

# The final follow-up plain radiograph is sufficient for clinical evaluation of polyethylene wear in total hip arthroplasty

## A study of validity and reliability

Maiken Stilling<sup>1</sup>, Kristian Larsen<sup>3</sup>, Niels T Andersen<sup>2</sup>, Kjeld Søballe<sup>1</sup>, Søren Kold<sup>1</sup>, and Ole Rahbek<sup>1</sup>

<sup>1</sup>Department of Orthopedics, Aarhus University Hospital; <sup>2</sup>Department of Biostatistics, School of Public Health, Aarhus University, Aarhus; <sup>3</sup>Orthopedic Research Unit, Department of Orthopedics, Hospital Unit West, Holstebro, Denmark

Correspondence: maiken.stilling@ki.au.dk

Submitted 10-01-10. Accepted 10-04-13

**Background and purpose** Radiostereometric analysis (RSA) is a highly accurate tool for assessment of polyethylene (PE) wear in total hip arthroplasty (THA); however, PE wear measurements in clinical studies are often limited to plain radiographs. We evaluated the agreement between PE wear measured with PolyWare software, which uses plain radiographs, and by model-based RSA, which uses stereo radiographs.

**Methods** Measurements of PE wear postoperatively and at final follow-up (after mean 6 years) on plain radiographs of 12 patients after cementless THA were evaluated with PolyWare software and the results were compared with those from RSA as the gold standard (Model-based RSA using elementary geometrical shape models; EGS-RSA). With PolyWare, we either used the final radiographic follow-up (PW1) only or both the postoperative follow-up and the final follow-up (PW2).

**Results** The 2D mean wear measured (in mm) was 0.80, 1.07, and 0.60 for the PW2, PW1, and RSA method. 2D intra-method repeatability was similar for PW1 and RSA with limits of agreement (LOAs, in mm) of  $\pm 0.22$ , and  $\pm 0.23$ , respectively. 2D inter-method concurrent validity was best between PW1 and EGS-RSA with LOAs of  $\pm 0.55$ . For 2D linear wear measurements, the PW1 method had a clinical repeatability similar to that of RSA.

**Interpretation** PW1 is sufficient for retrospective determination of 2D wear from medium-term wear measurements above 0.5 mm. It alleviates the need for baseline plain radiographs, has a clinical precision similar to that of RSA, and is easy and inexpensive to use.

Wear of polyethylene (PE) components is widely regarded as the main factor limiting longevity of total hip arthroplasty (THA) (Cooper et al. 1992). Clinical studies have shown that periprosthetic osteolysis and aseptic loosening is strongly related to wear rates of above 0.2 mm/year (Sochart 1999, Dowd et al. 2000).

Radiostereometric analysis (RSA) is the most accurate tool for in vivo assessment of PE wear (Kärrholm et al. 1997, von Schewelov et al. 2004, Bragdon et al. 2006), and it is regarded as the gold standard (Ilchmann et al. 1995). However, many radiographic in vivo studies of PE wear in THA are restricted to measurements on plain radiographs because the RSA set-up is expensive and not widely available. Measurement of PE wear on plain radiographs is often limited to 2D analysis because poor quality of cross-table lateral radiographs is a common problem (Sychterz et al. 1999b, 2001). Although PE wear is known to occur multidirectionally (Yamaguchi et al. 1997, Akisue et al. 1999), the bulk of the wear is detectable on the anterior-posterior radiographs alone (Sychterz et al. 1997, Hui et al. 2003, Martell et al. 2003). Based on the availability of radiographs and investigator preferences, some authors favor analysis of serial radiographs (Sychterz et al. 1997, Kim et al. 2001, Hernigou and Bahrami 2003) to describe the pattern of wear and the steady-state wear (Sychterz et al. 1999a, Bragdon et al. 2006), whereas others use 2 radiographic follow-ups (postoperative and latest) (Kraay et al. 2006), or only the latest radiographic follow-up with the assumption of zero wear at baseline (Norton et al. 2002).

Little is known about the conformity between PE wear results measured with RSA and computerized methods using plain radiographs (Ilchmann et al. 1995, von Schewelov et al. 2004, Bragdon et al. 2006). Our group has questioned the conformity of 2D PE wear measurements based on serial, 2, or 1 radiographic follow-up (Stilling et al. 2009b). We determined that there was a statistically significant difference between all approaches, but we were unable to determine which strategy best reflected the true extent of wear (Stilling et al. 2009b). In addition, we recently showed that model-based RSA is an accurate tool for measurement of PE wear in good agreement with the true wear (Stilling 2009).

We have now studied the intra-method repeatability and concurrent validity between 2 methods (PolyWare and EGS-

Table 1. Patient demographics

Input variable	Mean (range)
Age (years)	53 (44–65)
Height (cm)	172 (158–182)
Weight (kg)	84 (61–114)
Cup size (mm)	55 (50–62)
Polyethylene thickness (mm)	9.2 (6.8–11.8)
Follow-up (years)	6.1 (5.3–7.1)
Harris hip score at 5 years (points)	96 (84–100)
Gender ratio (males:females)	4:8
Hip side (right:left)	7:5

RSA) for measurement of PE wear in THA, in a group of patients with an average follow-up of 6 years. We wanted to determine (1) whether there would be a difference in repeatability between the methods, (2) whether there would be a difference in wear measured using 1 or 2 radiographic follow-ups with the PolyWare method, and (3) whether either of the 2 PolyWare measurement strategies (1 or 2 radiographic follow-ups) would give results similar to the wear measured by RSA (concurrent validity).

## Material and methods

The study was prepared in accordance with the Standards for Reporting of Diagnostic Accuracy (STARD) initiative (Bossuyt et al. 2003).

### Study design and patients (Table 1)

44 patients that were enrolled in an ongoing multicenter, randomized clinical trial (RCT) involving RSA had a primary THA between December 2001 and October 2005, and a subgroup of 18 patients had a minimum of 5 years of follow-up. These patients were invited for an additional clinical and radiographic double-examination follow-up linked to the present study, for measurement of 2D wear of the polyethylene liner by different methods. 12 patients with a mean follow-up of 6.1 (5.3–7.1) years responded and accepted. All investigations were conducted in accordance with ethical principles of research, informed consent was obtained from all participants, and the Central Denmark Region Committee on Biomedical Research reviewed and approved the study (Journal no. 20081096; issued December 15, 2008). Criteria for inclusion in the RCT were osteoarthritis of the hip and an age of > 18 and < 70 years. Criteria for exclusion from the RCT were osteoporosis (patients under medical treatment), neuromuscular or vascular leg disease, metabolic bone disorders, insufficient bone stock for total cementless THA, rheumatoid arthritis, malignant disease, planned pregnancy, and femoral neck fracture. 4 surgeons performed all the THAs using a posterolateral approach. Harris hip score was taken at the final follow-up.

## Implants

All components (femoral stems and acetabular cups) were cementless. The femoral component was a solid Ti6Al4V-alloy collarless, straight-stem Bi-Metric design (Biomet Inc, Warsaw, IN) with circumferential plasma-spray titanium and porous hydroxyapatite coating of the proximal one-quarter. The acetabular component was a plasma-sprayed titanium and hydroxyapatite-coated Mallory head, solid-finned Ringloc metal shell (Biomet). The cups were inserted using the same technique (approximately 2-mm press-fit by coating thickness, line-to-line reaming). The femoral stems were inserted by 2 alternative surgical techniques (bone rasping or bone compaction of the medullar canal) according to randomization in the RCT. The femoral heads (Biomet) were all of chrome-cobalt alloy, and they were 28 mm in diameter in 11 cases and 22 mm in diameter in one case. In all cases, the PE liners were of the Hi-Wall type and consisted of compression-molded, ultrahigh-molecular-weight PE (UHMWPE) resin, consolidated, packed, and sterilized by gamma irradiation in argon gas in the range of 2.5–4 Mrad (ArCom; Biomet).

## Radiographs

In the 2 follow-ups, all radiographs were obtained at the same hospital. The primary radiographs (stereo radiographs, anteroposterior pelvis, and cross-table lateral) were obtained during 2001 and 2003, within a week of surgery and after mobilization of the patients. The primary stereo radiographs were digital, but the plain radiographs were hard copy and were digitized to tagged image files at a resolution of 300 dots per inch at 100% scale in a high-resolution optical A3 scanner (Epson Expression 10000xl Pro A3). A standard RSA set-up of 2 synchronized ceiling-fixed roentgen tubes (Arco-Ceil/Medira; Santax Medico) angled toward each other at 40° and a uniplaner carbon calibration box (Box 24; Medis Specials, Leiden, the Netherlands) were used. At final follow-up, all radiographs were fully digital (FCR Profect CS; Fujifilm) and stored without compression. The anteroposterior and cross-table lateral radiographs had a size of 2,364 × 2,964 pixels (grayscale TIFF format) and the stereo radiographs had a size of 2,080 × 2,529 pixels (grayscale BMP format). The final radiographs were collected as double examinations by the same radiographer in January and February of 2009, with complete repositioning of the radiographic equipment and the leg of the patient between examinations (stereo radiographs, anteroposterior pelvis, and cross-table lateral). The quality of the digitized anteroposterior radiographs was generally good; however, in 3 patients the automatic circle fitting and edge detection with the PolyWare software was turned off and overruled by the manual digitizer tablet, as recommended to maintain reasonable reproducibility (Collier et al. 2003).

## Methods for PE wear measurement

In the non-weight-bearing pelvic radiographs, the location of the central ray was estimated by penciling diagonals between

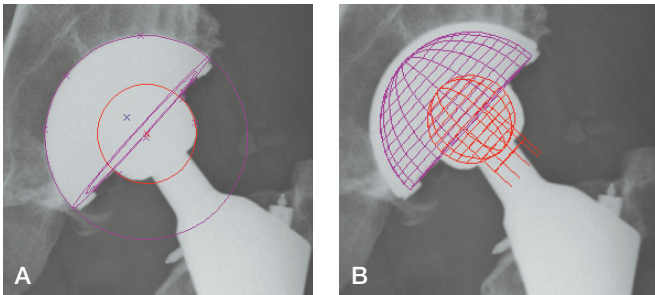


Figure 1. Polyethylene wear analysis with the PolyWare software. A. Circles are fitted to the peripheries of the cup and femoral head shadows B. At the end of the analysis a solid cup model is added to the output. Right hips are flipped about the vertical axis to look like left hips prior to analysis.

the corners of the rectangular exposure on the radiograph. Analysis was performed by an experienced observer (LLA) with a computerized method (PolyWare Pro 3D Digital vs. 5.10; Draftware Developers, Conway, SC). This technique, developed by Devane et al. (Devane et al. 1995a, b), is only applicable to uncemented acetabular cups, and it features a digital edge-detection algorithm to fit circles and ellipses to the peripheral shadows of the femoral head and acetabular component (Figure 1). 2D PE wear is measured in the plane of the anteroposterior radiograph. At first, both the postoperative and the final radiographs were used for measurement of 2D and 3D PE wear vectors (PW2), but later only the final radiographs (PW1) were used.

Both the postoperative and the final stereo radiographs were obtained without weight bearing and with the patient supine. The leg was positioned with the anatomical axis parallel to the y-axis of the calibration box. Analysis of all stereo radiographs was performed by an experienced observer (RM) with the software Model-Based RSA vs. 3.2 (Medis Specials, Leiden, the Netherlands) using elementary geometrical shape (EGS) implant models (EGS-RSA) (Kaptein et al. 2006). This is a newly developed RSA feature alleviating the need for tantalum bead marking of components or for reverse engineering of cup models (Kaptein et al. 2003). By use of the EGS mathematical algorithm in the software, software-generated sphere models were matched to the peripheries of the femoral head and cup with errors of 0.08 mm and 0.13 mm, respectively. PE wear was evaluated with the cup sphere as the reference and the femoral head sphere as the migrating (penetrating) object (Figure 2). The centers of the spheres are automatically defined by the software. The postoperative and final stereo radiographs were used for analysis. The output of EGS-RSA is a standard for RSA with 3 coordinate numbers (X, Y, and Z), and from these, 2D and 3D linear wear vectors can be calculated by Pythagoras' theorem (as the square root of  $(X^2 + Y^2)$  and the square root of  $(X^2 + Y^2 + Z^2)$ , respectively).

### Economic evaluation

A cost analysis comparing the PolyWare and RSA methods

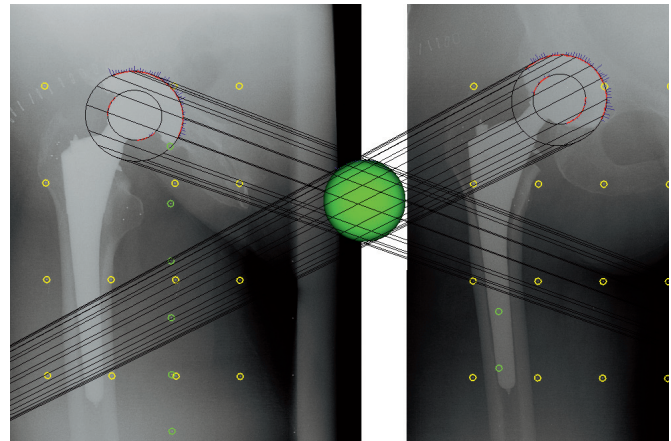


Figure 2. Polyethylene wear analysis in a right hip with model-based RSA, using elementary geometric shape models (EGS-RSA). Spheres are matched by an EGS algorithm to the peripheries of the femoral head and cup, and femoral head penetration is calculated as change in distance between their centers.

was performed with a marginal analysis (only differing costs) based on the present study; i.e., computer hardware that was necessary for both methods was not included. We defined 2 cost areas: investment costs and staff costs. The perspective of the analysis was that of the hospital. The investment costs consisted of additional X-ray equipment, calibration box, A3 transparency scanner, software, and education. X-ray equipment, calibration box, and software costs were calculated from actual costs. The staff costs consisted of the time used by the professions involved. The observed time for the radiographer to obtain 1 stereo radiograph was 30 min and 10 min for 1 AP pelvis plain radiograph. The observed time for retrieval and storage of 1 digital radiograph from the database at the radiology department was approximately 15 min, and the observed average time for finding 1 archived hard-copy radiograph and digitizing it in the transparency scanner was 45 min. RSA analysis took 90 min per patient (2 stereo radiographs) and PolyWare analysis took 30 min per patient (2 plain AP radiographs). Hourly salary for the radiographers (35 ) and for the research assistants (51 ) was obtained from the annual salary divided by 1,516 h, which was estimated by the administrative office to be the average number of effective working hours. All costs are based on 2010 prices.

### Statistics

**Repeatability.** The standard deviation of the difference ( $SD_{\text{dif-intra}}$ ) between the first and the second measurements (double examinations) within a method along with limits of agreement ( $LOA_{\text{intra}} = SD_{\text{dif-intra}} \times \pm 1.96$ ) were calculated according to Bland and Altman (1986). The systematic variation (bias) between the double examinations followed a normal distribution (Shapiro-Wilk test (Altman 1995)) and were tested with a paired t-test. The measures of repeatability ( $SD_{\text{dif-intra}}$  or equivalent the width of  $LOA_{\text{intra}}$ ) of the 3 methods were compared pairwise by looking at the ratios, and tested with an F-test.

**Table 2. Repeatability of radiographic double-examination wear measurements for the 3 methods**

Analysis method	Mean (range) (mm)	SD <sub>dif-intra</sub> <sup>a</sup>	Bias <sup>b</sup> ( $\pm$ LOA) <sup>c</sup> (mm)	95% CI <sup>d</sup> (mm)	p-value <sup>e</sup>
<i>2D measurements</i>					
PW2 <sup>f</sup>	0.80 (0.28–1.78)	0.26	-0.09 ( $\pm$ 0.53)	-0.26–0.08	0.25
PW1 <sup>g</sup>	1.07 (0.69–1.47)	0.11	0.04 ( $\pm$ 0.22)	-0.04–0.11	0.29
EGS-RSA <sup>h</sup>	0.60 (0.13–1.09)	0.11	0.06 ( $\pm$ 0.23)	-0.02–0.13	0.11
<i>3D measurements</i>					
PW2	1.12 (0.27–2.20)	0.31	-0.05 ( $\pm$ 0.62)	-0.25–0.15	0.61
PW1	1.48 (0.86–2.31)	0.44	-0.03 ( $\pm$ 0.87)	-0.31–0.25	0.82
EGS-RSA	0.75 (0.26–1.47)	0.16	0.05 ( $\pm$ 0.31)	-0.05–0.15	0.33

<sup>a</sup> SD<sub>dif-intra</sub> is the random variation within a method comparing double examinations.

<sup>b</sup> Bias: systematic variation within a method.

<sup>c</sup> LOA: limits of agreement around the bias (95% prediction interval = SD<sub>dif-intra</sub>  $\times$  1.96).

<sup>d</sup> 95% confidence interval for the bias.

<sup>e</sup> p value (paired t-test) bias between methods.

<sup>f</sup> PW2: PolyWare PE wear analysis using the postoperative and final follow-up radiographs.

<sup>g</sup> PW1: PolyWare PE wear analysis using only the final radiographic follow-up radiographs.

<sup>h</sup> EGS-RSA: radiostereometric analysis of PE wear using sphere models (the “gold standard”).

LOA<sub>intra</sub> provides the same measure as the bias  $\pm$  the 95% repeatability limit that is specified in the ASTM 177 standard practice for bias and precision (2008). For comparison of RSA precision with that in the literature, we calculated the 95% confidence interval (CI) for translation values of each axis.

**Concurrent validity.** Concurrent validity defines the chronological correlation between 2 measurement methods (International Epidemiological Association Inc. 1995). The RSA method was considered to be the “gold standard”. An average value from double examinations was calculated and used to estimate the bias between methods. The bias followed a normal distribution (Shapiro-Wilk test) and was tested with a paired t-test. Furthermore, the standard deviation of the difference (SD<sub>dif-inter</sub>) between methods and the agreement limits between methods (LOA<sub>inter</sub>) were calculated according to Altman (1995) (LOA<sub>inter</sub> = SD<sub>dif-inter</sub>  $\times$   $\pm$ 1.96).

Statistical significance was assumed at  $p < 0.05$ . Intercooled Stata software version 10.0 (StataCorp, College Station, TX) was used for statistical computations.

## Results

Repeatability evaluated within methods revealed no clinically relevant or statistically significant bias between any 2 pairs of radiographic double examinations of PE wear. The 2D intra-method repeatability (LOA, mm) was  $\pm 0.22$ ,  $\pm 0.23$ , and  $\pm 0.53$  for PW1, EGS-RSA, and PW2, respectively. The 3D intra-method repeatability (LOA, mm) was  $\pm 0.31$ ,  $\pm 0.62$ , and  $\pm 0.87$  for EGS-RSA, PW2, and PW1, respectively (Table 2 and Figure 3). The relative repeatability between 2D PW1 and 2D EGS-RSA (the “gold standard”) was 1.02 ( $p = 0.95$ ) (Table 3). Precision (95% CI, mm) was 0.14, 0.26, and 0.29 for the x-, the y-, and the z-axis, respectively.

Concurrent validity showed a statically significant ( $p < 0.04$ ) bias between all pairwise comparisons of methods, except between 2D PW2 and 2D EGS-RSA. The 2D inter-method concurrent validity (LOA, mm) was  $\pm 0.55$ ,  $\pm 0.89$ , and  $\pm 0.68$  for PW1 relative to EGS-RSA, PW2 relative to EGS-RSA, and PW1 relative to PW2, respectively. The 3D inter-method concurrent validity (LOA, mm) was  $\pm 1.13$ ,  $\pm 1.06$ , and  $\pm 0.90$  for PW1 relative to EGS-RSA, PW2 relative to EGS-RSA, and PW1 relative to PW2, respectively (Table 3 and Figure 4).

The total investment costs were 132,982 for RSA, 7,217 for PW2, and 2,052 for PW1 using only the final radiographic follow-up (Table 4). The total staff costs for the 12 patients in this study were 1,644 for RSA, 1,068 for PW2, and 612 for PW1 (Table 4).

## Discussion

Although RSA is considered to be the most accurate and precise analysis method for PE wear (the gold standard) (Ilchmann et al. 1995, von Schewelov et al. 2004), many radiographic in vivo studies, especially retrospective studies, have been restricted to wear measurements on plain radiographs. Several computer-assisted methods for assessment of PE wear on plain radiographs are available (von Schewelov et al. 2004, McCalden et al. 2005, Geerdink et al. 2008), but few have been compared clinically with RSA (Ilchmann et al. 1992, 1995, von Schewelov et al. 2004, Bragdon et al. 2006), and to our knowledge no previous studies have evaluated the concurrent validity of RSA and the commonly used PolyWare method for plain radiographs (Devane and Horne 1999). Specifically, we wanted to determine whether it was more accurate (in agreement with RSA) to use only the final radiographic follow-up or to use both the postoperative and

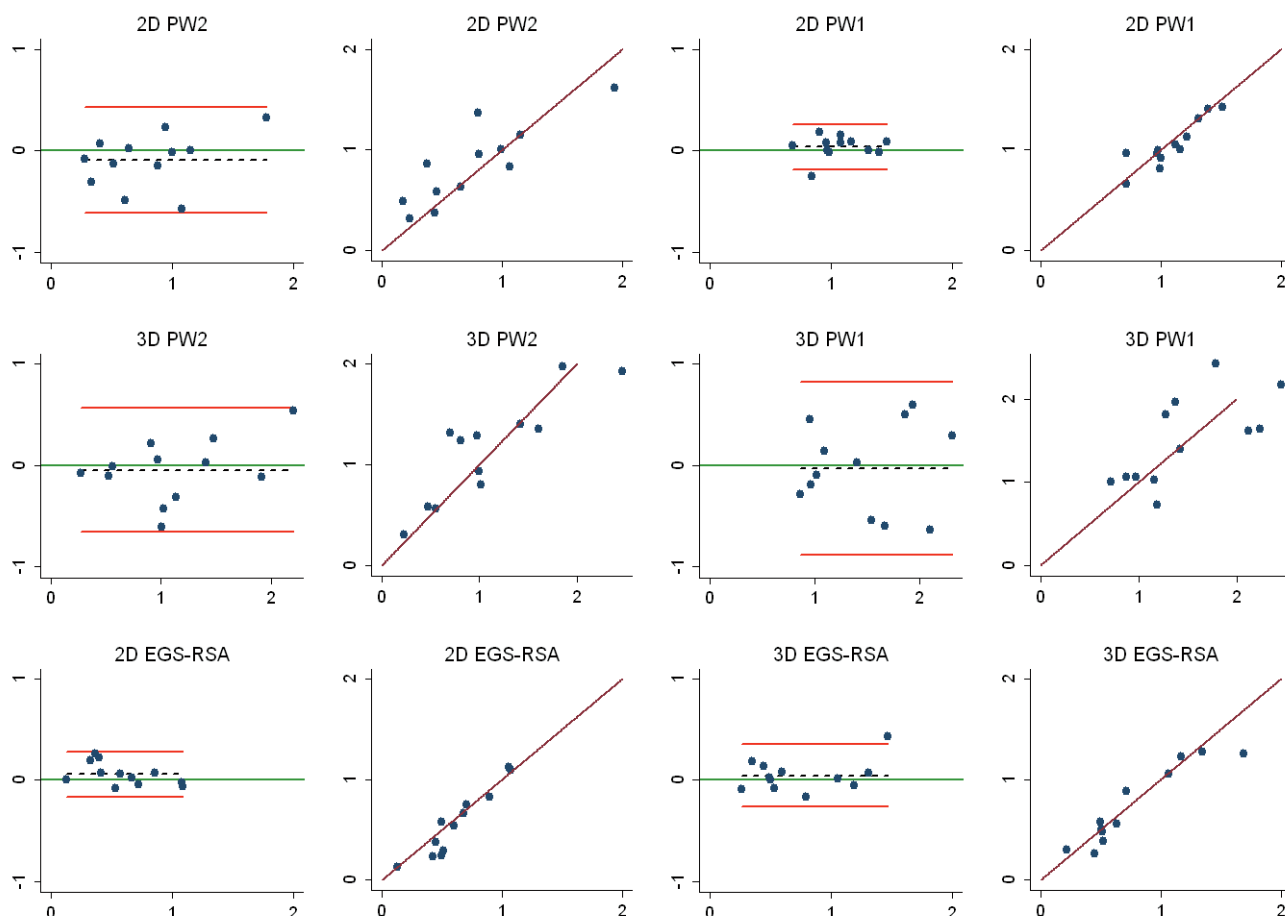


Figure 3. Bland-Altman plots and scatter plots, with lines of equality for repeatability measures for each of the 3 methods.

Bland-Altman plots (columns 1 and 3); x-axis: average of two measurements; y-axis: the difference between 2 measurements ( $y = \text{measurement 1} - \text{measurement 2}$ ); red lines: 95% limits of agreement; dashed line: bias from 0; solid green line:  $y = 0$  line; dots: individual double measures.

Scatter plots (columns 2 and 4); x-axis: first measurement; y-axis: second measurement; maroon lines: lines of equality; EGS-RSA: radiostereometric analysis using sphere models; PW1: PolyWare using only the final follow-up radiographs; PW2: PolyWare using the postoperative and final follow-up radiographs.

Table 3. Comparison of repeatability and concurrent validity between methods

Analysis method	Repeatability		$SD_{\text{dif-intr}}^c$ (mm)	Concurrent validity		
	Repeatability ratio <sup>a</sup>	p-value <sup>b</sup>		Bias <sup>c</sup> ( $\pm LOA$ ) <sup>e</sup> (mm)	CI 95% of true bias <sup>f</sup> (mm)	p-value <sup>g</sup>
<i>2D measurements</i>						
PW1 <sup>h</sup> relative to EGS-RSA <sup>i</sup>	1.02	0.95	0.27	0.48 ( $\pm 0.55$ )	0.30–0.65	< 0.001
PW2 <sup>j</sup> relative to EGS-RSA	2.32	< 0.001	0.44	0.21 ( $\pm 0.89$ )	-0.08–0.49	0.14
PW1 relative to PW2	2.36	< 0.01	0.34	0.27 ( $\pm 0.68$ )	0.05–0.49	0.02
<i>3D measurements</i>						
PW1 relative to EGS-RSA	2.80	0.002	0.56	0.73 ( $\pm 1.13$ )	0.37–1.09	< 0.001
PW2 relative to EGS-RSA	2.00	0.03	0.53	0.36 ( $\pm 1.06$ )	0.03–0.70	0.04
PW1 relative to PW2	1.40	0.28	0.45	0.36 ( $\pm 0.90$ )	0.08–0.65	0.02

<sup>a</sup> Repeatability ratio: ratios of variance.

<sup>b</sup> p-value: test of variance between methods (F-test).

<sup>c</sup>  $SD_{\text{dif-intr}}$ : random variation from the 2 different methods.

<sup>d</sup> Bias: systematic variation between methods.

<sup>e</sup> LOA: limits of agreement around the bias (95% prediction interval =  $SD_{\text{dif-intr}} \times 1.96$ ).

<sup>f</sup> 95% confidence interval for the bias.

<sup>g</sup> p-value (paired t-test) bias between methods.

<sup>h</sup> PW1: PolyWare PE wear analysis using only the final follow-up radiographs.

<sup>i</sup> EGS-RSA: radiostereometric analysis of PE wear using sphere models (the "gold standard").

<sup>j</sup> PW2: PolyWare PE wear analysis using the postoperative and final follow-up radiographs.

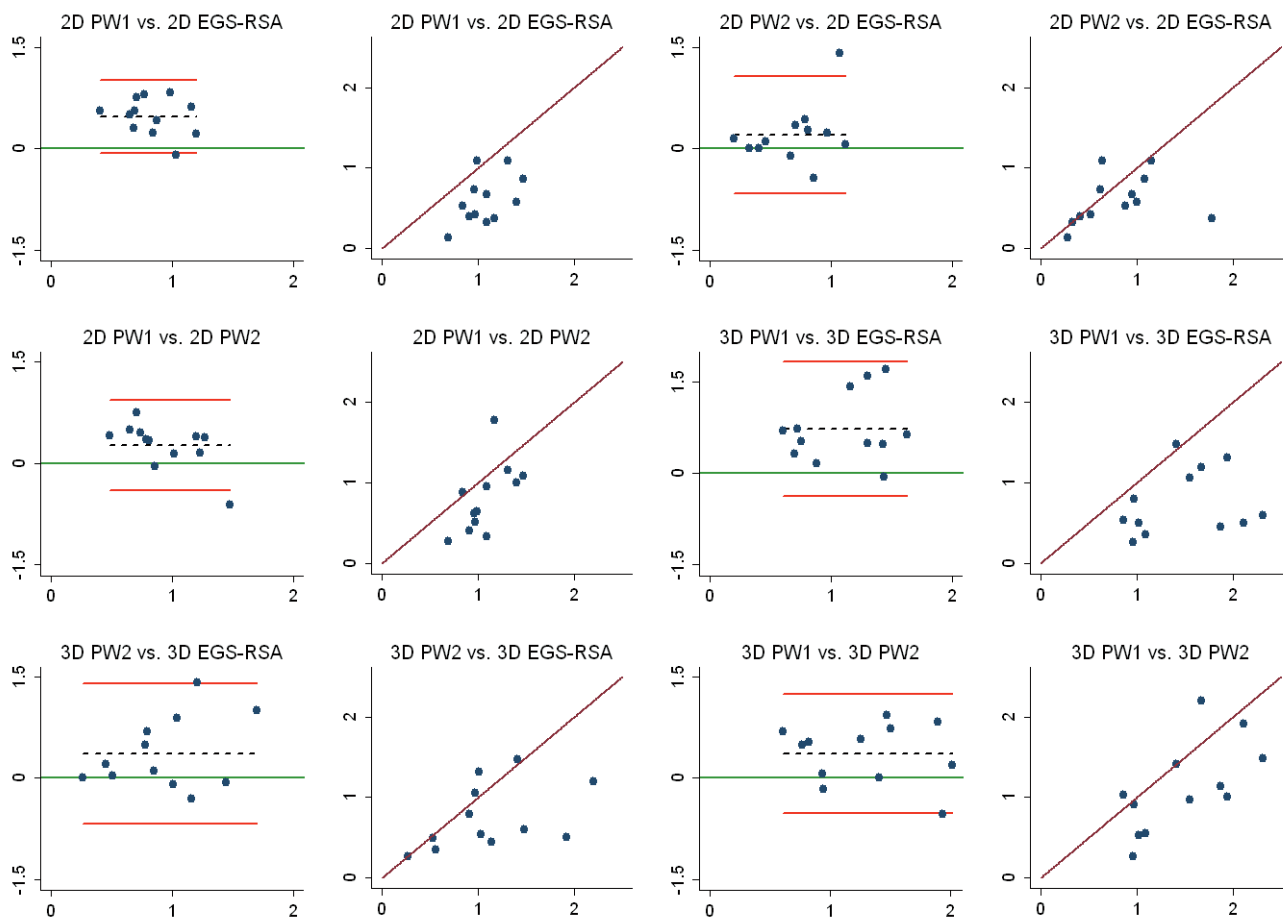


Figure 4. Bland-Altman plots and scatter plots with lines of equality for concurrent validity between the 3 methods. Bland-Altman plots (columns 1 and 3); x-axis: average of the measurements of 2 methods; y-axis: difference between measurements of 2 methods; red lines: 95% limits of agreement; dashed line: bias from 0; solid green line: y = 0 line; dots: individual double measures. Scatter plots (columns 2 and 4); maroon lines: lines of equality; EGS-RSA: radiostereometric analysis using sphere models; PW1: PolyWare using only the final follow-up radiographs; PW2: PolyWare using the postoperative and final follow-up radiographs.

Table 4. Marginal cost analysis in €(euros) for polyethylene wear analysis by EGS-RSA and with PolyWare as used in this study. For easy comparison, the staff costs are listed per patient

Cost area	EGS-RSA <sup>a</sup>	PW2 <sup>b</sup>	PW1 <sup>c</sup>
<i>Investment costs (1 time)</i>			
Additional X-ray equipment	80,604	0	0
Calibration box	10,075	0	0
A3 transparency scanner	0	5,165	0
Software	34,595	510	510
Education of radiographer	1,542	0	0
Education of research assistant	6,166	1,542	1,542
Total investment costs	132,982	7,217	2,052
<i>Staff costs per patient</i>			
Radiographer	35	12	12
Research assistant	102	77	39
Total staff costs per patient	137	89	51

<sup>a</sup> EGS-RSA: radiostereometric analysis of PE wear using sphere models (2 stereo radiographs were used).

<sup>b</sup> PW2: PolyWare PE wear analysis using the postoperative and final follow-up plain radiographs.

<sup>c</sup> PW1: PolyWare PE wear analysis using only the final radiographic follow-up plain radiographs.

the final radiograph follow-ups with the PolyWare method (Stilling et al. 2009b).

Several variables in the clinical set-up may, in theory, influence the amount of wear measured. Small changes in the radiographic set-up from follow-up to follow-up, under- or overexposure of radiographs that can affect the quality and sharpness of the component borders, patient position and leg rotation, body size and soft tissue mass of the patients, and angulations and size of components are just some of the variables that may affect clinical radiographs. Wear measurements based on uncalibrated plain radiographs would naturally be more sensitive to these changes than calibrated stereo radiographs. Despite all these potential problems with plain radiographs, we did not exclude any patients or radiographs because the border of the femoral head was sufficiently visible in all radiographs.

When only the final follow-up plain radiograph (PW1) is used to estimate wear, the primary position of the femoral head in each patient (zero wear) is assumed by the PolyWare software based on CAD-based knowledge of the cup and head,

and the keyed-in information on sizes. For PW2 and RSA, the postoperative radiographs provide the baseline. The algorithms for determination of wear by use of plain or stereo radiographs are not identical. Consequently, exact agreement between PE wear measurements based on different angle radiographs evaluated with different software packages cannot be anticipated, but some similarity can be expected. Both EGS-RSA and PolyWare are shadow-casting methods (Collier et al. 2003), and PolyWare relies on the marking of a beam center in the radiographs. We used only pelvic anteroposterior plain radiographs at postoperative and final follow-up; thus, the center of the ray should have been similar at different follow-ups. The resolution of the scanned primary hard copy plain radiographs and the follow-up digital radiographs we used followed the recommendations in the instruction manuals.

The radiographic set-up and the leg of the patient were repositioned between the double-examination radiographs in our study, and the calculated inter-method repeatability therefore reflects the contribution of variance from the radiographic set-up, the leg position, and the method of PE wear analysis. All 3 methods had small biases (range -0.09–0.06 mm), which were of no clinical or statistical significance. The best intra-method repeatability was obtained with 2D PW1 and 2D EGS-RSA, with approximate limits of agreement of  $\pm 0.22$  mm and  $\pm 0.23$  mm. Repeatability for all the 2D PE methods of wear measurement had limits of agreement below  $\pm 0.5$  mm, whereas repeatability of all 3D PE methods of wear measurement had limits of agreements above  $\pm 0.5$  mm.

In a clinical study, Digas et al. (2003) assessed double examinations of 45 patients and reported that precision absolute mean  $\pm 2.7$  SD (99% CI) for the 3D total was 0.22 mm. This is somewhat better than our observation for 3D EGS-RSA (LOA:  $\pm 0.31$  mm). These authors also reported translational precision of marker-based RSA to be 0.13 mm for the transverse axis, 0.10 mm for the longitudinal axis, and 0.22 mm for the sagittal axis. Röhrl et al. (2004) evaluated double examinations of patients with slight repositioning between exposures and found a longitudinal axis precision of 0.15 mm (95% CI). We used a model-based RSA method and observed a similar precision (95% CI) for the x-axis (0.14 mm) but a poorer precision for the y- and z-axis (Kaptein et al. 2006). It has already been emphasized that a 3D precision is mathematically difficult to present, as the precisions of the different directions cannot easily be added (Ryd 1986). Yet, this was necessary for a direct comparison of the repeatability of RSA and PolyWare.

In a retrieval study, PolyWare has been shown to underestimate 2D linear wear by 20% and dimensional 3D wear by 18% (Hui et al. 2003). We found the opposite tendency; that is, overestimation of wear by PolyWare in comparison to EGS-RSA as the gold standard. The relative mean difference between the 2D and 3D PE wear measured by PolyWare using two radiographic follow-ups (PW2) and EGS-RSA was 21% and 30%, respectively. Comparing PolyWare using one radiographic follow-up (PW1) and EGS-RSA, the relative difference for 2D and 3D

wear was even larger (40% and 46%, respectively). As a consequence of these large differences in measured mean wear, we only established concurrent validity of the mean bias with EGS-RSA and PW2 based on statistical testing. However, the systematic variation (bias) can be corrected for when known, whereas the random variation cannot, and thus the methods with the concurrent smallest LOA are the ones in closest agreement. In our study, this was EGS-RSA and PW1.

A clinical threshold of interest for the detection of PE wear that leads to long-term osteolysis and implant failure has been established to be 0.2 mm/year (Dowd et al. 2000, Sochart 2001). This is at the lower limit of clinically measureable wear with the best 2D wear methods used in our study. When total wear measurements close to 0.2 mm are of interest (i.e. cross-linked liners at medium-term follow-up), the images should be analyzed several times, with the average value representing the true value (Vickers 2003). For PE wear analysis, however, the number of repeat wear measurements that is optimal is not known at present. Using the most accurate method (EGS-RSA), the medium-term wear rate was 0.12 mm/year, which is in accordance with a recent report (Skoldenberg et al. 2009). We have previously determined the medium-term PE wear rate (0.25 mm/year) in similar ArCom PE liners articulated with 28-mm cobalt-chromium femoral heads by wear analysis on serial radiographs (Stilling et al. 2009a). Later, we were able to show that the use of serial radiographs for wear analysis with PolyWare results in an increased amount of measured wear (Stilling et al. 2009b), which explains the higher wear rate we found in ArCom PE.

Assessing concurrent validity, the mean PE wear measured with PolyWare (PW1 and PW2) was greater than wear measured by EGS-RSA. This is similar to the report of Bragdon et al. (2006) who compared marker-based digital RSA and the Martell method on plain radiographs. They suggested a calculation and comparison of the steady-state wear between methods. In our patient series no 1- or 2-year radiographic follow-ups were available, so this was not possible.

The accuracy of 2D PE wear measurement by the EGS-RSA method was recently shown to be in very good agreement with the true wear (Stilling et al. 2009b). Thus, based on the present results, the use of only the final plain radiographic follow-up with the PolyWare method (PW1) comes within  $\pm 0.55$  mm of the true value. This is sufficient for comparative studies assessing differences between 2 groups, and if desired, the systematic error can be corrected for. Furthermore, limiting the assessed plain radiographs to the final follow-up will improve repeatability and also provide the chance of good-quality digital radiographs. Also, it permits definition of a pre-study protocol for the last follow-up radiographs, thus ensuring that there is less projection variation between radiographs in a retrospective clinical study targeting PE wear. This could also reduce the number of patients needed for evaluation.

The marginal cost analysis favors PolyWare over RSA concerning both investment costs and staff costs; however, some

adjustments of the costs shown may be needed in another institution depending on the additional equipment needed. PE wear analysis with PolyWare, where only the final (and digital) radiograph is used, is the lowest priced method overall. However, because PolyWare is a less precise method than RSA, a 2–3 times larger sample size will be needed for this method (Stilling 2009), which evens out the staff costs for a prospective clinical study with the 2 methods. Yet, investment costs are 20 to 60 times more expensive for RSA, and to be cost-effective the RSA system should be used for more than 1 study. Furthermore, and something that was not included in the marginal analysis, plain radiographs are needed for documentation after surgery, whereas stereo radiographs are additional and therefore add to the total radiation dose per patient studied.

We expect our findings to have good external validity and to be applicable to good-quality radiographs of various brands of hemispheric metal shells with polyethylene liners and metal femoral heads. The PolyWare method using only final radiographic anteroposterior images is inexpensive and easy to use, is applicable for 2D wear measurements above 0.5 mm in total, and offers a simple and fast set-up that is applicable for the assessment of PE wear in most hospitals. The PolyWare method using only final radiographic anteroposterior images has a clinical repeatability similar to that of EGS-RSA (“the gold standard”) and is ideal for retrospective research because it alleviates the need for baseline images that are often lost, stored in hard copy, and of variable quality. For assessment of low PE wear (i.e. with new cross-linked liners), PolyWare software does not supply the accuracy required, and for such situations we recommend RSA. For assessment of medium-term or long-term wear measurements in larger groups of patients, the PolyWare method is optimal, simple, and in relatively close agreement with the gold standard of RSA.

Designed the study: KS, OR, MS, and SK. Operated on the patients: KS. Gathered the data: MS, OR, and SK. Analyzed the data: MS, NTA, and KL. Wrote the initial draft: MS and KL. Revised the initial draft: KS, SK, OR, NTA, and KL. Ensured the accuracy of the data and analysis: NTA and KL.

The authors thank Rikke Mørup for performing radiostereometric analyses on all patients and Lone Løvgren Andersen for performing wear analyses with PolyWare on all patients.

No competing interests declared.

ASTM: Standard Practice for Use of the Terms Precision and Bias in ASTM Test Methods E177. Subcommittee E11 20 on Test Method Evaluation and Quality Control, ASTM International, West Conshohocken, PA, USA 2008.

Akisue T, Bauer T W, Yamaguchi M, Matejczyk M B, Stulberg B N, Wilde A H. Multidirectional deformation in fully congruent acetabular components. *J Arthroplasty* 1999; 14 (8): 1011-8.

Altman D G. *Practical Statistics for Medical Research*. Chapman & Hall, 1995.

Bland J M, Altman D G. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1 (8476): 307-10.

Bossuyt P M, Reitsma J B, Bruns D E, Gatsonis C A, Glasziou P P, Irwig L M, Lijmer J G, Moher D, Rennie D, de Vet H C. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Clin Biochem* 2003; 36 (1): 2-7.

Bragdon C R, Martell J M, Greene M E, Estok D M, Thanner J, Karrholm J, Harris W H, Malchau H. Comparison of femoral head penetration using RSA and the Martell method. *Clin Orthop* 2006; (448): 52-7.

Collier M B, Kraay M J, Rimnac C M, Goldberg V M. Evaluation of contemporary software methods used to quantify polyethylene wear after total hip arthroplasty. *J Bone Joint Surg (Am)* 2003; 85 (12): 2410-8.

Cooper R A, McAllister C M, Borden L S, Bauer T W. Polyethylene debris-induced osteolysis and loosening in uncemented total hip arthroplasty. A cause of late failure. *J Arthroplasty* 1992; 7 (3): 285-90.

Devane P A, Bourne R B, Rorabeck C H, Hardie R M, Horne J G. Measurement of polyethylene wear in metal-backed acetabular cups. I. Three-dimensional technique. *Clin Orthop* 1995a; (319): 303-16.

Devane P A, Bourne R B, Rorabeck C H, MacDonald S, Robinson E J. Measurement of polyethylene wear in metal-backed acetabular cups. II. Clinical application. *Clin Orthop* 1995b; (319): 317-26.

Devane P A, Horne J G. Assessment of polyethylene wear in total hip replacement. *Clin Orthop* 1999; (369): 59-72.

Digas G, Karrholm J, Thanner J, Malchau H, Herberts P. Highly cross-linked polyethylene in cemented THA: randomized study of 61 hips. *Clin Orthop* 2003; (417): 126-38.

Dowd J E, Sychterz C J, Young A M, Engh C A. Characterization of long-term femoral-head-penetration rates. Association with and prediction of osteolysis. *J Bone Joint Surg (Am)* 2000; 82 (8): 1102-7.

Geerdink C H, Grimm B, Vencken W, Heyligers I C, Tonino A J. The determination of linear and angular penetration of the femoral head into the acetabular component as an assessment of wear in total hip replacement: A comparison of four computer-assisted methods. *J Bone Joint Surg (Br)* 2008; 90 (7): 839-46.

Hernigou P, Bahrami T. Zirconia and alumina ceramics in comparison with stainless-steel heads. Polyethylene wear after a minimum ten-year follow-up. *J Bone Joint Surg (Br)* 2003; 85 (4): 504-9.

Hui A J, McCalden R W, Martell J M, MacDonald S J, Bourne R B, Rorabeck C H. Validation of two and three-dimensional radiographic techniques for measuring polyethylene wear after total hip arthroplasty. *J Bone Joint Surg (Am)* 2003; 85 (3): 505-11.

Ilchmann T, Franzen H, Mjoberg B, Wingstrand H. Measurement accuracy in acetabular cup migration. A comparison of four radiologic methods versus roentgen stereophotogrammetric analysis. *J Arthroplasty* 1992; 7 (2): 121-7.

Ilchmann T, Mjoberg B, Wingstrand H. Measurement accuracy in acetabular cup wear. Three retrospective methods compared with Roentgen stereophotogrammetry. *J Arthroplasty* 1995; 10 (5): 636-42.

International Epidemiological Association Inc. *A Dictionary of Epidemiology*. Oxford University Press, 1995.

Kaptein B L, Valstar E R, Stoel B C, Rozing P M, Reiber J H. A new model-based RSA method validated using CAD models and models from reversed engineering. *J Biomech* 2003; 36 (6): 873-82.

Kaptein B L, Valstar E R, Spoor C W, Stoel B C, Rozing P M. Model-based RSA of a femoral hip stem using surface and geometrical shape models. *Clin Orthop* 2006; (448): 92-7.

Kärrholm J, Herberts P, Hultmark P, Malchau H, Nivbrant B, Thanner J. Radiostereometry of hip prostheses: Review of methodology and clinical results. *Clin Orthop* 1997; (344): 94-110.

Kim Y H, Kim J S, Cho S H. A comparison of polyethylene wear in hips with cobalt-chrome or zirconia heads. A prospective, randomised study. *J Bone Joint Surg (Br)* 2001; 83 (5): 742-50.



- Kraay M J, Thomas R D, Rimnac C M, Fitzgerald S J, Goldberg V M. Zirconia versus Co-Cr femoral heads in total hip arthroplasty: early assessment of wear. *Clin Orthop* 2006; (453): 86-90.
- Martell J M, Berkson E, Berger R, Jacobs J. Comparison of two and three-dimensional computerized polyethylene wear analysis after total hip arthroplasty. *J Bone Joint Surg (Am)* 2003; 85 (6): 1111-7.
- McCalden R W, Naudie D D, Yuan X, Bourne R B. Radiographic methods for the assessment of polyethylene wear after total hip arthroplasty. *J Bone Joint Surg (Am)* 2005; 87 (10): 2323-34.
- Norton M R, Yarlagadda R, Anderson G H. Catastrophic failure of the Elite Plus total hip replacement, with a Hylamer acetabulum and Zirconia ceramic femoral head. *J Bone Joint Surg (Br)* 2002; 84 (5): 631-5.
- Röhrh S M, Nivbrant B, Strom H, Nilsson K G. Effect of augmented cup fixation on stability, wear, and osteolysis: a 5-year follow-up of total hip arthroplasty with RSA. *J Arthroplasty* 2004; 19 (8): 962-71.
- Ryd L. Micromotion in knee arthroplasty. A roentgen stereophotogrammetric analysis of tibial component fixation. *Acta Orthop Scand (Suppl 220)* 1986: 1-80.
- Skoldenberg O, Salemyr M, Muren O, Johansson A, Ahl T, Boden H. The Ringloc liner compared with the Hexloc liner in total hip arthroplasty. *Orthop Rev* 2009; 1 (2): 21-6
- Sochart D H. Relationship of acetabular wear to osteolysis and loosening in total hip arthroplasty. *Clin Orthop* 1999; (363): 135-50.
- Sochart D H. True wear rates as predictors of risk for osteolysis after total hip arthroplasty. *J Bone Joint Surg (Am)* 2001; 83 (8): 1277.
- Stilling M. Polyethylene wear analysis. Experimental and clinical studies in total hip replacement. *Acta Orthop (Suppl 337)* 2009; 80: 1-74.
- Stilling M, Nielsen K A, Soballe K, Rahbek O. Clinical Comparison of Polyethylene Wear with Zirconia or Cobalt-Chromium Femoral Heads. *Clin Orthop* 2009a; (467) (10): 2644-50.
- Stilling M, Soballe K, Andersen N T, Larsen K, Rahbek O. Analysis of polyethylene wear in plain radiographs. The number of radiographs influences the results. *Acta Orthop* 2009b; 80 (6): 675-82.
- Sychterz C J, Engh C A, Jr., Shah N, Engh C A, Sr. Radiographic evaluation of penetration by the femoral head into the polyethylene liner over time. *J Bone Joint Surg (Am)* 1997; 79 (7): 1040-6.
- Sychterz C J, Engh C A, Jr., Yang A, Engh C A. Analysis of temporal wear patterns of porous-coated acetabular components: distinguishing between true wear and so-called bedding-in. *J Bone Joint Surg (Am)* 1999a; 81 (6): 821-30.
- Sychterz C J, Yang A M, McAuley J P, Engh C A. Two-dimensional versus three-dimensional radiographic measurements of polyethylene wear. *Clin Orthop* 1999b; (365): 117-23.
- Sychterz C J, Young A M, Engh C A. Effect of radiographic quality on computer-assisted head penetration measurements. *Clin Orthop* 2001; (386): 150-8.
- Vickers A J. How many repeated measures in repeated measures designs? Statistical issues for comparative trials. *BMC Med Res Methodol* 2003; 3: 22.
- von Schewelov T, Sanzen L, Borlin N, Markusson P, Onsten I. Accuracy of radiographic and radiostereometric wear measurement of different hip prostheses - An experimental study. *Acta Orthop Scand* 2004; 75 (6): 691-700.
- Yamaguchi M, Bauer T W, Hashimoto Y. Three-dimensional analysis of multiple wear vectors in retrieved acetabular cups. *J Bone Joint Surg (Am)* 1997; 79 (10): 1539-44.