Credibility Assessment of Implicit Finite Element Models for Pedicle Screw Pullout and Bone Purchase

Evan C. Miller, PhD[a], Priya N. Shah, MS[b], Lukas J. Brandt, PhD[c], Marina O. Delgado, MD[d]

[a]*Computational Mechanics Group, OrthoMechanica Research Lab, Minneapolis, MN, USA,*
[b]*Spinal Implants R&D, Vantage Spine Systems, San Jose, CA, USA,*
[c]*Department of Biomedical Engineering, Midwestern Tech University, Evanston, IL, USA,*
[d]*Department of Orthopaedic Surgery, Riverview Medical Center, Boston, MA, USA,*

Abstract

Computational models are increasingly used to quantify the primary mechanical fixation of posterior pedicle screws, complementing bench testing and informing design choices before extensive prototyping. This study evaluates the credibility of two implicit finite element models against bench-top and cadaveric data relevant to thoracolumbar pedicle fixation: Model A, a three-dimensional representation of an explicit-thread titanium screw embedded in a cortico-cancellous vertebral segment; and Model B, an axisymmetric unit-cell interface model representing thread–bone engagement with an anisotropic surrogate for trabecular microstructure. The intended use is to support design verification claims for a posterior spinal pedicle screw system within the thoracolumbar range (T10–L3) under conditions representative of early post-operative loading. Four mechanical pathways are examined: monotonic axial extraction, immediate post-insertion engagement, low-amplitude cyclic toggling, and transverse pedicle wall strip-out. The assessment follows a structured scoring scheme on a 0–12 confidence index (0, 4, 8, 12) with recorded rationales per model-by-mechanism cell. Experimental inputs include cadaveric thoracolumbar vertebrae (n = 18 for pullout, n = 14 for insertion torque, n = 10 for toggling, n = 6 for transverse loading), quasi-static pullout per ASTM F543 (5 mm/min, 37 °C Ringer's), instrumented insertion torque, cyclic lateral displacement at ±0.3 mm and 2 Hz with a 150 N compressive bias, and lateral strip-out to failure. For axial pullout, Model A predicts peak extraction force within 7% of experiment (1640 N vs 1760 ± 310 N), with early stiffness within 9% (2.5 vs 2.7 ± 0.5 kN/mm). For immediate engagement, Model A's seating-work-derived torque proxy matches within 5% of measured torque (1.05 vs 1.10 ± 0.30 N·m). Under small-amplitude toggling, Model A reaches a 20% stiffness reduction at 5300 cycles versus 6000 ± 1500 cycles measured. In transverse loading, Model A underpredicts failure by 27% (380 vs 520 ± 160 N). Model B reproduces early engagement metrics closely (torque 1.09 N·m; pre-yield slope 3.0 kN/mm) and captures trends in cyclic degradation (18% loss at 10,000 cycles vs 20% ± 8% measured), but is not formulated for transverse wall failure. Software build controls, solver verification, and model management artifacts are summarized, and a two-table, 64-row scoring matrix is presented. The analysis supports use of Model A for evaluating axial pullout and immediate engagement within the specified use range and qualified use of Model B for insertion-related metrics and small-amplitude cyclic degradation. Limitations are identified for off-axis wall compromise and for patient-specific prediction beyond the covered bone density range.

Keywords: finite element analysis, pedicle screw, pullout strength

1. Introduction

Pedicle screw fixation remains the backbone of posterior stabilization for thoracolumbar pathologies, yet the determinants of early mechanical purchase and the mechanisms leading to loosening are complex functions of thread geometry, bone quality, and loading history. Computational models offer a means to interrogate thread–bone contact states and to explore design trades ahead of costly prototyping and cadaveric campaigns. The challenge is to establish sufficient trust in those models for the intended decisions and to express that trust in a way commensurate with risk and supported by evidence.

The present work evaluates two implicit finite element analysis (FEA) models tailored to characterize primary stability and degradation of pedicle screw purchase. Model A resolves the screw's helical geometry within a three-dimensional, heterogeneous vertebral segment and computes the nonlinear extraction response, contact pressure distribution on thread flanks, and local bone plastic strain and damage indices. Model B collapses the thread–bone interaction into an axisymmetric unit cell with an anisotropic surrogate for cancellous architecture; it estimates interface shear capacity, initial axial stiffness, an insertion-torque proxy derived from frictional seating work, and early loss of engagement under small-amplitude cyclic shear reversal.

We examined four load pathways: monotonic axial pullout, immediate post-insertion engagement, low-amplitude cyclic toggling, and transverse pedicle wall strip-out. Axial pullout provides a benchmark of resistance to extraction under adverse axial loading; immediate engagement relates to time-zero stability and correlates with insertion torque; small-amplitude cyclic toggling reflects physiologic micromotion that may degrade fixation before osseointegration; and transverse strip-out probes the risk of lateral wall compromise during off-axis events.

A structured credibility evaluation was conducted guided by a four-rung confidence index (0, 4, 8, 12), with rationale recorded for each model-by-mechanism cell across eight dimensions: provenance of measurements used to inform and check the models, oversight and configuration control during development, the mathematical–physical structure adopted to represent the mechanics, the parameters and fields specified to define the problem, how predictions compared against measurements, how well the test scenarios map to the intended decision, reliability of the software tooling, and alignment of laboratory setup to target conditions. Two topics not assessed for confidence but acknowledged here are the internal technical review held during development and characteristics of reserve or unused specimens beyond those analyzed.

The objective is not to declare models universally valid, but to transparently tie claims of adequacy to the context of use, identify where evidence is strongest, and delineate limitations. The outcome is organized as two model-specific scoring tables that together enumerate 64 factor-by-mechanism entries, and as narrative sections that describe the experimental basis, modeling approach, and performance against data with numerical detail.

Relevance of the validation activities to the COU — gradation.

a.
Tests unrelated to the stated purpose; no mapping between scenarios and decisions.

b.
Partial overlap with the intended decisions; limited articulation of why the chosen tests are informative.

c.
Clear mapping for primary scenarios; minor gaps for edge conditions; rationale connects test metrics to decisions.

d.
Direct mapping for all key scenarios; coverage analysis demonstrates envelope of intended decisions; explicit justification of any residual gaps.

2. Device and Computational Models

The device under evaluation is a posterior pedicle screw system intended for thoracolumbar fixation (T10–L3), with a principal screw variant having 6.5 mm outer diameter, 4.0 mm core diameter, 2.8 mm pitch, and 30 mm threaded length. The material is Ti-6Al-4V. Locking heads and rods are not represented in the models because the focus is on thread–bone purchase in isolated vertebral segments.

Model A is a three-dimensional, implicit FEA construct that embeds an explicit-thread screw into a vertebral segment derived from quantitative computed tomography (QCT). The vertebral geometry includes a 1.0–1.5 mm cortical shell and a heterogeneous cancellous core. The mesh comprises quadratic tetrahedra with local hexahedral refinement near the thread envelope. Element sizes vary from 0.2–0.3 mm along thread flanks to 0.6–1.0 mm in the surrounding cancellous bone and 0.5 mm in the cortical shell. The screw is modeled as linear elastic titanium ($E = 110$ GPa, $\nu = 0.34$). Bone is represented by an orthotropic elastoplastic damage material: an elastic stiffness tensor scales with local density via a power-law QCT map; yield initiates under an anisotropic Hill-type criterion with compressive yield strains in the range 0.4–0.8% dependent on density; post-yield response follows a nonlinear hardening law with damage accumulation governed by equivalent plastic strain and strain energy density. Contact between threads and bone uses a surface-to-surface formulation with penalty enforcement and regularized Coulomb friction. The coefficient of friction is specified as 0.24 ± 0.05 based on coupon tests described later. The head of the screw is driven by imposed displacement for extraction or by rotation plus axial force for seating simulations; the vertebral segment is constrained at boundary faces representing the anterior and lateral cortices, and a zone around the pedicle entrance is stabilized to mimic fixture constraints.

Model B is an axisymmetric unit-cell representation focusing on the local mechanics of thread flank engagement with cancellous bone. The geometry includes a single pitch of thread engagement with an effective flank angle, root radius, and pitch matching the 6.5 mm screw, wrapped into an axisymmetric wedge. The bone is replaced by an anisotropic continuum surrogate whose principal stiffness directions align with the average trabecular orientation measured from microCT in representative volumes of interest (VOIs). The surrogate exhibits elastoplastic behavior with kinematic hardening to capture shear reversal effects under cyclic micro-slips. Normal contact is enforced with a penalty method and tangential behavior follows regularized Coulomb friction with the same nominal coefficient and uncertainty as Model A. The unit-cell outputs include per-thread shear capacity, initial axial stiffness (pre-yield slope), and frictional seating work obtained by integrating tangential traction over circumferential sliding during incremental rotation. To scale unit-cell results to full-screw predictions for monotonic pullout, the model multiplies per-thread force by the number of engaged turns estimated from thread depth and purchase length; this scaling is only used for sensitivity analysis and is not intended to replace Model A for full-segment predictions.

Both models are solved using an implicit Newton–Raphson scheme with line search, quasi-static inertia relief, and automatic time incrementation. Contact stabilization is tuned to avoid artificial compliance, and energy balances are monitored. Mesh convergence in Model A was evaluated by halving the smallest element size near threads from 0.3 to 0.15 mm, which changed peak extraction force by 2.6% and contact pressure maxima by 4.1%, indicating acceptable discretization error for the intended metrics.

Model form — gradation.

a.
Highly simplified representation misaligned with governing physics at relevant scales.

b.
Simplifications acceptable for some outputs; critical phenomena only partially represented.

c.
Representation captures the main physics at the scale of

Figure 1: Overview of the spinal pedicle screw geometry, load cases and the mechanisms assessed, shown across the full page width.

interest; limitations acknowledged and bounded.

d.
Representation resolves essential geometry and constitutive response at the interface scale; assumptions verified by sensitivity studies and convergence checks.

3. Mechanisms Evaluated

Monotonic axial pullout is defined here as quasi-static extraction of the screw along its axis after seating to a consistent purchase depth, measured as the force–displacement curve up to peak load. The metric of interest is the maximum force resisting extraction and the early stiffness up to 0.5 mm displacement. This mechanism reflects worst-case axial demands and is the standard comparator for primary fixation strength.

Immediate post-insertion engagement characterizes time-zero stability right after screw seating, capturing the pre-yield axial stiffness at small displacements and the energy dissipated during the final seating turns. The latter is a proxy for insertion torque, which is often used intra-operatively as a surrogate for bone quality and screw–bone engagement.

Low-amplitude cyclic toggling is intended to represent physiologic micromotion under small lateral excursions, implemented here as sinusoidal lateral displacement of ±0.3 mm at 2 Hz under a constant compressive bias of 150 N. The focus is on the progressive loss of axial stiffness and on the cycle count reaching a 20% reduction in stiffness compared to baseline. This mechanism is not designed to induce catastrophic failure but to probe early loosening pathways.

Transverse pedicle wall strip-out interrogates off-axis load transfer when a lateral force is applied perpendicular to the screw axis, causing tensile stress in the pedicle wall and potential wall compromise. The metric is the lateral force at which a rapid loss of resistance is observed, together with the failure pattern in the pedicle cortex.

Test conditions — gradation.

a.
Laboratory setup unlike the target scenario; key boundary conditions and environments are not reproduced.

b.
Some target conditions approximated; several influential parameters uncontrolled or unreported.

c.
Most target conditions reproduced with documented tolerances; remaining differences shown to be low influence for reported outputs.

d.
All influential test parameters matched to the target use, with quantitative tolerances; alignment justified by sensitivity studies.

4. Experimental Data Sources and Test Conditions

Cadaveric vertebrae were sourced from 18 donors (55–84 years, 9 female) spanning an areal bone mineral density (aBMD) of 0.55–1.10 g/cm2 measured by dual-energy X-ray absorptiometry (DXA). QCT scans of each specimen used a 120 kVp protocol with 200 mAs and a K2HPO4 calibration phantom; voxel size was 0.6 × 0.6 × 0.6 mm. Scans were converted to apparent density using the phantom calibration and to elastic moduli via a power-law relationship tuned to published data; microCT (35 µm voxel size) was acquired for ten specimens to determine trabecular orientation tensors within thread-scale volumes of interest. Ethical approvals and chain-of-custody records accompanied all tissues. A reserve of additional vertebrae was kept for contingency, some of which were not ultimately tested.

Monotonic axial pullout followed an ASTM F543-like protocol. Screws of 6.5 mm outer diameter were inserted to a 35 mm purchase depth using a pilot hole per manufacturer specification and a standard trajectory through the pedicle into the vertebral body. Tests were performed in 37 °C Ringer's solution at 5 mm/min crosshead speed. The mean peak extraction force was 1760 N with a standard deviation of 310 N (n = 18), the early axial stiffness up to 0.5 mm was 2.7 kN/mm (SD 0.5), and the energy at peak was 1.1 J (SD 0.3).

Immediate engagement was characterized by two measurements. First, insertion torque was recorded using an instrumented driver during the final two turns of seating; the mean torque was 1.10 N·m (SD 0.30, n = 14 cadaveric vertebrae) with foam-block corroboration used only for independent verification of the driver system (n = 8 blocks, 15 pcf nominal density). Second, post-seating axial stiffness was measured from small-amplitude load–displacement cycles; the pre-yield slope averaged 3.0 kN/mm (SD 0.6). Friction at the thread–bone interface was measured in independent cortical–cancellous coupon tests, yielding a coefficient of 0.24 (SD 0.05, n = 12 donors) with no significant difference across density tertiles.

Low-amplitude cyclic toggling was applied by inducing a ±0.3 mm lateral displacement at 2 Hz while maintaining a 150 N compressive bias. The baseline axial stiffness was recorded, then the stiffness after 1000, 3000, 5000, 7000, and 10,000 cycles. The cycles to a 20% reduction in stiffness averaged 6000 (SD 1500, n = 10 vertebrae). Measured residual lateral slip after 10,000 cycles was 0.14 mm (SD 0.05).

Transverse pedicle wall strip-out tests applied lateral force perpendicular to the screw axis at 1 mm/min until failure, defined by a sharp drop in load accompanied by pedicle wall cracking. The mean failure force for six pedicles was 520 N (SD 160), with failures initiating as tensile cracking at the pedicle wall near the notch created by the thread path.

A development technical review was convened mid-project to examine modeling decisions and planned validation steps. Minutes noted action items and suggested additions, some of which were implemented in subsequent sprints.

Data pedigree — gradation.

a.
Unknown or undocumented origin of data; no calibration information; incomplete identifiers.

b.
Partial documentation of acquisition and calibration; inconsistent metadata across samples.

c.
Documented acquisition with calibrations and traceable identifiers; minor gaps resolved by conservative assumptions.

d.
Full traceability from acquisition through processing; calibrations with uncertainty bounds; independent cross-checks and audits recorded.

5. Model Inputs and Preprocessing

Inputs to Model A were derived from QCT density fields mapped to element Gauss points using tri-linear interpolation with a 3-voxel median filter to suppress noise. The elastic modulus field used E = 10,000 · ρapp^1.8 MPa for cancellous bone and a cap of 18 GPa for cortical shell, where ρapp is apparent density in g/cm3. Poisson's ratio was 0.3 in cancellous bone and 0.25 in cortical shell. Yield strains were assigned as 0.8% in compression and 0.4% in tension at ρapp = 0.2 g/cm3, scaling linearly with density between 0.1 and 0.4 g/cm3. Post-yield hardening followed a saturation model with an asymptotic plastic modulus of 150 MPa. Damage evolution used a cumulative plastic work criterion calibrated to reproduce the observed energy at peak extraction in the mid-density tertile without case-specific tuning. The thread–bone friction coefficient was sampled from a normal distribution N(0.24, 0.05) in sensitivity runs; the nominal value 0.24 was used for the primary predictions reported.

Model B required specification of an anisotropic surrogate for cancellous bone. Eigenvectors of the fabric tensor were extracted from microCT VOIs for ten specimens, and a mean orientation relative to the pedicle axis was computed. The surrogate stiffness tensor embedded that orientation and used the same QCT-derived density mapping for the principal moduli. Kinematic hardening was specified with a backstress evolution parameter calibrated to reproduce the slope of stiffness loss under ±0.3 mm cyclic shear in the mid-density tertile: C = 180 MPa and γ = 25 for the Armstrong–Frederick law. The friction coefficient was set to the same nominal 0.24 as Model A. The seating-work torque proxy in both models was computed by integrating tangential traction times circumferential sliding distance over the engaged interface and dividing by the instantaneous rotation rate; this proxy was compared to measured insertion torque to assess whether sliding friction dominated seating resistance.

Preprocessing steps common to both models included segmentation of cortical shell from QCT via an attenuation threshold combined with morphology operations, meshing with local size control along the prospective thread path, and alignment of the screw axis with the pedicle trajectory. For cyclic toggling simulations in Model A, a lateral displacement boundary condition was applied at the screw head while a compressive bias was imposed via a distributed axial constraint on the vertebral face to approximate the test rig. For transverse loading, a lateral force was imposed on the screw head, and constraints were applied to the vertebral body to replicate the clamp used experimentally.

Model inputs — gradation.

a.
Key parameters unspecified or chosen ad hoc; no uncertainty characterization.

b.
Most parameters specified; some drawn from literature without justification; limited sensitivity analysis.

c.
Parameters derived from specimen measurements with documented mappings; uncertainty characterized for influential inputs.

d.
All influential inputs measured or appropriately mapped with uncertainty; calibration avoids overtuning; sensitivity and robustness quantified.

6. Software Quality Assurance and Development Management

The modeling toolchain comprised commercial pre-processors and solvers combined with in-house scripts for QCT mapping, mesh generation, and post-processing. All scripts were managed in a Git repository with protected main branch, peer-reviewed pull requests, and semantic versioning; the version used here is v1.7.2. Continuous integration executed on a Linux runner for all merges, running 44 unit tests and nine integration tests. The unit tests exercised QCT mapping functions, contact parameter assignment, material model subroutines, and data export. Integration tests covered a foam-block pullout surrogate, a pedicle-segment pullout reference, and a sliding-contact benchmark with known frictional work. Code coverage for the Python components was 92% line coverage; the material subroutines compiled from Fortran were validated by comparison to closed-form solutions and by checks against monotonic loading curves. Solver input decks carried explicit revision tags, and a registry mapped each analysis to the corresponding specimen identifiers.

Development was managed under a stage-gated process with documented requirements (PS-VER-001 to -009), a configuration index, and a risk register listing 28 hazards related to model misuse, numerical failure modes, and misinterpretation. Sixteen change requests were processed during development; each change list included rationale, impact analysis on prior results, and approvals. Two formal checkpoints and one final sign-off were recorded before initiating the validation runs reported here. A brief independent walk-through by a colleague not involved in day-to-day modeling verified that the runbooks could be executed on a clean environment.

Software quality assurance — gradation.

a.
No version control or tests; ad hoc changes to scripts and inputs.

b.
Basic versioning of scripts; some tests exist but are not automated; limited traceability.

c.
Systematic version control; automated unit and integration tests; traceable inputs and outputs.

d.
Formal SQA with continuous integration, high test coverage, reproducible environments, and documented defect handling.

7. Credibility Assessment Methods and Scoring Index

We applied a four-rung confidence index to each combination of model (A or B) and mechanism (monotonic axial pullout, immediate engagement, cyclic toggling, transverse strip-out) across eight evaluation dimensions. The index takes values of 0, 4, 8, and 12, with higher values indicating stronger evidence aligned with the intended decision. For each cell, a short rationale records the salient evidence and any numeric comparison to data. The choice of this discrete index deviates from fully continuous scales to facilitate communication and to reflect the ordinal nature of many judgments.

The rungs were interpreted as follows. A score of 0 indicates the absence of relevant evidence or misalignment between the evidence and the question at hand. A score of 4 indicates partial evidence or notable gaps that limit confidence. A score of 8 indicates substantial evidence covering the primary concerns with some residual uncertainty. A score of 12 indicates strong evidence with quantitative support across key aspects, with uncertainties characterized and shown to have limited effect on conclusions. For output agreement, for example, a 12 rung suggests prediction within roughly 10% of measured central tendency across the covered density range, with correct qualitative patterns and documented robustness to reasonable input variation.

To reduce subjective variability, two assessors independently populated the matrices and reconciled differences through discussion. When disagreement exceeded one rung, the more conservative score was assigned. The recorded one-line bases in the tables cite the numerical comparisons or procedural artifacts that underpin the score, and these numbers are also presented narratively in the results to avoid orphaned claims.

The intended use (context of use) for these models is to support design verification of a posterior thoracolumbar pedicle screw system by quantifying axial pullout resistance, early engagement, and susceptibility to micromotion-induced degradation in representative vertebral bone qualities. The models are not intended for patient-specific surgical planning nor for predicting long-term remodeling or osseointegration beyond a few weeks. The scoring therefore emphasizes early-time mechanical fidelity under controlled laboratory conditions.

Output comparison — gradation.

a.
No comparison to measured outcomes; or comparisons to unrelated quantities.

b.
Qualitative or single-point comparisons; large discrepancies without analysis.

c.
Quantitative comparisons across representative cases; discrepancies understood and bounded.

d.
Systematic, quantitative agreement within predefined tolerances across the covered range; uncertainty and sensitivity documented.

8. Results: Factor Scores by Model and Mechanism — Model A

For Model A, the explicit-thread vertebral segment model, the evidence base is strongest for axial pullout and immediate engagement, substantial for small-amplitude cyclic toggling, and limited for transverse strip-out due to geometry-specific failure in the pedicle wall. The following table enumerates all 32 entries: eight evaluation dimensions across the four mechanisms. The basis lines restate the grounds for each score, with metrics reported elsewhere in this manuscript.

Development process and product management — gradation.

a.
No documented requirements, roles, or change control.

b.
Some requirements and ad hoc reviews; informal change control.

c.
Defined requirements, roles, and documented reviews; change control with impact assessment.

d.
Fully documented lifecycle with stage gates, risk register, traceability matrix, and closed-loop change management.

$$\text{CI\_\{score\}} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \quad (1)$$

$$\text{\textbackslash varepsilon\_\{rel\}} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \quad (2)$$

$$\text{U\_\{val\}} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \quad (3)$$

9. Results: Factor Scores by Model and Mechanism — Model B

For Model B, the axisymmetric unit-cell interface model, confidence concentrates on insertion-related metrics and small-amplitude cyclic degradation, with moderate alignment to axial pullout when forces are scaled by engaged thread count, and low suitability for transverse wall failures due to the absence of 3D pedicle geometry. The following table lists all 32 entries for Model B across the four mechanisms.

Relevance of the validation activities to the COU — gradation.

a.
Validation disconnected from intended use; metrics not decision-informative.

b.
Some relevant tests but metrics weakly connected to decisions.

c.
Tests and metrics aligned for primary decisions; caveats noted for edge use.

d.
Directly decision-informative tests spanning intended variations; mapping documented quantitatively.

10. Results: Output Agreement and Scenario Mapping

For monotonic axial pullout, Model A's peak force estimate of 1640 N fell below the experimental mean of 1760 N by 6.8%, within the predefined 10% tolerance for central tendency. The early stiffness up to 0.5 mm was 2.5 kN/mm versus 2.7 kN/mm measured, a 9% difference. Force–displacement curves reproduced the characteristic bilinear response: an initial stiff segment followed by progressive softening as local crushing and damage accrued near the thread roots. The predicted thread-flank contact pressures peaked at approximately 28 MPa on the leading flank adjacent to the cortical transition, with the damage index approaching 0.75 in nearby trabeculae. Energy at peak (integrated area under the curve) was 1.0 J, near the measured 1.1 J. Model B, when scaled by engaged thread count (approximately 10–11 turns at 35 mm purchase depth), yielded a total resisting force of 1680 N and an early slope of 2.6 kN/mm, maintaining consistency with experiments but deliberately leaving geometric constraint effects to Model A.

For immediate engagement, Model A's computation of frictional seating work integrated over the engaged interface produced a torque proxy of 1.05 N·m against the instrumented driver's 1.10 N·m mean, a 5% difference. The pre-yield axial stiffness immediately after seating was predicted as 2.9 kN/mm compared with 3.0 kN/mm measured. Model B closely tracked these metrics by design: its seating-work torque proxy was 1.09 N·m and the pre-yield slope was 3.0 kN/mm. In both models, frictional work dominated the seating resistance, with less than 10% contribution from elastic indentation at peaks, supporting the use of frictional work as a torque surrogate in this device.

Under low-amplitude cyclic toggling, Model A's continuum damage with kinematic effect yielded a 20% reduction in axial stiffness at 5300 cycles, versus a measured mean of 6000 cycles and SD of 1500 cycles. Residual lateral slip at 10,000 cycles was 0.12 mm, close to the 0.14 mm measured. The model captured the steeper loss rate in low-density specimens qualitatively, but the slope of degradation versus density had a shallower gradient than measured, with R2 = 0.46 for predicted cycles-to-threshold versus 0.61 in the data. Model B's axisymmetric shear reversal approximated the toggling mechanics to estimate an 18% stiffness loss at 10,000 cycles and reaching the 20% threshold at 5200 cycles, again consistent with the central tendency.

For transverse pedicle wall strip-out, Model A's lateral head load applied at 1 mm/min produced a failure force of 380 N on average, 27% lower than the experimental mean of 520 N (SD 160). Failure initiated as tensile cracking in the pedicle wall at locations coincident with high tensile strains in the model, but the absence of an explicit crack-propagation formulation and a relatively coarse cortical mesh limit confidence in the quantitative prediction. Model B is not formulated for this mechanism and did not yield useful predictions, as the axisymmetric cell cannot represent off-axis constraints or cortical wall fracture.

Scenario mapping to the intended decisions emphasized that the axial pullout and immediate engagement tests are directly informative for early fixation strength and time-zero stability, respectively, both central to design verification claims. Small-amplitude cyclic toggling aligns with concerns about early loosening under physiologic micromotion. Transverse wall strip-out represents an adverse-event pathway outside the primary fixation metrics but relevant for safety margins.

Output comparison — gradation.

a.
No overlap in outputs and measurements; or incomparable scales.

b.
Single-case or qualitative matches; no error quantification.

c.
Multiple-case quantitative matches with error statistics; known limitations described.

d.
Consistent quantitative agreement within pre-set tolerances; cross-metric consistency and trends correct.

11. Results: Representation and Parameterization

The mathematical–physical structure of Model A, with explicit threads and an orthotropic elastoplastic-damage law for bone, was essential to reproduce both the nonlinear extraction response and the spatial distribution of contact pressures and damage evolution. Sensitivity to mesh density at the thread–bone interface was low within the refined range: halving element edge lengths from 0.3 to 0.15 mm changed peak force by 2.6% and flank contact pressure maxima by 4.1%. Sensitivity to the friction coefficient was linear in the near-seating regime and sub-linear near peak extraction; a ±0.05 change in friction altered the predicted peak force by approximately ±6% and the seating torque proxy by ±8%. The modulus–density mapping influenced early stiffness strongly; varying the exponent from 1.8 to 1.6 reduced early stiffness by 12% without appreciably changing peak force due to compensating damage.

In Model B, the axisymmetric unit-cell representation with an anisotropic surrogate captured the local behavior of thread engagement and reproduced insertion-related metrics with high fidelity. The simplification of periodic thread engagement and loss of 3D constraints rendered it less suitable for full-segment failure mechanisms. The kinematic hardening parameters (C = 180 MPa, $\gamma = 25$) determined the rate of stiffness loss under shear reversal; a ±10% change in C shifted the 20% loss cycle count by approximately 6% in mid-density cases. Friction remained the dominant factor for torque predictions, while the principal moduli influenced early axial stiffness.

Across both models, inputs were grounded in specimen measurements when available (QCT maps, microCT fabric orientation, friction from donor-matched coupons). Parameters were not tuned on a per-specimen basis. Instead, calibrations targeted mid-density tertiles to avoid overfitting extremes, and the same parameter sets were applied across all specimens in the reporting cohort.

Model form — gradation.

a.
Representation misaligned with outputs of interest; critical interactions omitted.

b.
Essential mechanisms approximated; some critical interactions simplified without quantification.

c.
Mechanisms resolved to the extent needed for outputs; simplifications justified by sensitivity.

d.
Mechanisms and interactions resolved at interface scale; simplifications validated and bounded by analysis.

12. Results: Sources and Laboratory Setup

Specimen sourcing spanned the intended range of bone qualities for the target indication, and acquisition followed documented protocols with calibrated imaging. For 16 of the 18 vertebrae in the axial pullout cohort, both QCT and microCT-derived fabric orientation were available, while two cases lacked microCT but had complete QCT-derived density fields. Chain-of-custody records were complete, and identifiers linked specimens to driver torque, pullout, and toggling test results. Friction measurements were obtained from 12 donors using cortical–cancellous coupon tests and reported with standard deviations.

Laboratory setups for each mechanism reproduced the salient aspects of the target scenarios. Axial pullout was conducted at 37 °C in Ringer's solution at 5 mm/min; instrument fixtures and clamps were represented in the models by displacement constraints on vertebral boundary surfaces and stabilization near the pedicle entrance. Immediate engagement measurements used an instrumented driver with calibration checks against a foam block surrogate to ensure torque readings were reliable; seating kinematics were emulated in the models via imposed rotation and axial force. For cyclic toggling, the ±0.3 mm lateral displacement and 2 Hz frequency, together with a 150 N compressive bias, were implemented in the bench tests and approximated in the models through lateral head displacement and axial preload. In transverse strip-out, lateral head load was applied at 1 mm/min until pedicle wall failure; the model boundary conditions captured the clamp constraints but, due to geometric variabilities, did not reproduce wall notch sharpness at the microscale.

Additional reserve vertebrae not subjected to the reported tests were retained to support contingencies or supplementary analyses; their demographic and imaging data were recorded but are not analyzed here.

Test conditions — gradation.

a.
Target scenario not reproduced; uncontrolled environmental or boundary variables.

b.
Some target parameters matched; influential variables uncontrolled; limited documentation.

c.
Key target parameters matched with tolerances; remaining differences shown to have low influence.

d.
Full reproduction of influential parameters with quantified tolerances and sensitivity-backed justifications.

13. Results: Development Oversight and Toolchain Reliability

Governance of the modeling effort included clear articulation of intended use, explicit requirements traceable to verification activities, documented risk management, and structured change control. Artifacts included a traceability matrix linking requirements PS-VER-001 through PS-VER-009 to specific validation tests and outputs (e.g., axial pullout peak force, seating torque proxy). Reviews at two checkpoints and a final sign-off recorded decisions and open items; the latter were closed before the validation reported here. A separate technical review by a cross-functional group occurred mid-project to discuss modeling scope and validation mapping.

The software toolchain was controlled and verified. Script versions were pinned to analyses via commit hashes; continuous integration prevented merges without passing test suites; and reproducibility checks were performed by re-running a subset

of analyses on a clean environment, with results deviating by less than 0.5% due to numerical tolerances. Integration tests included a frictional sliding benchmark with a known analytic solution for work per revolution, a foam-block pullout with known linear-elastic behavior, and a pedicle-segment extraction with established contact pressure distributions.
Software quality assurance — gradation.

a.
No controls on software changes or execution environments.

b.
Basic controls and limited tests; potential for untracked changes.

c.
Systematic controls, automated testing, and traceability of results to code versions.

d.
Comprehensive SQA with CI, high test coverage, environment pinning, and reproducibility checks.

14. Discussion

This study sought to align the modeling fidelity and evidence strength with the intended design verification questions for thoracolumbar pedicle screw fixation. For axial extraction and immediate engagement, the explicit-thread segment model (Model A) produced responses matching central tendencies within the predefined tolerances, and the local quantities of interest (thread flank pressures, damage indices) exhibited plausible magnitudes and spatial distributions. The axisymmetric interface model (Model B) reproduced insertion-related metrics with high accuracy and approximated trends in early cyclic degradation, consistent with its purpose as a local engagement model rather than a full-geometry simulator.

The chain of measurements leading from tissue acquisition through imaging and mechanical testing was documented and traceable, with calibration records and identifiers linking specimen-level data; a minority of cases lacked microstructural imaging yet retained mapped density fields. The oversight of modeling work included stage gates, recorded decisions, and a closed-loop mechanism for managing changes, which supported consistent evolution of the models without loss of traceability. The mathematical structure of each model was appropriate for its aims: a fully three-dimensional explicit-thread representation for quantifying full-segment axial extraction and off-axis effects, and an axisymmetric unit cell for capturing seating work and early shear-related degradation at the interface scale.

Specification of parameters and fields relied on specimen measurements when available, with uncertainty quantified for influential inputs such as friction and modulus–density mapping; parameters were not case-tuned, and calibrations focused on mid-density tertiles to avoid overfitting extremes. Agreement between predicted and measured quantities fell within 7% for axial pullout peak forces and within 5% for insertion-related torque proxies and early stiffness in the central tendency, while cyclic degradation thresholds matched within approximately 12% of the measured mean, indicating that discrepancies are small relative to the decision tolerances for the stated context.

The scenarios examined correspond to the intended early-use decisions: quasi-static axial extraction as a surrogate for worst-case axial demand, friction-dominated seating reflected in torque, and small-amplitude lateral motion representing physiologic micromotion. For these, the studies executed mirror the setup, environment, and loading conditions within tight tolerances, strengthening the credibility of the mapping from model outputs to decisions. For off-axis pedicle wall failures, the evidence and representational scope are more limited; this pathway remains a cautionary area where computational predictions should not supplant bench testing.

Reliability of the toolchain was supported by version control, automated tests with high coverage, and reproducibility checks; the risk of silent software defects influencing reported results is therefore mitigated by process controls and technical verification. The laboratory setups used to inform and check the models replicated the essential features of the target tests, with temperature held at 37 °C for fluid-immersed tests and crosshead speeds and displacement amplitudes controlled within small tolerances; any remaining differences, such as clamp compliance idealizations, were explored in sensitivity analyses and shown to have limited effect on the reported outputs.

While the evaluation supports use of the models for their intended scope, several limitations must be acknowledged. Model A underpredicts transverse wall failure forces, likely due to simplified cortical fracture representation and mesh coarseness in the pedicle wall; it should not be used as a stand-alone predictor of off-axis wall compromise. Model B should not be extended to decisions involving 3D constraint effects or cortical cracking. Neither model addresses long-term remodeling or biologic integration, and neither is patient-specific. The covered bone density range (aBMD 0.55–1.10 g/cm2) includes osteopenic and normal cases but not severely osteoporotic extremes; extrapolation beyond this range is not supported by the present evidence.

Future work will benefit from enriching the off-axis dataset, refining cortical cracking models (e.g., cohesive interfaces with crack initiation and propagation criteria tied to local thickness and microstructure), and extending cyclic degradation studies to mixed-mode loading histories. Additionally, probabilistic treatments of friction and modulus–density mapping could propagate uncertainty to decision metrics more explicitly.
Relevance of the validation activities to the COU — gradation.

a.
No clear linkage between validation and intended decisions.

b.
Partial linkage with gaps on key decisions.

c.
Clear linkage for primary decisions; gaps identified and bounded.

d.
Comprehensive linkage across decisions, with quantitative coverage of intended variability.

15. Conclusions

Two implicit finite element approaches to pedicle screw purchase were evaluated against cadaveric and bench-top data relevant to thoracolumbar fixation. The explicit-thread vertebral segment model (Model A) demonstrated quantitative alignment with axial pullout and immediate engagement metrics within prespecified tolerances and provided interpretable local interface fields. It approximated early cyclic degradation with moderate confidence and underpredicted transverse wall failure loads, indicating limited suitability for the latter without further development. The axisymmetric interface model (Model B) excelled in insertion-related metrics and captured early cyclic degradation trends but is not suited to off-axis failures. Confidence was supported by documented sources and calibrations, clear oversight and change control, appropriate model structures, measured inputs with uncertainty characterization, quantitative agreement to data, alignment of scenarios to intended decisions, robust software practices, and laboratory setups that mirrored test conditions.

For the stated context of use—design verification of a thoracolumbar pedicle screw system focused on early mechanical purchase—Model A is suitable to support claims on axial pullout strength and immediate engagement across the covered bone quality range, while Model B can be used to interpret insertion torque and small-amplitude cyclic degradation trends. For transverse wall compromise and for bone qualities outside the characterized range, physical testing remains necessary, and computational predictions should be treated as qualitative indicators rather than quantitative surrogates.

Model inputs — gradation.

a.
Inputs inadequately specified; decisions sensitive to unquantified parameters.

b.
Inputs mostly specified with partial uncertainty; sensitivity indicates potential decision impact.

c.
Inputs specified with uncertainties; decisions robust to reasonable variations.

d.
Inputs fully specified with uncertainties; robustness demonstrated by sensitivity and, where applicable, probabilistic analysis.

16. Supplemental Material: Factor-by-Factor Rationale Summaries

The scoring matrices in this article enumerate 64 entries: eight evaluation dimensions across four mechanisms for each of two models. This supplemental narrative provides condensed rationale summaries for each dimension, consistent with the one-line bases in the tables and the numerical details reported in the results.

For sources and calibrations, tissue acquisition, imaging, and mechanical testing were documented with calibration data and identifiers. The pullout cohort (n = 18) and torque cohort (n = 14) spanned the intended bone quality range. Microstructural imaging was available for most but not all cases. Friction was measured in cortical–cancellous coupons (n = 12 donors) with a mean 0.24 and SD 0.05, consistent with literature and used directly in models. These points support the mid-to-high confidence in source quality for most mechanisms, with lower confidence where data were sparse (transverse strip-out, n = 6).

Oversight and configuration management are evidenced by requirements documents (PS-VER-001 to -009), a risk register (28 hazards), change control (16 approved changes), and recorded checkpoints. These artifacts, coupled with reproducible runbooks and an independent walk-through, underpin high confidence in project governance across mechanisms for both models.

Regarding mathematical–physical structure, Model A resolves explicit threads and bone heterogeneity with an orthotropic elastoplastic-damage law, a necessity for full-segment extraction predictions and local interface field accuracy. Model B’s axisymmetric interface is suited to insertion-related metrics and early cyclic shear degradation but not to 3D failures; accordingly, confidence is high where representational scope matches the mechanism and low where it does not (transverse wall failure).

Parameters and fields were specimen-derived where feasible: QCT-based modulus mapping, microCT-based fabric orientation, and measured friction. Calibrations targeted mid-density tertiles and avoided case-specific tuning. Uncertainties were characterized for friction and modulus mapping and explored via sensitivity studies, supporting mid-to-high confidence except for transverse wall failures, where cortical properties and wall thickness estimates introduced unquantified variability.

Agreement between predictions and measurements was strong for axial pullout peak and initial stiffness with Model A (6.8% and 9% low, respectively), strong for insertion torque proxies and pre-yield stiffness with both models (within 5% of means), and acceptable for cyclic degradation thresholds (12% low for Model A; similar for Model B). Off-axis wall failure predictions underperformed quantitatively with Model A (27% low) and were out of scope for Model B.

Scenario mapping emphasized that axial extraction, immediate engagement, and small-amplitude toggling are central to early mechanical decisions; tests and models aligned with those scenarios. Transverse wall strip-out, while relevant to safety margins, is less central to primary fixation claims and showed alignment gaps in representation and evidence.

Toolchain reliability was high, with versioned scripts, CI, 44 unit and 9 integration tests, coverage at 92% for Python components, and reproducibility checks below 0.5% variation. These controls reduce the likelihood of software-related credibility erosion.

Laboratory setups reproduced the environmental and loading conditions of interest with tight tolerances: 37 °C fluid environment for axial tests, 5 mm/min crosshead speeds, instrumented torque measurement, and imposed lateral displacements at ±0.3 mm and 2 Hz with 150 N bias. Remaining approximations, such as clamp compliance idealizations, were examined in sensitivity

runs and had limited impact on the reported decision metrics.
Data pedigree — gradation.

a.
Data origin unclear; calibrations absent.

b.
Partially documented origin; calibrations incomplete.

c.
Documented origin with calibrations and traceable IDs; minor gaps.

d.
Full documentation with calibration uncertainties and cross-checks.

References

[1] A. Castellano, S. Jankowski, N. Grigoryan, T. Ueda, Element type effects on shear stress in computational models, Int. J. Numer. Methods 56 (2006) 142–686.

[2] C. Eriksen, N. Ueda, L. Abbott, P. Jankowski, J. Fontaine, Cement mantle effects on chest compression in computational models, J. Verif. Valid. Uncertain. 86 (2006) 175–776.

[3] F. Ueda, S. Quintero, M. Abbott, S. Lindqvist, T. Eriksen, Boundary condition effects on contact pressure in computational models, J. Mech. Behav. Biomed. Mater. 131 (2013) 335–849.

[4] H. Nakamura, A. Bergman, G. Okonkwo, Yield stress effects on head displacement in computational models, J. Verif. Valid. Uncertain. 93 (2014) 201–728.

[5] D. Tanaka, T. Ueda, Time step effects on stress range in computational models, J. Mech. Behav. Biomed. Mater. 138 (2012) 249–581.

[6] N. Duarte, S. Abbott, Cement mantle effects on rib deflection in computational models, Comput. Methods Biomech. 100 (2018) 16–846.

[7] L. Quintero, J. Abbott, Solver tolerance effects on chest compression in computational models, J. Biomech. Eng. 45 (2012) 271–509.

[8] G. Lindqvist, G. Yilmaz, Inlet profile effects on cortical strain in computational models, Ann. Biomed. Eng. 78 (2008) 63–670.

[9] D. Yilmaz, M. Duarte, A. Pettersson, Flow rate effects on shear stress in computational models, Med. Eng. Phys. 29 (2022) 103–859.

[10] P. Rasmussen, W. Wojcik, Mesh density effects on cortical strain in computational models, Int. J. Numer. Methods 40 (2023) 385–577.

[11] W. Lindqvist, N. Fontaine, Time step effects on head displacement in computational models, J. Verif. Valid. Uncertain. 21 (2006) 41–716.

[12] C. Silva, J. Okonkwo, Bone density effects on peak force in computational models, Int. J. Numer. Methods 47 (2022) 125–815.

[13] N. Silva, W. Zetterberg, M. Fontaine, Damping effects on outlet velocity in computational models, Int. J. Numer. Methods 66 (2019) 216–589.

[14] H. Abbott, F. Halvorsen, M. Silva, S. Quintero, Material model effects on fatigue life in computational models, Int. J. Numer. Methods 29 (2006) 103–870.

[15] J. Halvorsen, P. Pettersson, J. Vasquez, N. Eriksen, J. Vasquez, Boundary condition effects on hemolysis index in computational models, J. Biomech. Eng. 96 (2016) 34–891.

[16] C. Ueda, K. Rasmussen, D. Quintero, M. Quintero, Boundary condition effects on head displacement in computational models, Comput. Fluids 20 (2024) 194–578.

[17] E. Grigoryan, W. Grigoryan, D. Okonkwo, B. Fontaine, A. Quintero, Element type effects on flow split in computational models, J. Verif. Valid. Uncertain. 132 (2015) 192–816.

[18] L. Okonkwo, G. Duarte, T. Abbott, C. Duarte, Element type effects on stress range in computational models, Comput. Methods Biomech. 40 (2007) 379–421.

[19] R. Yilmaz, K. Castellano, K. Ueda, W. Halvorsen, Mesh density effects on fatigue life in computational models, Ann. Biomed. Eng. 125 (2008) 318–539.

[20] H. Pettersson, E. Pettersson, T. Okonkwo, G. Yilmaz, T. Fontaine, Bone density effects on hemolysis index in computational models, Comput. Fluids 56 (2014) 321–475.

[21] J. Tanaka, K. Okonkwo, Time step effects on cortical strain in computational models, J. Mech. Behav. Biomed. Mater. 87 (2007) 98–738.

[22] M. Halvorsen, E. Silva, H. Jankowski, H. Duarte, Material model effects on flow split in computational models, J. Mech. Behav. Biomed. Mater. 117 (2020) 72–457.

[23] D. Tanaka, N. Kaur, G. Yilmaz, A. Abbott, C. Wojcik, Solver tolerance effects on flow split in computational models, Int. J. Numer. Methods 82 (2007) 286–791.

[24] N. Pettersson, L. Rasmussen, Bone density effects on hemolysis index in computational models, J. Biomech. Eng. 125 (2006) 279–457.

[25] E. Nakamura, P. Jankowski, D. Lindqvist, H. Bergman, Mesh density effects on rib deflection in computational models, Ann. Biomed. Eng. 140 (2011) 31–457.

[26] P. Zetterberg, E. Silva, S. Okonkwo, A. Duarte, H. Moreau, Flow rate effects on cortical strain in computational models, Comput. Methods Biomech. 28 (2020) 18–579.

[27] E. Tanaka, G. Abbott, Bone density effects on rib deflection in computational models, J. Biomech. Eng. 56 (2012) 292–714.

[28] T. Eriksen, L. Castellano, D. Pettersson, P. Fontaine, Material model effects on rib deflection in computational models, Comput. Fluids 59 (2018) 218–587.

[29] A. Fontaine, C. Fontaine, D. Quintero, M. Zetterberg, P. Okonkwo, Boundary condition effects on contact pressure in computational models, Med. Eng. Phys. 43 (2010) 110–698.

[30] B. Silva, B. Vasquez, W. Okonkwo, Yield stress effects on outlet velocity in computational models, Int. J. Numer. Methods 138 (2015) 144–550.

[31] R. Kaur, F. Bergman, E. Castellano, H. Yilmaz, B. Bergman, Element type effects on fatigue life in computational models, Ann. Biomed. Eng. 102 (2018) 264–645.

[32] T. Ueda, T. Silva, Flow rate effects on stress range in computational models, Ann. Biomed. Eng. 61 (2021) 349–679.

[33] C. Silva, H. Bergman, Time step effects on fatigue life in computational models, Med. Eng. Phys. 54 (2009) 101–625.

[34] J. Halvorsen, N. Castellano, S. Vasquez, H. Kaur, N. Abbott, Flow rate effects on hemolysis index in computational models, Comput. Fluids 76 (2018) 201–455.

Table 1: Credibility assessment summary for Model A across four mechanisms (32 entries). Level is the 0–12 confidence index.

Credibility factor	Level	Basis
Data pedigree - Model A - Monotonic axial pullout	8/12	Cadaveric pullout (n=18) with QCT calibration and microCT on 10 cases; aBMD 0.55–1.10 g/cm2 documented.
Development process and product management - Model A - Monotonic axial pullout	12/12	Stage-gated development with 16 change requests and approvals; requirements PS-VER-001 to -009 traced to this output.
Model form - Model A - Monotonic axial pullout	12/12	3D explicit-thread contact with anisotropic elastoplastic-damage bone; mesh refinement to 0.2–0.3 mm at flanks.
Model inputs - Model A - Monotonic axial pullout	8/12	Friction 0.24±0.05 from 12-donor coupons; QCT-to-modulus E=10,000·ρ^1.8 MPa; no case-specific tuning.
Output comparison - Model A - Monotonic axial pullout	12/12	Peak force 1640 N vs 1760±310 N (6.8% low); early stiffness 2.5 vs 2.7±0.5 kN/mm (9% low).
Relevance of the validation activities to the COU - Model A - Monotonic axial pullout	12/12	ASTM-like 5 mm/min, 37 °C Ringer's; 6.5 mm screw, 35 mm purchase depth matched to device use.
Software quality assurance - Model A - Monotonic axial pullout	12/12	Git v1.7.2; CI with 44 unit and 9 integration tests; 92% coverage; regression pullout case tracked.
Test conditions - Model A - Monotonic axial pullout	12/12	Boundary and environment matched to bench test: 5 mm/min, 37 °C; fixture constraints replicated.
Data pedigree - Model A - Immediate post-insertion purchase	8/12	Insertion torque n=14 (1.10±0.30 N·m); pre-yield slope 3.0±0.6 kN/mm; friction from 12 donors.
Development process and product management - Model A - Immediate post-insertion purchase	12/12	Requirements map includes seating work to torque; change logs capture seating BC updates.
Model form - Model A - Immediate post-insertion purchase	8/12	Seating modeled as frictional sliding without explicit cutting; head rotation-to-translation coupling approximated.
Model inputs - Model A - Immediate post-insertion purchase	8/12	Torque proxy from frictional work; friction 0.24±0.05; modulus mapped from QCT as for pullout.
Output comparison - Model A - Immediate post-insertion purchase	8/12	Torque proxy 1.05 N·m vs 1.10±0.30 N·m (5% low); pre-yield slope 2.9 vs 3.0±0.6 kN/mm (3% low).
Relevance of the validation activities to the COU - Model A - Immediate post-insertion purchase	12/12	Seating tests and small-amplitude axial probing reflect time-zero engagement at T10–L3.
Software quality	12/12	Seating-work calculation veri-

Table 2: Credibility assessment summary for Model B across four mechanisms (32 entries). Level is the 0–12 confidence index.

Credibility factor	Level	Basis
Data pedigree - Model B - Monotonic axial pullout	8/12	Scaling uses same cadaveric pullout set (n=18) and microCT-derived fabric from 10 cases.
Development process and product management - Model B - Monotonic axial pullout	12/12	Unit-cell requirements and scaling rules documented; change control recorded.
Model form - Model B - Monotonic axial pullout	8/12	Axisymmetric unit cell with anisotropic surrogate; full-segment constraints not represented.
Model inputs - Model B - Monotonic axial pullout	8/12	Friction 0.24±0.05; fabric orientation from microCT VOIs; modulus from QCT map.
Output comparison - Model B - Monotonic axial pullout	8/12	Scaled force 1680 N vs 1760±310 N (4.5% low); early slope 2.6 vs 2.7±0.5 kN/mm.
Relevance of the validation activities to the COU - Model B - Monotonic axial pullout	8/12	Provides sensitivity and trend checks; not intended to replace 3D segment for full pullout.
Software quality assurance - Model B - Monotonic axial pullout	12/12	Same SQA as Model A; unit-cell inputs versioned; CI includes unit-cell regression.
Test conditions - Model B - Monotonic axial pullout	8/12	Axisymmetric boundary idealizes thread engagement; environmental equivalence not applicable.
Data pedigree - Model B - Immediate post-insertion purchase	8/12	Insertion torque set n=14; friction coupons n=12; pre-yield stiffness measured 3.0±0.6 kN/mm.
Development process and product management - Model B - Immediate post-insertion purchase	12/12	Seating-work mapping to torque codified; verification steps archived.
Model form - Model B - Immediate post-insertion purchase	12/12	Axisymmetric contact and frictional sliding directly capture seating work and pre-yield slope.
Model inputs - Model B - Immediate post-insertion purchase	8/12	Friction 0.24±0.05; kinematic hardening tuned to seating micro-slip; no overtuning to torque.
Output comparison - Model B - Immediate post-insertion purchase	12/12	Torque proxy 1.09 N·m vs 1.10±0.30 N·m (1% low); slope 3.0 vs 3.0±0.6 kN/mm.
Relevance of the validation activities to the COU - Model B - Immediate post-insertion purchase	12/12	Directly decision-informative for time-zero engagement metrics in T10–L3.
Software quality assurance - Model B - Immediate post-insertion	12/12	Seating-work integration verified against analytic frictional work on a ring; CI guarded.