

# A Comparison of Radiostereometric Analysis and Computed Tomography for the Assessment of Lumbar Spinal Fusion in a Sheep Model

Ali Humadi<sup>1</sup> Brian J. C. Freeman<sup>1</sup> Rob J. Moore<sup>2</sup> Stuart Callary<sup>3</sup> Klas Halldin<sup>4</sup> Vikram David<sup>5</sup>  
William Maclaurin<sup>6</sup> Paul Tauro<sup>7</sup> Mark Schoenwaelder<sup>6</sup>

<sup>1</sup> Department of Spinal Surgery, Royal Adelaide Hospital, University of Adelaide, Adelaide, Australia

<sup>2</sup> Adelaide Centre for Spinal Research, SA Pathology, Adelaide, Australia

<sup>3</sup> Department of Orthopaedics, Royal Adelaide Hospital, University of Adelaide, Adelaide, Australia

<sup>4</sup> Department of Orthopaedics, Spinal Unit, Sahlgrenska University Hospital, Gothenburg, Sweden

<sup>5</sup> Department of Orthopaedics, Bendigo Hospital, Victoria, Australia

<sup>6</sup> Department of Radiology, The Alfred Health, Victoria, Australia

<sup>7</sup> Department of Radiology, Northern Health, Victoria, Australia

**Address for correspondence** Ali Humadi, MB, ChB, Msurg, Department of Spinal Surgery, Level 4, North Wing, Royal Adelaide Hospital, Adelaide, SA 5000, Australia (e-mail: alidulami1974@yahoo.com).

Evid Based Spine Care J 2013;4:78–89.

## Abstract

**Study Design** Prospective animal study.

**Objective** The aim of this animal study is to evaluate the accuracy of radiostereometric analysis (RSA) compared with computed tomographic (CT) scan in the assessment of spinal fusion after anterior lumbar interbody fusion (ALIF) using *histology* as a gold standard.

**Methods** Three non-adjacent ALIFs (L1–L2, L3–L4, and L5–L6) were performed in nine sheep. The sheep were divided into three groups of three sheep. All the animals were humanely killed immediately after having the last scheduled RSA. The lumbar spine was removed and in vitro fine cut CT and histopathology were performed.

**Results** Using histological assessment as the gold standard for assessing fusion, RSA demonstrated better results (100% sensitivity and 66.7% specificity; positive predictive value [PPV] = 27.3%, negative predictive value [NPV] = 100.0%) compared with CT (66.7% sensitivity and 60.0% specificity [PPV = 16.7%, NPV = 93.8%]).

**Conclusions** RSA demonstrated higher sensitivity and specificity when compared with CT. Furthermore, RSA has the advantage of much lower radiation exposure compared with fine cut CT. Further studies are required to see if RSA remains superior to CT scan for the assessment spinal fusion in the clinical setting.

## Keywords

- ▶ radiostereometric analysis
- ▶ spinal fusion
- ▶ assessment
- ▶ computed tomography
- ▶ histology

received  
April 1, 2013  
accepted after revision  
July 18, 2013

© 2013 Georg Thieme Verlag KG  
Stuttgart · New York

DOI <http://dx.doi.org/10.1055/s-0033-1357359>.  
ISSN 1663-7976.

Assessment of Class of Evidence (CoE) for individual studies of diagnostic test evaluation

| Methodological principle                                   |    |
|--|----|
| Study design   |    |
| Prospective cohort design                                  | X  |
| Retrospective cohort design                                |    |
| Case-control design  |    |
| Broad spectrum of patients with expected condition         | a  |
| Appropriate reference standard used                        | X  |
| Adequate description of test and reference for replication | X  |
| Blinded comparison with appropriate reference              | X  |
| Reference standard performed independently of test         | X  |
| Evidence level   | II |

Note: Blank box indicates criterion not met, could not be determined, or information not reported by author or was not reported.

<sup>a</sup>This study contained nine animal subjects.

**Study Rationale and Context**

- This study is designed to explore the accuracy and practicality of using radiostereometric analysis (RSA) in assessment of spinal fusion. RSA has been claimed to have a high accuracy and has the advantage of low-radiation penalty.

**Objectives**

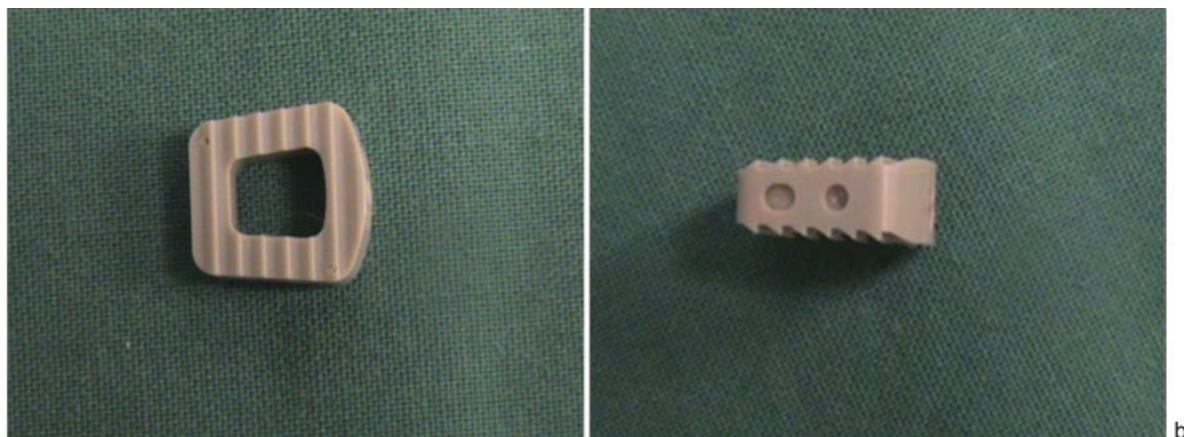
- To explore the accuracy and practicality of using RSA as an alternative to computed tomographic (CT) scan in assessment of spinal fusion after anterior lumbar interbody fusion (ALIF) in a sheep animal model.

**Materials and Methods**

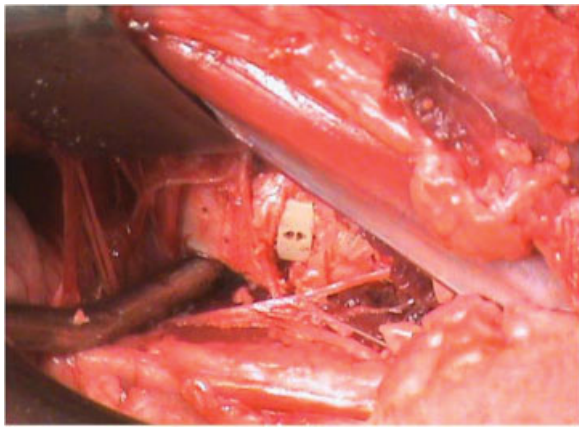
- Nine, healthy 18-month-old Merino sheep were included in this study; each underwent ALIF at three levels (L1-L2, L3-L4, and L5-L6). The sheep were placed in the left-lateral position with the left side upward. Two sheep were placed

in the right-lateral position because of surgeon preference. A longitudinal paraspinous incision was performed as a modification of the oblique incision described by Baramki et al.<sup>1</sup> A retroperitoneal approach is utilized to expose the disc space between psoas major and minor.

- The cage used in this study was the CORNERSTONE PSR cage (Medtronic, Memphis, Tennessee, United States). The cage is manufactured from polyethylene terephthalate (PET) and is radiolucent apart from its tantalum markers and measures 4 × 11 × 11 mm (► Fig. 1). It is normally used for cervical fusion in humans. The dimensions of the cage were considered appropriate for use in the sheep lumbar spine. Iliac crest bone graft was used for fusion. The cage containing the bone graft was then inserted into the intervertebral disc space after disc excision and rasping the cartilage of end plate (► Fig. 2).
- Tantalum markers were placed at five to seven anatomical sites in each vertebral body from L1-L6, allowing for the planned RSA. The holes were made using a special awl. The tantalum beads (1-mm diameter, UmRSA Biomedical, Umea, Sweden) were held in bone wax and inserted using a Penfield dissector.



**Fig. 1** CORNERSTONE PSR cage (Medtronic, Memphis, Tennessee, United States) that is used in this study.



**Fig. 2** Intraoperative picture showing the cage in situ.



**Fig. 3** The radiostereometric analysis (RSA) table used in our study. RSA setup with the roentgen tubes.

### RSA Assessment

- A uniplanar-type RSA setup with two radiographic tubes was used. Two mobile radiographic units (Toshiba KCD-10M-7; Toshiba Corp., Minatu-ku, Tokyo, Japan) were positioned with a 40 degree angle between the tubes. The calibration cage (Cage 43; RSA Biomedical, Umea, Sweden) contained two 35 × 43 cm high-resolution digital radiographic cassettes. The distance between each focus to film was 1.6 m. All radiographs were exposed at 60 kV and 10 mAs. The image plates were digitized with a processor (Point-of-Care CR 360 System; Kodak, New York, United States). The DICOM images were downloaded as tagged image format files images at 300-dots per square inch resolution. Each radiographic examination was analyzed using the UmRSA software (v6.0, RSA Biomedical, Umea, Sweden) by limiting the mean error of rigid body fitting to 0.3 mm. Any values *above* this were not included in the study.
- All RSA assessments were performed under general anesthesia. The anaesthetized sheep were positioned on a table above the RSA calibration cage (►Fig. 3).

### RSA Assessment Position

- Each RSA assessment was performed with the animal in full flexion and full extension. Flexion was achieved by shackling the legs together under general anesthesia (►Fig. 4a). Extension was achieved by pulling the fore and hind legs apart and tying them to the table in maximum extension (►Fig. 4b).
- The RSA results were reported by an independent RSA expert medical scientist (S.C.). The scientist was blinded to the results of CT scan and histology.

### Schedule of Sacrifice

- The nine sheep were divided into three groups; the first group underwent RSA at 6 months and were then sacrificed, the second group underwent RSA at 9 months and

were then sacrificed, and the last group underwent RSA at 12 months and were then sacrificed (see online supplementary material).

- Group 1 had RSA assessments at 12 months after the operation then was humanely killed. Group 2 had RSA assessments at 9 months after the surgical procedure then was humanely killed. Group 3 had RSA assessments at 6 months after the operation then was humanely killed.
- Two sets of RSA were performed in each position (two in flexion and two in extension) for each time point to minimize errors in the assessment. The termination of life was performed while the animals were fully anaesthetized, using the intravenous administration of an overdose (20 mL) of sodium pentobarbitone. The lumbar spine from each sheep was carefully transected at thoracolumbar junction and the lumbosacral junction and the entire lumbar spine was removed. The extracted specimens were kept in formalin.

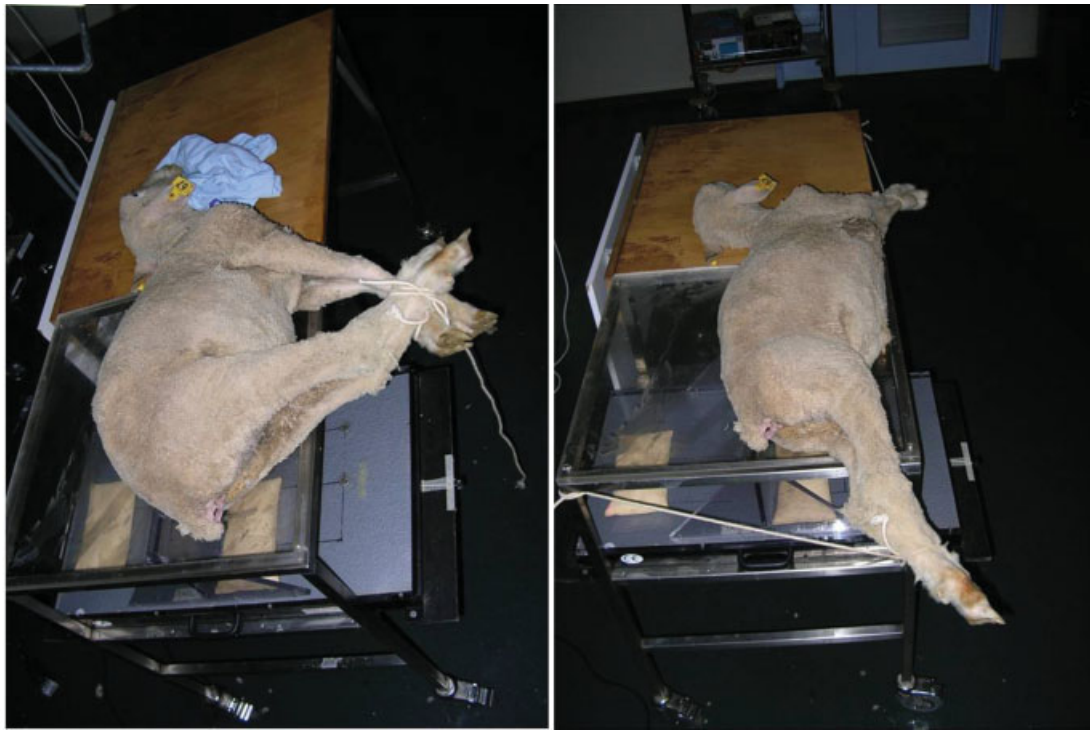
### Criteria of Successful Fusion by RSA

- The criteria for fusion were determined before starting the RSA assessment and were adapted from other study groups as cited. The presence of both of the following criteria was a prerequisite for diagnosis of successful fusion:

1. Sagittal translation of less than 0.7 mm in the z-axis.<sup>2-4</sup>
2. Sagittal rotation of less than 2 degrees.<sup>4,5</sup>

### CT Scan Assessment

- The extracted spines underwent fine cut (0.25 mm) CT scan assessment at the Radiology Department, Royal Adelaide Hospital (Aquilion, Toshiba Medical Systems, Tokyo, Japan), and images obtained were managed using Sorna DICOM Litebox software (Sorna Corporation, Eagan, Minnesota, United States). The scans were done on the same date by the same radiographer to minimize technique-related



**Fig. 4** (a) The radiostereometric analysis (RSA) assessment with the sheep’s lumbar spine in maximum flexion with legs tied together after flexion obtained. (b) RSA assessment with the sheep’s lumbar spine in extension with legs tied to the operative table after full extension obtained.

variations and errors. The scans were reconstructed into coronal, axial, and sagittal images (►Fig. 5).

**Reading of CT Scan**

- The results of the CT scans were reported by two independent radiologists (Royal Australian and New Zealand College of Radiologists accredited) who were blinded to the results of RSA and histology. They independently reported whether each level was *fused* or *not fused* on two separate occasions, 4 weeks apart to allow determination of the intraobserver error. The first reading of each radiologist was used to decide the status of fusion. When there was disagreement between the two radiologists, a third radiologist’s opinion was obtained for final diagnosis.

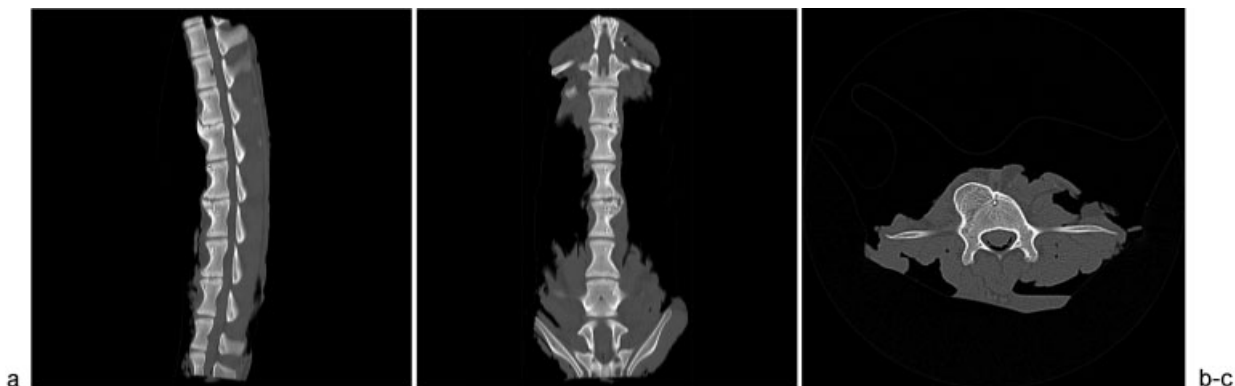
**Criteria for Assessing Fusion by CT Scan**

- The criteria for fusion were adapted from the evidence presented in the literature. The presence of all of the following criteria was a prerequisite for diagnosis of fusion:
  1. The presence of two or more bridging bony trabeculae passing from one vertebral end plate to the other in both the sagittal and coronal planes.<sup>6,7</sup>
  2. No evidence of lucency around the cage.<sup>8</sup>

**Histology**

**Histological Preparation**

- The excised lumbar spines were sectioned *midway* through each vertebral body in the sagittal plane using a



**Fig. 5** Computed tomographic scan of lumbar spine: (a) sagittal; (b) coronal; and (c) axial.

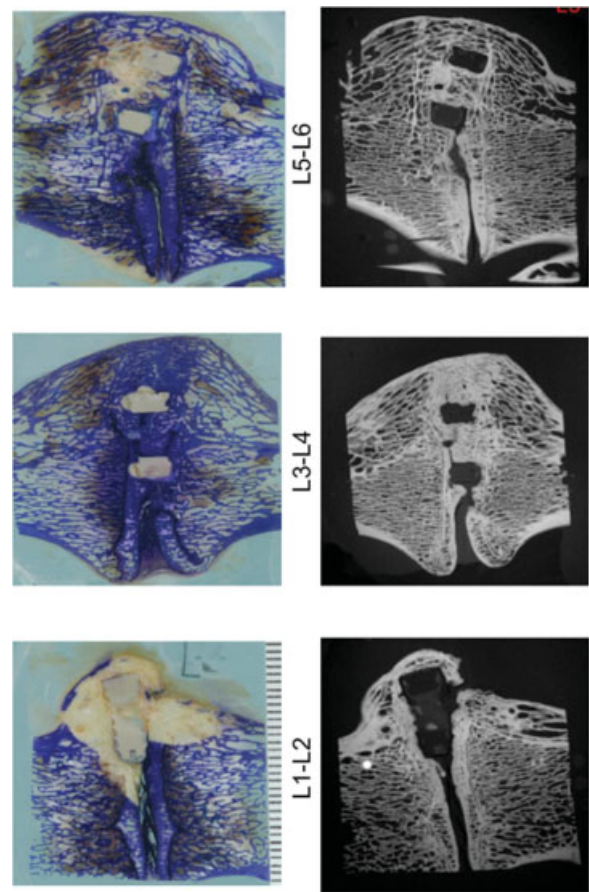
This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

butcher's saw. Specimens were fixed in 10% neutral buffered formalin for at least 2 days and transferred to 70% ethanol (1 × 3 days); 90% ethanol (1 × 4 days); 100% ethanol (2 × 2 days); methyl methacrylate (MMA) + 10% Solution B (plasticizer) (21 × days); MMA + Solution B (plasticizer) + 0.8% Powder C (initiator, peroxydicarbonate-perkadox 16) (2–4 days).

- At each stage, in this processing cycle, there was sufficient solution to cover the specimens and all solutions were prepared fresh on the day of use. The specimens were placed face down in clear plastic containers in a 37°C water bath until polymerization was complete (2–4 days).
- After polymerization, the embedded blocks were trimmed with a band saw and X-rayed to ascertain the orientation of the implanted cage. A straightened paper clip was used as a guide to mark the first cut (approximately 800 µm thick), using an Exakt diamond band saw (Exakt, Norderstedt, Germany), which in each case was sliced through the center of the implant to ensure the area of bony fusion was adequately represented. In addition to the initial section through the center of the implant, at least two further sections were taken approximately 1.5 mm distant from either side, and the *most representative section* was taken for analysis. The remainders were retained as spares.
- The exposed surface of each section was ground and polished with a Buehler MetaServ grinder/polisher (Buehler, a division of Illinois Tool Works, Inc, Lake Bluff, Illinois, United States) using progressively fine grades of sandpaper (100, 120, 240, 300, 400, and 600 grit) with continuous water irrigation. After the sections were polished on one side, they were attached to clear polycarbonate slides, using 5-minute epoxy resin. To ensure removal of air bubbles from between the block and the slide, gentle but firm pressure was applied. The slides were left overnight for maximum bonding. The other surface was prepared in the same way, and polishing was finished with 1-µm α alumina polishing suspension on a chamois cloth with only minimal irrigation to remove any remaining scratches until a mirror-like surface was produced.
- Sections were stained with 1% aqueous Toluidine blue to visualize cell detail and differentiate between tissue types.<sup>9</sup>

### Assessment of Fusion Using Histology

- The degree of fusion was assessed microscopically by two assessors: one laboratory scientist skilled in interpreting histology (R.J.M.) and one consultant histopathologist (D.W.E.). Both assessors were blinded to the results of the other modalities. Assessors were asked to determine if fusion was evident by the presence of continuous connecting trabecular elements passing from one vertebral end plate to another (► Fig. 6).
- As an adjunct to the histology, the polished sections were also subjected to contact digital microradiography (Faxitron LX60, Faxitron, Lincolnshire, Illinois, United States) to visualize the distribution of trabecular bone.



**Fig. 6** Examples of histology (left) and digital microradiography (right) (Faxitron LX60, Faxitron, Lincolnshire, Illinois, United States).

### Statistical Analysis

- The CT scan results were reported by two RANZCR-accredited radiologists (Royal Australian and New Zealand College of Radiologists), repeating their assessment 4 weeks later to determine the intrarater reliability. The two outcomes of the study were *fused* (F) and *not fused* (NF).
- The first and second readings of each radiologist were used to calculate the intrarater reliability. The first reading of the CT scan from each radiologist was then compared with assess the extent of interrater reliability (the first reading was considered because we normally obtained only one reading on the daily clinical practice). Stata v10.1 (Stata-Corp, College Station, TX, United States) was used to estimate Fleiss Kappa statistics for the intrarater and interrater reliability calculations.
- On occasions of discrepancy between the radiologists' first CT scan reading, a third independent radiologist assessed the CT scan to determine the final CT scan result. The results of the final CT scan were then compared with the RSA results through estimation of the Fleiss Kappa statistic. The results of the final CT scan and the RSA were compared with the histological results as the gold standard of successful spine fusion, by estimating the sensitivity, specificity, and positive and negative predictive values.

**Results**

**RSA Results**

- The majority of the movement observed in the RSA was in the sagittal plane or z-axis rotation. This rotation is a combination of x-axis (anterior-posterior) and y-axis (cranio-caudal) translations. Minor translations were observed in z-axis (left-right). The sagittal plane Range of Motion (ROM) (z-axis rotation) and the total movement (three-dimensional translations, which is calculated using

the vectorial sum of each independent axis) are presented in our results.

- The categorization of each motion segment into those that were *fused* and those that were *not fused* are summarized in ► **Table 1**. Eleven levels were fused and sixteen were not fused.

**CT Scan Results**

- The first radiologist (W.M.) reported 8 levels as *fused* in the first reading and 11 levels in the second reading. The

**Table 1** Summary of the fusion assessed by RSA

| Sheep ID     | Months postop | RSA Measurements     |                     | RSA result |
|--------------|---------------|----------------------|---------------------|------------|
|              |               | 3D rotation (degree) | 3D translation (mm) |            |
| <b>L1-L2</b> |               |                      |                     |            |
| 1            | 12            | 0.4                  | 0.4                 | F          |
| 2            | 12            | 2.6                  | 0.6                 | NF         |
| 3            | 12            | 1.1                  | 0.6                 | F          |
| 4            | 9             | 1.4                  | 0.4                 | F          |
| 5            | 9             | 3.1                  | 1.2                 | NF         |
| 6            | 9             | 2.3                  | 0.7                 | NF         |
| 7            | 6             | 1.6                  | 0.4                 | F          |
| 8            | 6             | 1.1                  | 0.3                 | F          |
| 9            | 6             | 1.5                  | 0.7                 | NF         |
| <b>L3-L4</b> |               |                      |                     |            |
| 1            | 12            | 1.5                  | 0.5                 | F          |
| 2            | 12            | 1.7                  | 1.0                 | F          |
| 3            | 12            | 0.9                  | 0.2                 | F          |
| 4            | 9             | 2.0                  | 0.6                 | NF         |
| 5            | 9             | 0.8                  | 0.4                 | F          |
| 6            | 9             | 0.8                  | 0.3                 | F          |
| 7            | 6             | 1.6                  | 0.7                 | NF         |
| 8            | 6             | 1.6                  | 0.7                 | NF         |
| 9            | 6             | 2.9                  | 0.9                 | NF         |
| <b>L5-L6</b> |               |                      |                     |            |
| 1            | 12            | 3.5                  | 2.6                 | F          |
| 2            | 12            | 8.5                  | 2.5                 | NF         |
| 3            | 12            | 1.5                  | 0.7                 | NF         |
| 4            | 9             | 4.6                  | 2.1                 | NF         |
| 5            | 9             | 2.4                  | 1.3                 | NF         |
| 6            | 9             | 8.7                  | 4.2                 | NF         |
| 7            | 6             | 5.5                  | 2.4                 | NF         |
| 8            | 6             | 3.2                  | 1.6                 | NF         |
| 9            | 6             | 6.1                  | 3.0                 | NF         |

Abbreviations: F, fused; NF, not fused; RSA, radiostereometric analysis; 3D, three dimensional.  
 Note: Of 27 levels, 11 were *fused* and 16 were *not fused*.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

**Table 2** The final CT scan results with the third observer's contribution to achieve a consensus result

| Sheep ID | Months | 1st reader (M.W.) | 2nd reader (P.T.) | 3rd reader (M.S.) | Final CT scan results |
|----------|--------|-------------------|-------------------|-------------------|-----------------------|
| L1–L2    |        |                   |                   |                   |                       |
| 1        | 12     | F                 | NF                | F                 | F                     |
| 2        | 12     | NF                | F                 | NF                | NF                    |
| 3        | 12     | F                 | F                 |                   | F                     |
| 4        | 9      | NF                | F                 | NF                | NF                    |
| 5        | 9      | NF                | NF                |                   | NF                    |
| 6        | 9      | NF                | F                 | F                 | F                     |
| 7        | 6      | NF                | F                 | F                 | F                     |
| 8        | 6      | NF                | NF                |                   | NF                    |
| 9        | 6      | NF                | NF                |                   | NF                    |
| L3–L4    |        |                   |                   |                   |                       |
| 1        | 12     | NF                | NF                |                   | NF                    |
| 2        | 12     | F                 | F                 |                   | F                     |
| 3        | 12     | F                 | F                 |                   | F                     |
| 4        | 9      | F                 | NF                | NF                | NF                    |
| 5        | 9      | F                 | NF                | F                 | F                     |
| 6        | 9      | F                 | NF                | F                 | NF                    |
| 7        | 6      | NF                | NF                |                   | NF                    |
| 8        | 6      | NF                | NF                |                   | NF                    |
| 9        | 6      | NF                | F                 | F                 | F                     |
| L5–L6    |        |                   |                   |                   |                       |
| 1        | 12     | F                 | F                 |                   | F                     |
| 2        | 12     | NF                | NF                |                   | NF                    |
| 3        | 12     | NF                | NF                |                   | NF                    |
| 4        | 9      | NF                | F                 | NF                | NF                    |
| 5        | 9      | NF                | NF                |                   | NF                    |
| 6        | 9      | NF                | F                 | F                 | F                     |
| 7        | 6      | NF                | F                 | F                 | F                     |
| 8        | 6      | NF                | NF                |                   | NF                    |
| 9        | 6      | NF                | NF                |                   | NF                    |

Abbreviations: CT, computed tomographic; F, fused; NF, not fused.

Note: Of 27 levels 11 were *fused* and 16 were *not fused*.

second radiologist (P.T.) reported his results with 12 *fused* in the first reading and 10 in the second reading. Where there was a discrepancy between the two readers, an additional opinion was obtained by a third radiologist to reach a consensus (M.S.). The second reading has been used to measure the interobserver error. In the final CT scan diagnosis, 11 levels were *fused* and 16 levels were *not fused* (►Table 2).

## Histology

- The histological assessment of fusion was performed by one research scientist skilled in histological assessment

(R.J.M.) and one consultant histopahtologist (D.W.E.). Both observers were blinded to the results of other modalities and blinded to the results of each other. ►Table 3 shows the results reported in this study. The results of histology assessment between both observers reached 100% agreement. There were 3 levels reported as *fused* and 25 levels reported as *not fused*.

## Summary Results

- The results of the three different modalities utilized for the assessment of spinal fusion are summarized below in ►Table 4.

**Table 3** The histology results of assessment of fusion

| Sheep ID | Months | Histology results (R.J.M.) | Histology results (D.W.E.) | Final histology results |
|----------|--------|----------------------------|----------------------------|-------------------------|
| L1-L2    |        |                            |                            |                         |
| 1        | 12     | F                          | F                          | F                       |
| 2        | 12     | NF                         | NF                         | NF                      |
| 3        | 12     | NF                         | NF                         | NF                      |
| 4        | 9      | NF                         | NF                         | NF                      |
| 5        | 9      | NF                         | NF                         | NF                      |
| 6        | 9      | NF                         | NF                         | NF                      |
| 7        | 6      | NF                         | NF                         | NF                      |
| 8        | 6      | NF                         | NF                         | NF                      |
| 9        | 6      | NF                         | NF                         | NF                      |
| L3-L4    |        |                            |                            |                         |
| 1        | 12     | F                          | F                          | F                       |
| 2        | 12     | NF                         | NF                         | NF                      |
| 3        | 12     | NF                         | NF                         | NF                      |
| 4        | 9      | NF                         | NF                         | NF                      |
| 5        | 9      | NF                         | NF                         | NF                      |
| 6        | 9      | NF                         | NF                         | NF                      |
| 7        | 6      | NF                         | NF                         | NF                      |
| 8        | 6      | NF                         | NF                         | NF                      |
| 9        | 6      | NF                         | NF                         | NF                      |
| L5-L6    |        |                            |                            |                         |
| 1        | 12     | F                          | F                          | F                       |
| 2        | 12     | NF                         | NF                         | NF                      |
| 3        | 12     | NF                         | NF                         | NF                      |
| 4        | 9      | NF                         | NF                         | NF                      |
| 5        | 9      | NF                         | NF                         | NF                      |
| 6        | 9      | NF                         | NF                         | NF                      |
| 7        | 6      | NF                         | NF                         | NF                      |
| 8        | 6      | NF                         | NF                         | NF                      |
| 9        | 6      | NF                         | NF                         | NF                      |

Abbreviations: F, fused; NF, not fused.

Note: Of 27 levels, 3 levels were fused and 24 were not fused.

**Statistical Analysis**

- The results of the intrarater and interrater reliability tests are indicated in ► **Table 5**. There was a moderate agreement between the two readings conducted 4 to 6 weeks apart by the first radiologist (kappa = 0.60, p value < 0.001) and no agreement for the second radiologist (kappa = 0.09, p value = 0.328).
- The assessment of the interrater reliability of the first CT scan of the two radiologists indicated no agreement (kappa = 0.07, p value = 0.353). As indicated in the methods section, when a discrepancy occurred, the third independent radiologist assessed the CT scans, with this review

providing the final CT scan result. The Fleiss Kappa statistic indicated no agreement between the first and third radiologist (kappa = 0.10, p value = 0.333) on the occasions where the results of the third radiologist were known (n = 12 of the 27 results). The Fleiss Kappa statistic between the third and third radiologist also indicated no agreement (kappa = -0.13, p value = 0.668).

- The sensitivity, specificity, positive predictive value, and negative predictive value were obtained for each method of assessment (► **Table 6**). The final CT scan results were compared with final RSA results with RSA considered as the standard technique of measuring movement between two bodies. CT scans have a 63.6% sensitivity and 75.0%



**Table 4** Summary of the fusion assessed by the three different modalities

| Sheep ID | Months | RSA result | CT scan result | Histology result |
|----------|--------|------------|----------------|------------------|
| L1–L2    |        |            |                |                  |
| 1        | 12     | F          | F              | F                |
| 2        | 12     | NF         | NF             | NF               |
| 3        | 12     | F          | F              | NF               |
| 4        | 9      | F          | NF             | NF               |
| 5        | 9      | NF         | NF             | NF               |
| 6        | 9      | NF         | F              | NF               |
| 7        | 6      | F          | F              | NF               |
| 8        | 6      | F          | NF             | NF               |
| 9        | 6      | NF         | NF             | NF               |
| L3–L4    |        |            |                |                  |
| 1        | 12     | F          | NF             | F                |
| 2        | 12     | F          | F              | NF               |
| 3        | 12     | F          | F              | NF               |
| 4        | 9      | NF         | NF             | NF               |
| 5        | 9      | F          | F              | NF               |
| 6        | 9      | F          | NF             | NF               |
| 7        | 6      | NF         | NF             | NF               |
| 8        | 6      | NF         | NF             | NF               |
| 9        | 6      | NF         | F              | NF               |
| L5–L6    |        |            |                |                  |
| 1        | 12     | F          | F              | F                |
| 2        | 12     | NF         | NF             | NF               |
| 3        | 12     | NF         | NF             | NF               |
| 4        | 9      | NF         | NF             | NF               |
| 5        | 9      | NF         | NF             | NF               |
| 6        | 9      | NF         | F              | NF               |
| 7        | 6      | NF         | F              | NF               |
| 8        | 6      | NF         | NF             | NF               |
| 9        | 6      | NF         | NF             | NF               |

Abbreviations: CT, computed tomography; F, fused; NF, not fused; RSA, radiostereometric analysis.

**Table 5** Results of intrarater and interrater reliability tests

| Fleiss Kappa test                   | Kappa statistic | p Value | Agreement (%) |
|-------------------------------------|-----------------|---------|---------------|
| Two tests for Radiologist 1         | 0.60            | < 0.001 | 80.0          |
| Two tests for Radiologist 2         | 0.09            | 0.328   | 55.6          |
| First test for Radiologists 1 and 2 | 0.07            | 0.353   | 55.6          |
| First test for Radiologists 1 and 3 | 0.10            | 0.333   | 50.0          |
| First test for Radiologists 2 and 3 | -0.13           | 0.668   | 50.0          |

**Table 6** Results of screening analysis

| Comparison       | Sensitivity (%) | Specificity (%) | PPV (%) | NPV (%) |
|------------------|-----------------|-----------------|---------|---------|
| CT <sup>a</sup>  | 63.6            | 75.0            | 63.6    | 75.0    |
| RSA <sup>b</sup> | 100.0           | 66.7            | 27.3    | 100.0   |
| CT <sup>b</sup>  | 66.7            | 60.0            | 16.7    | 93.8    |

Abbreviations: CT, computed tomography; NPV, negative predictive value; PPV, positive predictive value; RSA, radiostereometric analysis.

<sup>a</sup>Compare with RSA.

<sup>b</sup>Compare with histology.

specificity to detect a fusion (positive predictive value [PPV] = 63.6%, negative predictive value [NPV] = 75.0%).

- If one assumes histological assessment to be the gold standard for fusion, RSA showed better results (100% sensitivity and 66.7% specificity [PPV = 27.3%, NPV = 100.0%]) compared with CT assessment (66.7% sensitivity and 60.0% specificity [PPV = 16.7%, NPV = 93.8%], respectively) (► **Table 6**).

## Discussion

### Does CT Scan Provide an Accurate and Reliable Assessment of Spinal Fusion?

- CT scan is a static assessment of spinal fusion. For the assessment of anterior lumbar interbody fusion, it relies on the presence of bridging bony trabeculae crossing from one vertebral end plate to another.<sup>6</sup> Modern CT scanning allows assessment of fusion in multiple planes (i.e., coronal, sagittal, and axial). Imperative to the assessment is the thickness of the slice used when the investigation is performed. We employed a fine cut CT scan (0.25-mm slice), which is considered as one of the best available techniques for the assessment of fusion.<sup>10,11</sup>
- We employed histology as the gold standard in our study, which provided direct tissue diagnosis of fusion.<sup>12-15</sup> When we used histology as the gold standard, we found CT scan to have 66.7% sensitivity and 60.0% specificity for detecting FUSION. There was significant inter- and intraobserver error amongst radiologists (► **Table 5**). This questions the accuracy and reliability of CT scans in the assessment of fusion. The radiologists are experts in this field but still found it difficult to provide a diagnosis with high precision. In our study, Radiologist 2 (P.T.) had a low intraobserver agreement compared with Radiologist 2 (W.M.), and this may be explained by the fact that W.M. is a dedicated musculoskeletal radiologist whereas P.T. is a senior general radiologist.

### Does RSA Provide an Accurate and Reliable Assessment of Spinal Fusion?

- RSA has the advantage over CT scan in that it is a dynamic method for assessing fusion. It measures the “movement” across the tested motion segment when the spine is placed in a flexed and extended position. Once little or no movement has been detected by RSA, one can assume the segment is fused.

- In this study, RSA has 100% sensitivity and 66.7% specificity (PPV = 27.3%, NPV = 100.0%). This relatively low specificity may indicate RSA may overestimate fusion and miss cases of pseudarthrosis or incomplete fusion. This overestimation of fusion may be because of locked pseudarthrosis or stiffness of the tested disc space after discectomy and attempted fusion. This stiffness limits the movement and gives a false impression of achieving bony fusion. Fraser and coworkers first coined the term “locked pseudarthrosis” when bone graft within the cage had fused with the superior and inferior end plate but failed to fuse within the cage.<sup>16</sup>
- This low PPV may be attributed to error in the assessment. This may be related to the over- or underprovocation. RSA involved a variety of instruments, and error can arise in one or all of the instruments used (radiographs, calibration table, digital scanner, and software algorithms).<sup>17</sup> All these have been considered in the development of modern software, which efficiently eliminates most of these errors.<sup>17</sup>
- RSA is also considered an invasive technique as it requires a surgical operation to insert the beads. In assessment of spinal fusion, the spine is already exposed to do the fusion and the beads can be inserted in the vertebral body above and below the fused disc without much extra dissection.

### Histology As the Gold Standard?

- In this study, we assumed *histology* to be the gold standard assessment of fusion.<sup>12-15</sup> However, both the RSA method and the CT scan method vastly increased the number of levels that were assessed as fused. RSA detected 11 levels as fused, compared with 11 levels as fused for CT and 3 levels as fused for histology. Another possible conclusion from this could be that histology vastly underestimated the number of levels that were fused. A limitation of the histological methods used in this study is that a single midsagittal slice through the middle of the fusion cage that was thought to be representative was chosen. The chances of bony trabeculations passing through this exact slice would have to be really quite low. In other words, the tissue that was sampled by histology was a very small part of the potential fusion mass. CT scanning has the ability to interrogate multiple planes of the potential fusion mass in the coronal, the sagittal, and the axial plane. Furthermore, with fine cuts of 0.25 mm, the chances of missing a segment of trabeculation would seem slight.

- RSA detects movement in the potentially fused segment; as such, it is assessing the total fusion mass of the segment by degrees of stiffness. It is sampling 100% of the fusion mass rather than 1 to 2% of the fusion mass that histology might have assessed in this study. Clearly, it is important to set the parameters of fusion carefully in this situation. We chose a movement of less than 2 degrees angular rotation and less than 0.7-mm translation from the literature as prerequisite for diagnosis of fusion.<sup>2-5</sup> Broadly speaking, there was much more agreement in the levels of fusion between the CT assessment and the RSA. These observations would tend to suggest that histology may have incorrectly assessed the number of fused levels, or to put it in another way, histology in our study should not, perhaps, be considered as the gold standard.

### Limitations of This Study

- The sample size was small in terms of the number of animals used ( $N = 9$ ). However, a total of 27 levels were operated and all were assessed. The timing of each assessment was deliberately chosen to provide a *wide variation* in the fusion rates of these levels, thereby maximizing the range of fusion for each method to assess.
- The RSA was performed by an independent RSA expert. It would have been better to have *more than* one assessor to improve the accuracy of this assessment, reducing the possibility of observer error. RSA is still a research tool and a relatively new technique in Australia. There are limited people with the experience of operating and analyzing RSA results.
- The RSA was performed with the animal anesthetized and the front and back legs tied together to achieve flexion and hold the legs maximally apart for extension (► **Fig. 4a, b**). These positions were not quantitative and hard to reproduce with certainty. The positioning of all the sheep during the assessment was performed by one person as an attempt at uniformity.
- A limitation of the histological methods used in this study is that one midsagittal and two paracentral slices that were thought to be representative were chosen. The chances of bony trabeculations passing through this exact slice would have to be really quite low. In other words, tissue that was sampled by the histology was a very small part of the potential fusion mass. We should perhaps have used multiple slices to be more representative.
- Perhaps we could have employed *ex vivo* biomechanical testing of the extracted spines after they had undergone CT scanning. The stiffness and ROM for each motion segment could have been established and correlated with the *in vivo* RSA results.

### Conclusion

- This study concluded that in an animal model, RSA is at least as good as CT scan for the assessment of fusion. RSA has a fraction of the radiation penalty that CT scanning

currently has. Consideration should be given to further testing RSA, using a larger sample size. Because RSA is considered to be an invasive technique, the beads need to be inserted at the time of original surgery. In the case of spinal fusion, the spine is already exposed and the beads can be inserted in the vertebral body above and below the fused disc without much extra dissection. This makes RSA feasible to be tested in humans.

### Conflict of Interest

None of the authors have received any payment or consideration from any source for the conduct of this study.

### References

- 1 Baramki HG, Papin P, Steffen T. A surgical approach to the ventral aspect of the lumbar vertebrae in the sheep model. *Surg Radiol Anat* 2000;22(1):25-27
- 2 Pape D, Fritsch E, Kelm J, et al. Lumbosacral stability of consolidated anteroposterior fusion after instrumentation removal determined by roentgen stereophotogrammetric analysis and direct surgical exploration. *Spine (Phila Pa 1976)* 2002;27(3):269-274
- 3 Johnsson R, Selvik G, Strömquist B, Sundén G. Mobility of the lower lumbar spine after posterolateral fusion determined by roentgen stereophotogrammetric analysis. *Spine (Phila Pa 1976)* 1990;15(5):347-350
- 4 Axelsson P, Johnsson R, Strömquist B. Radiostereometry in lumbar spine research. *Acta Orthop Suppl* 2006;77(323):1-42
- 5 Zoëga B, Kärrholm J, Lind B. Mobility provocation radiostereometry in anterior cervical spine fusions. *Eur Spine J* 2003;12(6):631-636
- 6 Shah RR, Mohammed S, Saifuddin A, Taylor BA. Comparison of plain radiographs with CT scan to evaluate interbody fusion following the use of titanium interbody cages and transpedicular instrumentation. *Eur Spine J* 2003;12(4):378-385
- 7 Cunningham BW, Kanayama M, Parker LM, et al. Osteogenic protein versus autologous interbody arthrodesis in the sheep thoracic spine. A comparative endoscopic study using the Bagby and Kuslich interbody fusion device. *Spine (Phila Pa 1976)* 1999;24(6):509-518
- 8 Williams AL, Gornet MF, Burkus JK. CT evaluation of lumbar interbody fusion: current concepts. *AJNR Am J Neuroradiol* 2005;26(8):2057-2066
- 9 Dickson G. *Methods of Calcified Tissue Preparation*. New York, NY: Elsevier Science; 1984
- 10 Carreon LY, Glassman SD, Djurasovic M. Reliability and agreement between fine-cut CT scans and plain radiography in the evaluation of posterolateral fusions. *Spine J* 2007;7(1):39-43
- 11 Carreon LY, Djurasovic M, Glassman SD, Sailer P. Diagnostic accuracy and reliability of fine-cut CT scans with reconstructions to determine the status of an instrumented posterolateral fusion with surgical exploration as reference standard. *Spine (Phila Pa 1976)* 2007;32(8):892-895
- 12 Solchaga LA, Hee CK, Aguiar DJ, et al. Augment bone graft products compare favorably with autologous bone graft in an ovine model of lumbar interbody spine fusion. *Spine (Phila Pa 1976)* 2012;37(8):E461-E467
- 13 Grauer JN, Patel TC, Erulkar JS, Troiano NW, Panjabi MM, Friedlaender GE. 2000 Young Investigator Research Award winner. Evaluation of OP-1 as a graft substitute for intertransverse process lumbar fusion. *Spine (Phila Pa 1976)* 2001;26(2):127-133

- 14 Sidhu KS, Prochnow TD, Schmitt P, Fischgrund J, Weisbrode S, Herkowitz HN. Anterior cervical interbody fusion with rhBMP-2 and tantalum in a goat model. *Spine J* 2001;1(5):331–340
- 15 Nakajima T, Iizuka H, Tsutsumi S, Kayakabe M, Takagishi K. Evaluation of posterolateral spinal fusion using mesenchymal stem cells: differences with or without osteogenic differentiation. *Spine (Phila Pa 1976)* 2007;32(22):2432–2436
- 16 Santos ER, Goss DG, Morcom RK, Fraser RD. Radiologic assessment of interbody fusion using carbon fiber cages. *Spine (Phila Pa 1976)* 2003;28(10):997–1001
- 17 Derbyshire B, Prescott RJ, Porter ML. Notes on the use and interpretation of radiostereometric analysis. *Acta Orthop* 2009; 80(1):124–130

## Editorial Perspective

The reviewers found great value in this well-performed study for a number of reasons. While computed tomography (CT) scans have been accepted as the best available diagnostic medium, radiostereometric analysis (RSA) may indeed hold advantages over CT scan. Radiation exposure and need for reference markers, which may have to be implanted separately as well as device costs, have yet to be addressed prior to making RSA a preferred diagnostic medium. The other,

somewhat unexpected finding was that of histology being inferior to imaging in assessing fusion. While the methodology chosen to sample intervertebral tissue for fusion assessment may be questionable, this study makes a strong case against using this modality for future studies. As we continue to look for the ideal way to assess integrity of a fusion, we yet have to find any corollary between clinical symptoms and occult nonunions for fusion patients.