Methods for determining the accuracy of radiostereometric analysis (RSA)

Leif Ryd[1], Xunhua Yuan[1] and Håkan Löfgren[2]

Departments of [1]Orthopedic Surgery, University Hospital, SE-221 85 Lund, Sweden. Fax +46 46 13 07 32.
E-mail: leif.ryd@ort.lu.se, [2]Neuro-Orthopedics, Ryhovs Hospital, Jönköping, Sweden
Submitted 99-10-10. Accepted 00-03-07

ABSTRACT – The problem of determining the noise level in any measuring system remains urgent. Radiostereometric analysis (RSA) is a radiographic system of unique accuracy which has applications in areas where minute motions or no motion at all occurs. Examples are micromotion between endoprostheses and bone and in fracture healing.

We have determined the accuracy of the RSA system as applied to a clinical series of spinal fusions, where the conditions for RSA were not optimal. Using the usual test-retest methodology on a phantom, we showed that its accuracy can be grossly overestimated in the individual case.

We found considerable variations in the accuracy in the individual case, depending on the rigid-body configuration. The overall accuracy, expressed as 3-D "vectors" for rotation and translation, respectively, correlated with the condition number, a method for characterizing the marker configuration. Indeed, the condition number explained as much as 92% of the variation in overall rotation. This condition number, however, cannot be used to analyze the accuracy of one degree of freedom of rotation alone. Mathematical simulation of the accuracy in the individual case of the individual dimension, using in-house software, showed that the accuracy (95% confidence) varied between 0.4 and 4.6 degrees of rotation about the transverse axis, corresponding to a clinical stress series of extension and flexion.

■

During the past decade, radiostereometric analysis (RSA) has emerged as a valuable tool for studying various problems in the musculoskeletal system (Kärrholm 1989). RSA permits perfect characterization of motion in all 3 dimensions. This has been used to investigate the locomotion pattern of the shoulder joint (Högfors et al. 1991), kinematics of the knee (Nilsson et al. 1991b, Nilsson 1992) and the kinematics of the intact knee joint with ligamentous injuries (Kärrholm et al. 1988). More importantly, RSA has an accuracy which is about an order of magnitude better than conventional radiography. This fact has been utilized in a number of publications regarding the fixation of prosthetic components after hip and knee arthroplasty (Ryd et al. 1988, 1990, Nilsson et al. 1991a, Önsten et al. 1994, Toksvig-Larsen et al. 1995, Kärrholm et al. 1996), the healing patterns after spinal fusions (Johnsson et al. 1990, 1992) and the stability of the intact spine (Axelsson et al. 1992, 1997). All these latter investigations rely on high and well-controlled accuracy, since the presence or absence of motion often constitutes the crucial effect parameter used.

In any measuring system, there are always some input errors. These errors produce "noise" in the results and the magnitude of this noise defines an interval of the result parameter where the measuring system cannot be used, i.e., measurements of motions smaller than the resolution of the measuring system used are not possible. Determination of the accuracy of RSA has usually been done by test-retest investigations. In the absence of a gold standard, this is the only feasible method (Ranstam et al. 1999). Test-retest evaluations under conditions of no true motion is one way to mimic a gold standard, i.e., the true motion is assumed to be zero. In most RSA publications, this has been an important part of the study.

However, such test-retest procedures on a random sample of patients in an investigation will only yield an approximation of the truth hidden in the entire cohort of patients. As such it will represent a mean accuracy of a number of investigations of a number of patients. Given a reasonably consistent number of markers and rigid-body configuration, which is usually the case, this procedure can be sufficient and has been a standard procedure in RSA (Selvik 1974).

In principle, however, in any cohort of patients there will be individuals with marker configurations allowing an accuracy better than the statistical value and others with a value worse than that value. In investigations where the factors affecting the accuracy are hard to control, there will be large variations in the accuracies for individual cases. Such a scenario can be translated into a situation where patients will show motion values above or below a mean computed accuracy which does not apply to such individual cases. Faulty interpretations regarding the presence or absence of motion will result.

Our article concerns the cervical spine after fusion. The clinical data have previously been reported elsewhere (Löfgren et al. 1999). The accuracy of the RSA system was first determined by test-retest examinations of a phantom representing the typical marker configuration aimed at in this particular application. Subsequently, test-retest examinations were carried out on a number of randomly chosen patients. Finally, assessment of the accuracy was performed on all patients using in-house software (Yuan et al. 1997, Yuan and Ryd 1999) to arrive at a correct interpretation of the RSA results.

Material and methods

36 in a total cohort of 43 patients undergoing one-segment fusion in the cervical spine, using autografts, allografts or xenografts, were marked for RSA during surgery. Marking was undertaken by placing 3 or 4 tantalum markers (diameter = 0.8 mm) in the vertebrae adjacent to the fusion. RSA radiographs were obtained postoperatively and after 6, 12 and a mean of 36 months, using the convergent-ray set-up (Selvik 1974). The radiographic set-up was first calibrated with a calibration cage of Plexiglass using markers of known position in each of the 4 walls and with reference planes, also provided with markers, between the cage and the film cassettes. Subsequently, the cage was removed but the reference planes were retained during the patient examinations. The patients were first examined during maximum extension of the cervical spine when facing away from the cassettes. After this, exposures were obtained during maximum flexion while facing 90 degrees different from the first exposures in order not to have the teeth obscuring the tantalum markers. The radiographs were manually digitized using a digitizing table with an accuracy of about 0.02 mm (Hasselblad Engineering, Gothenburg, Sweden). The primary result variable was rotation about the transverse axis, as determined by RSA, between maximum extension and flexion as completed voluntarily by the patient (Löfgren et al. 1999). The aim of the study was to determine whether or not fusion had occurred—that is, was there motion at the fusion level or not during the flexion-extension exercise? The accuracy of the RSA system was thus of the utmost importance.

A Plexiglass rod was constructed and provided with 2 × 4 tantalum markers in a configuration resembling the optimum achieved in the patients. 12 RSA examinations were performed on this rod with and 10 without submersion in water, representative of soft tissue. Slight changes in position and X-ray tubes were done between exposures. Compilation of data was done according to the formula

$$\Sigma D^1_{\text{x-rot, y-rot...z-transl}} - D^2_{\text{x-rot, y-rot...z-transl}} = \text{mean} \pm \text{SD}$$

where the mean of an indefinite number of tests will approach zero and can therefore be neglected. The SD is a measure of the accuracy (Ranstam et al. 2000).

Test-retest examinations were performed on 8 patients. Double provocations were performed in the sequence max. extension – max. flexion – max. extension – max. flexion. Data were arranged so that the first pair provided the first result parameter and the second pair the second result parameter in the test-retest set-up. Compilation was performed as above.

A system for determination of the accuracy of RSA in the individual patient was developed (Yuan et al. 1997, Yuan and Ryd 2000). Here the marker configuration of the two vertebrae, as measured on the digitizing table and computed by the standard RSA technique, of each case was imported into MATLAB. In MATLAB, an identical KINEMA part of the RSA program was written. Measurement errors of a magnitude (zero mean, SD = 0.04 mm, unbiased) coincidental with those of the measuring table used were randomly generated and added to the error-free data imported into the program. Subsequently, KINEMA calculations were performed in MATLAB 1000 times with different errors added each time. 1000 motions were thus computed for each of the 6 degrees of freedom of motion in space (x-rot, y-rot,...z-transl.). The mean of these motions approached zero and the SD of each degree of freedom of motion was calculated for each patient individually. The SD represents the noise in the system. This individual accuracy was related (regression analysis) to the condition numbers as provided by the RSA program (UmRSA, Unicum, Umeå, Sweden).

$$\text{The condition number} = 1 / (d_1^2+d_2^2+......d_n^2)^{1/2},$$

where d = the distance between each marker and an arbitrary straight line passing through the cluster of markers. By a mathematical algorithm, the condition number is minimized, i.e., the "best" position and direction of the straight line are found. The larger the value, the more the markers align with the line, i.e., the less they lend themselves to the provision of high accuracy.

The condition number is thus a mathematical expression of how the markers in an object of interest, i.e., a "rigid body", relate to an arbitrary straight line passing through that rigid body (Söderkvist and Wedin 1993). To allow comparison between the arbitrary straight line and the condition number, the overall rotations and translations from the simulations were calculated by the 3D Pythagorean theorem and related to the condition numbers of the two rigid bodies involved. For "small" rotations, such overall computation is permissible (Selvik 1974).

Results

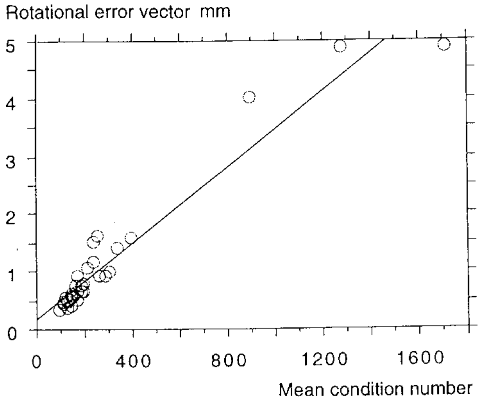
Determination of the accuracy of the RSA system used here by the phantom Plexiglass rod showed that the influence of water, i.e., "soft tissue", was negligible and the two sets of data were combined. The SD was found to be 0.18 degrees and 0.06 mm, respectively. Using the 95% confidence limit, the accuracy could thus be determined as 0.36 degrees and 0.12 mm for rotation and translation, respectively.

Table 1. Results from 8 double stress investigations pertaining to rotation about the transverse axis, corresponding to the motion from first maximum flexion followed by maximum extension

Case	First stress series	Second stress series	Difference
28	9.1	0.9	8.2
29	0.1	2.1	–2.0
30	0.3	0.5	–0.2
31	0.9	–0.6	1.5
32	0.2	1.1	–0.9
33	2.4	3.1	–0.7
34	–0.3	–0.7	0.4
35	–0.2	–0.3	0.1
			0.8 ± 3.2 (SD)

Double stress examinations were performed on 8 patients. The results are shown in Table 1. Formal compilation of the SD yielded a value of 3.2 degrees for rotation about the relevant transverse axis. There were large differences between the two stress examinations and it was obvious that the biological variability in performing a voluntary maximum extension followed by a maximum voluntary flexion was considerable.

The results of the simulations are shown in Table 2. For the relevant x-axis rotation, the SD ranged from 0.18 to 2.26 degrees. Since the condition number is a 3-D value giving the overall affinity to the straight line that passes through the rigid body in any direction, the mean value for the SD of all three rotations is given. The correlation between this mean rotational accuracy and the condition number was $r^2 = 0.93$ ($p < 0.0001$) (Figure). The correlation between the $SD_{x\text{-rot}}$ and the condition number was $r^2 = 0.20$ ($p < 0.0001$) (Table 2). As expected, there were large variations in the rotational accuracies about the individual axes in the same patient. Further, the condition number explained a much smaller portion of the variability of the mean translational accuracy,



$Y = 0.188 + 0.003*X; R^2 = 0.929$

Figure. Regression analysis between the 3-D vectorial SD of rotations and the condition number. The correlation is very significant ($p < 0.00001$).

$r^2 = 0.36$. Finally, marker instability was not a problem in this study and the mean error was well below the usual threshold of 200 μm in most cases (Table 2).

Table 2. Results of simulations. SD of transverse axis rotations and condition numbers. The columns correlate at p < 0.007

Case	Condition number [a]		SD, transverse axis rotation	Mean Error (segm.1/segm. 2) [b]
1	90	(90/90)	0.201	55/103
2	109	(97/121)	0.260	85/222
3	184	(108/260)	0.311	132/103
4	108	(94/121)	0.278	101/87
5	124	(95/154)	0.179	321/354
6	181	(189/173)	0.613	57/181
7	127	(102/152)	0.315	122/220
8	166	(215/116)	0.201	118/93
9	891	(105/1677)	2.262	255/241
10	138	(114/163)	0.367	83/190
11	138	(134/142)	0.304	133/98
12	143	(161/125)	0.196	51/224
13	300	(133/466)	0.50	98/159
14	194	(153/235)	0.560	124/83
15	122	(139/106)	0.299	168/114
16	1276	(2273/279)	1.052	50/148
17	144	(140/147)	0.359	33/105
18	331	(533/129)	1.136	61/73
19	116	(114/118)	0.374	130/110
20	148	(163/134)	0.302	156/151
21	189	(140/238)	0.381	101/46
22	233	(303/163)	0.872	97/91
23	130	(114/147)	0.317	82/47
24	233	(154/312)	1.369	138/41
25	133	(114/152)	0.303	265/255
26	284	(183/386)	0.578	163/142
27	391	(370/412)	1.366	104/57
28	166	(114/217)	0.528	72/133
29	166	(184/149)	0.332	110/110
30	1712	(430/2994)	0.580	93/26
31	210	(265/156)	0.710	100/119
32	122	(102/142)	0.249	75/95
33	246	(89/404)	1.475	175/31
34	122	(82/161)	0.291	70/101
35	256	(126/385)	0.841	129/108
36	156	(155/158)	0.511	96/119

[a] The mean of the two condition numbers were used for the regression analysis. The condition numbers for segment 1 and segment 2 are given in parentheses.
[b] The mean error is different at each point in time. Data given represent the mean error between the first and the last examination. This represents a worst-case scenario, the true mean error in the computations (extension-flexion) were often much smaller.

Discussion

As stated earlier, the use of RSA is commonest in applications where measurements of minute motions are performed. Often the question is whether or not there is any motion at all. In this study, the question whether the fusion had or had not healed was assessed. In these instances, perfect control of the accuracy of the system becomes imperative.

Usually, in RSA, accuracy is assessed by test-retest, or double, examinations. In a series of test-retest pairs, the variation of the measurements can be determined. The SD represents a measure of the reproducibility and translates into accuracy if the true value is known. If not, it represents the precision of the system. Assessment of the accuracy of the system using double examinations of the Plexiglass rod yielded results which were compatible with previous determinations. Thus in the knee, the accuracy has previously been determined to be about 0.3 degrees at the 95% confidence limit (Ryd et al. 1986). The slightly inferior accuracy here can be explained by the smaller size of the vertebral rigid bodies, as modelled by the Plexiglass rod. The rigid bodies in the patients, however, did not correspond to the models in the plexiglass. Their sizes and configurations were worse and often there were only 3 markers in each rigid body.

We therefore reverted to the double stress series investigation. The results here were erratic and reflect 2 basic problems in this part of the study.

First, motion could take place between the rigid bodies in the interval between the 2 stress series. Thus, there was no known true value.

Secondly, when asked to perform a motion twice, the patients were unable to do so with any degree of reproducibility. This biological variability was added to, and grossly overwhelmed the technical variability under scrutiny. This finding has considerable implications for the interpretation of the actual patient-related data, which can be disregarded here. In this methodological study on accuracy, double stress investigations did not yield usable information. Such paradoxical results due to erratic biological behavior have been found previously (Fridén et al. 1992) and the phenomenon constitutes a problem inherent to all measurements of dynamic activities of patients.

Regression analysis of the data from individual simulation showed that the condition number explained well over 90% of the variability in the mean rotational accuracy. This supports the conceptual fact that the condition number "by definition" represents the accuracy. This is only true, however, regarding rotation about the arbitrary straight line and only 20% of the variability of rotation about the x-axis was accounted for. In a perfect situation, where the markers in each individual case are isotropically distributed in the object of interest, that is located in the corners of a tetrahedron, and their centers of gravity are close (Yuan et al. 1997), the accuracy in all dimensions is linearly related to digitizing errors alone. The influence on the accuracy of an increasing "mean error of rigid body fitting", that is, of a decreasing marker stability, has been investigated previously (Ryd 1986). Söderkvist and Wedin (1993) showed that the error increased exponentially with increasing condition number, but an in vivo investigation of the matter has not been performed. A previous experimental study, however, has pointed to the effect of isotropically-situated markers and showed results compatible with ours (deLange et al. 1990).

An increasing condition number indicates that the marker configuration approaches that of a straight line which may have any direction. When this is the case, the accuracy of the rotational determination about the axis parallel to this straight line will decrease. Determinations of rotational accuracy in directions perpendicular to the line, however, may be very accurate. For this reason, assessment of rotational accuracies about individual axes from the condition number will be misleading by definition. In this study, the actual accuracy of measurements of rotations about the transverse axis ranged from 0.35 degrees, a fairly normal RSA accuracy, to as much as 4.4 degrees (at 95% confidence level) with, again, a poor correlation to the condition number. Simple measurements directly on the radiographs would have been more accurate. Finally and to conclude the issue, it should be noted that the rigid-body configuration has a much smaller effect on the translational accuracy, as confirmed by the findings in this study.

Despite the best of intentions, using a prospective and randomized protocol, the clinical investigation upon which this report was based represents a study where the use of RSA was problematic. The cervical vertebrae are small, it is hard to control the insertion of the markers during the operation and this resulted in large variations in shape, size and number of markers in the different rigid bodies of these patients. Other RSA studies of the cervical spine have had the same problems (Zoega et al. 1998). In this investigation, provocation in extension and flexion was performed. The primary result variable by RSA was thus rotation about the transverse axis. We have shown how the accuracy in the individual case for this particular dimension can be pin-pointed. Without the use of the simulation routine, the data from the clinical study could probably not have been used for scientific purposes. With simulation, more accurate interpretation of results should be attainable.

The accuracy of RSA depends on a large number of factors including radiographic equipment, RSA set-up, number of markers, size of and distance between marker configurations. In this investigation, we focused on the importance of rigid body configuration. The accuracy varies greatly also with reasonable RSA characteristics. We have shown how determinations of the accuracy of RSA, or of any other system for kinematic assessment dealing with noisy input data, can be performed using computer simulation. The method improves the consistency of RSA in clinical applications where the methodology, for practical reasons is compromised. The best way to ensure high accuracy in RSA is to insert the markers so that large, rigid bodies with many markers are arrived at. In such cases, simulations are superfluous.

Axelsson P, Johnsson R, Strömqvist B. Effect of lumbar orthosis on intervertebral mobility. A roentgen stereophotogrammetric analysis. Spine 1992; 17 (6): 678-81.

Axelsson P, Johnsson R, Strömquist B. The spondylotic vertebra and its adjacent segment. Mobility measured before and after posterolateral fusion. Spine 1997; 4: 414-7.

deLange A, Huiskes R, Kauer J M G. Measurement errors in roentgen-stereophotogrammetric joint-motion analysis. J Biomech 1990; 23 (3): 259-69.

Fridén T, Ryd L, Lindstrand A. Laxity and graft fixation after reconstruction of the anterior cruciate ligament. A roentgen stereophotogrammetric analysis of 11 patients. Acta Orthop Scand 1992; 63 (1): 80-4.

Högfors C, Peterson B, Sigholm G, Herberts P. Biomechanical model of the human shoulder joint. II: Shoulder rhythm. J Biomech 1991; 24 (8): 699-709.

Johnsson R, Selvik G, Strömquist B, Sundén G. Mobility of the lower lumbar spine after posterolateral fusion determined by roentgen stereophotogrammetry. Spine 1990; 15: 347-50.

Johnsson R, Strömquist B, Axelsson P, Selvik G. Influence of spinal immobilization on consolidation of posterolateral lumbosacral fusion. A roentgen stereophotogrammetric and radiologic analysis. Spine 1992; 17: 16-21.

Kärrholm J. Roentgen stereophotogrammetry. A review of orthopedic applications. Acta Orthop Scand 1989; 60 (4): 491-503.

Kärrholm J, Selvik G, Elmquist L-G, Hansson L-I. Active knee motion after cruciate ligament rupture: Stereoradiography. Acta Orthop Scand 1988; 59 (2): 158-64.

Kärrholm J, Thanner J, Malchau H, Herberts P. Migration of porous acetabular cups with hydroxyapatite and tricalcium-phosphate coating. Trans ORS 1996; 21 (1): 239.

Löfgren H, Johannsson V, Olsson T, Ryd L, Levander B. Rigid fusion after Cloward operation for cervical disc disease with either autograft, allograft or xenograft. A randomized study with radiostereometric follow-up. Spine 1999 (In press).

Nilsson K G. Kinematics and fixation of total knee arthroplasties.Thesis, University of Umeå, Sweden 1992.

Nilsson K G, Kärrholm J, Ekelund L, Magnusson P. Evaluation of micromotion in cemented vs. uncemented knee arthroplasty in osteoarthritis and rheumatoid arthritis. Randomized study using roentgen stereophotogrammetric analysis. J Arthroplasty 1991a; 6: 265-78.

Nilsson K G, Kärrholm J, Gadegaard P. Abnormal kinematics of the artificial knee. Roentgen stereophotogrammetric analysis of 10 Miller-Galante and five New Jersey LCS knees. Acta Orthop Scand 1991b; 62 (5): 440-6.

Ranstam J, Ryd L, Önsten I. Accurate accuracy assessment. Acta Orthop Scand 2000; 71 (1): 106-8.

Ryd L. Micromotion in knee arthroplasty. A roentgen stereophotogrammetric analysis of tibial component fixation. Acta Orthop Scand (Supplement 220) 1986: 57 .

Ryd L, Albrektsson B, Herberts P, Lindstrand A, Selvik G. Micromotion in uncemented Freeman-Samuelson knee prostheses. Clinl Orthop 1988; 229: 205-12.

Ryd L, Lindstrand A, Rosenquist R, Selvik G. Tibial component fixation in knee arthroplasty. An in-vivo roentgen stereophotogrammetric study of cemented metal-backed devices. Clin Orthop 1986; 213: 141-9.

Ryd L, Lindstrand A, Stenström A, Selvik G. PCA tricompartmental tibial components; relation between prosthetic position and micromotion. Clin Orthop 1990; 251: 189-97.

Selvik G. Roentgen stereophotogrammetry. A method for the study of the kinematics of the skeletal systems. Thesis, University of Lund, Sweden 1974. Reprint in: Acta Orthop Scand (Suppl. 232) 1989: 60.

Söderkvist I, Wedin P-Å. Determining the movement of the skeleton using well-configured markers. J Biomech 1993; 26 (12): 1473-7.

Toksvig-Larsen S, Ryd L, Lindstrand A. Fixation of the tibial component in knee arthroplasty after six weeks. Int Orthop SICOT 1995; 19: 89-93.

Yuan X, Ryd L. Accuracy analysis for RSA. A computer simulation study on 3D marker reconstruction. J Biomech 2000; 33 (4): 493-8.

Yuan X, Ryd L, Blankevoort L. Error propagation for relative motion determined from marker positions. J Biomech 1997; 30 (9): 989-92.

Zoega B, Kärrholm J, Lind B. One-level cervical spine fusion.A randomized study, with or without plate fixation, using radiostereometry in 27 patients. Acta Orthop Scand 1998; 69 (4): 363-8.

Önsten I, Carlsson Å , Ohlin A, Nilsson J Å. Migration of acetabular components inserted with and without cement in one-stage bilateral hip arthroplasty. J Bone Joint Surg (Am) 1994; 76 (2): 185-94.