A comparison of four systems for calibration when templating for total hip replacement with digital radiography

M. Franken, 
B. Grimm, 
I. Heyligers

From Atrium Medical Centre, Heerlen, The Netherlands

We have investigated the accuracy of the templating of digital radiographs in planning total hip replacement using two common object-based calibration methods with the ball placed laterally (method 1) or medially (method 2) and compared them with two non-object-based methods. The latter comprised the application of a fixed magnification of 121% (method 3) and calculation of magnification based on the object-film-distance (method 4). We studied the post-operative radiographs of 57 patients (19 men, 38 women, mean age 73 years (53 to 89)) using the measured diameter of the prosthetic femoral head and comparing it with the true value.

Both object-based methods (1 and 2) produced large errors (mean/maximum: 2.55%/17.4% and 2.04%/6.46%, respectively). Method 3 applying a fixed magnification and method 4 (object-film-distance) produced smaller errors (mean/maximum 1.42%/5.22% and 1.57%/4.24%, respectively; p < 0.01). The latter results were clinically relevant and acceptable when planning was allowed to within one implant size. Object-based calibration (methods 1 and 2) has fundamental problems with the correct placement of the calibration ball. The accuracy of the fixed magnification (method 3) matched that of object-film-distance (method 4) and was the most reliable and efficient calibration method in digital templating.

Pre-operative templating is useful in planning total hip replacement (THR) and is an effective tool for training young surgeons. In addition to deciding pre-operatively the type and size of prosthesis likely to be required for anatomical restoration,1 it is thought to reduce intraoperative complications since it helps to anticipate challenges such as the resection of osteophytes, the need for a trochanteric osteotomy, bone grafting and the use of acetabular reinforcement devices.2

Traditionally, templating has been performed using transparent acetate sheets with images of the prosthesis at a fixed magnification of typically 115% or 120%, which were manually superimposed on the radiographs. Prospective studies using this method to compare the agreement between the pre-operative assessment and the size of the component implanted showed clinically acceptable results.3 The size of the acetabular component was predicted exactly in 83% of all procedures and to within ± one size in 99%, with that of the femoral component in 78%.

With the introduction of electronic Picture and Communication Systems (PACS), radiographs are no longer available as acetate films at a fixed image size. In a PACS system digital radiographs can be produced from scanned acetates, computed radiography or direct digital radiography systems. Such digital images are displayed on a computer screen of any size and can be freely scaled without reference to the true dimensions. Thus, fixed magnification templates are no longer suitable. In digital pre-operative planning, the digital image must be calibrated in order to scale the dimensions shown on the radiograph and the digital templates.4 Software manufacturers have developed various methods of calibration all of which are based on the placing of an object of known dimensions near the region of interest. In theory this should ensure the highest possible accuracy and should exceed the levels attainable by applying one fixed magnification to all images as was previously done with analogue templating. Calibration with an object at the region of interest should overcome problems resulting from intrapatient variability because of inconsistencies in the positioning of the patient for radiography or to obesity. However, the reliability and clinical relevance of digital templating remain unproven.

A validated geometric model of calculating the calibration of digital radiographs using spheres of known dimensions has shown that...
the correct magnification factor can theoretically be derived within a ± 1.5% maximum margin of error. In the clinical setting the 95% confidence interval for the determination of the correct magnification found an error of ± 3%. Using a coin as an alternative calibration object one study found a mean error of ± 1.12% while another identified a mean undersizing of 0.9%, but had a lower reproducibility than that using conventional templates on analogue radiographs. A comparison of the use of a steel ball as a calibration object with measurement of the object-film-distance revealed a similar accuracy for the two methods, with a mean absolute measurement error of 2.6% for the calibration object and 2.8% for measurement of the object-film-distance (p = 0.75).

In two studies the size of the implant was compared with the size digitally templated in retrospect on post-operative radiographs by a blinded observer. In the first study the correct sizing was found in 27% of the femoral components with 79% to within ± one size and 23% of acetabular components with 61% to within ± one size, with low intra- and interobserver reliability values. In the second study these values were similarly poor with 72% of stems and 86% of acetabular components digitally templated to within ± one size. In a prospective study comparison of the size of the implanted prosthesis with that pre-operatively templated, using both analogue and digital techniques, showed a small preference for analogue planning.

We have undertaken an investigation of two commonly used object-based and two self-developed non-object-based methods of calibration for digital templating to explore their reliability and to identify how common mistakes in using calibration objects in clinical practice can be prevented. Our study also aimed to quantify the threshold for the error of calibration which is clinically acceptable.

**Patients and Methods**

We performed a prospective study of four different methods of calibration in which the post-operative antero-posterior (AP) pelvic radiographs of 57 patients were studied. The study group comprised 38 women and 19 men with a mean age of 73 years (53 to 89) who had received either an ABG-II (Stryker, Kalamazoo, Michigan), Exeter (Stryker) or a Bi-metric (Biomet, Bridgend, United Kingdom) THR with a femoral head of 28 mm in diameter. The radiographs were taken with the patient standing using a Philips Diagnost H bucky system (Philips Healthcare, Eindhoven, The Netherlands) with a fixed focal distance of 1150 mm with the image captured on to a plate 430 mm × 350 mm in size. The images were digitised by phosphor plate technology (ADC Solo reader, AGFA, Wilmington, Massachusetts) to a resolution of 5.1 megapixels and sent to PACS. The radiographers had all been trained in the correct placement of the calibration objects. Image calibration and comparison between the calibrated and the true magnification were performed (by MF, BG) using the digital pre-operative planning software Endomap VA20A (Siemens/Hectec GmbH, Erlangen, Germany). Since the process and problems of calibrating digital radiographs are inherent to the digital image format and are the same regardless of whether they are derived from a scanned acetate or a computed or digital radiography system, our study used images from a computed system only. The patients’ height, weight and body mass index (BMI) were noted to avoid confounding factors of pre-operative templating and offered the availability of the prosthetic head as a calibration reference, avoiding systematic errors which can arise from problems such as rotation of the femur.

**Method 1, calibration ball, positioned laterally.** Most commercial software for pre-operative planning of THR incorporates the use of a metal ball of known diameter, 30 mm in our study, as a calibration object placed lateral to the patient at a height of the greater trochanter, which can be easily palpated and is close to the region of interest. In our study the calibration ball was fixed to a moveable stand so that it could be accurately adjusted in the mediolateral (AP and ML) directions of the standing patient. With the templating software the calibration object was enlarged to maximum size to increase accuracy when a circle was interpolated from three points selected on the margin of the ball using the cursor. The diameter of the circle was matched to the true dimensions of 30 mm and the software scaled the image to millimetres per pixel.

**Method 2, calibration ball positioned medially.** The method was identical to method 1 except that the calibration ball was placed medially between the patient’s legs so that it was positioned centrally on the image.

**Method 3, fixed magnification.** For this method the fixed magnification was derived from a pilot study using the digital AP pelvic radiographs of ten patients who had undergone THR. Our method of radiography with a fixed-focus-film-distance of 1150 mm produced a mean magnification of 121% when comparing the measured diameter of the head of the femoral component on the scanned film with the true value of 28 mm. This value was used as a constant magnification factor for all images used in our study.

**Method 4, individually measured object-film-distance.** This method required the manual measurement of the object-film-distance for each patient. The distance between the region of interest, the palpable greater trochanter, and the x-ray plate was measured using a simple caliper. The focus-plate distance of 1095 mm and the plate-film distance of 55 mm were known fixed specifications. From the known and measured distances the magnification for each exposure was calculated.

After calibration, the diameter of the prosthetic femoral head was measured using each of the methods described above and compared with the real value of 28 mm to calculate the calibration error (Fig. 1).
**Statistical analysis.** The accuracy of the calibration was evaluated by calculating the following errors: 1) the absolute error (AE = DC - DR) defined as the difference between the prosthetic femoral head diameter measured after image calibration (DC) and the real value DR (28 mm); 2) the relative error (RE = (DC - DR)/DR) defined as the ratio of the absolute error and the real dimension DR; 3) the minimum (RE_{\text{min}}) and maximum (RE_{\text{max}}) values of the relative error; and 4) the SD of the relative error. These error values were analysed in four ways as follows: 1) statistical comparison of the different methods using a two-tailed Mann-Whitney test; 2) the relationship to the patient’s weight, height and BMI using Spearman’s correlation coefficient; 3) the best and worst performing method was identified for each individual radiograph; and 4) comparison with the clinically required error limits for correct sizing and positioning calculated from the incremental sizes of the femoral and acetabular components of the ABG-II prosthesis.

**Results**

The different methods of calibration for the 57 images produced a range of 26.38 mm to 32.87 mm for the size of the 28 mm modular femoral head, with the minimum and maximum values both found using method 1 with the calibration ball placed laterally (Table I). With this method the image was oversized with a mean relative error of 2.55% and an (RE_{\text{max}}) of 17.40%. The relative error was significantly higher than that with method 3 (fixed magnification; relative error 1.42%, Mann-Whitney test, p = 0.004) and method 4 (object-film-distance; relative error 1.57%, Mann-Whitney test, p < 0.001). In method 1 the laterally placed calibration ball was often incompletely visible on the image and distorted to an oval shape due to the divergence of the x-ray beam caused by obesity. As a result, 12 of the 57 images were deemed to be unmeasurable and had to be excluded from the error calculations. Inclusion of these images would have worsened the accuracy of method 1 even further. No correlation of error with clinical factors was found (BMI $r = -0.02$, weight $r = -0.15$, length $r = -0.18$), although exclusion of unmeasurable images from very obese patients probably hid this effect.

Method 2 (calibration ball placed medially) gave slightly better results than method 1 with a mean diameter (27.78 mm vs 28.45 mm) significantly closer to the real value (Mann-Whitney test, $p < 0.01$). Both the mean relative error (2.04%) and the (RE_{\text{max}}) (6.46%) were less (Mann-Whitney test, $p < 0.01$). While method 1 generally underestimated the true dimensions, method 2 mainly caused overestimation. The error distribution of method 2 was the only one which showed two peaks towards under- and overestimating dimensions. At the same time, the relative error was significantly correlated with body-weight ($r = -0.30$, $p < 0.03$).

With method 3 the mean diameter of the femoral head as measured was equal to the exact diameter of 28 mm, a significantly better estimation than with method 1 (Mann-Whitney test, $p = 0.004$). Also the mean relative error (1.42%) and RE_{\text{max}} (5.22%) were less than those with method 1 (2.55% and 17.40%; Mann-Whitney test, $p = 0.004$) and method 2 (2.04% and 6.46%; Mann-Whitney test, $p = 0.11$). The relative error was significantly correlated with the BMI ($r = 0.51$, $p < 0.01$) and weight ($r = 0.27$, $p = 0.04$), causing undersizing of the templated implants for slim and oversizing for obese patients.

Method 4 produced estimations for the 28 mm femoral head in a similarly small range as was found with method 3 (26.81 mm to 28.93 mm). Also the mean relative error (1.57%) and RE_{\text{max}} (4.24%) were similar to those with method 3 (Mann-Whitney test, $p = 0.14$) and accordingly better than those with methods 1 and 2. The relative error of method 4 was correlated with height ($r = -0.29$, $p = 0.03$).

Finally a comparison of the most and least accurate calibration method for each individual radiograph was undertaken (Table II). Method 3 performed best in 21 (37%) cases followed by method 4 (15 cases, 26%), together almost twice as many as method 2 (12 cases, 21%) and method 1 (9 cases, 16%). Method 1 gave the worst measurement in 26 cases (45%). Exclusion of method 1 left method 2 as the worst in 27 cases (47%). The individual best method was selected from the two remaining methods, method 4 (object-film-distance) 33 cases (58%) and method 3 (fixed magnification) 24 cases (42%). Methods 3 and 4 produced the best results and smallest errors. The clinically tolerable margin of error depends on the steps between implant sizes within the system being considered for implantation, the parameter to be templated, such as the size of the acetabular component or length of the femoral neck, the absolute dimension to be measured, such as large stem length or shortstem width and the sizing error that a clinician would...
A COMPARISON OF FOUR SYSTEMS FOR CALIBRATION WHEN TEMPLATING FOR TOTAL HIP REPLACEMENT WITH DIGITAL RADIOGRAPHY

Tolerate, for example the exact size or width ± one size. Table III gives these values for the ABG-II implant.

Table III. Required limits of error for exact and within ± one size templating of several sizing parameters

<table>
<thead>
<tr>
<th>ABG-II cementless standard components</th>
<th>Size (mm)</th>
<th>Difference in size steps (mm)</th>
<th>Required error limit for exact size (small components) (%)</th>
<th>Required error limit for exact size (large components) (%)</th>
<th>Required error limit for method one size (small components) (%)</th>
<th>Required error limit for method one size (large components) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetabular</td>
<td>40 to 60</td>
<td>2.0</td>
<td>2.50</td>
<td>1.67</td>
<td>750</td>
<td>5.00</td>
</tr>
<tr>
<td>Stem length</td>
<td>100 to 145</td>
<td>5.0 to 10.0</td>
<td>2.50</td>
<td>3.45</td>
<td>750</td>
<td>10.34</td>
</tr>
<tr>
<td>Neck offset</td>
<td>372 to 53.2</td>
<td>1.1 to 3.2</td>
<td>1.48</td>
<td>3.01</td>
<td>4.44</td>
<td>9.02</td>
</tr>
<tr>
<td>Neck height</td>
<td>27 to 39</td>
<td>1.0 to 3.0</td>
<td>1.85</td>
<td>3.85</td>
<td>5.56</td>
<td>11.54</td>
</tr>
<tr>
<td>Distal diameter</td>
<td>8.5 to 15.8</td>
<td>0.7 to 1.5</td>
<td>4.12</td>
<td>4.75</td>
<td>12.35</td>
<td>14.24</td>
</tr>
<tr>
<td>Width proximal</td>
<td>21 to 28</td>
<td>1.0</td>
<td>2.38</td>
<td>1.79</td>
<td>7.14</td>
<td>5.36</td>
</tr>
</tbody>
</table>

The acceptable limits of error were the lowest for the neck offset and neck height of small stems, the diameter of large acetabular components and the width of large stems. To template exactly for a specific implant size would require error limits < 2%. If two sizes were considered to be clinically acceptable, the error limit increased to approximately 5%.

Table I. Measured ball diameters and errors for the different methods of calibration

<table>
<thead>
<tr>
<th>Method</th>
<th>Calibration object</th>
<th>No calibration object</th>
<th>Calibration object</th>
<th>No calibration object</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure values (mm)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Mean</td>
<td>28.45</td>
<td>27.78</td>
<td>28.00</td>
<td>27.87</td>
</tr>
<tr>
<td>SD</td>
<td>1.00</td>
<td>0.65</td>
<td>0.54</td>
<td>0.52</td>
</tr>
<tr>
<td>Maximum</td>
<td>32.87</td>
<td>29.11</td>
<td>29.46</td>
<td>28.93</td>
</tr>
</tbody>
</table>

* OFD, object-film-distance

Table II. The best and worst success rates for all four methods, by number and percentage

<table>
<thead>
<tr>
<th>Method</th>
<th>Individual best (%)</th>
<th>Individual worst (%)</th>
<th>Individual worst excluding individual best of lateral object (%)</th>
<th>Individual best of methods 3 and 4 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Lateral)</td>
<td>9 (16)</td>
<td>26 (46)</td>
<td>27 (47)</td>
<td>26 (45)</td>
</tr>
<tr>
<td>2 (Medial)</td>
<td>12 (21)</td>
<td>14 (24)</td>
<td>14 (24)</td>
<td>24 (42)</td>
</tr>
<tr>
<td>3 (Fixed magnification)</td>
<td>21 (37)</td>
<td>9 (16)</td>
<td>17 (29)</td>
<td>33 (58)</td>
</tr>
<tr>
<td>4 (object-film-distance)</td>
<td>15 (26)</td>
<td>9 (16)</td>
<td>17 (29)</td>
<td>33 (58)</td>
</tr>
</tbody>
</table>

Discussion

Large errors were sometimes found depending on the method used with the least accuracy provided by method 1 (a laterally placed calibration ball). Application of this method resulted in 12 radiographs (21%) which did not show the calibration ball since it was outside the field of view because of obesity or a suboptimally framed image. Even when these 12 images were excluded from further analysis the results were still unacceptably poor producing errors of up to 17%, an equivalent of up to four stem sizes in the ABG system. One reason for this error is that laterally placed objects are only partially visible reducing the area which can be used for calibration. Another is that the divergence of the x-ray beam causes the laterally placed objects, whether fully or partially visible, to project as an ellipse while the calibration routine requires a circular image. Contralaterally placed calibration balls produce an additional error margin of the order of 1% and can reduce accuracy even further. A properly centred radiograph of only the hip instead of a full pelvic radiograph would reduce the exclusion rate and the error of method 1, possibly to levels similar to those found for method 2, but for certain steps in pre-operative planning, such as leg-length correction, a full pelvic radiograph is required.

In method 2 the medial ball was visible and centrally located in all 57 images, but its correct positioning was more difficult since no clear anatomical landmark was available. The radiographer must palpate the trochanter to identify its level and transfer, with error, this level to the groin. However, the proximity to the genitalia makes this step difficult for patients and staff. This seems to have resulted in either a more dorsal or frontal placement of the ball as revealed by the two-peak distribution of error found. The correlation of relative error with weight showed that obesity further complicated the correct
positioning. Method 3 (fixed magnification) was by far the easiest to implement and is the standard in analogue templating. Thus it seems a good option for the calibration of digital radiography in clinical practice. This method gave results which were significantly better than those with method 1 and method 2. Fixed magnification was comparable with method 4 which required the radiographer to measure and document the object-film-distance for each radiograph. Although the error in the fixed magnification system was the smallest of all the methods, it correlated with the patient’s BMI. This indicated that the position of the hip changes in the x-ray field when patients become more obese. A correction factor based on this correlation could reduce such an error. In our study a fixed magnification of 121% was used. This value depended on the settings of the specific radiology equipment and the techniques used in the radiology department such as standing or lying and the use of support cushions, and must be determined individually for each institution and each anatomical region of interest.

Method 4 gave results which were also better than both methods using calibration objects but was not superior to method 3. Although the object-film-distance is measured for each patient which allows for differences in patient anatomy, positioning and radiological set-up, method 4 still failed to demonstrate better accuracy than method 3. One reason may have been that measurements of the object-film-distance tend to overestimate the true distance. The correlation of the relative error with patient height may be further evidence for this effect. Recalculating the results after subtracting 5 mm as compensation for the detector thickness from the measured object-film-distance improved the mean diameter measurement from 27.87 mm to 28.02 mm and reduced the mean error from 1.57% to 1.46%, but the result was still not better than that of method 3.

Wimsey et al. used a coin as a calibration object and reported a mean accuracy (mean relative error) within 1.12% (0% to 2.38%). This was better than the best results in our study (1.42%), but our images were taken standing, as opposed to lying, to allow leg-length correction. Standing provides a less standardised and less stable position than lying and thus would probably produce larger errors even when using a coin for calibration. The use of a coin has some advantages over a ball since it can be taped to the patient’s skin near the region of interest with less likelihood of relative movement between the calibration object and patient after palpation of the landmark. Any posteroanterior shift between the calibration object and the region of interest is more critical for determination of the magnification than a mediolateral or caudal-cranial shift which only produces distortion. However, the ball as a calibration object still seems to be better than a ruler which is provided with some calibration software. Without rotational symmetry like a coin or ball, the ruler suffers from projection errors because of rotation in the direction of the x-ray beam.

Investigation of the dimensions between different sizes of acetabular and femoral component show that the smallest relative size increment was found when two different femoral neck offsets were compared for small stem sizes. Correct sizing would require an error limit of 1.48%. This shows that calibration errors are more critical with regards to anatomical reconstruction (e.g. leg-length corrections, soft-tissue tension) but less with regards to fit and fill of the stem (stem length, width and positioning). If a difference within ± of one component size is accepted when templating, the limit of the relative error for calibration increases to 4.4% for the ABG II system.

None of the four methods of calibration was consistently able to measure within the required lowest limit of the relative error (1.48%). Increasing this to tolerating differences of within ± of one component size make methods 3 and 4 suitable for reliable templating.

Image calibration became a critical subject in radiography when fixed-size acetate radiographs were replaced by digital images which could be freely enlarged on any size of screen without dimensional reference. Thus the principle problem and findings of our study apply to any radiograph in digital format independent of its original source from scanned acetate, computed or digital radiography. In some modern digital radiograph systems the theoretical pixel size is calculated based on the focus-table-plate distances and stored into the header information of the DICOM file. Digital templating software could use this information to calibrate the image automatically. However, the pixel size calculated in this manner is a fixed magnification based on the set-up of the apparatus only. It does not consider the application-specific issues such as the region of interest (hip or knee), or patient positioning which is accounted for in the fixed magnification calculated in our study by averaging the true magnification over a pilot series.

We have shown that calibration of digital images with a reference ball does not necessarily lead to higher accuracy in digital templating when compared with fixed magnification as used with the analogue technique. In clinical practice a fixed magnification leads to more accurate results because fundamental problems with the correct placement of the calibration object are avoided. Fixed magnification reliably produces clinically relevant accuracy if within ± of one component size is tolerated when templating. For exact-size templating (± zero component size), even fixed magnification is not always accurate enough for templating certain dimensions such as the neck offset of small stems. Software manufacturers should encourage fixed magnifications to be easily set in their programs.

We thank Mr H. Saes and Mr M. Junghans and the Department of Radiology at the Atrium Medical Centre for their valuable contribution to our investigation. We also wish to thank the Department of Medical Physics which helped, supported and facilitated our study.

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

THE JOURNAL OF BONE AND JOINT SURGERY
References