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

C. H. Geerdink, MD
Orthopaedic Surgeon
Department of Orthopaedic Surgery
Iksia Hospital Rotterdam, Postbus 5009, 3008 AA, Rotterdam, The Netherlands.

B. Grimm, PhD, Research Manager
W. Vencken, Research Assistant
I. C. Heyligers, MD, PhD, Orthopaedic Surgeon
A. J. Tonino, MD, PhD, Orthopaedic Surgeon
Department of Orthopaedic Surgery
Atrium Hospital Heerlen, P.O. Box 4446, 6401 CK, Heerlen, The Netherlands.

Correspondence should be sent to Dr. C. H. Geerdink; e-mail: carelgeerdink@hotmail.com

We have compared four computer-assisted methods to measure penetration of the femoral head into the acetabular component in total hip replacement. These were the Martell Hip Analysis suite 7.14, Rogan HyperOrtho, Rogan View Pro-X and Roman v1.70. The images used for the investigation comprised 24 anteroposterior digital radiographs and 24 conventional acetate radiographs which were scanned to provide digital images. These radiographs were acquired from 24 patients with an uncemented total hip replacement with a follow-up of approximately eight years (mean 8.1; 6.3 to 9.1). Each image was measured twice by two blinded observers. The mean annual rates of penetration of the femoral head measured in the eight-year single image analysis were: Martell, 0.24 (SD 0.19); HyperOrtho, 0.12 (SD 0.08); View Pro-X, 0.12 (SD 0.06); Roman, 0.12 (SD 0.07). In paired analysis of the six-month and eight-year radiographs: Martell, 0.35 (SD 0.22); HyperOrtho, 0.15 (SD 0.13); View Pro-X, 0.11 (SD 0.06); Roman, 0.11 (SD 0.07). The intra- and inter-observer variability for the paired analysis was best for View Pro-X and Roman software, with intraclass correlations of 0.97, 0.87 and 0.96, 0.87, respectively, and worst for HyperOrtho and Martell, with intraclass correlations of 0.46, 0.13 and 0.33, 0.39, respectively.

The Roman method proved the most precise and the most easy to use in clinical practice and the software is available free of charge. The Martell method showed the lowest precision, indicating a problem with its edge detection algorithm on digital images.

Wear of the polyethylene acetabular component is widely regarded as the primary factor limiting the longevity of total hip replacements (THR). Particulate polyethylene debris invokes an inflammatory tissue response in the adjacent bone, leading to peri-prosthetic osteolysis and loosening of the implant. Various techniques have been developed to measure polyethylene wear in vivo, to monitor critical polyethylene wear in individual patients in the clinical setting and to compare the wear performance of different polyethylene inserts. All the methods of measurement of wear are based on quantifying the displacement (penetration) of the femoral head with reference to the acetabular component between two radiographs obtained at different times of follow-up (paired analysis), or on a single radiograph (single analysis) assuming a concentric position of the femoral head, and acetabular component directly after implantation before any wear could have occurred. Initially this was done by hand, on conventional radiographs using a compass, ruler and calliper, as described by Livermore, Ilstrup and Morrey, and more recently by Kang et al. These techniques can result in a high variability between different users, and they lack the precision to yield useful information over shorter periods or in cases in which relatively small amounts of displacement of the femoral head have occurred. In the 1990s computer-assisted techniques were developed to reduce the variability in measurement, to improve repeatability, and to increase the accuracy of measurement of penetration of the femoral head into the acetabular component compared with manual techniques. These programs combine the use of image analysis software with the determination of bone and prosthesis landmarks by hand or automated by a computer using an edge detection algorithm. Computer methods can determine wear two-dimensionally on anteroposterior (AP) radiographs, such as in the manual methods, or some can calculate three-dimensional wear.
vectors combining information from matching AP and cross-table lateral radiographs. All geometrical methods to assess wear only measure linear penetration. When volumetric wear is given, values are estimated from the linear penetration and a formula correlating both, based on simple geometric assumptions, substantiated in some cases by geometrical implant data.22

Standard radiological assessment of penetration by the femoral head into the polyethylene liner does not enable clinicians to distinguish between the two processes which cause movement of the head, namely true wear (the removal of polyethylene particles) and so-called bedding-in (creep and settling of the liner) which occurs in the early post-operative phase.11,23,24 In the literature, measurements of penetration are correlated directly with wear.

Röntgen stereophotogrammetric analysis (RSA), developed by Selvik,25 is considered the most accurate method for determining the magnitude of relative displacements, such as penetration of the femoral head, from radiological images.26,27 The requirement for tantalum beads to be placed into the implants and bone generally limits the use of this method to a small group of patients, and it can only be used in a prospective setting.

There is a tendency for cemented acetabular components to be used in elderly, less active patients with less exposure to the risks of polyethylene wear, and for uncemented acetabular components to be used in relatively young and still active patients with a higher chance of suffering the consequences of wear. Thus, polyethylene wear measurement is of greater importance in patients who have uncemented metal-backed components.

Several computer-assisted methods to determine linear penetration are commercially-available. From the available literature it is unclear whether these methods produce comparable results. The aim of our study was to find the easiest, most precise and most practical of four computer-assisted programs of wear measurement, and to see whether the values derived from different programs could be compared with each other.

**Materials and Methods**

We performed *in vivo* wear measurement by determining the linear penetration of the femoral head into the acetabular component on radiographs of a cohort of 24 patients who had undergone THR using uncemented hydroxyapatite (HA)-coated components (ABG-II, Stryker, Mahwah, New Jersey). In 13 patients conventional air-sterilised polyethylene inserts had been used and in 11 patients moderately cross-linked Duration polyethylene inserts (Stryker) were employed. The mean time of follow-up was 8.1 years (6.3 to 9.1). The primary diagnosis was osteoarthritis in all cases. The mean age at operation was 64 years (48 to 74). In 15 patients the left hip was replaced and in nine the right. An AP standing pelvic radiograph was obtained at six months and eight years post-operatively.

The six-month image was scanned using a flatbed scanner (VIDAR Sierra plus, Ampronix Inc., Irvine, California) at a resolution of 300 dots per inch. The result was a 20 megapixel (MPix) image (4375 × 5375 pixels). The eight-year digital image was made using a digital imager (AGFA Solo ADC, AGFA Gevaert NV, Peissenberg, Germany), producing a 5 Mpix image (4300 × 3500 pixels). The radiographs were originally DICOM and transformed into TIFF greyscale without compression, as required for the Martell software.

The penetration of the femoral head as a measure of polyethylene wear was assessed using four different computer-assisted methods: Martell Hip Analysis Suite version 7.14 (University of Chicago, Chicago, Illinois), HyperOrtho version 4.1 (Rogan Delft, Veenendaal, The Netherlands), View Pro-X version 2.0.1.11 (Rogan Delft), and Roman free to share software version V1.70 (Robert Jones & Agnes Hunt Orthopaedic Hospital, Oswestry, United Kingdom).28

In all four programs the edges of the femoral head and the metal-backed acetabular component were determined manually by mouse-clicking on the edges. In the Martell Hip Analysis Suite clicking the edges of the femoral head and acetabular shell is used to guide the automated edge detection.

The distance between the centre of the femoral head and the centre of the acetabular shell was after calibration of the image, using the known diameter of the femoral head to correct for magnification.

Only the AP radiograph was measured (2D). In general, 2D measurements have been considered to underestimate the 3D wear value.19,29-31 Although there is controversy as to whether to use standing or supine radiographs, we used only standing radiographs for all our measurements.32-37

**Martell Hip Analysis Suite version 7.14.** This software uses TIFF greyscale images only. Automated edge detection software calculates the displacement of the femoral head. The penetration of the head and the angle of penetration are reported as the wear vector (mm) and vector angle (°). The software uses a coordinate system which is similar to that of Livermore et al.13 The y axis is a line drawn through the left and right ischial tuberosities. The x axis is perpendicular to the y axis. The centre of the coordinate system is the centre of the acetabular component. Medial and lateral cranially directed wear values are described as positive. Value angles are described between -90° and +90°; laterally-directed angles are presented as a negative value. A 2D single measurement mode was chosen as the developer has reported that no accuracy is gained from the 3D analysis which the software also provides.29

**HyperOrtho version 4.1.** This software uses DICOM images only. The software calculates displacement of the femoral head by giving the cranial and medial displacement in millimetres according to the edge of the computer screen. The x axis is the horizontal edge of the image; the y axis is perpendicular to this line and thus independent of the patient’s...
anatomy or position on the image, as well as the acetabular orientation. Medial and cranial directions of wear result in positive values. A wear angle can be calculated applying the Pythagoras’ theorem.\textsuperscript{38} View Pro-X 2.0.1.11. This software uses DICOM images only. The software calculates femoral head displacement (mm) and the angle (°) as a representation of wear. The coordinate system used is different for each side of the patient. In the right hip, medial displacement is positive. Angles are given between -180° and +180°. In the left hip lateral displacement is positive and angles are also given between -180° and +180°. Roman free to share software version V1.70. This software accepts almost all image formats. We used TIFF greyscale images because they had to be converted in any case for the Martell method. Roman is a radiological measurement program designed for orthopaedic application in general, and the linear penetration of the head can be measured by applying a digital equivalent of the manual Livermore method using the compass and ruler function of the software. Circles of best fit are drawn around the femoral head and the acetabular component after identifying a minimum of three points for interpolation on the contours of both components, and using the ruler function to measure the displacement of the centres of each circle. An angle-measuring tool is provided in the software to quantify the angle of displacement of the head. The coordinate system can be chosen as preferred by the user by determining the x and y axes.

**Coordinate system.** As all four methods use different coordinate systems, it was necessary to transform them into one single system to compare the outcomes of the measurements. A cartesian coordinate system with reference to the acetabular component was considered more relevant for measurement of wear than those based on bony landmarks, which were more difficult to define and changed over time, or than the image itself, due to errors in positioning of the patient. The x axis was set as the line of inclination of the acetabular component as defined by its medial and lateral margins. The y axis was the perpendicular line on the x axis and travels through the centre of the acetabular component (Fig. 1).

The result of a measurement is a vector which consists of a wear distance measured as the penetration of the head in mm and a wear penetration angle defined as ranging between 0° and 360°, starting from the medial part of the x axis. In this way, vectors for the left and right hips can be compared and potential confusion about negative wear is eliminated.

**Measurements and calculation.** Intra-observer variability was determined for each method by determining the correlation coefficient of two measurements performed by one observer. Inter-observer variability was determined for each method by determining the correlation coefficient of two measurements performed by two different observers (CG, WV). Repeated measurements were performed after an interval of two weeks, and observers were blinded as to their first measurement. The average time to measure one AP radiograph was calculated for each of the methods. Correlation coefficients were determined to compare the rates of linear penetration of the head given by the four methods. The annual rates of penetration were calculated with single-image analysis using only the measurement of the eight-year radiographs, and in paired analysis using these at the six-month and the eight-year stage. Paired analysis was performed by combining the length of penetration and the direction in the six-month and eight-year images.

**Statistics.** Pearson’s correlation coefficient (r) was used to assess the agreement of the absolute values of linear penetration of the head in the four methods used. The intra-class correlation coefficient was used to address the intra- and inter-observer reliability of the four methods. We used SPSS 12.0 (SPSS Inc., Chicago, Illinois) for statistical analysis.

**Results**

When comparing the mean rates of linear penetration of the head measured for the entire group (Table I), both View Pro-X and Roman gave the same values, with a low standard deviation (SD) and small spread both for the single (0.12 mm/yr) and paired analysis (0.11 mm/yr). Hyper-Ortho gave the same mean rate in single-image analysis but a slightly higher result in paired analysis (0.15 mm/yr), but the range and SD were larger for both analyses. The Martell Hip Analysis Suite produced quite different results, with a single measure analysis being as good as the other methods and approximately three times greater for the paired ana-
yses. It also produced the greatest spread and SD of all the methods. Going from single-image analysis to paired analysis, the spread increased for all methods; in the Martell Hip Analysis Suite and HyperOrtho the SD also increased.

Intra-observer reliability for individual measurements of penetration was high and best for View Pro-X and Roman, with an intraclass correlation coefficient of 0.97 and 0.96, respectively in paired analysis at six months and eight years (n = 48). This was much lower for the Martell Hip Analysis Suite (0.33) and for HyperOrtho (0.46). The intra-class correlation coefficient was similar to these values when looking only at the eight-year images (n = 24), where total linear penetration is relatively high. Measuring the six-month images, when penetration is still low, the intra-class correlation coefficient was lower for all methods but remained highest for View-Pro-X and Roman (n = 24) (Table II).

Inter-observer reliability was generally less than intra-observer reliability, but was high and the best for View Pro-X and Roman, with an intra-class correlation coefficient of 0.87 for both at paired analysis at six months and eight years (n = 48). The inter-observer intra-class correlation coefficient was much lower for the Martell Hip Analysis Suite (0.39) and for HyperOrtho (0.43). Analysing only the eight-year images with high total penetration (n = 24, single-image analysis), the inter-observer intra-class correlation coefficient was higher for all methods than for paired analysis, with View Pro-X and Roman still providing the best results and the Martell Hip Analysis Suite and HyperOrtho remaining low. When looking at the six-month images only, when total penetration is still small, the inter-observer correlation was very low for all methods.

Comparing individual measurements of linear penetration of the head between the different programs showed a good correlation only between the Roman and View Pro-X for all measurements combined and for the eight-year images only (Pearson’s r = 0.90 and 0.91, respectively). Measuring penetration of the head at six months showed a low correlation between all four programmes (Pearson’s r = -0.13 to 0.30) (Table III).

For paired analysis correlation was fair between the Roman and View Pro-X (Pearson’s r = 0.61), but a low correlation was found between the other programs (Pearson’s r = 0.05 to 0.38).

The results of paired analysis and those of single-image analysis of the final follow-up at eight years for distance of penetration and the penetration angle were correlated (Table IV). Correlation between the distance measurements was high for View Pro-X (intraclass correlation coefficient 0.90) and Roman (intraclass correlation coefficient 0.87), and lower for Martell (intraclass correlation coefficient 0.43) and HyperOrtho (intraclass correlation coefficient 0.49). The correlation of measurements of the penetration angle was low for all programs.

Evaluation of the handling features in clinical practice revealed that the Roman software was best, apart from taking the longest time to make the measurements, with a mean of 2.63 minutes (2.51 to 2.72). The Roman system can handle all common file formats and image qualities. Uniquely, it has the option to save images and analysis for later review or sharing, offers the adjustment of contrast and brightness, has a zoom function, and is free to download from the internet (Table V).

**Discussion**

We studied the variation in the measurement of the distances and angles of linear penetration of the femoral head into the polyethylene liners of acetabular components as a
measure of wear in THR using four computer-assisted techniques. All these methods measured the distance between the centre of the femoral head and the centre of the metal acetabular shell with an accuracy and error limited partly by the individual pixel size. The digital images in this study were produced either by scanning the acetate radiographs at six months to provide images of 20 MPix, or at eight years by a digital imager, leading to images of 5 MPix. For an AP radiograph of the pelvis at six months this resulted in a pixel size of approximately 0.07 mm, with 0.14 mm for the eight-year images setting a methodological limit to the accuracy of the measurement. In order to reduce error of measurement by 50%, an image resolution four times greater is required. As this effect was the same for all four

| Table III. Correlation coefficients (Pearson’s r) between the linear penetration results for each method |
|---------------------------------------------------|-----------------|-----------------|
| Correlation coefficient (Pearson’s r) |
|         | Martell | Roman | HyperOrtho |
| 6 months (n = 24) |       |       |             |
| View Pro-X | 0.30  | 0.03  | 0.26        |
| Martell    | -0.13 |       | 0.19        |
| Roman      |       |       | 0.13        |
| 8 years (n = 24) |       |       |             |
| View Pro-X | 0.34  | 0.91  | 0.24        |
| Martell    | 0.33  |       | 0.36        |
| Roman      |       |       | 0.16        |
| Total (n = 48) |       |       |             |
| View Pro-X | 0.38  | 0.90  | 0.26        |
| Martell    | 0.32  |       | 0.30        |
| Roman      |       |       | 0.20        |
| Paired (n = 24) |       |       |             |
| View Pro-X | 0.27  | 0.61  | 0.05        |
| Martell    | 0.15  |       | 0.38        |
| Roman      |       |       | 0.17        |

| Table IV. Correlation of eight-year linear penetration distance and angle in paired vs single analysis |
|---------------------------------------------------|-------------------|
| Intra-class correlation |
| Linear penetration | Penetration angle |
| View Pro-X | 0.90  | 0.68 |
| Martell    | 0.43  | -0.11|
| Roman      | 0.87  | 0.58 |
| HyperOrtho | 0.49  | -0.15|

<table>
<thead>
<tr>
<th>Table V. Evaluation of software</th>
</tr>
</thead>
<tbody>
<tr>
<td>View Pro-X</td>
</tr>
<tr>
<td>---------------------------------------------------</td>
</tr>
<tr>
<td>Image requirement</td>
</tr>
<tr>
<td>Automated edge detection</td>
</tr>
<tr>
<td>Saving the image measured</td>
</tr>
<tr>
<td>Saving measurement results</td>
</tr>
<tr>
<td>Contrast/brightness adjustment</td>
</tr>
<tr>
<td>Region of interest zoom function</td>
</tr>
<tr>
<td>Mean time to measure one image (min)</td>
</tr>
<tr>
<td>Cost of the program</td>
</tr>
</tbody>
</table>

* PACS, Picture archive and communications system
methods, it will not influence the comparison of the rates of linear penetration produced by each method. With a pixel size of approximately 0.14 mm in a 5 MPix AP radiograph and the opinion that clinical linear penetration rates become important at approximately 0.05 to 0.10 mm/year, the image resolution for studies of clinical wear should not be less.

Comparing the annual rates of linear penetration, three of the four methods gave the same mean value and similar SD in single-image analysis. Only the Martell Hip Analysis Suite produced a much higher mean rate of linear penetration, as well as the largest SD and the widest range compared with the other three methods (Table I). The intra- and inter-observer reliabilities were also very low for this technique (Table II), as well as the correlations with the other programs (Table III). This disappointing result for the Martell method is in contrast to previous studies and suggests a systematic error, which we believe was caused by the inability of the automatic edge detection algorithm to correctly identify the mean between the femoral head and the polyethylene insert on direct digital images, which are less smooth than the scanned conventional radiographs used in the past. This explanation was validated when artificial smoothing of the direct digital images to emulate the appearance of scanned images improved the accuracy of edge detection and the reliability of the measurements of penetration. However, such image post-processing may alter the diagnostic quality of the radiograph, is time-consuming, not standardised, and too complex to be recommended for routine measurements of clinical wear in practice, where direct digital images are increasingly becoming the standard format. Thus, Martell analysis was not repeated on the manipulated images. Methods relying on the observer to identify the edge were not affected by the difference between scanned and direct digital images, as the human eye can easily identify the borders of the femoral head and the acetabular component on both image types.

Although Roman and View Pro-X produced measurements with the lowest SD, the best intra- and inter-observer reliability and the best correlations between the four methods, HyperOrtho performed less well regarding these parameters, despite giving identical mean rates of linear penetration in the single-image analysis. The reason for this is that HyperOrtho interpolates ellipses instead of circles, to be capable of measuring non-metal-backed acetabular components. Elliptical interpolation causes large shifts of the centre of circles, reducing intra- and inter-observer reliability as well as the correlation of individual measurements with other methods.

In theory, paired analysis of penetration of the femoral head should give more accuracy than single-image analysis, as in paired analysis direct observation of the post-operative position of the head and the acetabular component is compared with their position at follow-up. In single-image analysis the head and the acetabular component are assumed to be concentric, with an initial displacement of zero. However, the single six-month measurement showed that for all methods the standard deviation was high and intra- and inter-observer reliability was low (Table II). Displacement of the femoral head at six months is not dominated by wear but distributed around zero, with an SD so high that it can only influence the accuracy of paired analysis in an adverse manner, resulting in a lower overall accuracy of a paired over single image analysis.

In this study, View Pro-X and Roman were the most accurate and reliable methods, producing nearly identical average rates of linear penetration with equally low SD and ranges in both paired and single-image analyses. The intra- and inter-observer reliability and the correlation between the two methods were better in single-image analysis, but only for the single eight-year measurement and not for those at six-months.

Measurements at six months had very low correlation coefficients even between the most accurate methods (Table III), proving that the baseline measurement is too erratic to improve on the assumption of zero in a single-image analysis. In addition, The et al have shown that a significant error in measurement in a paired analysis is introduced by inevitable differences in the positioning between two follow-up radiographs, leading to inaccurate measurement of wear up to 0.4 mm. Although correction algorithms can compensate for this error, it is currently not implemented in any commercial software. Considering the positioning error for paired images and the low accuracy of the baseline measurement, it seems that in clinical practice single-image analysis can be sufficient or even superior.

Measurement of the angles of penetration of the head was far less reliable than the measurement of the distance of penetration (Table IV). With linear penetration often mea-
suring only a few multiples of the pixel size, the angle between two points can alter more in repeated measures than can the distance. As there are various definitions of wear angles and wear directions in the literature, comparison between published values is more difficult than for wear distance.

The definition of wear angles in this study and applied to the measured angles of penetration avoids the misleading term ‘negative wear’. In a recent publication of Wan, Bou-
tary and Dorr,42 almost 50% of measurements performed with the Martell Hip Analysis Suite32 gave negative wear values and were attributed to the errors in mea-
surements.1 However, negative wear distance or negative wear angles are only a function of the definition of the coordinate system, and these differ between the software methods used. Our coordinate system defines a vector con-
sisting of a positive length and a positive angle ranging between 0° and 360°. If, on the AP radiograph, the centre of the femoral head is located outside the acetabular com-
ponent, this will not be the result of negative wear but can occur for several reasons including positioning of the patient in the radiation field,41 the linear penetration path-
way pushing the head out of the insert, radiographs acquired with the patient supine, with microseparation between the head and acetabular component and owing to the design of the acetabular component, which may be eccentric or hooled. The tolerances in the manufacturing might lead to an eccentric baseline position of the head and insert. An error could arise in measurement itself, especially when penetration is very small; and finally, there may be 2D projection of a 3D penetration pathway (Fig. 2).

Computer-aided measurement of linear penetration of polyethylene can be accurate and reliable within the limits of image resolution, and a 5 MPix pelvic radiograph should be the minimum requirement. Edge detection by a human observer is more dependable with the new direct digital images than with automated software, and circular inter-
polation is superior for metal-backed components. We found that the Roman and View Pro-X software were the best for the assessment of linear penetration of the head as a measure of polyethylene wear in clinical practice. Although measurement with the Roman software takes slightly more time (Table V), its compatibility with all com-
mon image formats and its free availability make it an attractive option.

No benefits in any form have been received or will be received from a commer-
cial party related directly or indirectly to the subject of this article.

References

1. Cooper RA, McCullister CM, Borden LS, Bauer TW. Polyethylene debris-induced oste-
olysis and loosening in uncemented total hip arthroplasty: a cause of late failure. J Arthro-
8. Schmalzried TP, Jasty M, Harris WH. Periprosthetic bone loss in total hip arthro-
10. Orihime KF, Claus AM, Sychtzer CJ, Engh CA. Relationship between polyethyl-
11. Dowd JE, Sychtzer CJ, Young AM, Engh CA. Characterization of long-term femo-
13. Livemore J, Istrup DM, Morrey B. Effect of femoral head size on wear of the poly-
15. Kabo JM, Gehbard JS, Loren G, Amstutz HC. In vivo wear of polyethylene acetab-
18. Martell JM, Berdia S. Determination of polyethylene wear in total hip replace-
20. Devane PA, Bourne RB, Rorabeck CH, MacDonald S, Robinson EJ. Measure-
21. Shaver SM, Brown TD, Hillis SL, Callaghan JJ. Digital edge-detection measure-
23. Sychtzer C, Engh CA Jr, Yang A, Engh CA. Analysis of temporal wear patterns of porous-coated acetabular components: distinguishing between true wear and so-
24. Sychtzer CJ, Engh CA Jr, Shah N, Engh CA Sr. Radiographic evaluation of pene-
Keele.ac.uk/depts/rjah/ (date last accessed 21 May 2008).
29. Martell JM, Berksen E, Berger R, Jacobs C. Comparison of two and three-dimen-
sional radiographic techniques for measuring polyethylene wear after total hip arthro-
31. Sychtzer CJ, Yang AM, MacAuley JP, Engh CA. Two-dimensional versus three-


