

Inter- and Intra-Observer Reliability of Measurement of Pedicle Screw Breach Assessed by Postoperative CT Scans

William F. Lavelle, Ashish Ranade, Amer F. Samdani, John P. Gaughan, Linda P. D'Andrea and Randal R. Betz

Int J Spine Surg 2014, 8 ()
doi: <https://doi.org/10.14444/1011>
<http://ijssurgery.com/content/8/11>

This information is current as of April 17, 2021.

Email Alerts Receive free email-alerts when new articles cite this article. Sign up at:
<http://ijssurgery.com/alerts>

INTERNATIONAL
JOURNAL
of
SPINE
SURGERY

This article generously published free of charge by the International Society for the Advancement of Spine Surgery.



INTERNATIONAL
SOCIETY *for the* ADVANCEMENT *of*
SPINE SURGERY

Inter- and Intra-Observer Reliability of Measurement of Pedicle Screw Breach Assessed by Postoperative CT Scans

William F. Lavelle, MD,¹ Ashish Ranade, MD,² Amer F. Samdani, MD,³ John P. Gaughan, PhD,⁴ Linda P. D'Andrea, MD,⁵ Randal R. Betz, MD⁶

¹Department of Orthopaedic Surgery, State University of New York Upstate Medical University, Syracuse, NY; ²Department of Orthopaedics, SKN Medical College & Hospital, Deenanath Mangeshkar Hospital, Pune, India; ³Shriners Hospitals for Children—Philadelphia, Philadelphia, PA; ⁴Biostatistics Consulting Center, Temple University School of Medicine, Philadelphia, PA; ⁵Brandywine Institute of Orthopaedics, Pottstown, PA; ⁶Institute for Spine & Scoliosis, Lawrenceville, NJ

Abstract

Background

Pedicle screws are used increasingly in spine surgery. Concerns of complications associated with screw breach necessitates accurate pedicle screw placement. Postoperative CT imaging helps to detect screw malposition and assess its severity. However, accuracy is dependent on the reading of the CT scans. Inter- and intra-observer variability could affect the reliability of CT scans to assess multiple screw types and sites. The purpose of this study was to assess the reliability of multi-observer analysis of CT scans for determining pedicle screw breach for various screw types and sites in patients with spinal deformity or degenerative pathologies.

Methods

Axial CT scan images of 23 patients (286 screws) were read by four experienced spine surgeons. Pedicle screw placement was considered 'In' when the screw was fully contained and/or the pedicle wall breach was ≤ 2 mm. 'Out' was defined as a breach in the medial or lateral pedicle wall > 2 mm. Intra-class coefficients (ICC) were calculated to assess the inter- and intra-observer reliability.

Results

Marked inter- and intra-observer variability was noticed. The overall inter-observer ICC was 0.45 (95% confidence limits 0.25 to 0.65). The intra-observer ICC was 0.49 (95% confidence limits 0.29 to 0.69). Underlying spinal pathology, screw type, and patient age did not seem to impact the reliability of our CT assessments.

Conclusion

Our results indicate the evaluation of pedicle screw breach on CT by a single surgeon is highly variable, and care should be taken when using individual CT evaluations of millimeters of breach as a basis for screw removal. This was a Level III study.

Keywords: inter/intra observer reliability, CT imaging, pedicle screw breach
Volume 8 Article 11 doi: 10.14444/1011

Introduction

Pedicle screws are used increasingly in spine surgery. In particular, pedicle screws in the thoracic spine have offered surgeons an attractive alternative to hook and wire constructs, with the potential of rigid three column spinal fixation and improved coronal and axial correction.^{1,2,3,4} Unfortunately, along with the potential of added stability comes an increased risk of injury due to the close proximity of screw trajectories to critical neurological and vascular structures.⁵ Postoperative imaging helps in detecting screw malposition and assessing its severity.

Primarily, pedicle screw position has been assessed by x-ray or CT imaging, with CT imaging currently considered the preferred imaging modality.⁶ CT scans have been reported to be more accurate than x-rays in assessing pedicle screw placement; however, the same investigations have also reported a broad range for the accuracy, with up to 40% of screws reported as “misplaced” on CT scans.^{7,8,9,10} Inter- and intra-observer variability in interpreting screw position on CT imaging could affect the outcome of these studies that report screws as misplaced. Various factors, including the type of screw used, associated scatter, and difficult visualization of anatomical landmarks may affect precise measurement of a breach. Specifically, titanium pedicle screws have been reported to facilitate CT analysis, as they leave less artifact during scanning than other metals such as stainless steel or cobalt-chrome. Yoo et al. showed that the scanning artifact created by cobalt-chrome screws made identification of the screw more difficult than titanium screws, but easier than stainless screws which have been reported to hinder CT analysis.¹¹ Choma et al. revealed that assessment of the correct position of stainless steel screws was more difficult than titanium screws.¹² In addition, particular screw sites have been found by other investigators to have a higher propensity for pedicle breach, perhaps requiring that screws at these fixation sites be more carefully placed and more closely scrutinized upon CT review.¹³ Pedicle breach was significantly higher in the thoracic spine compared to the lumbosacral spine (31.6% and 10.6%, respectively).¹³ Though reliability has been previously studied,⁸ the purpose of our study was to investigate the rater reliability of pedicle screw breaches as interpreted by multiple experienced surgeon raters. We also specifically investigated screw type as well as the type of spinal pathology, either degenerative or deformity, as factors that may affect rater reliability.

Materials and Methods

After obtaining IRB approval, 268 screws were placed in 23 patients as a part of a prospective multicenter study evaluating efficacy of a pedicle drilling device. Surgeries were performed by different surgeons at various locations. Postoperative CT scans were obtained for all patients to evaluate the accuracy of placement. Appropriate review by the radiation safety committee was completed at each of the institutions. Axial images were blinded and assessed by at least two independent observers. The number of observers varied between two and four depending on availability for reading the scans. These observers were experienced spine surgeons skilled in evaluating pedicle screw placement by means of CT scan.

The criteria for evaluation were as follows: screws were graded "in" when the screw was fully contained and/or the pedicle wall breach was ≤ 2 mm. "Out" was defined as a breach in the medial or lateral pedicle wall > 2 mm. Thoracic pedicle screw placement using the in-out technique was considered "out" if the lateral breach was more than 2 mm. All of the observers for this study followed the same criteria for defining the screw position.

Of the twenty-three patients, fifteen were diagnosed with degenerative spine pathology and eight patients were diagnosed with spinal deformity. Twelve patients received 193 stainless steel pedicle screws and eleven patients received 93 titanium pedicle screws. (Table 1) Standard CT sequences were utilized. We optimized pedicle screw visualization with respect to the pedicle using 3mm fine axial cut CT images with bone windows. It was also determined that our ability to discern a pedicle breach was 2 mm, and this became our aforementioned criterion for our categorization of "in" and "out" breaches. Figure 1 and Figure 2 demonstrate breaches of greater than and less than 2 mm respectively. In addition, previous studies utilized a similar 2 mm incitement in their analysis. Two millimeters is often considered a critical breach as described by Belmont et al.¹⁴ Further, Reynolds et al. previously demonstrated radiographic evidence of a 2 mm of lateral epidural space from T7 to L4.¹⁵ This was confirmed by Gertzbein and Robbins who examined 71 thoracic screws (T8–T12) with a 26% incidence of medial cortical breaches.¹⁶ These authors again noted a 2 mm epidural space and the 2 mm subarachnoid space. All of these studies consider screws with a 2 mm breach as clinically acceptably and believed to be accompanied by cortical expansion and benign pedicle wall fracture.

Table 1. Distribution of Screws.

| | # of patients | # of screws |
|-----------------|---------------|-------------|
| Total | 23 | 286 |
| Titanium | 11 | 93 |
| Stainless steel | 12 | 193 |

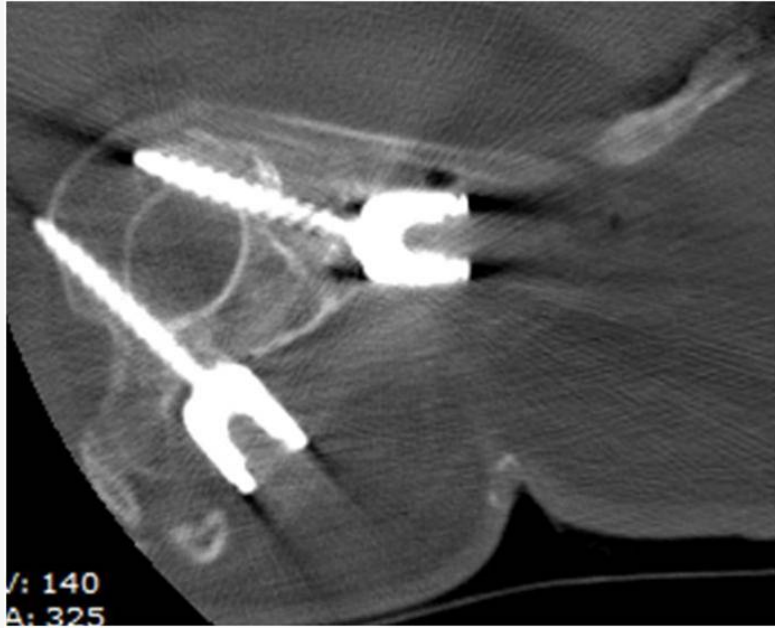


Figure 1. Example demonstrating a less than 2 mm breach representative of best agreement.

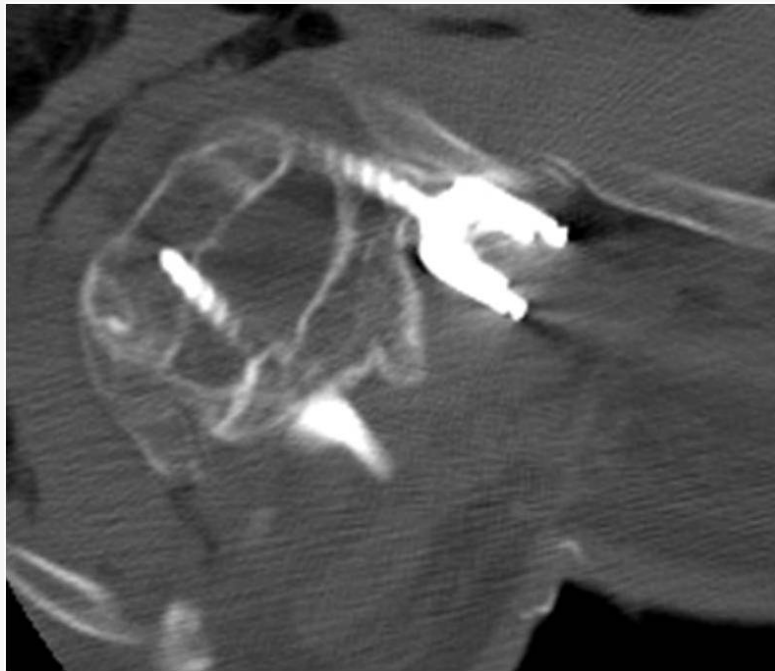


Figure 2. Example demonstrating a greater than 2 mm breach representative of worst agreement.

All CT scans were measured on the computer screen.

Statistical Methods

Binary categories (i.e., breach, no breach) are binomially distributed outcome data. In order to use traditional parametric statistics (ANOVA) usually based on normally distributed data to calculate the inter- and intra-rater reliability of binary outcomes.^{17,18} via the intraclass correlation coefficient (ICC) (Shrout and Fleiss, models 2k and 3k¹⁹), it was necessary to transform the data to normalized ranks.^{17,18} The ICC was then calculated using analysis of variance (ANOVA) for repeated measures with a nested observer effect and multiple screws per patient. All statistical analyses were carried out using SAS V9.1 statistical software (SAS Institute, Cary, NC). An ICC value of 0.90 and above reflects excellent reliability. Values between 0.75 and 0.89 suggest moderate reliability, and those falling below 0.75 suggest poor agreement.²⁰ Each ICC is accompanied by the 95% confidence interval (CI). The 95% CI provides an indication of the level of precision of the coefficients such that a wide CI is considered low precision.

Results

Twenty-three patients underwent placement of 286 pedicle screws. All patients included in the study had a postoperative CT scan and there were no exclusions. Marked inter- and intra-observer variability was noticed. The exact breach rate was not calculated since the purpose of this study was to assess the reliability of multi-observer analysis of CT scans for determining pedicle screw breach. The overall inter-observer ICC was 0.45 (95% confidence limits 0.25 to 0.65). The intra-observer ICC was 0.49 (95% confidence limits 0.29 to 0.69) suggesting poor inter- and intra-observer reliability. Several data observations were not available for the effect of age and the effect of diagnosis (deformity versus degenerative) categories. We also did not calculate the ICC separately for medial or lateral breaches. Although medial breaches are clinically relevant, we believed that lateral breaches were more important because they can injure vascular structures nearby.

Of the 286 screws, only 262 screws were accessible for studying the effect of diagnosis (degenerative versus deformity) on inter-rater ICC. While analyzing the effect of age on inter-rater ICC, only 261 screws were available. The disparity among these numbers may be attributed to loss of data points between the multi-centers.

Degenerative versus deformed spine

See Table 2. There were 15 patients (128 screws) with degenerative pathologies and 8 patients (134 screws) with a spinal deformity. Inter-rater ICC for deformity was 0.38 and for degenerative 0.21. The underlying spinal pathology did not appear to impact the reliability of CT assessment.

Table 2. Reliability of CT Reading by Diagnosis.

| Diagnosis | # of patients | # of screws | Inter-observer ICC (UCL-LCL) | Intra-observer ICC (UCL-LCL) |
|--------------|---------------|-------------|------------------------------|------------------------------|
| Degenerative | 15 | 128 | 0.21 (0.44-0.05) | 0.26 (0.50-0.07) |
| Deformity | 8 | 134 | 0.38 (0.61-0.18) | 0.41 (0.63-0.21) |

ICC = Intraclass correlation coefficient; UCL = 95% upper confidence limit; LCL = 95% lower confidence limit. Only 262/286 screws were available for evaluating the effect of diagnosis on Inter- and Intrarater ICC.

Type of screw

See Table 3. The screw type did not appear to affect the reliability of the CT assessment. There were 12 patients (193 screws) in the stainless steel group and 11 patients (93 screws) in the titanium group. ICC was similar for titanium and stainless steel screws, 0.36 and 0.34 respectively.

Table 3. Reliability of CT Reading: Titanium Versus Stainless Steel.

| Type of screw | # of patients | # of screws | Inter-observer ICC (UCL-LCL) | Intra-observer ICC (UCL-LCL) |
|-----------------|---------------|-------------|------------------------------|------------------------------|
| Titanium | 11 | 93 | 0.306 (0.58-0.16) | 0.36 (0.58-0.16) |
| Stainless steel | 12 | 193 | 0.34 (0.58-0.14) | 0.40 (0.63-0.19) |

ICC = Intraclass correlation coefficient; UCL = 95% upper confidence limit; LCL = 95% lower confidence limit. All 286 screws were available for this comparison.

Effect of age

See Table 4. Patient age was also investigated. After stratifying the study patients into age groupings (younger than 18 years, age between 18 and 60 years, and older than 60 years), no appreciable difference in intra-observer reliability was noticed. Twelve patients (156 screws) were younger than 18 years, 4 patients (53 screws) were between 18 and 60 years, and 7 patients (52 screws) were older than 60 years. All ICC scores were below the 0.75 benchmark, making them unreliable by definition (0.24, 0.37 and 0.16, respectively).

Table 4. Effect of Age on Reliability of CT Read

| Age | # of patients | # of screws | Inter-observer ICC (UCL-LCL) | Intra-observer ICC (UCL-LCL) |
|-------------|---------------|-------------|------------------------------|------------------------------|
| <18 years | 12 | 156 | 0.24 (0.51-0.04) | 0.25 (0.51-0.39) |
| 18-60 years | 4 | 53 | 0.37 (0.58-0.17) | 0.61 (0.60-0.19) |
| >60 years | 7 | 52 | 0.16 (0.42-0.39) | 0.19 (0.47-0.45) |

ICC = Intraclass correlation coefficient; UCL = 95% upper confidence limit; LCL = 95% lower confidence limit; 261 of 286 screws were available for studying the effect of age on the inter- and intrarater ICC.

Discussion

In this study, we observed poor reliability of CT scan assessment of pedicle screw placement among experienced inter- and intra-observers. It had been our assumption that senior surgeons would have had much higher agreement. However, one limitation of the study was that despite utilizing senior raters, we would have ideally used a greater number of raters for each scan. The specific intent of this study was to focus on patient and instrumentation factors that have been previously called into question as limiting the reliability and accuracy of CT analysis of pedicle screw placement. CT scans are considered to be the most accurate methods for assessing the accuracy of pedicle screw

placement.¹ CT imaging, however, does pose risks of radiation exposure and are typically reserved for patients who have experienced surgically related complications. Our study looked at screw placement in normal postoperative patients. Ideally, the study would have been improved with a larger number of patients, but this must be weighed against the risks of radiation exposure. With that being said, the long-term outcome of screw breaches that are potentially small and initially clinically silent is still unknown. The purpose of our study was to appreciate the rate and extent of screw misplacements, not to advocate for the need for CT scans after surgery. In a meta-analysis looking at pedicle screw placement accuracy, Kosmopoulos and Schizas identified 35 different pedicle screw placement assessment methods.²¹ In this study, the authors identified 130 studies incorporating 37,337 pedicle screw implantations. The authors stated a need for a standardized method for assessment of pedicle screw placement. The study does not endorse one particular method or assessment criterion.

Several studies look at the variability associated with CT scan-based assessment of pedicle screw accuracy. Rao et al. compared the position of a screw with direct visualization of the instrumented specimen.²² There was moderate agreement (mean kappa score of 0.51). The inter-observer kappa value for titanium screws was higher than the one for stainless steel screws. The Rao and Kosmopoulos studies have also demonstrated that the accuracy rate for CT imaging and higher inter-observer reliability occur when titanium pedicle screws are utilized. Intra-observer agreement was substantial (mean kappa score of 0.63). The study showed that artifact and flare from stainless steel can affect the reliability of CT scans in determining the accuracy of pedicle screw placement. Yoo et al. have reported similar findings.¹¹ In their study, the sensitivity of CT scanning in assessing the accuracy of pedicle screw placement in the lumbar spine was 86±5% for titanium screws and 67±6% for cobalt-chrome screws. In a cadaveric study by Fayyazi et al., CT scans were read for assessing intra- and inter-observer reliability.²³ In this study, screw placement in the rib head was not considered a malposition. The average sensitivity and specificity for assessment of malpositioned screws for all observers was 76±16% and 75±13%, respectively. Inter-observer kappa values showed large variability. Three observers correctly identified 8 of 20 screws (40%) with medial malposition. Four of 19 (21%) were correctly identified with lateral malposition, but they were unable to identify any of the six screws (0%) with inferior malposition.

In another study by Kosmopoulos et al., 59 titanium screws were evaluated blindly by two radiologists.⁸ Coronal and axial reconstructed images were blindly assessed according to criteria established by Farber et al.²⁴ Three categories were defined: in, out, and questionable. "Out" was further subclassified into "medial" or "lateral," depending on the direction of the perforation. Inter-observer agreement was substantial for both axial and coronal images (kappa value 0.78 and 0.78, respectively). Intra-observer agreement was excellent for both observers using either axial or coronal images. All screws in this study were titanium, which might have resulted in the high level of agreement between the observations. Another reason for the high agreement could have been the use of simplified criteria to define the accuracy on CT scans. None of the studies have compared consensus versus single observer.

As demonstrated in this study, there was poor agreement among experienced spine surgeons in the interpretation of postoperative CT scans regarding pedicle screw breach for various screw types and sites in patients with spinal deformities or degenerative pathologies. It was difficult to define a significant breach due to the scatter associated with the screw. Precise measurement may prove to be difficult. Previously, it has been shown that a medial breach less than 2 mm and a lateral breach less than 6 mm are acceptable measures.²⁵ As the technology of CT imaging progresses, computer methods to reduce scatter may reduce the technical limitations related to interpreting screw placement and improve the reliability seen in future studies.

An obvious limitation of this study is that the screws could not be directly visualized as they were placed into living patients. This hinders the study, as only reliability statistics can be examined; as opposed to accuracy statistics, which can only be investigated in cadaver studies. However, we believe the opportunity to review spines in patients would lend our study greater clinical applicability. The question of asymptomatic breaches remains unsolved.^{26,27} Surgeons should take these factors into consideration before deciding to reposition or remove a screw.

References

1. Kim YJ, Lenke LG, Cho SK, et al. Comparative analysis of pedicle screw versus hook instrumentation in posterior spinal fusion of adolescent idiopathic scoliosis. *Spine* 2004;29:2040-2048.
2. Lee SM, Suk SI, Chung ER. Direct vertebral rotation: a new technique of three dimensional deformity correction with segmental pedicle screw fixation in adolescent idiopathic scoliosis. *Spine* 2004;29:343-349.
3. Luhmann SJ, Lenke LG, Kim YJ, et al. Thoracic adolescent idiopathic scoliosis curves between 70 degrees and 100 degrees: Is anterior release necessary? *Spine* 2005;30:2061-2067.
4. Suk SI, Lee SM, Chung ER, et al. Selective thoracic fusion with segmental pedicle screw fixation in the treatment of thoracic idiopathic scoliosis: more than 5-year follow-up. *Spine* 2005;30:1602-1609.
5. Kakkos SK, Shepard AD. Delayed presentation of aortic injury by pedicle screws: report of two cases and review of the literature. *J Vasc Surg* 2008;47:1074-1082.
6. Kim YJ, Lenke LG, Cheh G, et al. Evaluation of pedicle screw placement in the deformed spine using intraoperative plain radiographs: a comparison with computerized tomography. *Spine* 2005;30:2084-2088.
7. Castro WH, Halm H, Jerosch J, et al. Accuracy of pedicle screw placement in lumbar vertebrae. *Spine* 1996;21:1320-1324.
8. Kosmopoulos V, Theumann N, Binaghi S, et al. Observer reliability in evaluating pedicle screw placement using computed tomography. *Internat Orthop* 2007;31:531-536.
9. Learch TJ, Massie JB, Pathria MN, et al. Assessment of pedicle screw placement utilizing conventional radiography and computed tomography: a proposed systematic approach to improve accuracy of interpretation. *Spine* 2004;29:767-773.

10. Schizas C, Theumann N, Kosmopoulos V. Inserting pedicle screws in the upper thoracic spine without the use of fluoroscopy or image guidance. Is it safe? *Eur Spine J* 2007;16:625-629.
11. Yoo JU, Ghanayem AJ, Petersilge C, et al. Accuracy of using computed tomography to identify pedicle screw placement in cadaveric human lumbar spine. *Spine* 1997;22:2668-2671.
12. Choma TJ, Denis F, Lonstein JE, et al. Stepwise methodology for plain radiographic assessment of pedicle screw placement: a comparison with computed tomography. *J Spinal Disord Tech* 2006;19:547-553.
13. Rampersaud YR, Pik JH, Salonen D, et al. Clinical accuracy of fluoroscopic computer assisted pedicle screw fixation: a CT analysis. *Spine* 2005;30:E183-E190.
14. Belmont PK Jr, Klemme WR, Dhawan A, et al. In vivo accuracy of thoracic pedicle screws. *Spine* 2001 Nov 1; 26(21):2340-2346.
15. Reynolds AF, Roberts A, Pollay M, et al. Quantitative anatomy of the thoracolumbarepidural space. *Neurosurgery* 1985 Dec;17 (6):905-7.
16. Gertzbein SD, Robbins SE. Accuracy of pedicular screw placement in vivo. *Spine* 1990 Jan;15 (1):11-4.
17. Conover WJ, Iman RL. Rank transformations as a bridge between parametric and nonparametric statistics. *American Statistics* 1981;35:124-129.
18. Harter HL. Expected values of normal order statistics. *Biometrika* 1961;48:151-165.
19. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psych Bull* 1979;86:420-428.
20. Portney LG, Watkins MP. *Foundations of Clinical Research. Applications to Practice*. Upper Saddle River, NJ: Prentice Hall Health, 2000.
21. Kosmopoulos V, Schizas C. Pedicle screw placement accuracy. A meta-analysis. *Spine* 2007;32:E111-E120.
22. Rao G, Brodke DS, Rondina M, et al. Comparison of computerized tomography and direct visualization in thoracic pedicle screw placement. *J Neurosurg* 2002;97:223-226.
23. Fayyazi AH, Hugate RR, Pennypacker J, et al. Accuracy of computed tomography in assessing thoracic pedicle screw malposition. *J Spinal Disord Tech* 2004;17:367-371.
24. Farber GL, Place HM, Mazur RA, et al. Accuracy of pedicle screw placement in lumbar fusions by plain radiographs and computed tomography. *Spine* 1995;20:1494-1499.
25. Belmont Jr PJ, Klemme WR, Robinson M, et al. Accuracy of thoracic pedicle screws in patients with and without coronal plane spinal deformities. *Spine* 2002;27:1558-1566.
26. Lehman Jr RA, Lenke LG, Keeler KA, et al. Computed tomography evaluation of pedicle screws placed in the pediatric deformed spine over an 8-year period. *Spine* 2007;32:2679-2684.
27. Polly Jr DW, Potter BK, Kuklo T, et al. Volumetric spinal canal intrusion: a comparison between thoracic pedicle screws and thoracic hooks. *Spine* 2004;29:63-69.

Funding

This study was supported by a research grant from SpineGuard, Inc.

Disclosures

Amer Samdani is a paid consultant for DePuy Synthes Spine, Stryker, and Zimmer. Randal Betz receives royalties from DePuy Synthes Spine and Medtronic, has received speaking fees from DePuy Synthes Spine, is a paid consultant for DePuy Synthes Spine, Orthocon, SpineGuard, Medtronic, and Zimmer, and owns stock in Advanced Vertebral Solutions, SpineGuard, MiMedx, Orthocon, Orthobond, and SpineZ. All other authors declare no financial disclosures.

IRB approval

Temple University School of Medicine, #4727 has been obtained for this study.

Corresponding Author

William F. Lavelle, MD, SUNY Upstate Medical University, 6620 Fly Road, Suite 200, East Syracuse, NY, 13057, email: lavellew@upstate.edu

Copyright © 2014 ISASS - International Society for the Advancement of Spine Surgery. To see more or order reprints or permissions, see <http://ijssurgery.com>.