clinical practice and further clinical validation is required.

1. Introduction

Polyethylene (PE) wear is an important cause of implant failure of total knee arthroplasty (TKA), as it can lead to instability and aseptic loosening (Naudie et al., 2007; Sharkey et al., 2002; Sundfeldt et al., 2006). Therefore, an accurate and precise method is required to assess the in vivo progression of PE wear in vivo, which can be used to predict instability and loosening so as to initiate a timely intervention.

The current method to assess the progression of PE wear in vivo is measuring the minimum distance between the femoral condyles and the tibial plateau using radiographic and fluoroscopic imaging (Collier et al., 2003; Duryea et al., 2001; Miller, 2005; Sanzén et al., 1996; van IJsseldijk et al., 2012). However, this 2-dimensional measurement does not reflect the total volume of wear debris that has been released. Therefore, Gill et al. (2006) presented a method to measure the in vivo wear volume using 3-dimensional (3-D) geometric models of the implant components, by estimating their 3-D poses (positions and orientations) from stereo X-ray images and calculating the overlap volume with the insert.

For the most part the accuracy and precision of this measurement method have not been validated. The goal of this study was to determine the influence of important sources of variation on the accuracy and precision of the volumetric wear measurement. Amongst others, these depend on the 3-D pose estimation and deviations in the original insert thickness as a result of the manufacturing process. Simulation studies were conducted in which the isolated influences of these sources on the measurement were determined.

In practice, wear is often caused by the sliding motion of the femoral component relative to the insert. Therefore, the accuracy

* Corresponding author. Biomechanics and Imaging Group, Department of Orthopaedics, Leiden University Medical Center, Room J9-127, P.O. Box 9600, 2300 RC Leiden, The Netherlands. Tel.: +31 71 5265358; fax: +31 71 5266743.
E-mail address: e.a.van_ijsseldijk@lumc.nl (E.A. van IJsseldijk).

0021-9290/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jbiomech.2013.02.021

and precision of the measurement will also relate to the flexion angle at which the measurement is conducted and the variation in the femoral contact location on the insert. A phantom experiment was done to determine the influence of these sources, using inserts with abrasive wear.

2. Materials and methods

The volumetric wear measurement was conducted based on image pairs that were acquired using a röntgen stereophotogrammetric analysis (RSA) setup with the calibration box in vertical orientation (Kaptein et al., 2003). The image pairs were analyzed with model-based RSA software (v3.32, Medis Specials, Leiden, The Netherlands) to estimate the poses of the prosthesis components, which are described with triangulated surface models (Kaptein et al., 2003). Since the insert component does not produce clear image contours, its pose was derived from the pose of the tibia model, as they have a fixed spatial relationship.

Volumetric wear was detected by calculating the 3-D overlap region between the femoral and insert component models. A regular 2-D grid was defined (0.8×0.8 mm cell size) that coincided with the tibial plateau. For each grid point the overlap distance between the femoral component's surface and the insert surface was calculated. The wear volume was computed using a numerical integration of these distance values based on Simpson's rule.

3. Simulation experiments

The influences of pose estimation and insert thickness deviations were determined in simulation experiments. We calculated the difference in the detected volumetric wear as a function of the relative pose of the femoral component with respect to the tibial component. This pose is expressed as $p=(x, y, z, \alpha, \beta, \gamma)^T$, where x, y, and z are the medial, caudal and anterior position parameters and α, β and γ are the corresponding Euler angles (Fig. 1).

The experiments were repeated with eight initial poses $p_{0,j}$ ($j=1\ldots8$), which were obtained from eight RSA data of patients with size 4 Triathlon PS total knee prostheses (Stryker Europe, Raheen, Ireland).

The effect of pose estimation error was computed in a Monte Carlo Simulation. For each initial pose the detected volumetric wear $w_{0,j}$ was calculated and 500 new poses were generated as $p_{i,j}=p_{0,j}+d_{i,j}$. The pose errors $d_{i,j}=(d_{xi,j}, d_{yi,j}, d_{zi,j}, d_{\alpha i,j}, d_{\beta i,j}, d_{\gamma i,j})^T$ were drawn from a normal distribution with zero mean and a standard deviation (SD) of (0.085 mm, 0.085 mm, 0.22 mm, 0.343°, 0.414°, 0.23°)^T. These SDs were derived from a clinical validation study (Kaptein et al., 2007). For each pose the detected wear volume $w_{i,j}$ and measurement error $e_{i,j}=w_{i,j}-w_{0,j}$ were calculated.

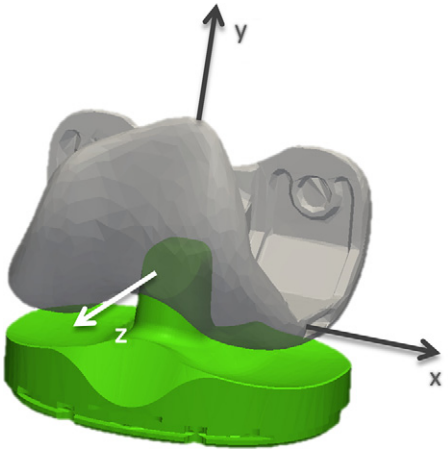

Fig. 1. The coordinate system that was used in the simulation study.

The variation in insert thickness was simulated by varying the caudal position parameter of the relative pose with Δd, resulting in $p_j=p_{0,j}+(0, \Delta d, 0, 0, 0, 0)^T$. The parameter was varied between +0.12 mm and −0.12 mm, which is in the range of the 95% prediction interval assuming that the thickness among insert components of the same type and size vary with an SD of 0.06 mm (Collier et al., 2003; Edwards et al., 2002; Psychoyios et al., 1998).

4. Phantom experiment

The phantom experiment was conducted to assess the influence of variation in the femoral-tibial contact location and the knee angle to the volumetric wear measurement. We used a knee prosthesis (size 4 Triathlon PS) with inserts containing a predefined wear pool and determined how accurately these wear pools could be reconstructed by the volumetric wear measurement.

The wear in the inserts was designed in SolidWorks CAD software (Dassault Systemes, Paris, France). A femoral component model (size 5 Triathlon PS) was placed in bearing contact with the insert model and subsequently moved downward (into the insert). This produced a 3-D overlap volume between the models, which was removed from the insert model. Different sizes and shapes of the wear pool were created ($N=6$) by varying the flexion angle of the femoral component and the distance over which it was moved into the insert. We used a larger size femur component to simulate wear caused by the sliding motion of the femoral component. The physical insert was manufactured by a computer controlled milling device (Stryker Europe, Raheen, Ireland). We selected an insert for which the femoral-tibial contact location was consistently found inside the wear pool in the volumetric wear measurement (see Fig. 2). The data of all other inserts is presented in Appendix A.

A total knee prosthesis was assembled with the selected insert placed in the tibial component. For analysis and pose estimation 3-D scans of the insert, femoral and tibial components were generated by means of reversed engineering (Introtech, Nuenen, the Netherlands). Based on the insert scan, the shape and volume of the true (predefined) wear pool were determined.

This especially prepared prosthesis was fixed into sawbones. The tibia sawbone was placed in a vertical position on a tripod. The femur sawbone could be positioned on top of the tibia in any flexion angle, as a 7 kg balancing weight was used to stabilize the set-up (see Fig. 3). The sawbones were placed in a horizontally oriented RSA imaging setup. Five consecutive RSA image pairs were obtained for three flexion angles (0°, 30° and 60°) resulting in 15 image pairs totally. Before obtaining each of these image pairs, the femoral component was remounted in such a position that the predefined flexion angle was set (verified by a

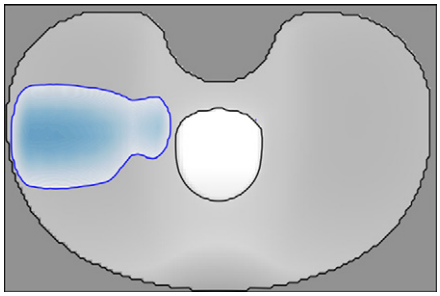

Fig. 2. Illustration of the predefined wear pool (size=230 mm^3). The shading intensity of the blue area corresponds to the depth of the wear pool with respect to the insert surface. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

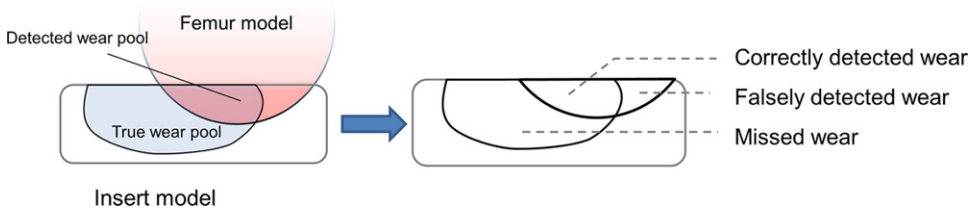

Fig. 4. Schematic cross-section of an insert with wear and the femoral-insert overlap measurement. This figure shows how correctly detected wear, falsely detected wear and missed wear are defined.

goniometer) and its contact location resided inside the wear pool. By this operation, variation in the femoral-tibial contact location was introduced.

For each RSA image pair, the volumetric wear was assessed and the detected wear pool was compared to the true wear pool, defining both the correctly and falsely detected wear (Fig. 4). The part of the true wear pool that was not detected was defined as missed wear. The volumes of these quantities were calculated and the means and SDs over the flexion angles were compared.

5. Results

5.1. Simulation experiments

Table 1 shows the results related to the pose estimation error and the insert thickness variation. For the pose estimation error, the mean wear was slightly larger than w_0 (8 mm^3, $p=0.001$, paired samples t test). This difference is caused by the non-linear relation between the wear volume detected and the y-position. The sizes of the prediction intervals (PI) ranged between 111 and 283 mm^3 and were positively and significantly correlated with w_0 (Pearson's $\rho=0.96$).

The effect of varying the thickness of the insert (Δd) on the detected wear volume can be seen in Fig. 5. Their relation is not entirely linear, as the size of the slope (measurement error as a function of Δd) declined for increasing Δd. The 95% PIs ranged between 74 and 174 mm (Table 1).

5.2. Phantom experiment

The bar plot in Fig. 6 presents the correctly and falsely detected wear volumes for the 15 image pairs. Below the figure, typical examples of these wear pools are shown per flexion angle. As a reference, the leftmost bar shows the volume of the true wear pool.

A comparison of the results per flexion angle is shown in Table 2. The mean of both the correctly and falsely detected wear volumes showed a significant difference between the flexion angles ($p<0.05$, one-way ANOVA). The mean detected volume for the flexion angle that was used to generate the wear pool (30°) was only 50% of the true wear volume.

In all cases, the volume of falsely detected wear was small (<15 mm^3) compared to the true wear volume (230 mm^3). At a flexion angle of 30°, no falsely detected wear was found for all RSA image pairs.

The standard deviation at 0°, 30° and 60° of knee flexion were 8 mm^3, 7 mm^3 and 18 mm^3, respectively. The corresponding 95% prediction intervals, which are a measure for influence of variation in the femoral positioning, ranged between 12% and 33% of the volume of the true wear pool.

Table 1
Results of the simulations. w_0 is the wear volume corresponding to the initial pose p_0. For the pose estimation error the mean and 95% prediction intervals of the 500 detected wear volumes $w_{i,j}$ are presented. For the insert thickness variation, the 95% prediction intervals of the wear volume are shown, which are defined as the wear volume measured after adding ± 0.12 mm to the y-position of the relative pose. The size and relative size of the PIs with respect to the original wear are presented.

Original wear (w_0)	**Pose estimation error**				**Insert thickness variation**		
mm^3	Mean wear mm^3	95% PI mm^3	PI size mm^3	size/ w_0 –	95% PI mm^3	PI size mm^3	size/ w_0 –
51	60	(11–122)	111	(2.18)	(21–95)	74	(1.45)
67	78	(21–154)	133	(1.99)	(29–121)	92	(1.37)
84	93	(30–178)	148	(1.76)	(42–138)	96	(1.14)
125	139	(60–235)	175	(1.40)	(74–189)	15	(0.92)
157	166	(64–282)	218	(1.39)	(95–234)	139	(0.89)
163	171	(87–273)	186	(1.14)	(114–225)	111	(0.68)
172	181	(93–290)	197	(1.15)	(110–246)	136	(0.79)
304	310	(185–468)	283	(0.93)	(222–396)	174	(0.57)

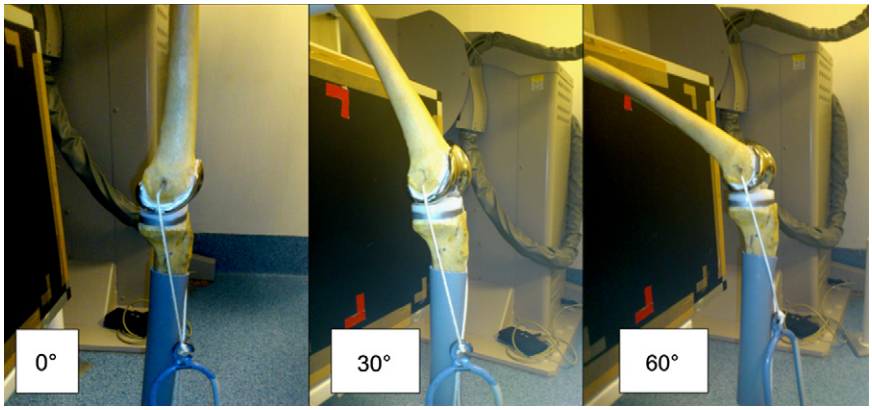

Fig. 3. The set-up of the phantom experiment during the image acquisition. For each angle of knee flexion an image is shown.

6. Discussion

We investigated the influence of the pose estimation error, insert thickness deviation and variation of the femoral-tibial contact location on the accuracy and precision of the volumetric wear measurement We found that pose estimation was the largest source of variation, producing a variation between 111 and 283 mm^3 (95% prediction interval). This equaled 100–200% for smaller wear pools relative to the detected wear volume.

The 95% prediction interval due to insert thickness deviation was between 74 and 174 mm^3, which was smaller than the effect of pose estimation error. An important difference between these error sources is that pose estimation error influences each measurement randomly, whereas the error due to insert thickness deviation is constant per patient. So in relative measurements to determine the wear progression, the error due to insert thickness variation is negligible. Concerning variation due to femoral positioning, the repeated measurements in the phantom experiment ($n=5$) showed an average SD of 10 mm^3, which is equivalent to a PI of 40 mm^3, i.e. 17% of the true wear pool volume.

The measurement accuracy in the phantom experiment was very limited as even in the best case only 56% of the wear pool volume was detected (129 of 230 mm^3, 30° of flexion). Moreover, for some of the inserts we were unable to detect any wear (Appendix A). For some cases the low accuracy may be caused by a large distance between the femoral-tibial contact location and the center of the wear pool. A positive finding was that the falsely detected volumes were low (<15 mm^3), resulting in a low risk of overestimating the wear pool.

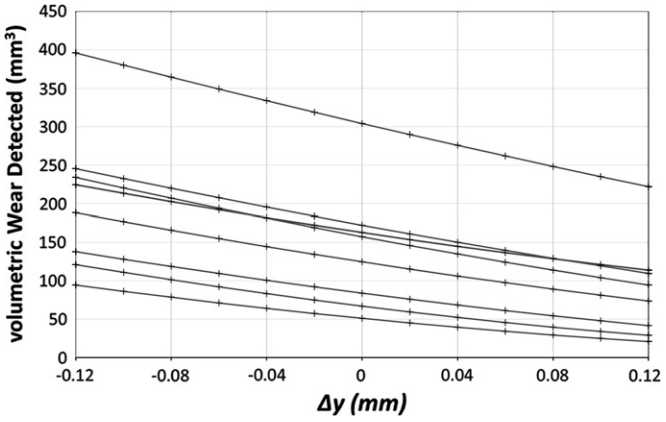

Fig. 5. The volumetric wear detected as a function of changing the insert thickness (Δd). The eight inputs are presented with separate lines. The limits $\Delta d = \pm 0.12$ mm equal the 95% PI interval of the manufacturing process.

Table 2
The volumes (vol.) detected in mm^3 per flexion angle ($N=5$). The means and standard deviations (SD) for the total detected volume, correctly detected, falsely detected volume and missed volume are shown. The 95% prediction intervals give the expected variation in practice and are calculated as 4 × SD of the total detected wear volume.

	Vol. total detected	Vol. correctly detected	Vol. falsely detected	Vol. missed	Prediction interval	
	Mean (SD) in mm^3	Mean (SD) in mm^3	Mean (SD) in mm^3	Mean (SD) in mm^3	Size in mm^3	Relative size
Flexion=0°	42 (8)	38 (7)	3 (5)	192 (7)	32	(14%)
Flexion=30°	122 (7)	122 (7)	0 (0)	109 (7)	28	(12%)
Flexion=60°	17 (19)	11 (20)	7 (3)	220 (20)	76	(33%)

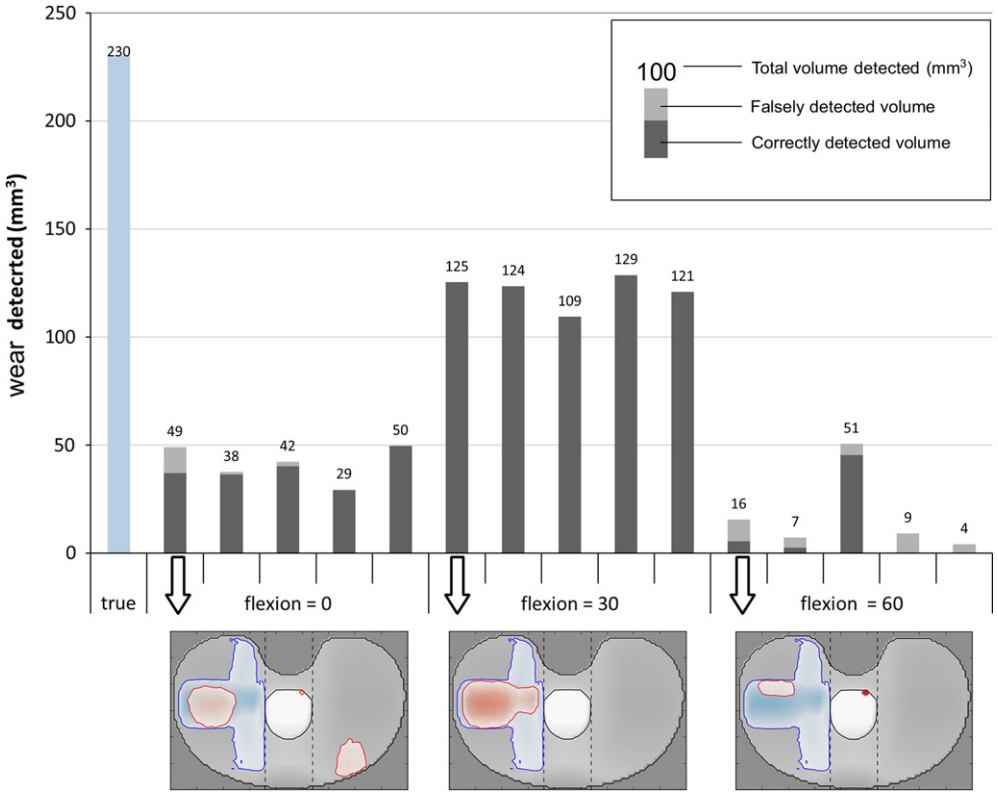

Fig. 6. The wear volumes detected for each of the fifteen RSA image pairs, compared to the true wear volume (bar on the left). The images on the bottom of the figure are examples of the wear pools per flexion angle. The blue overlay indicates the true wear pool and the red overlay on top indicates the detected wear pool. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

A limitation of our study is that the validation is based on phantom and in silico data only, whereas the ideal validation would be based on RSA data from patients shortly before insert revision, ensuring that both the shape of the wear pool in the retrieved inlay and the femoral-insert contact location in the pre-op RSA image are representative. As such data was not available a phantom experiment was utilized in which both the shapes of the wear pool and the freedom of the femoral-tibial contact location could be controlled to mimic clinical conditions. It is likely that the underestimation of the wear pool size and limited reliability found in this study are representative for clinical practice, as the created wear pools were a reasonable reproduction of abrasive wear.

Gill et al. (2006) suggested superimposing assessments with volumetric wear measurements at different flexion angles to get a better detection of the wear pool. The findings in our phantom study confirm that superimposing assessments can be beneficial, as large differences were found in the detected wear volume among the flexion angles. However, our simulation study also showed that a single assessment already has a variation of 111–283 mm^3. When several (almost) disjoint wear pools detected in alternate flexion angles are superimposed, the total variation will further increase. In practice, we expect a tradeoff between the accuracy (underestimation) and the precision of the measurement. Repeated measurements for each flexion angle could be used to improve the precision, but then the required number of RSA acquisitions quickly becomes impractical.

In summary, the accuracy of the volumetric wear measurement was limited, as at most 56% of the true wear volume was detected. In addition, the precision of the measurement was low, mainly caused by the pose estimation. The prediction interval of the volumetric wear measurement is at least as large as the measurement outcome itself. This is an important limitation to the applicability of the volumetric wear measurement in clinical practice and further clinical validation is required.

Conflict of interest statement

The authors declare that none of them have any financial or personal relationships with people or organizations that could inappropriately influence their work or the conclusions drawn from this work.

Acknowledgment

Funding for this study was provided by the Dutch Arthritis Foundation (no. 08-1-303) and Stryker.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.jbiomech.2013.02.021.

Bibliography

Collier, M.B., Jewett, B.A., Engh, C.A., 2003. Clinical assessment of tibial polyethylene thickness. Comparison of radiographic measurements with as-implanted and as-retrieved thicknesses. The Journal of Arthroplasty 18, 860–866.

Duryea, J., Grainger, A.J., Hide, I.G., Genant, H.K., Campbell, R.S., 2001. Fully automated software to monitor wear in prosthetic knees using fluoroscopic images. European Radiology 11, 2184–2187.

Edwards, S.A., Pandit, H.G., Ramos, J.L., Grover, M.L., 2002. Analysis of polyethylene thickness of tibial components in total knee replacement. The Journal of bone and joint surgery. American volume 84-A, 369–371.

Gill, H.S., Waite, J.C., Short, A., Kellett, C.F., Price, A., Murray, D.W., 2006. In vivo measurement of volumetric wear of a total knee replacement. The Knee 13, 312–317.

Kaptein, B.L., Valstar, E.R., Stoel, B.C., Reiber, H.C., Nelissen, R.G., 2007. Clinical validation of model-based RSA for a total knee prosthesis. Clinical Orthopaedics and Related Research 464, 205–209.

Kaptein, B.L., Valstar, E.R., Stoel, B.C., Rozing, P.M., Reiber, J.H.C., 2003. A new model-based RSA method validated using CAD models and models from reversed engineering. Journal of Biomechanics 36, 873–882.

Miller, T.T., 2005. Imaging of knee arthroplasty. European Journal of Radiology 54, 164–177.

Naudie, D.D.R., Ammeen, D.J., Engh, G.A., Rorabeck, C.H., 2007. Wear and osteolysis around total knee arthroplasty. The Journal of the American Academy of Orthopaedic Surgeons 15, 53–64.

Psychoyios, V., Crawford, R.W., O'Connor, J.J., Murray, D.W., 1998. Wear of congruent meniscal bearings in unicompartmental knee arthroplasty. The Journal of Bone and Joint Surgery 80, 976–982.

Sanzén, L., Sahlström, A., Gentz, C., Johnell, I., 1996. Radiographic wear assessment in a total knee prosthesis. Journal of Arthroplasty 11, 738–742.

Sharkey, P.F., Hozack, W.J., Rothman, R.H., Shastri, S., Jacoby, S.M., 2002. Why are total knee arthroplasties failing today? Clinical Orthopaedics and Related Research 404, 7–13.

Sundfeldt, M., Carlsson, L.V., Johansson, C.B., Thomsen, P., Gretzer, C., 2006. Aseptic loosening, not only a question of wear: a review of different theories. Acta Orthopaedica 77, 177–197.

van IJsseldijk, E.A., Valstar, E.R., Stoel, B.C., Nelissen, R.G.H.H., Kaptein, B.L., 2012. A model-based approach to measure the minimum joint space width of total knee replacements in standard radiographs. Journal of Biomechanics 45, 2171–2175.