Three-dimensional assessment of robot-assisted pedicle screw placement accuracy and instrumentation reliability based on a preplanned trajectory

*Bowen Jiang, MD,¹ Zach Pennington, BS,¹ Alex Zhu, PA-C,¹ Stavros Matsoukas, MD,² A. Karim Ahmed, BS,¹ Jeff Ehresman, BS,¹ Smruti Mahapatra, BE,³ Ethan Cottrill, MS,¹ Hailey Shepell,³ Amir Manbachi, PhD,³ Neil Crawford, PhD,⁴ and Nicholas Theodore, MD¹

¹Department of Neurosurgery, Johns Hopkins School of Medicine; ²Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland; ³Aristotle University of Thessaloniki School of Medicine, Thessaloniki, Greece; and ⁴Globus Medical, Audubon, Pennsylvania

OBJECTIVE Robotic spine surgery systems are increasingly used in the US market. As this technology gains traction, however, it is necessary to identify mechanisms that assess its effectiveness and allow for its continued improvement. One such mechanism is the development of a new 3D grading system that can serve as the foundation for error-based learning in robot systems. Herein the authors attempted 1) to define a system of providing accuracy data along all three pedicle screw placement axes, that is, cephalocaudal, mediolateral, and screw long axes; and 2) to use the grading system to evaluate the mean accuracy of thoracolumbar pedicle screws placed using a single commercially available robotic system.

METHODS The authors retrospectively reviewed a prospectively maintained, IRB-approved database of patients at a single tertiary care center who had undergone instrumented fusion of the thoracic or lumbosacral spine using robotic assistance. Patients with preoperatively planned screw trajectories and postoperative CT studies were included in the final analysis. Screw accuracy was measured as the net deviation of the planned trajectory from the actual screw trajectory in the mediolateral, cephalocaudal, and screw long axes.

RESULTS The authors identified 47 patients, 51% male, whose pedicles had been instrumented with a total of 254 screws (63 thoracic, 191 lumbosacral). The patients had a mean age of 61.1 years and a mean BMI of 30.0 kg/m². The mean screw tip accuracies were 1.3 ± 1.3 mm, 1.2 ± 1.1 mm, and 2.6 ± 2.2 mm in the mediolateral, cephalocaudal, and screw long axes, respectively, for a net linear deviation of 3.6 ± 2.3 mm and net angular deviation of $3.6^\circ \pm 2.8^\circ$. According to the Gertzbein-Robbins grading system, 184 screws (72%) were classified as grade A and 70 screws (28%) as grade B. Placement of 100% of the screws was clinically acceptable.

CONCLUSIONS The accuracy of the discussed robotic spine system is similar to that described for other surgical systems. Additionally, the authors outline a new method of grading screw placement accuracy that measures deviation in all three relevant axes. This grading system could provide the error signal necessary for unsupervised machine learning by robotic systems, which would in turn support continued improvement in instrumentation placement accuracy.

<https://thejns.org/doi/abs/10.3171/2020.3.SPINE20208>

KEYWORDS robot assistance; pedicle screw accuracy; three-dimensional accuracy; surgical technique

ROBOTIC technology is one of the most rapidly expanding areas in spine surgery, with increases in both the number of systems available nationwide and the number of publications on the subject.¹ Because the main purpose of these robotic systems is to assist with the accurate placement of pedicle screws, the most widely described endpoint is instrumentation accuracy. Numerous retrospective and prospective studies have demonstrated that robot-assisted instrumentation is as accurate

as,² if not more than, the freehand placement of screws.^{3,4} However, these studies have two critical shortcomings: 1) they only evaluate screw accuracy in two dimensions, and 2) they do not assess whether the trajectory of the placed screw is coincident with the planned trajectory.

Conventional systems of assessing the accuracy of pedicle screw placement—such as those of Gertzbein and Robbins,⁵ Heary,⁶ Wiesner,⁷ and Rampersaud⁸—are limited in that they capture only the degree to which screws

SUBMITTED February 13, 2020. ACCEPTED March 30, 2020.

INCLUDE WHEN CITING Published online May 29, 2020; DOI: 10.3171/2020.3.SPINE20208.

* B.J. and Z.P. contributed equally to this study.

violate the pedicular cortex in the mediolateral and cephalocaudal directions. They treat violations in both directions as equivalent and lack the granularity necessary to refine robot-assisted spine surgery systems. Identifying placement accuracy along the screw axis as well as within the mediolateral and cephalocaudal directions would deliver a better understanding of the true placement accuracy of the robotic system. Additionally, this information would greatly enhance feedback sent to the robotic system for error-based learning and accuracy refinement.

Thus, we sought to create a novel method of categorizing pedicle screw accuracy in three dimensions. Our specific goals were 1) to define a system capable of providing accuracy data along all three screw placement axes, that is, cephalocaudal, mediolateral, and screw long axes; and 2) to use the system to evaluate the mean accuracy of thoracolumbar pedicle screws placed using a single commercially available robotic system.

Methods

After obtaining approval from the institutional review board, we retrospectively reviewed a prospectively gathered database of all robot-assisted spine procedures performed at a single academic tertiary care center between October 1, 2017, and August 31, 2019. All patients enrolled in the database were adults (age ≥ 18 years) who underwent instrumented fusion operations of the thoracic and/or lumbar spine. From this database, we identified all patients who had preoperative screw trajectory plans and postoperative CT studies of the instrumented levels. All included patients had undergone instrumented fusion using a single robotic spine system (ExcelsiusGPS, Globus Medical Inc.), which has been approved by the FDA. Instrumentation was performed using either an open or a minimally invasive approach based on surgeon preference. Similarly, surgeon preference dictated whether CT or plain radiographs were used to assess postoperative instrumentation accuracy in the consecutive series, though all included patients had postoperative CT scans.

Screw Accuracy

The primary endpoint of the study was screw placement accuracy, which was measured using three techniques. As a gold-standard definition, accuracy was assessed using the system of Gertzbein and Robbins.⁵ Under this system, screws are classified as grades A–E based on the following criteria: grade A, no cortical breach; grade B, 0 to < 2 mm of cortical breach; grade C, 2 to < 4 mm of cortical breach; grade D, 4 to < 6 mm of cortical breach; and grade E, ≥ 6 mm of cortical breach. Screws graded as either A or B were classified as clinically acceptable, in keeping with prior studies.^{9–12} Screws were graded by two independent reviewers (Z.P. and A.K.A.), with disagreements in classification systems resolved by the lead author (B.J.).

To assess screw placement accuracy in three dimensions, we used two techniques: vector component analysis and volumetric overlap. The vector component analysis was performed for all placed screws and was conducted in a manner similar to that described by Jiang et al.¹³ Measurements were obtained by two authors (A.Z., S.M.), with

the lead author (B.J.) confirming measurement in cases of discrepancy. Preoperative image volumes with planned trajectories (Fig. 1) were superimposed on and coregistered with postoperative image volumes (Fig. 2), vertebra by vertebra, using custom software, and planned versus actual screw trajectories were compared. Deviation of the screw tip from the planned tip placement and deviation of the screw tail from the planned tail placement were assessed for linear deviation along the following three axes: radially away from the screw centerline in the craniocaudal direction, radially away from the screw centerline in the mediolateral direction, and longitudinally along the screw axis (gauge of error in depth of screw placement). Figure 3A and B illustrates how these measures were made using the actual software, and Fig. 4A and B serves as graphic representations of how these measurements were made. Three-dimensional accuracy was then assessed using the Euclidean norm. Angular deviation was calculated as the angle formed by the intersection of the vectors defining the planned and actual screws (Fig. 4C).

As a second method for assessing 3D screw accuracy, we looked at volumetric overlap in a randomly selected subset of the planned and actual screws. Using AutoCAD (Autodesk Inc.), we modeled each screw as a right circular cylinder, defined by a diameter equivalent to the outer diameter of the screw and a length equivalent to that of the placed screw. Models of both the planned and the actual screws were placed in a 3D space according to the angular and net tail deviations measured using the methodology above. Volumetric overlap was defined as the percentage of 3D space occupied by both the cylinder defining the planned screw and the cylinder defining the actual screw; accuracy varied on a scale from 0% overlap (screw completely off planned trajectory) to 100% (planned and actual screws share exact same trajectory and placement). Similar to above, angular deviation in this analysis was defined as the angle formed by the vector defining the long axes of the planned and actual screws. To analyze the impact of tail linear deviation (error in entry point placement) and angular deviation on volumetric overlap, we determined the volumetric overlap of two screw volumes using a multiple-angle linear deviation combination for a hypothetical 50-mm screw.

Statistical Analysis

All data were recorded using Excel (Microsoft Corp.) and analyzed using Statistica 13.3.0 (TIBCO). All values were reported as the mean $\pm$ standard deviation. Screw accuracy was reported for all placed screws and subdivided by the spinal region—thoracic, lumbar, sacral, and pooled lumbosacral. Comparison of linear and angular displacements between the thoracic, lumbar, and sacral regions was done using one-way ANOVA, with post hoc comparisons performed using the Tukey honestly significant difference (Tukey HSD) test. Direct comparison between the thoracic and lumbosacral regions was performed using the Mann-Whitney U-test for nonparametric parameters. A p value < 0.05 was considered statistically significant.

Results

From 93 total cases, we identified 47 that met the

FIG. 1. Example of screw trajectories plotted on a preoperative surgical plan.

inclusion criteria. The reason for exclusion in the overwhelming majority of cases was the absence of postoperative CT for evaluating screw accuracy (40 cases). The other reasons for exclusion were the need to revise screw trajectories intraoperatively (1 case) and corruption of the electronic file containing the plotted screw trajectories (5 cases).

Among the included patients (Table 1), the mean age was 61.1 years and the mean BMI was 30.0 kg/m². Fifty-one percent of the patients were men. Forty-four patients (242 screws) underwent open procedures, while 3 patients (12 screws) underwent minimally invasive procedures. The indication for surgery was degenerative pathology (42/47 [89.4%]), trauma (2/47 [4.3%]), tumor (2/47 [4.3%]),

FIG. 2. Preoperative planned trajectories were superimposed onto actual screw trajectories on the postoperatively acquired CT volume using the analysis software designed for the ExcelsiusGPS device.

FIG. 3. Planned and actual screw trajectories are compared in the analysis software. In this example, the accuracy of a left L5 screw is examined. **A:** Measurement of the tip and tail linear deviations along the cephalocaudal/craniocaudal axis using a sagittal plane reconstruction. **B:** Measurement of the tip and tail linear deviations along the medio-lateral and screw long axis using an axial plane reconstruction. *Blue dots* and *green dots* represent the tip of the placed screw and the tip of the planned screw, respectively. *Red dots* and *orange dots* represent the tail of the placed screw and the tail of the planned screw, respectively.

FIG. 4. Graphic representation of the methodology for measuring screw accuracy. **A:** Measurement of the tip and tail linear deviations along the cephalocaudal/craniocaudal axis. **B:** Measurement of the tip and tail linear deviations along the mediolateral and screw long axis. The planned screw trajectory (*green screw*) overlaps the actual screw (*purple screw*) in the analysis software. *Blue dots* and *green dots* represent the tips of the placed and planned screws, respectively. *Red dots* and *orange dots* represent the tails of the placed and planned screws, respectively. **C:** Screws were transformed into vectors of finite length, and linear deviation was calculated along the craniocaudal/cephalocaudal (*green*), mediolateral (*blue*), and screw long (*red*) axes. Overall linear deviation was calculated as the length of the net vector resulting from the addition of the component vectors. Angular deviation was determined by the angle formed by the intersection of the vectors defining the planned and actual screws.

and a tethered cord that was being treated with vertebral column resection (1/47 [2.1%]). Across all surgical procedures, the mean estimated blood loss was 357 ml, and the mean length of stay was 4.8 days.

TABLE 1. Demographics of 47 patients who underwent pedicle screw placement using robotic assistance

Variable	Value
Mean age in yrs	61.1 ± 14.2
Race, no. (%)	
White	35 (74.5)
Black	6 (12.8)
Other	6 (12.8)
Male sex, no. (%)	24 (51.1)
Mean weight in kg	85.4 ± 20.3
Mean height in m	1.685 ± 0.117
Mean BMI in kg/m ²	30.0 ± 6.0
Mean LOS in days	4.79 ± 4.0
Mean EBL in ml	357.3 ± 485.9
Surgical indication, no. (%)	
Degenerative	42 (89.4)
Trauma	2 (4.3)
Tumor	2 (4.3)
Other	1 (2.1)

EBL = estimated blood loss; LOS = length of stay.

Screw Accuracy Based on Gertzbein-Robbins

A total of 254 screws were placed using robotic assistance (Table 2). One hundred eighty-four screws (72%) were classified as Gertzbein-Robbins grade A, while the remainder (28%) were classified as Gertzbein-Robbins grade B. All screws met the definition of clinically acceptable as defined by prior groups.^{9–12} A comparison of screw accuracy by region showed a significantly higher proportion of grade A screws in the lumbosacral vertebrae than in the thoracic vertebrae (78% vs 56%, $p < 0.01$).

3D Screw Accuracy

The 3D accuracy of screws placed at each vertebral level is reported in Table 3. Screw accuracies are summarized by spine region in Table 4, which shows that the overall accuracy of the placed screws was quite good,

TABLE 2. Gertzbein-Robbins grading

Region	Total No.	Grade, No. (%)					
		A	B	C	D	E	A or B
T	63	35 (56)	28 (44)	0 (0)	0 (0)	0 (0)	63 (100)
LS	191	149 (78)	42 (22)	0 (0)	0 (0)	0 (0)	191 (100)
L	167	125 (75)	42 (25)	0 (0)	0 (0)	0 (0)	167 (100)
S	24	24 (100)	0 (0)	0 (0)	0 (0)	0 (0)	24 (100)
All	254	184 (72)	70 (28)	0 (0)	0 (0)	0 (0)	254 (100)

L = lumbar; LS = lumbosacral; S = sacral; T = thoracic.

with a mean angular deviation of $3.6^\circ \pm 2.8^\circ$ and a mean screw tip deviation of 3.6 ± 2.3 mm. Accuracy was best in the thoracic spine and lowest in the sacrum. One-way ANOVA showed significant differences between groups in 2D tip deviation along the superoinferior axis ($F = 3.70$, $p = 0.03$), 2D tail deviation along the mediolateral axis ($F = 8.05$, $p < 0.01$), angular offset ($F = 3.55$, $p = 0.03$), 3D tail deviation ($F = 3.88$, $p = 0.02$), and 3D overall accuracy ($F = 3.16$, $p = 0.04$; Table 5). Post hoc comparisons using the Tukey HSD test showed that 2D linear accuracy along the superoinferior access was superior for thoracic compared to sacral screws (1.2 ± 1.2 vs 1.8 ± 1.5 mm, $p = 0.03$) and for lumbar screws compared to sacral screws (1.2 ± 1.1 vs 1.8 ± 1.5 mm, $p = 0.03$). Thoracic screws were also found to have smaller mediolateral tail deviation (1.2 ± 1.1 vs 1.9 ± 1.5 mm, $p < 0.01$), smaller 3D tail deviation (3.2 ± 2.1 vs 4.2 ± 2.5 mm, $p = 0.02$), and smaller overall 3D deviation (3.2 ± 2.1 vs 4.0 ± 2.4 mm, $p = 0.04$), compared to lumbar screws. Angular deviation was also smaller for thoracic relative to sacral screws ($2.9^\circ \pm 2.6^\circ$ vs $4.6^\circ \pm 3.5^\circ$, $p = 0.03$). When lumbar and sacral screws were grouped as a single region, significant differences relative to thoracic screws were only found in tail linear deviation along the mediolateral axis (1.9 ± 1.5 vs 1.2 ± 1.1 mm, $p < 0.01$).

Using the volumetric analysis described, we analyzed 14 of the 254 screws. We found a mean overlap of $59.0\% \pm 21.5\%$ and a mean angular deviation of $5.3^\circ \pm 3.2^\circ$. Investigation of the impact of angular deviation and entry point linear deviation on screw volume overlap demonstrated that the majority of placement error could be attributed to angular deviation (Supplemental Fig. 1).

Discussion

In this investigation, our objectives were twofold: 1) to devise a method for evaluating the 3D accuracy of pedicle screw placement, and 2) to use this method to evaluate the accuracy of an FDA-approved spinal surgery robotic system. With regard to the former, we formulated a simple vector system for plotting 3D accuracy composed of angular deviation and the overall linear deviation. Using this method, we found that the robotic system under study was able to place screws within 4 mm of the planned position on average, with a mean angular deviation of less than 4° . Furthermore, placement accuracy was highest in the thoracic spine, which is noted to have the least tolerance for screw misplacement given the presence of the thoracic cord medially and the lung parenchyma laterally. Examination of the individual accuracy components showed that accuracy was greatest in the highest-risk dimensions: the mediolateral and craniocaudal axes. Screw deviation along these axes can result in spinal cord or nerve root injury, respectively. In both cases, the mean screw deviation was only 25% of the screw diameter. As confirmation of the high degree of accuracy of the placed screws, we found that all screws were clinically acceptable according to previously established interpretations of the Gertzbein-Robbins grading system.^{9–12} However, in a volumetric analysis of 14 randomly selected screws, we found only 59% overlap between planned and placed screw trajectories, suggesting the Gertzbein-Robbins system lacks the

granularity necessary to evaluate robotic surgical systems and support machine learning algorithms for automatic screw trajectory planning.

Prior Classification Systems

In the paradigm of freehand instrumentation, several classification systems have been developed to describe the accuracy of pedicle screw placement, including those of Gertzbein and Robbins,⁵ Heary,⁶ Wiesner,⁷ and Ramper-saud.⁸ Of these, the system of Gertzbein and Robbins⁵ and that of Heary⁶ have gained the greatest momentum in the spine literature. The older system—that of Gertzbein and Robbins—was developed to assess the accuracy of screw placement in thoracolumbar spines using the degree of cortical bony breach in the axial plane as the indicator of accuracy. Screws placed completely within the pedicle are classified as grade A and are considered to have perfect placement, those with less than 2 mm breach of the pedicle are grade B, etc., and screws more than 6 mm outside the pedicle are categorized as grade E. Since the introduction of the Gertzbein-Robbins system, grade A and B screws have generally been deemed clinically acceptable. This system has been used in numerous studies^{2–4} to assess the accuracy of robotic systems, but the ordinal nature of the grading system and the lack of a direction component prevent its use in an error-based feedback role for the improvement of robotic spine surgery systems.

Subsequently, Heary et al.⁶ described a classification system for thoracic pedicle screws that graded screws based on the direction of the cortical bone breach, giving different designations based on whether the breach existed along the long axis of the screw, the cephalocaudal axis, or the mediolateral axis. However, the degree of deviation was not quantified, again precluding its use in an error-based feedback role for the improvement of robotic spine surgery systems.

Prior Quantitative Descriptions of 3D Accuracy

To our knowledge, only a handful of studies have described the 3D accuracy of screws placed with robotic assistance. In the first, Togawa et al.¹⁴ described the placement of 43 screws in cadaveric spines as part of the proof-of-concept work for the bone-mounted SpineAssist robot (Mazor Surgical Technologies). The authors found that the average accuracy was 0.65 ± 0.57 mm along the cranio-caudal axis and 1.01 ± 0.64 mm along the mediolateral axis, with an overall average deviation of 0.96 ± 0.62 mm. Though these findings are more accurate than the present results, the screws in that study were placed with the assistance of Kirschner wires, which may account for the apparent differences in accuracy. Additionally, unlike in the present study, neither deviation along the long axis of the screw nor angular deviation was reported, limiting the ability of their results to provide the spine robot with error-based learning. Nevertheless, this preliminary effort provided the precedent upon which the present study is based and laid the foundation for robot-assisted spine surgery.

Subsequent to the study of Togawa and colleagues, Stürer et al.¹⁵ described the placement of 100 lumbar pedicle screws using the same robotic system. They reported an average deviation of 0.34 ± 1.49 mm from the preopera-

TABLE 3. Screw accuracy by level, among 254 screws

Level	No. of Screws	2D Accuracy: Tip			2D Accuracy: Tail			3D Accuracy			
		Rt/Lt	S/I	L/S	Rt/Lt	S/I	L/S	AD (°)	Tip	Tail	Overall
T1											
Rt	2	1.7 ± 0.3	1.3 ± 0.9	1.0 ± 0.6	1.8 ± 0.8	1.0 ± 0.3	0.9 ± 0.5	5.6 ± 1.6	2.5 ± 0.6	2.3 ± 0.7	2.4 ± 0.6
Lt	2	0.6 ± 0.8	1.5 ± 2.1	0.4 ± 0.6	1.7 ± 0.6	1.6 ± 1.9	0.5 ± 0.6	2.0 ± 2.3	1.6 ± 2.3	2.8 ± 0.8	2.2 ± 1.6
T2											
Rt	2	0.8 ± 1.2	0.0 ± 0.1	0.2 ± 0.3	1.1 ± 1.1	0.8 ± 0.5	0.2 ± 0.3	1.8 ± 1.6	0.9 ± 1.2	1.4 ± 1.2	1.1 ± 1.2
Lt	2	0.1 ± 0.0	0.9 ± 0.5	1.4 ± 0.2	0.7 ± 0.2	0.7 ± 0.1	1.4 ± 0.2	1.6 ± 0.0	1.7 ± 0.1	1.7 ± 0.1	1.7 ± 0.1
T3											
Rt	2	0.6 ± 0.8	0.4 ± 0.6	0.4 ± 0.5	0.4 ± 0.5	0.5 ± 0.4	0.4 ± 0.5	0.6 ± 0.4	1.1 ± 0.5	0.9 ± 0.1	1.0 ± 0.3
Lt	2	0.2 ± 0.2	0.7 ± 0.4	0.1 ± 0.1	0.2 ± 0.1	1.6 ± 1.7	0.1 ± 0.0	2.1 ± 1.7	0.8 ± 0.3	1.6 ± 1.7	1.2 ± 1.0
T4											
Rt	2	0.6 ± 0.3	0.4 ± 0.4	0.3 ± 0.1	0.4 ± 0.3	0.7 ± 0.3	0.3 ± 0.1	1.9 ± 0.6	0.9 ± 0.1	0.9 ± 0.4	0.9 ± 0.6
Lt	2	0.4 ± 0.1	0.5 ± 0.2	0.6 ± 0.2	0.2 ± 0.1	0.3 ± 0.1	0.6 ± 0.2	1.5 ± 1.0	1.0 ± 0.1	0.7 ± 0.2	0.8 ± 0.0
T5											
Rt	1	1.4	0.1	1.4	2.0	1.0	1.6	6.9	2.0	2.8	2.4
Lt	1	0.0	0.0	0.0	1.0	0.1	0.0	1.9	0.0	1.0	0.5
T6											
Rt	1	1.8	0.1	0.6	0.0	0.0	0.7	2.9	0.6	1.9	1.3
Lt	1	0.0	0.0	0.0	0.3	0.0	0.0	0.6	0.0	0.3	0.2
T7											
Rt	2	0.9 ± 0.4	2.0 ± 2.8	0.2 ± 0.2	0.7 ± 0.5	2.9 ± 4.1	0.8 ± 0.8	9.5 ± 9.5	2.4 ± 2.4	3.5 ± 3.4	3.05 ± 2.9
Lt	1	1.0	0.0	1.0	0.5	0.7	1.0	1.4	1.4	1.3	1.3
T8											
Rt	4	0.8 ± 0.5	1.8 ± 1.5	1.8 ± 0.8	1.2 ± 0.8	2.1 ± 0.3	1.8 ± 0.7	2.8 ± 1.4	3.0 ± 0.9	3.1 ± 0.6	0.1 ± 0.6
Lt	4	2.9 ± 1.7	1.9 ± 2.7	3.7 ± 2.6	2.0 ± 0.9	0.9 ± 0.7	3.7 ± 2.6	3.0 ± 1.6	5.8 ± 2.6	4.5 ± 2.4	5.1 ± 2.4
T9											
Rt	4	0.9 ± 0.9	1.3 ± 1.5	2.6 ± 1.6	0.6 ± 0.8	2.0 ± 0.7	2.6 ± 1.7	3.1 ± 2.5	3.5 ± 1.4	3.6 ± 1.1	3.6 ± 1.1
Lt	4	1.2 ± 1.1	1.4 ± 1.2	5.0 ± 3.6	0.6 ± 0.3	2.3 ± 2.9	5.0 ± 3.5	2.8 ± 2.6	5.5 ± 3.6	6.0 ± 3.7	5.7 ± 3.6
T10											
Rt	3	1.6 ± 1.3	1.3 ± 1.1	3.4 ± 2.4	1.1 ± 1.3	1.4 ± 1.9	3.4 ± 2.4	2.1 ± 1.5	4.6 ± 1.2	4.6 ± 1.2	4.6 ± 1.2
Lt	3	1.5 ± 2.1	0.8 ± 1.1	3.0 ± 2.2	0.5 ± 0.7	1.3 ± 0.5	3.0 ± 2.1	3.6 ± 3.0	4.1 ± 1.9	3.6 ± 1.4	3.9 ± 1.6
T11											
Rt	4	2.5 ± 2.3	1.6 ± 0.4	3.6 ± 2.2	2.7 ± 1.3	0.9 ± 0.7	3.6 ± 2.2	2.6 ± 2.3	5.2 ± 1.7	4.1 ± 1.4	5.1 ± 1.5
Lt	4	2.2 ± 1.3	1.4 ± 1.4	4.6 ± 2.0	1.0 ± 1.1	1.9 ± 0.9	4.4 ± 2.1	4.0 ± 3.0	5.5 ± 1.7	5.1 ± 2.0	5.3 ± 1.7
T12											
Rt	5	0.7 ± 0.8	1.6 ± 0.2	1.4 ± 1.2	2.1 ± 1.4	1.2 ± 1.0	1.5 ± 1.2	3.4 ± 1.2	2.5 ± 0.8	3.2 ± 1.3	2.9 ± 0.9
Lt	5	1.0 ± 0.9	0.9 ± 0.5	2.0 ± 1.0	1.4 ± 1.6	0.9 ± 0.6	2.0 ± 0.9	2.3 ± 1.8	2.6 ± 0.6	3.0 ± 1.2	2.8 ± 0.8
L1											
Rt	6	1.9 ± 1.7	1.4 ± 1.2	4.2 ± 5.4	1.5 ± 1.4	2.5 ± 2.8	4.4 ± 5.8	4.7 ± 4.4	5.8 ± 4.6	5.8 ± 6.0	5.8 ± 5.3
Lt	5	0.7 ± 0.7	0.9 ± 1.0	3.0 ± 2.7	1.1 ± 0.5	1.2 ± 0.8	3.0 ± 2.8	2.1 ± 2.1	3.5 ± 2.5	3.5 ± 2.7	3.5 ± 2.6
L2											
Rt	8	2.0 ± 2.1	1.1 ± 1.1	1.9 ± 2.1	1.6 ± 1.2	1.1 ± 0.9	1.9 ± 2.0	3.8 ± 3.5	3.5 ± 2.3	3.2 ± 1.8	3.3 ± 1.7
Lt	8	1.2 ± 1.1	1.3 ± 0.9	1.4 ± 2.0	1.1 ± 0.7	1.6 ± 1.5	1.4 ± 2.0	3.6 ± 1.6	2.8 ± 1.7	3.0 ± 1.6	2.9 ± 1.5
L3											
Rt	15	1.1 ± 1.1	1.1 ± 1.1	2.43 ± 2.3	1.5 ± 0.9	1.2 ± 1.2	2.4 ± 2.3	2.4 ± 1.1	3.3 ± 2.3	3.4 ± 2.3	3.4 ± 2.2
Lt	15	1.6 ± 1.7	1.3 ± 1.4	3.5 ± 2.5	2.7 ± 1.9	1.9 ± 1.5	3.4 ± 2.3	4.5 ± 2.9	4.6 ± 2.3	5.2 ± 2.4	4.9 ± 2.3
L4											
Rt	29	1.1 ± 1.0	1.3 ± 1.1	2.7 ± 2.2	2.4 ± 1.9	1.7 ± 1.3	2.5 ± 2.1	4.7 ± 2.8	3.6 ± 2.0	4.5 ± 2.2	4.1 ± 1.9
Lt	29	1.6 ± 1.4	1.1 ± 1.1	3.3 ± 2.3	2.0 ± 1.4	1.7 ± 2.0	3.1 ± 2.2	3.7 ± 3.1	4.1 ± 2.4	4.6 ± 2.5	4.3 ± 2.4

CONTINUED ON PAGE 526 »

» CONTINUED FROM PAGE 525

TABLE 3. Screw accuracy by level, among 254 screws

Level	No. of Screws	2D Accuracy: Tip			2D Accuracy: Tail			3D Accuracy			
		Rt/Lt	S/I	L/S	Rt/Lt	S/I	L/S	AD (°)	Tip	Tail	Overall
L5											
Rt	26	1.2 ± 1.3	1.2 ± 1.0	2.3 ± 3.0	1.8 ± 1.5	1.4 ± 1.1	2.3 ± 2.9	3.2 ± 2.1	3.4 ± 2.9	3.9 ± 2.7	3.6 ± 2.7
Lt	26	1.3 ± 1.2	1.0 ± 0.9	2.7 ± 2.0	1.8 ± 1.4	1.6 ± 1.4	2.5 ± 1.9	3.2 ± 2.7	3.6 ± 1.8	4.2 ± 1.7	3.9 ± 1.6
S1											
Rt	11	1.4 ± 1.2	2.3 ± 1.8	3.4 ± 2.7	1.7 ± 1.5	2.2 ± 2.5	3.2 ± 2.4	6.3 ± 3.9	4.8 ± 2.7	4.7 ± 3.0	4.8 ± 2.7
Lt	11	1.5 ± 1.2	1.6 ± 1.1	2.0 ± 1.6	1.3 ± 1.0	2.0 ± 1.8	2.0 ± 1.5	3.6 ± 2.6	3.4 ± 1.3	3.6 ± 1.9	3.5 ± 1.4
S2											
Rt	1	1.0	1.1	4.8	0.1	1.4	4.8	0.6	5.0	5.0	5.0
Lt	1	0.4	0.1	2.2	1.5	1.4	2.2	1.3	2.2	3.0	2.6

AD = angular deviation; L/S = long/short; S/I = superior/inferior.

Values expressed as the mean ± standard deviation, in millimeters unless indicated otherwise.

tive plan, with a mean angular offset of 2.78°. This translated to an overall accuracy of 90% perfect screws per the Gertzbein-Robbins system and 97% clinically acceptable screws. Unlike in the present study, however, accuracy was only considered within the mediolateral and cephalocaudal axes; accuracy along the long axis of the screw was not reported.

Around the same time, Lieberman and colleagues¹⁶ described their experience placing 197 thoracolumbar pedicle screws in 10 cadavers. They reported an overall accuracy of 1.1 ± 0.4 mm, which translated to a rate of 66.2% perfect placements and 92.4% clinically acceptable placements according to the Gertzbein-Robbins scale. Similar to prior reports, this study failed to report mean errors along the three component axes and therefore gave an incomplete assessment of 3D screw accuracy.

Our group¹³ published a case report that described the potential use of a 3D accuracy system for evaluating screw placement accuracy using 2 single-level lumbar fusion cases. We found a mean tip deviation of 2.1 mm and an angular offset of 2.4°, which corresponded to clinically acceptable screw placement (according to the Gertzbein-Robbins system) in all cases and perfect placement (per the Gertzbein-Robbins system) in 7 of 8 screws. However, unlike the present report, deviation along all three major screw axes was not reported—a detail we have refined in the present report.

The second study to describe the use of a 3D accuracy grading system was that of Godzik et al.,¹⁷ who also used

postoperative CT to assess the 3D accuracy of percutaneous lumbosacral pedicle screw placement. As we noted in the present study, they found that the mean screw error was on average less than the outer diameter of the placed screw, with a mean lumbar 3D accuracy of 5 ± 2.4 mm (angular offset 5.6° ± 4.3°), as compared to 4.0 ± 2.3 mm in the present study (angular offset 3.8° ± 2.8°). Unlike Godzik et al., however, we found that this overall level of accuracy is preserved even for thoracic screws, which are noted to be less tolerant of placement inaccuracy because of smaller pedicles and the presence of the spinal cord within the canal at this level. Additionally, Godzik and colleagues did not report the errors along each of the three primary screw axes, which would be a necessary component for error-based learning by the robotic system.

Merits of the Present Grading System and Applicability for Future Surgeons

In order to provide the best outcomes for patients, robotic spine surgery systems must be capable of learning from their mistakes. This requires an error-based learning system whereby the robotic system can compare its actual screw placement to that which was intended for the purposes of identifying the nature of any error. The screw accuracy system outlined in this paper fulfills the requirements for such a system, as it defines linear deviation along all three screw axes, as well as the net 3D linear and angular deviations. Moreover, as the present accuracy

TABLE 4. Screw accuracy by spine region

Region	No. of Screws	2D Accuracy: Tip			2D Accuracy: Tail			3D Accuracy			
		Rt/Lt	S/I	L/S	Rt/Lt	S/I	L/S	AD (°)	Tip	Tail	Overall
T	63	1.2 ± 1.2	1.2 ± 1.2	2.1 ± 2.1	1.2 ± 1.1	1.3 ± 1.3	2.1 ± 2.1	2.9 ± 2.6	3.1 ± 2.2	3.2 ± 2.1	3.2 ± 2.1
LS	191	1.3 ± 1.3	1.2 ± 1.1	2.8 ± 2.5	1.9 ± 1.5	1.7 ± 1.6	2.7 ± 2.4	3.8 ± 2.8	3.8 ± 2.4	4.2 ± 2.5	4.0 ± 2.3
L	167	1.3 ± 1.3	1.2 ± 1.1	2.7 ± 2.5	1.9 ± 1.5	1.6 ± 1.5	2.7 ± 2.5	3.7 ± 2.7	3.7 ± 2.4	4.2 ± 2.5	4.0 ± 2.4
S	24	1.4 ± 1.1	1.8 ± 1.5	2.8 ± 2.2	1.4 ± 1.2	2.1 ± 2.0	2.7 ± 2.4	4.6 ± 3.5	4.1 ± 2.1	4.1 ± 2.4	4.1 ± 2.1
All	254	1.3 ± 1.3	1.2 ± 1.1	2.6 ± 2.2	1.7 ± 1.4	1.6 ± 1.5	2.5 ± 2.3	3.6 ± 2.8	3.6 ± 2.3	4.0 ± 2.4	3.8 ± 2.3

Values expressed as the mean ± standard deviation, in millimeters unless indicated otherwise.

TABLE 5. Statistical comparisons of screw accuracy by spine region

Region	2D Accuracy: Tip			2D Accuracy: Tail			3D Accuracy			
	Rt/Lt	S/I	L/S	Rt/Lt	S/I	L/S	AD	Tip	Tail	Overall
T vs LS (p)	0.59	0.04†	0.77	0.11	0.31	0.65	0.33	0.37	0.73	0.68
ANOVA (F)	0.29	3.70†	1.59	8.05†	2.56	1.16	3.55†	2.27	3.88†	3.16†
T vs L (p)	0.76	0.98	0.20	<0.01†	0.32	0.30	0.16	0.15	0.02†	0.04†
L vs S (p)	0.99	0.03†	0.99	0.19	0.31	0.99	0.26	0.78	0.99	0.96
T vs S (p)	0.83	0.03†	0.47	0.69	0.07	0.59	0.03†	0.18	0.26	0.20

† p < 0.05.

grading system was developed using electronic files from the CT scanner and the robotic system, it has the potential to be digitized, imbuing the robotic surgery system with the intrinsic ability to learn from its mistakes without an external teacher.

In addition to providing technologists with the data necessary to improve the robot's anatomical mapping, such a grading system—when applied to screws placed using a freehand technique—can also provide surgeons with quantitative assessment of their own instrumentation errors. By presenting surgeons with quantifiable measures of their instrumentation-misplacement tendencies, the present grading system can empower surgeons to improve their own freehand techniques for superior patient outcomes.

Study Limitations

As with all studies, there are several limitations to the present investigation that potentially limit the generalizability of the results. First, although patients were recruited into the study prospectively, the analysis was performed retrospectively, creating the possibility that statistical bias affected the study results. We attempted to avoid such bias by having raw data collection and data analysis performed by independent members of the study team without financial conflicts. Second, the present series is relatively small (47 patients and 254 screws placed). This small series means that the accuracy results reported herein may not reflect the accuracy realized by other users of the surgical system. Third, the majority of cases were performed by the senior author—the co-inventor of the surgical spine system—who is likely more facile with the device described herein and therefore perhaps able to achieve better results than the average user. We do not believe this to be the case, given that the accuracy studied is between the plotted trajectory and the actual screw trajectory. Nevertheless, we allow for the possibility of this confounding.

Another limitation to the screw accuracy grading system presented here is that the quality of the “error” signal it provides is limited by the intrinsic accuracy of the imaging systems used to generate the pre- and postinstrumentation volumes, as well as the navigation system used intraoperatively, which inevitably injects some error into the navigation. However, the limits of present technology suggest that these are likely minor contributors, as the CT scanners at our institution allow for volumes to be obtained with slices as thin as 0.75 mm. Additionally, the robotic system described allows for automatic image registration, which has been shown by other groups to reduce registra-

tion errors to 0.74 mm.¹⁸ Even combined to produce maximal error, the imaging and navigation inaccuracies total 1.5 mm, which is small enough to detect differences in screw accuracy within a single Gertzbein-Robbins grade. It is likely that these errors will continue to be reduced with greater refinement of CT imaging techniques and volume reconstruction algorithms, as has already been demonstrated in the radiation oncology literature.^{19–22} Advances in technology that enable automatic screw trajectory planning are also likely to reduce the inherent error of robot-assisted instrumentation. At present, our screw accuracy system utilizes surgeon-plotted trajectories as the gold standard for accuracy. These are subject to human error, though, and in the future are likely to be replaced by computer-plotted trajectories that are confirmed by the operator. The feasibility of this approach has already been demonstrated by multiple groups,^{23–27} including Vijayan et al.,²⁶ who demonstrated its use with the current robotic system. For these automatic registration systems to move into the clinical realm, it will be necessary to have a formalized, detailed accuracy grading system, such as that presented here, which has the granularity necessary for error-based learning by the computer algorithms. Lastly, although the overwhelming majority of included patients were treated for a degenerative pathology, a small fraction were treated for tumor or trauma, which can be associated with altered morphology of the pedicle. As the pedicles are used as reference points for determining the 3D geometry of the plotted and placed screws, altered pedicle morphology may prevent accurate assessment of true screw position.

Conclusions

Herein we describe a novel means of evaluating pedicle screw accuracy that assesses screw dimension in all three linear planes along with net angular deviation. Using this system, we found that robotic spine surgery systems can facilitate highly accurate pedicle screw placement, with mean deviations of only 1 mm or < 25% of the diameter of the average pedicle screw. We believe the present grading system allows for more accurate assessment of screw placement accuracy and can be used across robotic spine surgical systems to understand the true degree to which these systems can facilitate accurate screw placement.

References

1. Ghasem A, Sharma A, Greif DN, et al. The arrival of robot-

- ics in spine surgery: a review of the literature. *Spine (Phila Pa 1976)*. 2018;43(23):1670–1677.
2. Gao S, Lv Z, Fang H. Robot-assisted and conventional freehand pedicle screw placement: a systematic review and meta-analysis of randomized controlled trials. *Eur Spine J*. 2018;27(4):921–930.
  3. Fan Y, Du JP, Liu JJ, et al. Accuracy of pedicle screw placement comparing robot-assisted technology and the free-hand with fluoroscopy-guided method in spine surgery: an updated meta-analysis. *Medicine (Baltimore)*. 2018;97(22):e10970.
  4. Li H-M, Zhang R-J, Shen C-L. Accuracy of pedicle screw placement and clinical outcomes of robot-assisted technique versus conventional freehand technique in spine surgery from nine randomized controlled trials: a meta-analysis. *Spine (Phila Pa 1976)*. 2020;45(2):E111–E119.
  5. Gertzbein SD, Robbins SE. Accuracy of pedicular screw placement in vivo. *Spine (Phila Pa 1976)*. 1990;15(1):11–14.
  6. Heary RF, Bono CM, Black M. Thoracic pedicle screws: postoperative computerized tomography scanning assessment. *J Neurosurg*. 2004;100(4)(Suppl Spine):325–331.
  7. Wiesner L, Kothe R, Rütther W. Anatomic evaluation of two different techniques for the percutaneous insertion of pedicle screws in the lumbar spine. *Spine (Phila Pa 1976)*. 1999;24(15):1599–1603.
  8. Rampersaud YR, Pik JH, Salonen D, Farooq S. Clinical accuracy of fluoroscopic computer-assisted pedicle screw fixation: a CT analysis. *Spine (Phila Pa 1976)*. 2005;30(7):E183–E190.
  9. Fan Y, Peng Du J, Liu JJ, et al. Radiological and clinical differences among three assisted technologies in pedicle screw fixation of adult degenerative scoliosis. *Sci Rep*. 2018;8(1):890.
  10. Han X, Tian W, Liu Y, et al. Safety and accuracy of robot-assisted versus fluoroscopy-assisted pedicle screw insertion in thoracolumbar spinal surgery: a prospective randomized controlled trial. *J Neurosurg Spine*. 2019;30(5):615–622.
  11. Solomiichuk V, Fleischhammer J, Molliqaj G, et al. Robotic versus fluoroscopy-guided pedicle screw insertion for metastatic spinal disease: a matched-cohort comparison. *Neurosurg Focus*. 2017;42(5):E13.
  12. Zhang Q, Han X-G, Xu Y-F, et al. Robot-assisted versus fluoroscopy-guided pedicle screw placement in transforaminal lumbar interbody fusion for lumbar degenerative disease. *World Neurosurg*. 2019;125:e429–e434.
  13. Jiang B, Ahmed AK, Zygorakis CC, et al. Pedicle screw accuracy assessment in ExcelsiusGPS robotic spine surgery: evaluation of deviation from pre-planned trajectory. *Chinese Neurosurg J*. 2018;4:23.
  14. Togawa D, Kayanja MM, Reinhardt MK, et al. Bone-mounted miniature robotic guidance for pedicle screw and transaminar facet screw placement: part 2—evaluation of system accuracy. *Neurosurgery*. 2007;60(2)(suppl 1):ONS129–ONS139.
  15. Stürer C, Ringel F, Stoffel M, et al. Robotic technology in spine surgery: current applications and future developments. *Acta Neurochir Suppl*. 2011;109:241–245.
  16. Lieberman IH, Hardenbrook MA, Wang JC, Guyer RD. Assessment of pedicle screw placement accuracy, procedure time, and radiation exposure using a miniature robotic guidance system. *J Spinal Disord Tech*. 2012;25(5):241–248.
  17. Godzik J, Walker CT, Hartman C, et al. A quantitative assessment of the accuracy and reliability of robotically guided percutaneous pedicle screw placement: technique and application accuracy. *Oper Neurosurg (Hagerstown)*. 2019;17(4):389–395.
  18. Zhao J, Liu Y, Fan M, et al. Comparison of the clinical accuracy between point-to-point registration and auto-registration using an active infrared navigation system. *Spine (Phila Pa 1976)*. 2018;43(22):E1329–E1333.
  19. Elstrøm UV, Muren LP, Petersen JBB, Grau C. Evaluation of image quality for different kV cone-beam CT acquisition and reconstruction methods in the head and neck region. *Acta Oncol*. 2011;50(6):908–917.
  20. Elstrøm UV, Olsen SRK, Muren LP, et al. The impact of CBCT reconstruction and calibration for radiotherapy planning in the head and neck region—a phantom study. *Acta Oncol*. 2014;53(8):1114–1124.
  21. Garayoa J, Castro P. A study on image quality provided by a kilovoltage cone-beam computed tomography. *J Appl Clin Med Phys*. 2013;14(1):3888.
  22. Marchant TE, Joshi KD, Moore CJ. Accuracy of radiotherapy dose calculations based on cone-beam CT: comparison of deformable registration and image correction based methods. *Phys Med Biol*. 2018;63(6):065003.
  23. Goerres J, Uneri A, De Silva T, et al. Spinal pedicle screw planning using deformable atlas registration. *Phys Med Biol*. 2017;62(7):2871–2891.
  24. Knez D, Likar B, Pernus F, Vrtovc T. Computer-assisted screw size and insertion trajectory planning for pedicle screw placement surgery. *IEEE Trans Med Imaging*. 2016;35(6):1420–1430.
  25. Knez D, Nahle IS, Vrtovc T, et al. Computer-assisted pedicle screw placement planning: towards clinical practice. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. April 4–7, 2018:249–252. Accessed April 14, 2020. <https://ieeexplore.ieee.org/document/8363566/>
  26. Vijayan R, De Silva T, Han R, et al. Automatic pedicle screw planning using atlas-based registration of anatomy and reference trajectories. *Phys Med Biol*. 2019;64(16):165020.
  27. Xiaozhao C, Jinfeng H, Baolin M, et al. A method of lumbar pedicle screw placement optimization applied to guidance techniques. *Comput Assist Surg*. 2016;21(suppl):142–147. Accessed April 14, 2020. <https://www.tandfonline.com/doi/full/10.1080/24699322.2016.1240301>

Disclosures

Dr. Jiang is a consultant for Longevity Neuro Solutions and receives grant funding from DePuy Synthes. Dr. Crawford receives royalties from Globus Medical. Dr. Theodore receives royalties from and has ownership interest in Globus Medical and is a consultant for Globus Medical and DePuy Synthes.

Author Contributions

Conception and design: Theodore, Jiang, Crawford. Acquisition of data: Pennington, Zhu, Matsoukas, Ahmed, Mahapatra. Analysis and interpretation of data: Theodore, Jiang, Pennington. Drafting the article: Jiang, Pennington, Ehresman. Critically revising the article: Theodore, Jiang, Pennington, Ahmed, Ehresman, Cottrill. Reviewed submitted version of manuscript: all authors. Statistical analysis: Pennington, Matsoukas, Mahapatra, Sheppell. Study supervision: Theodore.

Supplemental Information

Online-Only Content

Supplemental material is available with the online version of the article.

Supplemental Figure 1. <https://thejns.org/doi/suppl/10.3171/2020.3.SPINE20208>.

Correspondence

Nicholas Theodore: Johns Hopkins School of Medicine, Baltimore, MD. theodore@jhmi.edu.