Adherence of hip and knee arthroplasty studies to RSA standardization guidelines
A systematic review

Rami Madanat1, 3, Tatu J Mäkinen1, Hannu T Aro2, Charles Bragdon3, and Henrik Malchau3

Background and purpose — Guidelines for standardization of radiostereometry (RSA) of implants were published in 2005 to facilitate comparison of outcomes between various research groups. In this systematic review, we determined how well studies have adhered to these guidelines.

Methods — We carried out a literature search to identify all articles published between January 2000 and December 2011 that used RSA in the evaluation of hip or knee prosthesis migration. 2 investigators independently evaluated each of the studies for adherence to the 13 individual guideline items. Since some of the 13 points included more than 1 criterion, studies were assessed on whether each point was fully met, partially met, or not met.

Results — 153 studies that met our inclusion criteria were identified. 61 of these were published before the guidelines were introduced (2000–2005) and 92 after the guidelines were introduced (2006–2011). The methodological quality of RSA studies clearly improved from 2000 to 2011. None of the studies fully met all 13 guidelines. Nearly half (43) of the studies published after the guidelines demonstrated a high methodological quality and adhered at least partially to 10 of the 13 guidelines, whereas less than one-fifth (11) of the studies published before the guidelines had the same methodological quality. Commonly unaddressed guideline items were related to imaging methodology, determination of precision from double examinations, and also mean error of rigid-body fitting and condition number cutoff levels.

Interpretation — The guidelines have improved methodological reporting in RSA studies, but adherence to these guidelines is still relatively low. There is a need to update and clarify the guidelines for clinical hip and knee arthroplasty RSA studies.

Since the introduction of radiostereometry (RSA) in 1974, the method has continuously proven to be a valuable tool in the assessment of migration following hip and knee arthroplasty. Although the importance of a stepwise introduction of new orthopedic implants was presented in 1995 (Malchau 1995), only recently has it been fully acknowledged that deletion of some of these crucial steps can have catastrophic consequences (Kärrholm 2012). The phased or stepwise introduction process is based on the hypothesis that a more precise and careful evaluation of new implants will reduce the number of patients at risk of unexpected failures. There are 4 suggested steps to this process (Malchau 1995). The initial step involves preclinical implant testing. After this, the first clinical step consists of prospective randomized studies that use RSA. The second clinical step involves multicenter studies and the final step is composed of registry studies.

The detection of early migration or accelerated wear using RSA are ways of predicting early implant failure, and are therefore important components of this stepwise introduction process (Kärrholm 2012). Although RSA would probably not have been able to forecast some of the problems related to the metal-on-metal articulation, it could have prevented the premature introduction of Boneloc cement. When Boneloc was introduced, the Norwegian Arthroplasty Register needed more than 1,000 patients before it could be proven to be an inferior cement based on a higher revision rate compared to regular bone cement (Havelin et al. 1995). Using RSA only, 30 patients were included in a study that—already after 6 months—showed that Boneloc-cemented implants were clearly less stable than those with conventional bone cement (Thanner et al. 1995). If Boneloc had been studied with RSA prior to being released onto the market, thousands of patients would have been spared from early revision. There are several other examples of poor orthopedic implants or materials that should have been tested with RSA prior to widespread release (Muirhead-Allwood 1998, Norton et al. 2002). In the past, many serious problems initially manifested as low-incidence events but only later proved to be the tip of the iceberg (Mal-
chau et al. 2011). Despite the fact that the algorithm for introduction of new implant designs was established more than 15 years ago, many new and untested implants still appear on the market. The orthopedic community has acknowledged that the fast spread of undocumented new implants is no longer acceptable (Kärholm 2012).

With the increasing importance of RSA, it has also been recognized that reporting of RSA outcomes should be standardized. The guidelines for standardization of radiostereometry of implants were published in August 2005, to facilitate comparison of RSA outcomes reported from different centers (Valstar et al. 2005). In that paper, an appendix of 13 main items that should be used to account for the results of a clinical RSA study is outlined. Furthermore, since RSA is a highly technical method, the reliability and reproducibility of a study can only be assessed when all standardized output items have been adequately addressed.

We investigated whether the RSA guidelines have resulted in improved methodological reporting in clinical RSA studies of hip and knee arthroplasty. We also assessed the overall level of adherence to each of the individual guidelines following their introduction.

**Material and methods**

We carried out a literature search using OVID Medline, the OVID Cochrane central register of controlled trials, and SCOPUS Embase to identify all articles published between January 2000 and December 2011 that used RSA in the evaluation of hip or knee prosthesis migration or wear. The search strategy included the following terms: radiostereometric analysis, radiostereometry, roentgen stereophotogrammetric analysis, RSA, hip, and knee. Only English language clinical studies with at least 2 follow-up time points were included. Case reports were excluded. Study identification was performed according to the PRISMA statement (Figure 1). 2 investigators (RM and TJM) independently evaluated each of the studies for adherence to the 13 individual guideline points (Table 1). Since most of the 13 guideline points included more than 1 issue that needed to be addressed, the individual guidelines were divided into subcategories (A to D) to facilitate the evaluation process. Studies were then assessed on whether each guideline was fully met, partially met, or not met. Other information collected from the studies regarding publication were the name of the journal, geographical region based on first author (Nordic countries, rest of Europe, USA/Canada, Australia) and the year of publication. We also collected data on the main study question (whether the study assessed hip or knee arthroplasty migration or wear), study design (retrospective, prospective cohort, or RCT), the period of data collection, the percentage of patients for whom RSA was not successful, and the formula used to calculate confidence intervals. Disagreement was resolved by consensus.

**Table 1. List of the 13 guidelines and subcategories used to evaluate study adherence (adapted from Valstar et al. 2005)**

<table>
<thead>
<tr>
<th>Guideline number</th>
<th>Subcategory</th>
<th>Guideline</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 A</td>
<td>Translation units</td>
<td>in millimeters</td>
</tr>
<tr>
<td>1 B</td>
<td>Rotation units</td>
<td>in degrees</td>
</tr>
<tr>
<td>2 A</td>
<td>Accuracy</td>
<td>presented</td>
</tr>
<tr>
<td>2 B</td>
<td>Precision</td>
<td>presented</td>
</tr>
<tr>
<td>2 C</td>
<td>Measurement interval</td>
<td>mentioned</td>
</tr>
<tr>
<td>2 D</td>
<td>Window of tolerance</td>
<td>mentioned</td>
</tr>
<tr>
<td>3 A</td>
<td>Type of calibration</td>
<td>cage</td>
</tr>
<tr>
<td>4 A</td>
<td>Fixed or portable</td>
<td>X-ray</td>
</tr>
<tr>
<td>5 A</td>
<td>Subject positioning</td>
<td>described</td>
</tr>
<tr>
<td>6 A</td>
<td>Orientation of</td>
<td>global</td>
</tr>
<tr>
<td>6 B</td>
<td>Method of image</td>
<td>acquisition</td>
</tr>
<tr>
<td>6 C</td>
<td>Scanner or</td>
<td>system</td>
</tr>
<tr>
<td>7 A</td>
<td>Software name</td>
<td>mentioned</td>
</tr>
<tr>
<td>7 B</td>
<td>Software version</td>
<td>mentioned</td>
</tr>
<tr>
<td>8 A</td>
<td>Marker size</td>
<td>mentioned</td>
</tr>
<tr>
<td>8 B</td>
<td>Marker size validated</td>
<td>(or manufacturer</td>
</tr>
<tr>
<td>9 A</td>
<td>RSA method</td>
<td>mentioned</td>
</tr>
<tr>
<td>10 A</td>
<td>Cutoff level</td>
<td>for condition</td>
</tr>
<tr>
<td>11 A</td>
<td>Cutoff level</td>
<td>for mean error</td>
</tr>
<tr>
<td>11 B</td>
<td>Rigid-body</td>
<td>fixed coordinate</td>
</tr>
<tr>
<td>12 A</td>
<td>Angular rotation sequence</td>
<td>mentioned</td>
</tr>
<tr>
<td>12 B</td>
<td>Precision</td>
<td>determined</td>
</tr>
<tr>
<td>13 A</td>
<td>All 6 degrees of freedom</td>
<td>reported</td>
</tr>
<tr>
<td>13 B</td>
<td>Point or segment</td>
<td>motion</td>
</tr>
<tr>
<td>13 C</td>
<td>Point(s) used</td>
<td>to measure</td>
</tr>
</tbody>
</table>
Results

153 studies that met our inclusion criteria were identified (Figure 1). 61 of these were published before the guidelines (2000 to 2005) and the remaining 92 after the guidelines (2006 to 2011). 96 studies evaluated hip arthroplasty and 57 studies evaluated knee arthroplasty. Most of the studies (121) used RSA to assess migration and 22 studies used RSA to evaluate wear. Only 10 studies used RSA to measure both wear and migration. Over half of the studies were randomized controlled trials (81 studies), the remainder consisted largely of prospective cohort studies (66 studies), and a few were retrospective studies (6 studies).

None of the studies fully met all 13 guidelines. Only 7 studies published after the guidelines showed partial adherence to all guidelines. The highest number of guidelines fully met was 10 and this was achieved in 8 studies, all of which were published after the guidelines. Almost half of the studies published after the guidelines (43) showed partial adherence to 10 of 13 guidelines, and accordingly we selected this adherence level to mean high adherence. As expected, the methodological quality of the RSA studies published before the guidelines was low, with less than one-fifth (11) of the studies partially fulfilling 10 of 13 guideline items.

The methodological quality of studies clearly improved from 2000 to 2011 (Figure 2). The proportion of studies with high adherence increased almost 3-fold when we compared the means of time periods before and after the guidelines. The annual number of publications included in the analysis varied from 6 to 26. The largest number of publications was from 2006, during which a special Radiostereometric Analysis in Orthopaedic Surgery symposium issue of Clinical Orthopaedics and Related Research was published.

Following introduction of the RSA guidelines, the items with highest adherence were guideline 1 (measurement units), 9 (RSA method), and 11 (rigid-body fixed coordinate frames and angular rotation sequence) whereas those with poorest adherence were 3 (type of calibration cage), 4 (fixed or portable X-ray source), 6 (image acquisition method and system details), and 10 (cutoff level for condition number or CN and mean error of rigid-body fitting or ME) (Figure 3). Guideline 2 (accuracy, precision, measurement interval, and window of tolerance) generally only had partial adherence. Although the level of precision was presented in 60 studies (65%), the level of accuracy was only mentioned in 12 studies (13%). The measurement interval was mentioned in all studies, but the window of tolerance was generally only mentioned for the postoperative examination and was usually a one-week window. Surprisingly, 35 studies did not mention the window of tolerance for any of the RSA examinations. Only 2 studies mentioned the window of tolerance for RSA examinations after the postoperative examination and used a 5–10% tolerance limit.

Most of the studies evaluated were from Nordic countries, and these also had the highest level of adherence to the RSA guidelines when compared to the other geographical regions (Table 2). Similarly, the majority of the studies were published in Acta Orthopaedica (Table 3).
The proportion of patients for whom RSA was not successful was difficult to extract from most of the studies, especially from studies that included patients with bilateral arthroplasty procedures. From the 75 studies published after the guidelines with available data, RSA was not successful in 8% of knee arthroplasty patients and in 12% of hip arthroplasty patients. Furthermore, there was no mention of how the confidence interval (CI)—and hence precision of the method—was calculated in 34 studies. The remaining studies that did report these data used heterogeneous but accepted methods for calculating the confidence interval, with some studies reporting the 99% CI while others used the 95% CI. More than half of the studies published from 2006 to 2011 clearly mentioned the time period of patient recruitment, but this information was lacking in 43 studies.

### Discussion

The guidelines for standardization of radiostereometry of implants have clearly improved the methodological reporting in RSA studies during the last decade. Nevertheless, the overall level of adherence to the guidelines is still low. We believe that low adherence to some of the guidelines was due to oversight—or to the assumption that the reader is already aware of the methodological details. Examples of this include not mentioning the type of calibration cage, the image acquisition method, or whether a fixed or portable X-ray source was used. However, some guidelines such as the cutoff level for condition number and mean error of rigid-body fitting are more crucial to understanding the validity of the results and consequently their clinical significance. Furthermore, the stability and distribution of RSA markers within a rigid body will influence the accuracy of the motion calculation and may affect the conclusions drawn from the study.

Guideline 2 (accuracy, precision, measurement interval, and window of tolerance) generally only had partial adherence. Although the level of precision was presented in 65% of the studies published after the guidelines, double examinations of all patients were performed in only 18% of these. This raises the question of whether it is reasonable to demand this, or if it would be sufficient to perform double examinations on a random selection of subcohort of the patients in a study. Similarly, the level of accuracy was only mentioned in 13% of the studies published after 2005. It is also noteworthy that several studies used the term accuracy to mean clinical precision. The measurement of true RSA accuracy requires data from phantom studies and although this information is important, it is not as clinically relevant as precision data from double examinations, which define the true clinical accuracy.

As expected, most of the RSA studies were from Nordic countries where the adherence to guidelines was also highest, probably for historical reasons. Journals with fewer RSA studies tended to have higher adherence. This could be due to more stringent enforcement during the peer review process, since the RSA method is less familiar to the reviewers of these journals. On the other hand, studies published in Clinical Orthopaedics and Related Research also showed a high level of adherence. This was clearly influenced by the RSA symposium issue published in 2006, which may have undergone a more stringent review process due to the special focus of the issue, and also as it was published soon after the guidelines.

Some information that is currently not included in the RSA guidelines is the proportion of patients for whom RSA was not successful. We feel that these data are important for 2 reasons: firstly, because this information is crucial in understanding the clinical relevance of the data; and secondly, because this information is valuable to other researchers when planning similar studies.

As model-based RSA is becoming more popular, it is necessary to also include parameters that give an indication of the quality of the data obtained. An example of this could be the difference value (DIF) that is used in MBRSA software (RSAcore, Leiden, the Netherlands), which is a surrogate for the difference of the actual implant contour and the virtual contour. Inclusion of such quality control parameters would further standardize new developments in the RSA methodology.

The strengths of this systematic review include the large number of studies included (153) over a 12-year period. This enabled us to gain a general understanding of the trends in adherence over time. To our knowledge, this type of study has not been published previously. One limitation of the study is that no attempt was made to contact the authors of the papers. This might have allowed us to obtain more information about some aspects of guideline adherence that could not be assessed from the papers directly. On the other hand, our aim was to assess adherence to guidelines based on the information available from publications. Another limitation of the present study is that it only assessed hip and knee arthroplasty studies and excluded studies related to other applications such as spinal fusion or fracture studies. Although these other applications

### Table 3. Adherence of studies published during 2006–2011 to the RSA guidelines according to journal

<table>
<thead>
<tr>
<th>Journal</th>
<th>Total no. of studies</th>
<th>No. of studies with high adherence a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acta Orthopaedica</td>
<td>23</td>
<td>9</td>
</tr>
<tr>
<td>Clinical Orthopaedics and Related Research</td>
<td>18</td>
<td>12</td>
</tr>
<tr>
<td>Journal of Arthroplasty</td>
<td>13</td>
<td>4</td>
</tr>
<tr>
<td>Journal of Bone and Joint Surgery (British Volume)</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>International Orthopaedics</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>The Knee</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Other b</td>
<td>16</td>
<td>11</td>
</tr>
</tbody>
</table>

a Partial or full adherence to at least 10 of 13 RSA guidelines.
b 11 journals with 2 publications or less that met the inclusion criteria during the study period.
Table 4. Suggested presentation of standardized output for a clinical RSA study

<table>
<thead>
<tr>
<th>Items relevant to the Materials and Methods section of the study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time period of patient recruitment</td>
</tr>
<tr>
<td>Positioning of subject and calibration cage</td>
</tr>
<tr>
<td>Orientation of the global coordinate system (consider a diagram)</td>
</tr>
<tr>
<td>Rigid body fixed coordinate frames defined</td>
</tr>
<tr>
<td>RSA method (marker-based, geometrical or model-based)</td>
</tr>
<tr>
<td>version of RSA software</td>
</tr>
<tr>
<td>Units used for translation millimeters and the units for rotations degrees</td>
</tr>
<tr>
<td>Measurement interval</td>
</tr>
<tr>
<td>The points used to measure translation indicated (consider an x-ray or diagram)</td>
</tr>
<tr>
<td>Translations and angular rotations with all six degrees of freedom if applicable</td>
</tr>
<tr>
<td>Clinical precision</td>
</tr>
<tr>
<td>Number of patients for whom RSA was not successful</td>
</tr>
</tbody>
</table>

Items to be mentioned in an appendix of the study

- Accuracy of the arrangement (if phantom studies have been performed)
- Size and manufacturer of RSA markers
- Type and manufacturer of calibration cage
- Fixed or portable x-ray sources and manufacturer
- Radiographic exposure settings (kV and mAs)
- Window of tolerance for both postoperative and follow-up RSA examinations (days/weeks)
- Method of image acquisition and scanner or system details as well as image resolution
- Number of patients for whom double examinations were performed
- Cut-off level for CN and ME
- Formula used to calculate confidence interval for precision
- Angular rotation sequence defined (if relevant)

of RSA are also relevant, we decided to focus on the 2 main applications of RSA.

In summary, RSA is important to ensure a safe yet effective stepwise introduction of new implants. New developments in RSA have enabled faster analysis and even measurement of marker-free implants. The RSA guidelines are an integral part of the methodology. Our study demonstrated that the guidelines have clearly had a positive effect on the methodological quality of RSA publications, but there is still room for improvement. The next step is to update, simplify, and clarify the guidelines and also promote their use in the peer review process. In future RSA publications, we suggest that an appendix, similar to the one shown in Table 4, be included in all future RSA publications.

RM and TJM collected the data and carried out the data analysis. All authors participated in the design of the study and in preparation of the manuscript.

We thank information specialist Tiina Heino of Helsinki University Library for assistance with the literature search.

HM is a stockowner and board member in RSA Biomedical Inc. This study was supported by grants from the Swedish Cultural Foundation in Finland, the Jane and Aatos Erkko Foundation, the Paulo Foundation, and the Finnish Medical Foundation.


Decking J, Schuett, Decking R, Puhl W. The migration of femoral compo


Hansson U, Toksvig-Larsen S, Ryd L, Aspenberg P. Once-weekly oral medica


Li M G, Nilssom K G. No relationship between postoperative changes in bone density at the proximal tibia and the migration of the tibial component 2 years after total knee arthroplasty. J Arthroplasty 2001; 16 (7): 893-900.


Palm L, Olofsson J, Aström S E, Ivarsson I. No difference in migration or wear between cemented low-profile cups and standard cups : a randomized radiostereographic study of 53 patients over 3 years. Acta Orthop 2007; 78 (4): 479-84.


