This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Confidence-Aware Visualization for Pedicle Screw Guidance in Virtual Reality

Hamideh Hosseini Toudeshky[1*], Anil Ufuk Batmaz[1], and Marta Kersten-Oertel[2]

[1] Concordia University, Montréal, Canada
h_hamide@live.concordia.ca
[2] University of Ottawa, Ottawa, Canada

Abstract. Pedicle screw placement requires accurate reasoning about complex three-dimensional anatomy. While virtual reality has become a promising platform for surgical planning and guidance, existing systems typically communicate only a single recommended trajectory, providing limited information about the confidence associated with that recommendation. As artificial intelligence (AI) becomes increasingly integrated into surgical navigation, communicating confidence may improve human–AI collaboration. We compare conventional deterministic guidance with two uncertainty-inspired guidance corridors that communicate confidence using blur and colour. As a proof of concept, we conducted a user study to evaluate performance, user perception, and visualization preference using parameterized rather than model-derived confidence. No significant differences were observed in entry-point error, tip-point error, task completion time, or screw adjustments between guidance representations. Angular deviation differed significantly, with deterministic guidance producing the smallest deviation. Subjective trajectory-understanding ratings showed the largest numerical difference across conditions, although this difference was not statistically significant. Nevertheless, 11 of 14 participants preferred a confidence-aware representation. These findings support further investigation of confidence-aware visualization for future AI-assisted surgical navigation systems.

Keywords: Virtual reality · Image-guided surgery · Pedicle screw placement · AI-assisted surgical guidance · Uncertainty visualization

1 Introduction

Pedicle screw fixation is the standard technique for stabilizing the spine in the treatment of trauma, deformity, degenerative disease, and spinal instability. Accurate screw placement is essential, as cortical breach may compromise fixation or result in neurological or vascular injury [1, 2]. Achieving accurate placement requires surgeons to reason about complex three-dimensional vertebral anatomy while selecting a safe and biomechanically appropriate screw trajectory [3].

* Corresponding author

Computer-assisted planning and navigation have substantially improved pedicle screw placement through patient-specific planning and trajectory guidance [4]. Recently, virtual and mixed reality (VR/MR) systems have emerged as promising platforms for surgical planning, navigation, and education by enabling immersive visualization and intuitive interaction with patient-specific anatomy [5–8]. Despite these advances, existing systems almost universally communicate guidance using a single recommended trajectory [9]. As surgical planning becomes increasingly data-driven, communicating confidence alongside a recommended trajectory may become essential for AI-assisted guidance [10].

Artificial intelligence (AI) is rapidly being integrated into surgical planning, navigation, and clinical decision support. AI-based systems are increasingly being used to recommend surgical trajectories, implant positioning, and operative plans. However, these recommendations inevitably contain uncertainty arising from medical imaging, segmentation, registration, tracking, and planning algorithms [11]. Improving predictive accuracy alone is therefore insufficient; clinicians must also develop an appropriate level of trust in AI-generated recommendations. Overtrust may encourage automation bias and inappropriate reliance, whereas undertrust may cause clinicians to disregard useful guidance [12]. Consequently, AI-assisted navigation systems need to communicate not only a recommended trajectory but also the confidence associated with that recommendation.

Visualization research has shown that uncertainty can be effectively communicated using visual encodings such as colour, transparency, blur, and probabilistic representations, improving users' understanding of system reliability without substantially increasing cognitive workload or degrading task performance [13–15]. More recently, uncertainty visualization has been shown to improve trust calibration in AI-assisted decision making by helping users appropriately interpret model confidence and limitations [16, 12]. However, most prior work has focused on medical image analysis, visualization, or diagnostic decision support, while comparatively little is known about how uncertainty should be communicated during immersive surgical guidance. Before uncertainty estimates can be effectively incorporated into AI-assisted surgical guidance, it is important to understand how different visualization strategies influence user perception, trust, and decision making during interactive guidance.

In this work, we investigate confidence-aware visualization for pedicle screw guidance in VR. Specifically, we compare conventional deterministic guidance with two uncertainty-inspired representations that communicate confidence using blur and colour. Through a controlled within-subjects user study, we examine their effects on objective performance, trust, confidence, and perceived understanding of the range of acceptable trajectories. The study provides an initial empirical characterization of how alternative confidence visualization strategies are perceived and used during an immersive surgical guidance task.

2 Methods

A VR environment was developed in Unity and deployed on a Meta Quest 3 head-mounted display. The system enables users to inspect patient-specific vertebral anatomy and position a virtual pedicle screw using natural mid-air hand interaction. Figure 1 provides an overview of the system and the three evaluated guidance representations.

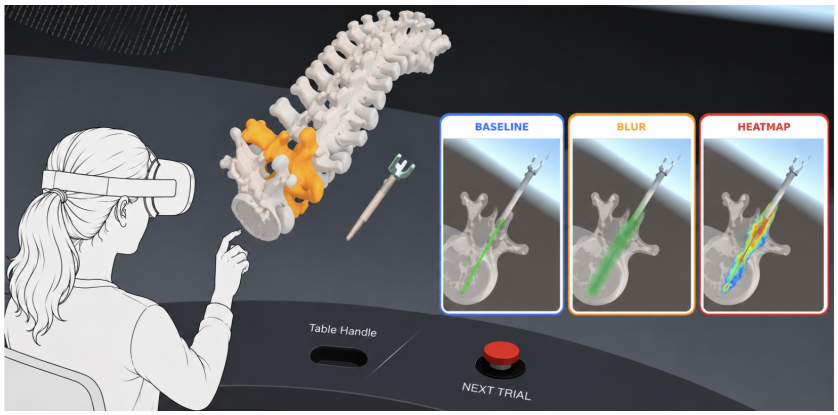


Fig. 1. Overview of the VR user study and the three guidance representations: deterministic (Baseline), Blur Corridor, and Heatmap Corridor.

2.1 VR Environment

Segmented vertebrae (L1, L2, and T10) from a single CT scan of a real patient from the Cancer Imaging Archive [17] were imported into Unity as anatomical-scale 3D meshes. A virtual pedicle screw was represented as a rigid-body object that could be grasped and manipulated using the Quest 3's built-in hand tracking. Two reference points located at the screw entry and tip recorded the final screw position and orientation for subsequent analysis.

2.2 Guidance Representations

Pedicle screw guidance was based on a reference trajectory connecting a predefined entry point and target point within the vertebral body. The trajectory was generated offline using anatomical heuristics based on pedicle geometry and accepted surgical planning guidelines [4, 18]. The entry point was selected on the posterior elements of the vertebra, while the target point was positioned within the vertebral body to maximize screw purchase while avoiding cortical breach.

Three guidance representations were evaluated (Fig. 1). The *deterministic* condition represented conventional surgical guidance using a single line indicating the recommended screw trajectory. The remaining two conditions represented the recommendation as a guidance corridor inspired by uncertainty visualization techniques, with confidence encoded around the reference trajectory. To evaluate the visualization independently of a specific uncertainty-estimation method, the displayed confidence was intentionally parameterized rather than estimated from an underlying model and should be interpreted as a representative visualization rather than a measure of true planning uncertainty. The corridor radius was selected empirically to remain clearly visible while minimizing occlusion of the vertebral anatomy and was held constant across both corridor conditions.

In the *Blur Corridor*, confidence was communicated using image blur, with sharper regions indicating higher confidence and progressively blurred regions indicating lower confidence. The corridor was wider at the entry point and the target point and narrower at the pedicle, allowing participants to explore the range of acceptable trajectories by varying the angular deviation while remaining within the constrained pedicle region. In the *Heatmap Corridor*, confidence was communicated using colour, with warmer colours indicating higher confidence and cooler colours indicating lower confidence. Blur- and colour-based encodings were selected because they are widely used techniques for visualizing uncertainty while preserving spatial context [11, 13, 15].

2.3 Study Design

A within-subjects study compared three guidance representations: deterministic, Blur Corridor, and Heatmap Corridor. Participants completed 18 pedicle screw placement trials (three vertebrae $\times$ two pedicles $\times$ three guidance representations). Guidance representation order was counterbalanced, while vertebral level and pedicle side followed a fixed sequence within each condition.

2.4 Procedure

Participants first signed a consent form and completed a pre-questionnaire. The researcher then explained the study objectives before providing a tutorial introducing pedicle screw placement, the VR environment, and the guidance representations, followed by a practice trial. Participants subsequently completed the experimental trials for each guidance representation. After each condition, they completed a questionnaire assessing confidence in their placement, trust in the guidance, visualization clarity, understanding of the recommended trajectory, and understanding of safe screw placement. Following all three conditions, a post-test questionnaire was completed.

2.5 Measures

Objective measures included entry-point error, tip-point error, angular deviation, task completion time, and the number of screw readjustments. Subjective

Table 1. Participant demographics and prior experience. Experience ratings were collected on a 7-point Likert scale (1 = none/not at all familiar, 7 = extensive/extremely familiar).

Demographics	**Value**	**Experience/Familiarity**	**Value**
Participants	14	VR/MR experience	4.64 ± 2.06
Age (years)	27.1 ± 4.1	Video game experience	5.36 ± 1.74
Female / Male	6 / 8	Spinal anatomy familiarity	2.43 ± 1.60
Medical background	1 / 13	Pedicle screw familiarity	1.86 ± 1.03
		AI familiarity	5.71 ± 0.83
		Image-guided surgery familiarity	2.57 ± 1.60

measures included confidence, trust, visualization clarity, subjective trajectory understanding, and understanding of safe screw placement corridors, each rated on a 7-point Likert scale. The post-study questionnaire additionally assessed overall preference, perceived usefulness for learning, communication of confidence, and qualitative feedback.

3 Results

The guidance representations were evaluated through objective performance measures, post-condition questionnaires, and post-study preference rankings.

3.1 Participants

A priori power analysis (G*Power) for a repeated-measures ANOVA ($\alpha = 0.05$, $f = 0.40$, three conditions) indicated a minimum sample size of 12 participants. Fourteen participants (age: 27.1 ± 4.1 years; 8 male, 6 female) were recruited. One participant had a medical background, while the remaining participants were recruited from computer science and related disciplines. All participants provided informed consent under institutional ethics. Participant demographics and prior experience are summarized in Table 1.

3.2 Objective Performance

Performance measures were averaged across the six trials for each guidance representation. Table 2 summarizes the performance measures. Performance did not differ significantly across guidance representations for task completion time ($\chi^2(2) = 4.43$, $p = 0.109$), entry-point error ($\chi^2(2) = 1.29$, $p = 0.526$), tip-point error ($F(2, 26) = 1.70$, $p = 0.203$), and the number of screw adjustments ($\chi^2(2) = 3.37$, $p = 0.185$). Guidance representation significantly affected angular deviation ($\chi^2(2) = 13.0$, $p = 0.002$, $W = 0.46$). Post hoc Durbin–Conover comparisons with Holm correction showed that the Heatmap Corridor differed significantly from both the Blur Corridor ($p = 0.011$) and the deterministic representation ($p < 0.001$). However, no significant difference was found between

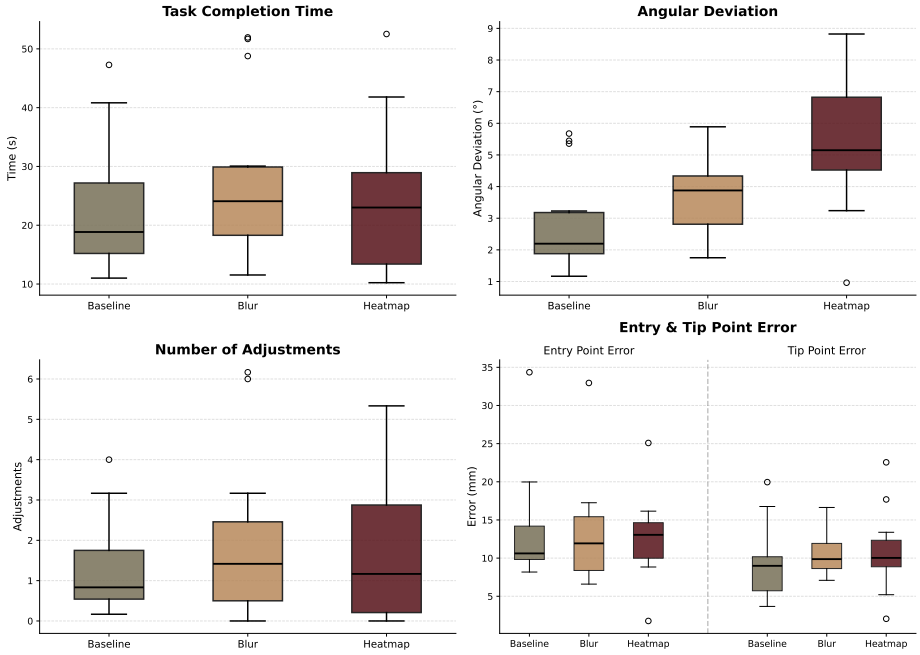


Fig. 2. Distribution of objective performance measures across the three guidance representations.

the Blur Corridor and the deterministic representation ($p = 0.057$), with the deterministic representation yielding the lowest median angular deviation (2.20°), followed by the Blur Corridor (3.88°) and the Heatmap Corridor (5.15°).

3.3 Perception of the Guidance

Following each guidance representation, participants rated confidence in their final screw placement, trust in the guidance, visualization clarity, subjective trajectory understanding, and understanding of safe screw placement (Table 3).

Several subjective questionnaire measures violated the normality assumption; therefore, Friedman tests were used to compare ratings across the three guidance representations. No significant effect of guidance representation was found for confidence ($\chi^2(2) = 0.61$, $p = 0.739$, $W = 0.02$), trust ($\chi^2(2) = 1.00$, $p = 0.607$, $W = 0.04$), visualization clarity ($\chi^2(2) = 2.00$, $p = 0.368$, $W = 0.07$), subjective trajectory understanding ($\chi^2(2) = 5.38$, $p = 0.068$, $W = 0.19$), or perceived safe screw placement ($\chi^2(2) = 0.71$, $p = 0.701$, $W = 0.03$). Subjective trajectory understanding showed the largest numerical difference across conditions ($W = 0.19$), although the omnibus test was not statistically significant ($p = 0.068$).

Table 2. Descriptive statistics and repeated-measures test results by guidance representation ($N = 14$). Medians are reported except for Tip Point Error (mean; repeated-measures ANOVA). Shared superscripts indicate no significant pairwise difference (Holm-corrected, $\alpha = 0.05$).

Metric	Deterministic	Blur	Heatmap	Test Statistic	p
Task Completion Time (s)	23.1^a	27.7^a	24.7^a	$\chi^2(2) = 4.43$	0.109
Entry Point Error (mm)	13.4^a	13.0^a	12.6^a	$\chi^2(2) = 1.29$	0.526
Tip Point Error (mm)	9.0^a	10.5^a	10.9^a	$F(2, 26) = 1.70$	0.203
Angular Deviation (°)	2.20^a	3.88^a	5.15^b	$\chi^2(2) = 13.0$	**0.002**
Number of Adjustments	0.83^a	1.42^a	1.17^a	$\chi^2(2) = 3.37$	0.185

Table 3. Participant ratings following each guidance representation (mean ± SD, 7-point Likert scale).

Measure	Deterministic	Blur	Heatmap	p
Confidence	5.50 ± 1.16	5.71 ± 1.38	5.79 ± 1.31	0.739
Trust	5.86 ± 0.95	5.78 ± 1.25	6.21 ± 1.05	0.607
Clarity	6.29 ± 0.91	5.64 ± 1.23	6.21 ± 0.93	0.368
Trajectory understanding	4.64 ± 1.86	5.93 ± 0.96	6.21 ± 0.86	0.068
Safe placement	5.64 ± 1.39	6.07 ± 0.96	5.92 ± 1.22	0.701

3.4 Visualization Preference

Participants completed a post-study questionnaire comparing the three guidance representations. Consistent with the post-condition ratings, 11/14 participants preferred one of the confidence-aware guidance representations over the conventional deterministic trajectory representation. The blur corridor was ranked as the overall preferred representation by 6 participants, followed by the heatmap corridor (5) and the deterministic representation (3).

The confidence-aware representations were preferred for different reasons. The Heatmap Corridor was most often selected as best communicating confidence in the guidance (7/14), inspiring confidence in screw placement (5/14), and supporting learning of pedicle screw placement (5/14). In contrast, the Blur Corridor was most often preferred for pedicle screw guidance (6/14) and for understanding where screws could be safely placed (6/14).

Qualitative feedback highlighted complementary strengths of the two confidence-aware representations. Participants described the Blur Corridor as intuitive because it clearly conveyed the region surrounding the recommended trajectory, whereas the Heatmap Corridor effectively communicated gradual changes in confidence along the trajectory. In contrast, the deterministic representation was appreciated for providing a precise reference path but was frequently criticized for conveying little information beyond the recommended trajectory. Several participants suggested combining the Blur and Heatmap Corridors, indicating that

future confidence-aware guidance may benefit from simultaneously communicating a clear reference trajectory and confidence information.

4 Discussion

This work serves as a proof-of-concept investigation of how confidence-aware visualization influences pedicle screw placement in virtual reality. Confidence was intentionally parameterized rather than derived from an AI planning algorithm, allowing the visualization strategies to be evaluated independently of a specific uncertainty-estimation method. The findings therefore do not demonstrate how confidence-aware guidance would behave with real AI-derived uncertainty (e.g., from segmentation, registration, tracking, or planning algorithms), but provide initial insight into visualization designs that future AI-assisted navigation systems might build upon.

Subjective ratings did not differ significantly across guidance representations, including for subjective trajectory understanding. While not statistically significant, trajectory understanding showed the largest numerical difference between conditions. Post-study preferences further indicated that most participants (11/14) favored one of the confidence-aware representations. Together, these findings motivate further investigation of whether corridor-based confidence visualization can help users interpret the flexibility of guidance.

No significant differences were observed for entry-point error, tip-point error, task completion time, or number of screw adjustments across guidance representations. The deterministic representation produced significantly smaller angular deviations than the confidence-aware conditions. As participants were instructed to align with the recommended trajectory rather than to explore the guidance corridor, this larger deviation is best interpreted as reduced adherence to the recommended trajectory rather than deliberate, acceptable exploration. Because direct safety outcomes (e.g., cortical breach, minimum cortical distance) were not assessed, we cannot determine the clinical consequence of this reduced adherence.

This study has several limitations. The design does not isolate confidence encoding from the effect of displaying a spatial corridor, as both experimental conditions differed from the deterministic baseline in geometry as well as confidence encoding; future work should include a uniform corridor control to separate these effects. Confidence was parameterized rather than estimated from an AI planning algorithm, allowing the visualization strategies to be evaluated independently of a specific uncertainty-estimation method. Future work should derive confidence from sources such as segmentation, registration, tracking, or AI-based planning uncertainty. Direct safety outcomes (e.g., cortical breach, minimum cortical distance) were not reported, limiting interpretation of the observed angular deviation; future work should incorporate such measures. In addition, the study was conducted in a virtual environment with primarily non-expert participants using vertebral models from a single patient, and did not include a detailed usability analysis (e.g., workload, system latency). Evaluating confidence-aware

guidance with surgeons and residents, larger cohorts, more diverse anatomy, and detailed usability measures will be important for assessing clinical applicability.

5 Conclusion

As AI-assisted surgical planning and navigation become increasingly common, future guidance systems will need to communicate not only recommended trajectories but also the confidence associated with those recommendations. This proof-of-concept study identifies confidence-aware visualization as a promising direction for immersive surgical navigation, with most participants favoring confidence-aware representations. Establishing whether confidence-aware visualization improves trajectory understanding, and whether these findings generalize to calibrated AI-derived uncertainty, expert users, and safety-relevant outcomes, will require further study with a uniform-corridor control and clinically representative participants and anatomy.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Mason, A., Paulsen, R., Babuska, J.M., Rajpal, S., Burneikiene, S., Nelson, E.L., Villavicencio, A.T.: The accuracy of pedicle screw placement using intraoperative image guidance systems. *Journal of Neurosurgery: Spine* **20**(2), 196–203 (2014)
2. Papalia, G.F., Vadalà, G., Russo, F., Marcello, G., Nardi, N., Papalia, R., Denaro, V.: Higher accuracy and better clinical outcomes in navigated thoraco-lumbar pedicle screw fixation versus conventional techniques: A systematic review and meta-analysis. *Spine* **49**(19), 1370–1380 (2024)
3. Liu, H., Zhou, Z.-Y., Wei, J.-X., Zhang, M., Bai, M., Huang, A.-B.: Comprehensive analysis of pedicle screw implantation in the C7 vertebra using computed tomography-based three-dimensional models. *BMC Surgery* **22**, 91 (2022)
4. Knez, D., Likar, B., Pernuš, F., Vrtovec, T.: Computer-assisted screw size and insertion trajectory planning for pedicle screw placement surgery. *IEEE Transactions on Medical Imaging* **35**(6), 1420–1430 (2016)
5. McKechnie, T., Levin, M., Zhou, K., Freedman, B., Palter, V.N., Grantcharov, T.P.: Virtual surgical training during COVID-19: Operating room simulation platforms accessible from home. *Annals of Surgery* **272**(2), e153–e154 (2020)
6. Bui, T., Ruiz-Cardozo, M.A., Dave, H.S., Barot, K., Kann, M.R., Joseph, K., Lopez-Alviar, S., Trevino, G., Brehm, S., Yahanda, A.T., Molina, C.A.: Virtual, augmented, and mixed reality applications for surgical rehearsal, operative execution, and patient education in spine surgery: A scoping review. *Medicina* **60**(2), 332 (2024)
7. Queisner, M., Eisenträger, K.: Surgical planning in virtual reality: A systematic review. *Journal of Medical Imaging* **11**(6), 062603 (2024)
8. Móga, K., Hölgyesi, Á., Zrubka, Z., Péntek, M., Haidegger, T.: Augmented or mixed reality enhanced head-mounted display navigation for in vivo spine surgery: A systematic review of clinical outcomes. *Journal of Clinical Medicine* **12**(11), 3788 (2023)

9. Scherer, M., Kausch, L., Bajwa, A., Neumann, J.-O., Ishak, B., Naser, P., Vollmuth, P., Kiening, K., Maier-Hein, K., Unterberg, A.: Automatic planning tools for lumbar pedicle screws: Comparison and validation of planning accuracy for self-derived deep-learning-based and commercial atlas-based approaches. *Journal of Clinical Medicine* **12**(7), 2646 (2023)
10. Sangalli, S., Sarwin, G., Erdil, E., Carretta, A., Staartjes, V., Serra, C., Konukoglu, E.: Conformal forecasting for surgical instrument trajectory. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2025, LNCS **15968**, pp. 117–127 (2025)
11. Gillmann, C., Saur, D., Wischgoll, T., Scheuermann, G.: Uncertainty-aware visualization in medical imaging: A survey. *Computer Graphics Forum* **40**(3), 665–689 (2021)
12. Tomsett, R., Preece, A., Braines, D., Cerutti, F., Chakraborty, S., Srivastava, M., Pearson, G., Kaplan, L.: Rapid trust calibration through interpretable and uncertainty-aware AI. *Patterns* **1**(4), 100049 (2020)
13. Simpson, A.L., Ma, B., Vasarhelyi, E.M., Borschneck, D.P., Ellis, R.E., Stewart, A.J.: Computation and visualization of uncertainty in surgical navigation. *The International Journal of Medical Robotics and Computer Assisted Surgery* **10**(3), 332–343 (2014)
14. Kay, M., Kola, T., Hullman, J., Munson, S.: When (Ish) Is My Bus? User-Centered Visualizations of Uncertainty in Everyday, Mobile Predictive Systems. In: *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 5092–5103 (2016)
15. Padilla, L., Kay, M., Hullman, J.: Uncertainty visualization. In: Piegorsch, W., Levine, R., Zhang, H., Lee, T. (eds.) Computational Statistics in Data Science, pp. 405–421. Wiley (2022)
16. Reyes, J., Batmaz, A.U., Kersten-Oertel, M.: Trusting AI: Does uncertainty visualization affect decision-making? *Frontiers in Computer Science* **7**, 1464348 (2025)
17. The Cancer Imaging Archive (TCIA): Spine-Mets-CT-SEG: Spine metastatic bone cancer: pre and post radiotherapy CT. Available at: https://www.cancerimagingarchive.net/collection/spine-mets-ct-seg/ (Accessed July 2026)
18. Vaccaro, A., Kandziora, F., Fehlings, M., Shanmughanathan, R.: Percutaneous posterior approach for pedicle screw placement. AO Surgery Reference, AO Foundation. https://surgeryreference.aofoundation.org/spine/trauma/thoracolumbar/approach/percutaneous-posterior-approach-for-pedicle-screw-placement (accessed July 2026)