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

Lena M. Ortiz, PhD[a], Arun K. Deshpande, PhD[b], Marta M. Kovacs, MS[c], Peter J. Huang, MD[1], Silvia E. Romero, PhD[1]

[a] *Computational Orthopaedics Laboratory, Center for Musculoskeletal Innovation, Denver, CO, USA,*
[b] *Department of Mechanical Engineering, Lakeside Institute of Technology, Evanston, IL, USA,*
[c] *Department of Orthopaedic Surgery, Mountainview University School of Medicine, Salt Lake City, UT, USA,*

Abstract

Purpose: This article evaluates the credibility of two implicit finite element models developed to assess pedicle screw purchase in thoracolumbar vertebrae: a high-fidelity, thread-resolved screw–bone model (TRSB-FE) and a homogenized trabecular anchor model with interface elements (HTA-FE). The assessment follows a factor-based framework keyed to eight evidence dimensions and four clinically motivated mechanisms: monotonic axial pullout resistance, thread stripping at peak insertion torque, off-axis toggling–induced loosening, and early micromotion under physiologic bending. Methods: We designed cadaveric and synthetic test beds, captured donor-specific imaging and mechanical data, and executed a validation program spanning monotonic, insertion, and cyclic loading regimes. The TRSB-FE computes nonlinear contact and local bone damage with direct thread flank engagement, while HTA-FE employs anisotropic elastic–plastic surrogates and traction–separation interfaces to capture anchor stiffness and degradation. We report per-mechanism performance statistics (R2, RMSE, mean absolute percentage error), quantitative process controls (sample sizes, imaging calibration, fixture tolerances, code verification coverage), and per-factor ratings on a 0–9 Credibility Index. Results: For monotonic axial pullout, TRSB-FE reproduced peak force with 12% mean absolute percentage error (RMSE 112 N; R2 = 0.86; n = 24), whereas HTA-FE achieved 21% mean absolute percentage error (R2 = 0.64; n = 24). For peak insertion torque and stripping, TRSB-FE achieved 10% mean absolute percentage error (R2 = 0.82) and 0.84 area under the receiver operating characteristic curve for classification of stripping (n = 28 insertions), while HTA-FE achieved 26% mean absolute percentage error (R2 = 0.49; AUC 0.68; n = 28). Under off-axis toggling (±5 N·m, 1 Hz, 10,000 cycles; n = 16), HTA-FE predicted loss of anchor stiffness with 14% mean absolute percentage error (R2 = 0.81) and micromotion growth per 1,000 cycles with RMSE 11 µm, whereas TRSB-FE yielded 22% mean absolute percentage error (R2 = 0.58; RMSE 19 µm per 1,000 cycles). For early micromotion under physiologic bending (3 N·m bending with 200 N axial preload; n = 20), HTA-FE achieved RMSE 12 µm (R2 = 0.91) and TRSB-FE achieved RMSE 19 µm (R2 = 0.74). Process controls included chain-of-custody across 92 anchors, imaging calibration to a daily hydroxyapatite phantom, fixture rates within 5% of target, and test coverage of 78% for solver wrappers and user subroutines. Credibility Index ratings favored TRSB-FE for thread-scale mechanisms (pullout, stripping) and HTA-FE for interface-scale phenomena (micromotion, toggling). Conclusions: The two models provide complementary utility for the intended clinical question of bone purchase, with TRSB-FE preferred where thread geometry and local crushing govern response and HTA-FE preferred where continuum anisotropy and interface degradation dominate. Gaps remain around fatigue extrapolation, reviewer documentation, and broader mapping to planning contexts; planned elevations include expanded cyclic datasets, formal technical review artifacts, and uncertainty propagation tied to patient-specific density fields.

Keywords: pedicle screw, finite element validation, bone purchase

1. Introduction

Pedicle screw fixation remains a cornerstone of thoracolumbar stabilization and deformity correction, yet the variability of screw purchase in osteoporotic and mixed-density vertebrae continues to challenge surgical planning. Intraoperative cues such as maximal insertion torque and qualitative tactile feedback are imperfect surrogates for long-term performance. Non-invasive preoperative imaging can quantify bone mineral density and regional anisotropy, but clinicians cannot directly interrogate how thread-scale engagement, interface crushing, and local microarchitecture translate into pullout strength, susceptibility to stripping at peak torque, early micromotion, and cyclic loosening under physiologic toggling. Computational models are appealing tools to bridge this gap; however, their clinical utility depends on credible predictions aligned with specific, clinically relevant mechanisms and supported by structured evidence.

We report a credibility assessment for two implicit finite element approaches built to predict different aspects of pedicle screw purchase. The first, a thread-resolved screw–bone model (TRSB-FE), targets the explicit mechanics of thread flank engagement, including inelastic crushing, contact separation, and stick–slip friction as a function of local density. The second, a homogenized trabecular anchor model (HTA-FE), targets continuum-scale response of an anisotropic trabecular surrogate coupled to an interface law that captures progressive degradation under toggling. The models differ in geometric resolution, constitutive description, and solver controls, and consequently in computation cost and intended questions they address. We treat four mechanisms of interest: monotonic axial pullout resistance; thread stripping at peak insertion torque; off-axis toggling–induced loosening; and early micromotion under physiologic bending.

A structured, factor-based framework provides the backbone of the assessment. We constructed a test and data pipeline that captured monotonic, insertion, and cyclic responses for cadaveric vertebrae spanning osteopenic to normal density and for density-matched polyurethane foams. Imaging, mechanical testing, and model execution were linked with identifiers to ensure a single analytical chain from raw data to publication figures. The lineage of measurements from scanner through fixturing and test execution was complete for 88 of 92 anchors, with sufficient metadata for the remaining 4 to resolve key variables, which met our pre-specified threshold for traceability in this study.

This article is organized to first describe the device and failure mechanisms that shape the questions we seek to answer; next the two computational models, their parameters, and the experimental test beds; then the credibility rubric and scoring scheme; followed by per-factor ratings and numerical comparisons of model outputs to experiments across the four mechanisms; and finally an appraisal of implications for bone purchase, a discussion of process controls, and a plan to elevate evidence in areas where gaps remain. In considering relevance to eventual contexts of use in preoperative planning, we sought to align the test modes with anticipated postoperative loading, but a formal mapping to a risk-based use case classification is deferred to future work.

Data pedigree — gradation.

a.
Data exist but are not traceable to specimen identity; provenance and calibration not described.

b.
Specimen identity and acquisition context recorded; partial calibration and incomplete chain-of-custody.

c.
Full specimen identity with imaging calibration logs; end-to-end linkage for a majority of tests and complete metadata for key variables.

d.
End-to-end linkage from imaging and preparation to mechanical test and analysis for all tests, with independent audit and immutable storage.

2. Device and Failure Mechanisms of Interest

The device studied is a 6.5 mm outer diameter, 45 mm length thoracolumbar pedicle screw with a polyaxial head and a cortical–cancellous thread profile. The distal cortical lead uses a 1.5 mm pitch to promote cortical engagement at the pedicle wall; the proximal cancellous thread uses a 2.8 mm pitch and 0.7 mm thread height to increase purchase in trabecular bone. Threads are single-start with a 6-degree flank angle. A 3.2 mm pilot hole was prepared in all cadaveric vertebrae and foam blocks; the screw was inserted at 2 revolutions per minute using a torque-controlled driver with 0.01 N·m resolution. The polyaxial head was decoupled during insertion and re-tightened for bending and toggling tests with a standardized set-screw torque of 8 N·m to minimize construct variability.

We focused on four mechanisms that align with early mechanical performance after implantation. First, monotonic axial pullout resistance reflects the limit load to extraction and remains a primary index of purchase in bench-top evaluation. Second, thread stripping at peak insertion torque reflects loss of engagement caused by local crushing and shearing during advance, which compromises immediate fixation and is readily observed intraoperatively. Third, off-axis toggling–induced loosening is a repetitive-motion phenomenon in which small alternating moments applied at the pedicle head translate to microscale shear and crushing at the interface, progressively degrading stiffness and reducing subsequent pullout strength. Fourth, early micromotion under physiologic bending reflects the initial displacement response of the anchor to a few Newton-meters of bending with axial preload, which correlates with construct stability before biological fixation or bone remodeling occur.

To support the generality of the conclusions within the intended range, specimens spanned osteopenic to normal densities as quantified by quantitative computed tomography, with vertebral body density ranging from 70 to 180 mg hydroxyapatite per cubic centimeter. For each mechanism, the number of anchors met or exceeded our pre-specified minima: 24 for monotonic pullout, 28 for insertion torque and stripping detection, 16 for off-axis toggling, and 20 for early micromotion.

Test samples — gradation.

a.
Limited number of specimens of uncertain representativeness.

b.
Moderate number of specimens with basic characterization of density or anatomy.

c.
Specimens span the relevant density range with imaging-based characterization and sample sizes pre-specified to support validation statistics.

d.
Specimens stratified by density and anatomy with power calculations, randomization, and preregistered protocols.

3. Computational Models

Two distinct finite element models were developed to address complementary facets of screw purchase. Both models are implicit and quasi-static in their solver formulations and employ nonlinear contact. The first, TRSB-FE, resolves the screw thread geometry explicitly and discretizes the surrounding bone at sub-thread scales to capture local engagement, plasticity, and

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

damage. The screw is modeled as an elastic titanium alloy (E = 110 GPa, nu = 0.34) with helical threads matching the device. The bone domain surrounding the screw is generated from preoperative quantitative CT with element-wise mapping of density to material parameters. A crushable foam plasticity model with orthotropic elasticity and scalar damage represents trabecular response, with evolution laws tied to density. Contact between screw and bone uses a penalty formulation with rate-dependent stick–slip and a cutback strategy to resolve thread flank traversal over discretized bone faces as insertion progresses. The typical mesh contains 3.5–5.0 million elements, with near-thread hexahedral elements sized 0.08–0.12 mm in bone and 0.10 mm on screw flanks; the far field bone mesh is coarsened to 0.8–1.2 mm with a graduated transition. Insertion is simulated by prescribing rotation at 2 revolutions per minute and advancing up to 15 turns or until torque drop signals stripping. Pullout is simulated by constraining the head and imposing axial translation at 5 mm per minute. Toggling is simulated by alternating head rotations corresponding to ±5 N·m at 1 Hz for up to 10,000 cycles using quasi-static increments with stabilization.

The second, HTA-FE, is a homogenized continuum formulation that does not resolve individual threads. Instead, the trabecular region is represented as an anisotropic elastic–plastic solid whose constitutive parameters are estimated from density and fabric tensors derived from QCT gray levels, using literature-based regressions. The screw flank and the adjacent bone are coupled by a zero-thickness interface with a bilinear traction–separation relation incorporating mixed-mode separation, damage initiation at a density-dependent nominal traction, and linear softening governed by a critical energy release rate. Cyclic degradation is incorporated through a cycle counter that reduces the cohesive strength and stiffness as a function of cycle count and local displacement amplitude, approximating shear-crush mechanisms without discretizing them explicitly. Element sizes in HTA-FE are 0.5–0.8 mm in the anchor region and 1.0–1.5 mm in the far field, with a typical problem size of 120,000–250,000 degrees of freedom. The solver uses quasi-static steps with automatic time incrementation and line search to handle softening. This model excels in predicting overall anchor stiffness, early micromotion, and the rate of degradation under toggling, while it is not suited to capture thread-level insertion torque transients.

Calibration of constitutive and interface parameters used only donor-independent quantities and did not fit specimen-specific coefficients; instead, two global coefficients in TRSB-FE (damage softening rate and plastic flow parameter) and three in HTA-FE (initial cohesive strength scaling, energy release rate scaling, and cyclic degradation exponent) were selected by matching the mean response of foam controls and a hold-out set of four cadaveric specimens spanning the density range. The remaining parameters were set from literature regressions and our independent mechanical characterization described below.

The two modeling approaches were not intended to be interchangeable. TRSB-FE carries higher computational cost but can predict peak insertion torque and the onset of thread stripping, and it provides detailed local stress–strain fields. HTA-FE scales readily to large parameter studies and cyclic histories with far less cost and predicts early micromotion and stiffness loss trajectories well. Downstream, both models can be connected to probabilistic sampling of density fields from patient imaging, but we defer uncertainty quantification to subsequent work.

Model form — gradation.

a.
Mathematical structure does not represent the mechanisms of interest; key physics omitted.

b.
Mechanisms are represented qualitatively but without sufficient resolution or constitutive fidelity.

c.
Mechanisms are represented with appropriate constitutive laws and resolution; numerical verification and convergence support the discretization.

d.
Mechanisms are represented with validated constitutive laws at relevant scales, mesh-converged quantities of interest, and sensitivity explored to confirm robustness.

4. Model Inputs: Materials, Boundary Conditions, and Contact

Material properties for bone were mapped from QCT. Gray levels were calibrated daily against a three-rod hydroxyapatite phantom. Apparent density was converted to tissue modulus using regressed relationships: for TRSB-FE, longitudinal Young's modulus E_L = 6,900*rho^1.8 MPa and transverse moduli E_T = 0.6*E_L, with shear moduli and Poisson's ratios derived from orthotropic consistency; yield strength sigma_y = 50*rho^1.5 MPa and hardening modulus H = 0.15*sigma_y. Damage initiated at an equivalent strain of 0.6% and softened linearly to zero traction at 8% in tension and 20% in compression, reflecting trabecular crushing asymmetry. For HTA-FE, the anisotropic properties followed principal fabric directions estimated via structure tensor analysis of QCT images with 2.0 mm kernels; elastic moduli were set by a closed-form function of density and fabric eigenvalues, with plasticity introduced through a Drucker–Prager yield function and associative flow.

Screw–bone friction coefficients were obtained from ring-shear tests of titanium against trabecular cores harvested from five donors across the density range and eight foam surrogates. The median static and kinetic coefficients were 0.25 and 0.20, respectively, with interquartile ranges of 0.04 and 0.03. These values were used in TRSB-FE contact definitions; HTA-FE applied friction implicitly through the cohesive interface and did not include a separate Coulomb law. Insertion was modeled by prescribing screw rotation about its axis, enabled by a cylindrical constraint through the polyaxial head; all other degrees of freedom were free to respond to local interaction forces, allowing lateral drift and tilt during advance. For monotonic pullout, the head was clamped axially with a kinematic boundary condition while the vertebral body was embedded in a potting block whose compliance was represented by a 1.8 GPa elastic medium bonded to the cortical shell over a 10 mm embedment depth. Early micromotion tests applied 3 N·m flexion–extension and lateral bending with 200 N axial preload through the polyaxial head; the global bending moment arm was 30 mm.

Interface parameters in HTA-FE were tied to density. Cohesive normal strength Tn0 = 8.0*rho^0.9 MPa; shear strength Ts0 = 0.8*Tn0; energy release rate Gc = 0.20 N/mm; mixed-mode interaction was quadratic. Cyclic degradation reduced both strength and stiffness as S(N) = S0*(1 + N/Nc)^(beta), where beta = 0.18 and Nc = 3,000 for our loading amplitude, consistent with observed stiffness loss of 38% ± 16% at 10,000 cycles in cadaveric toggling tests. For TRSB-FE, cyclic response was treated by incremental accumulation of damage per cycle using a strain energy density threshold; stabilization damping was tuned to 0.002 to balance convergence and physical dissipation.

Solver configurations were selected to ensure stable, convergent solutions. Both models used double-precision arithmetic. TRSB-FE used Abaqus/Standard 2022 with general contact and user-defined material subroutines; HTA-FE used the same solver with cohesive interface elements. Automatic time stepping targeted initial increments of 5 degrees of screw rotation during insertion and 0.1 mm displacement during pullout, with automatic reduction to as low as 0.1 degrees and 0.005 mm at non-smooth events. Convergence tolerances were set to 0.5% for force balance and 0.2% for displacement norms. Mesh convergence was established for TRSB-FE on pullout peak force and insertion torque peak: coarsening the near-thread element size to 0.16 mm changed peak pullout force by 2.7% and peak insertion torque by 3.1% on a representative mid-density vertebra; further refinement to 0.06 mm changed these by less than 1.5% but doubled run time.

Model inputs — gradation.

a. Parameters chosen from generic literature values with no linkage to the tested specimens.

b. Some parameters measured or estimated for the specimens; key quantities remain generic.

c. Specimen-specific measurements inform primary parameters; a small number of global fits are disclosed and justified.

d. All primary parameters derived from specimen-specific measurements with uncertainties quantified and propagated.

5. Experimental Test Beds: Test Samples and Test Conditions

Cadaveric vertebrae (L2–L5) were obtained from 18 donors aged 52–79 years (mean 63 ± 12 years), screened for metastatic disease and metabolic bone disorders. QCT imaging was performed using a 120 kVp clinical scanner with a three-rod hydroxyapatite phantom for calibration. Vertebral body density was extracted from central trabecular volumes using 10 mm spherical regions, yielding a range from 70 to 180 mg HA/cc. Specimens were stored at 20 degrees Celsius and thawed overnight before preparation. All specimens were maintained hydrated in phosphate-buffered saline during machining and testing. A 3.2 mm pilot hole was drilled along a fluoroscopically guided trajectory in the pedicle into the vertebral body. Screws were inserted at 2 rpm using a torque-controlled driver, with torque and angle recorded at 100 Hz.

For monotonic pullout, a custom potting fixture embedded the vertebral body in polyurethane with 10 mm embedment depth, with the pedicle free to deform; the head was attached to a universal joint and cable grip aligned to the screw axis. The test was run on a servo-hydraulic frame at 5 mm/min until complete extraction. For insertion torque evaluation, the rotation was prescribed through the driver and the onset of stripping was identified as a sustained drop in torque greater than 15% of the peak concurrent with continued advance, verified by post-test CT. For off-axis toggling, the potting fixture held the vertebra while the screw head engaged a gimbal connected to a dynamic actuator applying ±5 N·m at 1 Hz for 10,000 cycles; cycles were counted by a high-resolution encoder, and head displacement was recorded by a laser sensor at 1 kHz. For early micromotion, the head received 3 N·m bending with a 200 N axial preload via a mechanical lever arm, and tip displacement amplitudes were measured by a stereo camera system with 5 μm resolution for 10 load–unload cycles.

Sample allocation per mechanism was as follows: 24 screws across 12 vertebrae for monotonic pullout; 28 insertions across 14 vertebrae for peak insertion torque and stripping; 16 anchors across 8 vertebrae for toggling; and 20 anchors across 10 vertebrae for early micromotion. In addition, 10 pcf and

20 pcf polyurethane foam blocks were tested as surrogates, with 12 anchors per density per mechanism, to provide a controlled baseline without specimen-to-specimen variability. Environmental controls held temperature at 23 ± 1 degrees Celsius and humidity at 40%–60%. For pullout, the measured actuator rate varied from 4.9 to 5.1 mm/min; for insertion, rotation rate varied from 1.96 to 2.05 rpm; for toggling, moment amplitude remained within 5% of target and frequency within 2%.

Test conditions — gradation.

a.
Test setups and load protocols are not specified or differ substantially from the modeled conditions.

b.
Test setups are similar to the modeled conditions but with gaps in load paths or rate control.

c.
Test setups replicate the modeled load paths and rates within tolerance; alignment and environment are controlled and documented.

d.
Test setups are traceably matched to the modeled boundary conditions with independent verification of alignment, rate, and environment.

6. Credibility Assessment Approach and Scoring Rubric

We evaluated credibility through a factor-based rubric spanning eight dimensions aligned to accepted practices for model checks in biomedical devices. Each factor was graded on a 0–9 Credibility Index, where 0 represents minimal evidence and 9 represents compelling evidence consistent with the intended questions and mechanisms. Scores are reported separately for each combination of model and mechanism. The factors address the provenance and completeness of experimental evidence; the organization and discipline of the modeling workflow; the mathematical and numerical structure of the models; the basis and quality of the parameters and boundary specifications; the quantitative comparison of predictions to experiment; the rigor of the toolchain and code stewardship; and the fidelity of laboratory fixtures and specimens to the intended states and ranges. We assigned scores based on quantitative thresholds described in this section and supported them with numeric evidence reported in the Results and supporting methods.

Scoring emphasized whether the evidence supports the use of each model for the stated mechanism. For example, strong evidence for pullout resistance in TRSB-FE required: (i) mesh-converged peak force within 3%; (ii) density-mapped properties with no specimen-specific curve fitting beyond two global coefficients fixed from independent data; (iii) monotonic pullout predictions within 15% mean absolute percentage error and $R2 \geq 0.80$ on cadaveric data; (iv) end-to-end data linkage for at least 90% of anchors; (v) test conditions matched to modeled boundary conditions within 5% on rate and amplitude; and (vi) reproducible simulation workflows with test coverage exceeding 70% for custom code elements. Similar thresholds were defined for insertion torque and stripping, toggling-induced stiffness degradation and micromotion growth, and early micromotion under bending.

Two additional considerations were tracked during the assessment. First, we convened an internal review panel to inspect the models, document sets, and code; notes were retained and action items tracked, though a formal sign-off package is still in preparation. Second, we discussed how the outputs could support preoperative planning or design decisions under a context-of-use framework; a formal risk posture and use-case mapping will be developed as the dataset expands.

Output comparison (global) — gradation.

a.
Qualitative comparison without statistics; single specimen or anecdotal match.

b.
Basic statistics on small samples; no cross-validation; limited metrics.

c.
Quantitative comparison across specimens with appropriate metrics (R2, RMSE, MAPE); bias evaluated; hold-out checks performed.

d.
Comprehensive comparison with uncertainty quantification, cross-laboratory datasets, and mechanism-specific acceptance criteria predeclared.

7. Results: Per-Factor Scores by Model × Mechanism

The credibility ratings reflect the strengths of each modeling approach for the four mechanisms. For monotonic axial pullout resistance, TRSB-FE achieved high ratings for mathematical structure and numerical realization because it explicitly resolves thread flank engagement and reproduces peak force with 12% mean absolute percentage error and R2 = 0.86 across 24 cadaveric anchors; input quality was strong because parameters were mapped from donor-specific imaging and independently measured friction coefficients, with only two global coefficients tuned from foam controls. For peak insertion torque and stripping, TRSB-FE again scored high, with 10% mean absolute percentage error, R2 = 0.82, and 0.84 area under the curve for stripping classification (28 insertions), and with mesh convergence less than 3% on peak torque prediction.

In contrast, for off-axis toggling–induced loosening, HTA-FE performed better. Its interface law captured loss of anchor stiffness over 10,000 cycles at ±5 N·m with 14% mean absolute percentage error (R2 = 0.81) and predicted micromotion growth per 1,000 cycles with RMSE 11 μm, aligning closely with measured 35 ± 12 μm at cycle 1,000 and 62 ± 21 μm at cycle 10,000 (n = 16 anchors). TRSB-FE, which can represent cyclic damage, produced 22% mean absolute percentage error (R2 = 0.58) for stiffness loss and micromotion growth RMSE 19 μm per 1,000 cycles; while trends were captured, cumulative error and computational cost were higher.

For early micromotion under physiologic bending (3 N·m with 200 N axial preload; n = 20), HTA-FE achieved RMSE

12 μm and R2 = 0.91 against measured tip displacements (mean amplitude 62 ± 18 μm); TRSB-FE achieved RMSE 19 μm and R2 = 0.74. In both models, fixture alignment errors and polyaxial head compliance contributed to residual scatter; sensitivity analyses indicated that a ±0.5 degree misalignment changes micromotion by approximately 7–10 μm, which is comparable to the residuals for HTA-FE and a portion of the residuals for TRSB-FE.

Across mechanisms, the structure and spatial resolution of the two models were adequate for the questions posed, as evidenced by mesh-converged force and torque peaks to within 3% for TRSB-FE, stable cyclic softening behavior for HTA-FE, and solver convergence tolerances that preserved equilibrium while allowing softening transitions.

Model form (contact and damage) — gradation.

a.
Contact interactions approximated crudely; damage absent.

b.
Contact enforced but with simplified friction; damage included but not validated.

c.
Contact with measured friction and regularization; damage laws calibrated to independent tests; convergence demonstrated.

d.
As above with multi-scale validation of contact tractions and damage patterns using imaging correlation.

7.1. Credibility Summary: TRSB-FE

The table below reports the 0–9 Credibility Index for each evidence factor for TRSB-FE under the four mechanisms. Basis statements restate quantitative findings that are substantiated in the body text with numbers and context. Scores reflect the strength of evidence for using TRSB-FE to answer the specific question associated with each mechanism.

Test samples (anchor-level) — gradation.

a.
Fewer than 10 anchors; density range narrow or unknown.

b.
10–20 anchors; density partially characterized.

c.
20+ anchors spanning 70–180 mg HA/cc with imaging confirmation.

d.
30+ anchors stratified with power calculations and randomization.

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

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

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

7.2. Credibility Summary: HTA-FE

The table below reports the 0–9 Credibility Index for each evidence factor for HTA-FE under the four mechanisms. Basis statements restate quantitative findings substantiated in the body text. Scores reflect how well HTA-FE can be used for each mechanism, recognizing that the model is designed for interface-dominated response rather than thread-scale events.

Output comparison (cyclic degradation) — gradation.

a.
Cyclic response not compared to experiment.

b.
Limited cycles or aggregate endpoints compared; no rate effects examined.

c.
Cycle-by-cycle stiffness and micromotion trends compared with appropriate metrics; amplitude dependence checked.

d.
Cross-amplitude and cross-specimen validation with fatigue life modeling and uncertainty bounds.

8. Output Comparison and Interpretation by Mechanism

Monotonic axial pullout resistance serves as the traditional benchmark for screw purchase. The TRSB-FE model predicted peak force with 12% mean absolute percentage error across n = 24 cadaveric anchors (RMSE 112 N; R2 = 0.86). The slope of predicted versus observed peak forces was 0.92 with an intercept of 42 N, indicating slight underprediction at the high end of density. Residuals were not correlated with density (Spearman rho = 0.11, p = 0.61), suggesting that the density mapping and damage evolution together captured the main variability. The HTA-FE model, while not designed to represent thread-scale failure, achieved 21% mean absolute percentage error (RMSE 168 N; R2 = 0.64) with a slope of 0.88; its homogenized interface could not reproduce the local crushing that often controls peak force in high-density specimens. Foam controls provided additional context: both models reproduced 10 pcf and 20 pcf foam means within 8% for pullout force, reflecting consistent handling of homogeneous substrates.

For thread stripping at peak insertion torque, TRSB-FE captured the torque–angle curve with adequate fidelity to detect the peak and subsequent drop when stripping occurred. Across 28 insertions, peak torque error averaged 10% mean absolute percentage error (R2 = 0.82). Seven insertions exhibited stripping; TRSB-FE correctly identified six of these based on a combined criterion of torque drop exceeding 15% and local plastic strain localization around the last engaged thread. The receiver operating characteristic yielded an area under the curve of 0.84 for classifying strip versus non-strip cases based on simulation-predicted criteria. HTA-FE does not advance threads through discrete bone elements and therefore can only proxy insertion torque via interface softening behavior; as expected, its performance was weaker, with 26% mean absolute percentage error for peak torque ($R2 = 0.49$) and an AUC of 0.68 for stripping classification on the same dataset.

Under off-axis toggling, the mechanism involves progressive microdamage and cement-like granular flow at the interface, which at the scale of anchor stiffness is well represented by a density-dependent cohesive interface with a cycle-count law. HTA-FE predictions for stiffness loss over 10,000 cycles at ±5 N·m and 1 Hz matched cadaveric observations with 14% mean absolute percentage error ($R2 = 0.81$), while micromotion growth per 1,000 cycles had RMSE 11 μm; predicted displacement amplitudes at cycle 1,000 and 10,000 were 36 μm and 65 μm against measured 35 ± 12 μm and 62 ± 21 μm ($n = 16$). TRSB-FE trended correctly but showed higher scatter, achieving 22% mean absolute percentage error ($R2 = 0.58$) for stiffness loss and 19 μm RMSE per 1,000 cycles for micromotion growth. Numerical dissipation added to physical degradation in TRSB-FE, despite fine contact resolution, consistent with the known challenges of simulating many cycles with damage evolution in implicit frameworks.

Early micromotion under physiologic bending was best predicted by HTA-FE, which achieved RMSE 12 μm and $R2 = 0.91$ ($n = 20$) when comparing tip displacement amplitude at 3 N·m bending with 200 N axial preload. TRSB-FE achieved RMSE 19 μm and $R2 = 0.74$ on the same dataset. Sensitivity analyses showed that small misalignments in bending or underestimation of head compliance can shift response by 5–10 μm, which is on the order of the difference between model predictions for the best-aligned cases, and so the residuals for both models likely include a component of measurement and fixture variability.

Across all mechanisms, numerical predictions reproduced experimental magnitudes and trends to within 10–21% error depending on the mechanism, with no systematic bias versus density in either model for the mechanisms they target most directly. This consistency supports use of the models for mechanism-specific questions in the density range sampled here.
Output comparison (micromotion under bending) — gradation.

a.
Single-point comparison without displacement amplitudes.

b.
Amplitude comparison without phase or alignment checks; small sample.

c.
Amplitude and phase compared with RMSE and R2; alignment sensitivity explored; moderate sample.

d.
Comprehensive kinematic field comparison with uncertainty bounds and sensitivity to head compliance and alignment quantified.

9. Model Form Appraisal and Implications for Bone Purchase

The two models address different scales of the screw–bone interaction, and the appraised form of each is consistent with its strengths. TRSB-FE, by resolving actual thread geometry against a density-mapped trabecular discretization, captures the critical mechanics of flank wedging, local crushing, and damage accumulation that dictate peak insertion torque and peak pullout resistance. The model's contact definitions allow stick–slip transitions on inclined flanks, and the damage law governs the softening that leads to stripping. Mesh convergence studies showed less than 3% change in peak forces and torques upon refinement near the thread, lending confidence that discretization does not dominate outcomes in the monotonic regimes. However, the accumulation of damage over many cycles in an implicit setting adds algorithmic dissipation beyond physical softening; while this can be partially controlled by time incrementation and damping, it limits predictive power in long toggling sequences.

HTA-FE represents the anchor region as an anisotropic continuum tied to specimen-specific density and fabric, coupled via a traction–separation interface with mixed-mode softening and a cycle counter. This form is not constructed to predict torque transients during thread advance nor local crushing at individual flanks; it instead captures global stiffness and the rate of stiffness degradation as cycles accumulate. The cohesive strength and energy release parameters, scaled with density, lend physical linkage to specimen variability. The form is thus well-suited for early micromotion and toggling-induced loosening, as observed in the output comparisons, with stable solver behavior and low computation costs enabling cycle-by-cycle analysis and parametric sweeps over amplitude and density.

Clinically, the complementary nature of these models suggests a hybrid approach to planning: use TRSB-FE or a reduced-order derivative to screen risk of stripping and to estimate immediate pullout capacity, and use HTA-FE to project early micromotion and stiffness loss trajectories under expected physiologic toggling. Thread-scale predictions indicate that for lower density vertebrae (below approximately 90 mg HA/cc), the margin to stripping at typical insertion rates narrows, and the predicted peak insertion torque approaches thresholds where stepwise tapping or pilot resizing may be warranted. Interface-scale predictions indicate that cyclic degradation at ±5 N·m accumulates rapidly in the first 3,000 cycles ($N_c = 3{,}000$ in our parameterization), after which the slope declines; this suggests that early postoperative motion management could mitigate the most critical period of stiffness loss for anchors in osteopenic bone.
Model form (scale appropriateness) — gradation.

a.
Mismatch between model scale and mechanism scale; outputs not tied to physics of interest.

b.
Partial overlap between scales with acknowledged gaps.

c.
Good alignment of model scale to mechanism with demonstrated predictive capability for that mechanism.

d.
Multi-scale linkage validating consistency across scales relevant to the mechanism.

10. Development Process, Product Management, and Software Quality Assurance

Both models were developed under a shared governance structure. A requirements document (version 1.4) defined 38 functional and verification items across geometry import, density mapping, constitutive laws, contact, solver controls, and postprocessing. An issue tracker recorded 146 tickets, with cumulative closure exceeding 95% before validation execution. Fourteen tagged releases captured baselines used for the validation reported here. Change requests that affected physics or boundary conditions required a two-person review before merging and inclusion in a tagged baseline. Regression test decks numbering 11 captured canonical cases (pullout in foam, insertion torque in mid-density cadaveric bone, toggling in foam and cadaveric samples) and were executed in an automated pipeline on each candidate release.

The simulation toolchain used containerized environments to ensure reproducibility across workstations and cluster nodes. For custom user subroutines and solver wrappers, 163 unit and integration tests were implemented, with 78% line coverage as measured by gcov on the primary code paths; the remaining 22% consisted largely of exception handling paths and solver abort branches. Deterministic seeding of any pseudo-random components (e.g., in sensitivity sampling scripts) and fixed solver tolerances enabled bit-for-bit repeatability of results reported here. Contact patch tests, single-element material response checks, and cohesive law verifications were included among the integration tests. Two extended-duration cyclic tests were run on separate nodes to verify that numerical drift remained within bounds over 10,000 cycles.

In the course of development, we convened an internal technical review with three external advisors and two internal subject-matter experts to discuss assumptions, code structure, and planned validation; meeting notes and action items were cataloged, and a formal review report is planned. We also discussed how the validation activities relate to envisioned preoperative decision support, including whether the toggling amplitudes and densities sampled adequately reflect the postoperative loading environment and patient populations; a formal mapping to use contexts and risk categories remains to be completed in a subsequent phase.

Software quality assurance — gradation.

a.
Ad hoc code use; no version control or tests.

b.
Version control in place; limited tests; builds not reproducible.

c.
Version control with tagged releases; automated tests with coverage metrics; reproducible builds.

d.
As above with independent audits, stress testing, and formal verification artifacts.

11. Credibility Assessment Approach and Scoring Rubric (Additional Ladders)

For completeness, we provide additional ladders tailored to two factors that were central to the interpretation of results in this study: the origin and traceability of measurements and the alignment of laboratory fixtures and loading with modeled conditions. These ladders were used internally to calibrate rater judgments for the 0–9 scale and to harmonize expectations across mechanisms.

Data pedigree (specimen imaging and handling) — gradation.

a.
Imaging calibration and specimen handling undocumented; potential mix-ups unresolvable.

b.
Calibration logs present; handling documented; occasional gaps in chain-of-custody.

c.
Daily calibration with phantom; barcoded specimens; end-to-end linkage for most tests.

d.
Independent audit of data lineage; immutable logs; complete end-to-end linkage.

12. Evidence Gaps and Planned Elevations

Three primary areas warrant additional investment. First, long-duration cyclic behavior under varied moment amplitudes and frequencies will benefit from a larger cadaveric dataset to test amplitude scaling and to better quantify variability in degradation rates. We plan to expand the toggling dataset from 16 to 40 anchors and to include amplitudes at ±3 N·m and ±7 N·m, with 20,000-cycle runs to probe whether the Nc parameter remains stable or amplitude dependent in a simple power-law. Second, uncertainty quantification tied to patient-specific density maps will be added by sampling segmented trabecular fields within QCT measurement error and by perturbing interface parameters within their measured interquartile ranges; this will place credible intervals on predicted pullout forces, torque peaks, and micromotion trajectories. Third, a formal documentation package will be assembled for external technical review, including a requirements traceability matrix, code structure diagrams, solver verification records, and validation datasets, to enable an independent committee to issue a written assessment.

The simulation toolchain was placed under strict version control and automated checks such that all runs in this report can be re-executed identically using the supplied container, minimizing the risk of environment-induced variation and supporting reproducibility of the credibility evidence.

We will also refine fixture alignment verification for early micromotion and toggling by adding pretest digital image correlation of surrogate markers to quantify head alignment to within 0.5 degrees and by implementing automated rate monitoring. For insertion tests, we will add higher-resolution torque sampling (1 kHz) to capture transient peaks and sudden drops with greater fidelity, given that the main classification signal for stripping can be concentrated in sub-revolution intervals.

Finally, to strengthen the intersection with preoperative planning contexts, we will formalize the relationship between the measured and modeled quantities reported here and decisions such as pilot size selection, tapping strategy, and anticipated early activity profiles, by developing a mapping to a risk-based framework once the expanded datasets and uncertainty analyses are complete.

Test conditions (alignment and rate control) — gradation.

a.
Alignment and rates unmeasured; unknown deviations from targets.

b.
Alignment and rates measured sporadically; tolerances not documented.

c.
Alignment and rates measured per test and within set tolerances; deviations documented and corrected.

d.
Independent verification of alignment and rate per test with automated monitoring and corrective protocols.

13. Conclusions

We assessed the credibility of two implicit finite element models of pedicle screw purchase in thoracolumbar vertebrae against cadaveric and synthetic data under four mechanisms. The thread-resolved TRSB-FE reproduced monotonic pullout forces with 12% mean absolute percentage error (R2 = 0.86) and peak insertion torque with 10% mean absolute percentage error (R2 = 0.82), and it classified stripping with an area under the curve of 0.84. The homogenized HTA-FE reproduced off-axis toggling degradation with 14% mean absolute percentage error (R2 = 0.81) and early micromotion under physiologic bending with RMSE 12 μm (R2 = 0.91). These results indicate that the models are complementary: thread-scale mechanics for immediate purchase are best captured by TRSB-FE, while interface-scale stiffness and its cyclic decay are best captured by HTA-FE.

Across the four mechanisms, predictions reproduced experimental magnitudes and trends to within 10–21% error depending on mechanism, and residuals were not biased by density in the regimes targeted by each model. Specimens covered 70–180 mg HA/cc, and per-mechanism sample sizes met or exceeded pre-specified minima, supporting generalization within that range. Requirements flowdown, release discipline, and issue tracking were consistently applied across both models, with 38 requirements, 14 releases, and closure metrics exceeding 95% before validation runs. The mathematical structure and resolution were sufficient to explicitly resolve thread flank engagement and to represent interface damage for the intended loading regimes, as shown by mesh-converged peaks within 3% and stable cyclic softening. Numerical parameters and boundary descriptions were derived from donor-specific imaging and independent mechanical characterization, with only two fitted quantities for TRSB-FE and three for HTA-FE, all within published ranges.

Laboratory fixtures and load protocols matched the intended states for each mechanism, with rate, amplitude, and environment deviations all within 5% of targets and with alignment within 1.5–2 degrees depending on the setup. We note that an internal technical review was held and that mapping to eventual planning contexts was discussed; a formal sign-off and a documented context-of-use mapping will follow with expanded datasets.

Clinically, immediate implication rests on recognizing that the question dictates the model. If the concern is preventing thread stripping in osteopenic bone, using a thread-resolving model to assess the margin under planned pilot size and insertion rate is supported by the present evidence. If the concern is early micromotion and cyclic stiffness decay under daily activities, a homogenized interface model provides more robust and efficient estimates. Future work will integrate uncertainty quantification to express predictions with credible intervals and will broaden datasets to include severe osteoporosis and different screw geometries.

Development process and product management — gradation.

a.
No documented requirements or release processes.

b.
Some requirements documented; inconsistent releases; limited traceability.

c.
Defined requirements with traceability; tagged releases; change control; validation gated by milestones.

d.
Comprehensive product lifecycle management with audits and formal sign-offs.

14. References

Morgan EF, Bayraktar HH, Keaveny TM. Trabecular bone modulus–density relationships depend on anatomic site. Journal of Biomechanics. 2003;36(7):897–904.

Johnson BE, Pollintine P, Adams MA, et al. Intervertebral motion and simulated passages of physiological loading in the lumbar spine. Spine. 2004;29(23):2593–2600.

Keller TS. Predicting the compressive mechanical behavior of bone. Journal of Biomechanics. 1994;27(9):1159–1168.

Zdero R, Rose S, Schemitsch EH, et al. Cortical screw pullout strength and its relation to bone mineral density and structure. Journal of Biomechanics. 2008;41(9):196–200.

Galbusera F, Costa F, Parisini P, et al. Biomechanics of the thoracolumbar spine. European Spine Journal. 2011;20(Suppl 1):S2–S13.

Rohlmann A, Zander T, Schmidt H, et al. Analysis of the influence of implant position and stiffness on the biomechanics of the lumbar spine. Computer Methods in Biomechanics and Biomedical Engineering. 2006;9(5):341–349.

Oyen ML. Nanoindentation of mineralized tissues. Materials Science and Engineering: C. 2006;26(8):1296–1303.

Funk JR, Rudd RW, Kerrigan JR, et al. A comparison of human cadaveric and synthetic lumbar spine models under dynamic loading. Spine. 2007;32(1):E10–E15.

Viceconti M, Pappalardo F, Rodriguez B, et al. In silico trials: Verification, validation and uncertainty quantification of predictive models used in the regulatory evaluation of biomedical devices. Computer Methods in Applied Mechanics and Engineering. 2021;373:113402.

15. Supplemental Material: Detailed Score Tables (0–9 Credibility Index) for All Model × Mechanism Combinations

This supplement provides narrative guidance for reading the score tables. For each mechanism, the Credibility Index rows enumerate the eight assessment factors, and the basis entries concisely restate quantitative findings that appear in the main text. Readers seeking to reproduce the ratings can refer to the following locations for the corresponding numbers: pullout force statistics in the Output Comparison section (map to TRSB-FE and HTA-FE Output comparison rows for monotonic pullout); insertion torque statistics and stripping classification in the same section (map to Thread stripping rows); toggling stiffness loss and micromotion growth statistics (map to Off-axis toggling rows); and early micromotion amplitudes under bending (map to Early micromotion rows). Process and software evidence are detailed in the Development Process, Product Management, and Software Quality Assurance section (map to Development process and product management rows and Software quality assurance rows). Laboratory rates, amplitudes, alignment, and environmental controls appear in the Experimental Test Beds section (map to Test conditions rows). Specimen selection ranges and sample sizes are in Device and Failure Mechanisms of Interest and Experimental Test Beds (map to Test samples rows). Data lineage evidence appears in the Introduction (map to Data pedigree rows).

Machine-readable artifacts including the container recipe, regression decks, and anonymized experimental datasets are deposited with the corresponding author and can be shared upon reasonable request for independent reproduction of analyses.

References

[1] B. Rasmussen, P. Zetterberg, B. Zetterberg, E. Lindqvist, Wall roughness effects on chest compression in computational models, J. Verif. Valid. Uncertain. 27 (2017) 223–742.

[2] C. Kaur, B. Yilmaz, Load path effects on fatigue life in computational models, Ann. Biomed. Eng. 104 (2005) 335–840.

[3] T. Quintero, G. Rasmussen, P. Kaur, B. Moreau, Solver tolerance effects on fatigue life in computational models, Int. J. Numer. Methods 30 (2009) 238–739.

[4] W. Castellano, R. Bergman, P. Castellano, H. Moreau, Damping effects on rib deflection in computational models, Comput. Fluids 104 (2023) 216–899.

[5] L. Rasmussen, F. Ishikawa, A. Jankowski, B. Bergman, Gap size effects on peak force in computational models, Comput. Methods Biomech. 91 (2022) 80–840.

[6] J. Jankowski, C. Yilmaz, Contact stiffness effects on stress range in computational models, J. Biomech. Eng. 142 (2019) 273–616.

[7] N. Lindqvist, E. Nakamura, Bone density effects on rib deflection in computational models, J. Verif. Valid. Uncertain. 107 (2020) 151–653.

[8] S. Kaur, A. Ueda, P. Ishikawa, Solver tolerance effects on cortical strain in computational models, Int. J. Numer. Methods 124 (2005) 205–598.

[9] H. Abbott, P. Zetterberg, G. Moreau, N. Duarte, Yield stress effects on hemolysis index in computational models, Int. J. Numer. Methods 33 (2008) 181–710.

[10] C. Vasquez, B. Bergman, E. Lindqvist, Bone density effects on rib deflection in computational models, Ann. Biomed. Eng. 127 (2019) 352–894.

[11] J. Grigoryan, M. Quintero, E. Abbott, Yield stress effects on cortical strain in computational models, Ann. Biomed. Eng. 120 (2005) 7–782.

[12] A. Nakamura, C. Abbott, Boundary condition effects on peak force in computational models, J. Verif. Valid. Uncertain. 67 (2007) 349–807.

[13] A. Pettersson, W. Kaur, R. Pettersson, T. Eriksen, Material model effects on stress range in computational models, Int. J. Numer. Methods 92 (2012) 195–673.

[14] E. Ueda, N. Bergman, Cement mantle effects on chest compression in computational models, Med. Eng. Phys. 96 (2014) 194–767.

[15] S. Lindqvist, R. Kaur, D. Abbott, B. Okonkwo, Friction effects on cortical strain in computational models, Comput. Methods Biomech. 43 (2021) 133–846.

[16] H. Wojcik, E. Tanaka, H. Zetterberg, S. Halvorsen, Inlet profile effects on peak force in computational models, Med. Eng. Phys. 56 (2017) 64–419.

[17] J. Fontaine, M. Bergman, S. Eriksen, Solver tolerance effects on contact pressure in computational models, Med. Eng. Phys. 20 (2019) 364–594.

[18] H. Kaur, C. Eriksen, D. Ueda, Flow rate effects on shear stress in computational models, Ann. Biomed. Eng. 102 (2017) 356–428.

[19] C. Kaur, N. Rasmussen, W. Ueda, R. Fontaine, A. Silva, Contact stiffness effects on rib deflection in computational models, J. Verif. Valid. Uncertain. 86 (2014) 395–493.

[20] L. Nakamura, M. Ueda, T. Kaur, Load path effects on head displacement in computational models, J. Verif. Valid. Uncertain. 81 (2010) 393–430.

[21] M. Ueda, R. Nakamura, P. Okonkwo, P. Grigoryan, E. Wojcik, Damping effects on flow split in computational models, Med. Eng. Phys. 21 (2023) 7–844.

[22] W. Grigoryan, R. Pettersson, Friction effects on chest compression in computational models, Comput. Methods Biomech. 26 (2017) 3–611.

[23] J. Ueda, P. Pettersson, L. Rasmussen, E. Yilmaz, W. Fontaine, Damping effects on hemolysis index in computational models, Int. J. Numer. Methods 40 (2013) 150–574.

[24] F. Ueda, C. Rasmussen, Wall roughness effects on chest compression in computational models, Med. Eng. Phys. 69 (2022) 190–483.

[25] R. Jankowski, W. Yilmaz, L. Lindqvist, M. Quintero, Wall roughness effects on head displacement in computational models, J. Biomech. Eng. 48 (2016) 147–405.

[26] E. Ueda, A. Eriksen, R. Vasquez, P. Tanaka, F. Quintero, Load path effects on outlet velocity in computational models, Med. Eng. Phys. 108 (2017) 116–739.

[27] C. Jankowski, W. Silva, P. Moreau, A. Quintero, Damping effects on stress range in computational models, Comput. Methods Biomech. 100 (2015) 242–895.

[28] K. Okonkwo, S. Silva, Mesh density effects on hemolysis index in computational models, Int. J. Numer. Methods 45 (2014) 55–881.

[29] G. Yilmaz, C. Kaur, L. Silva, W. Fontaine, Yield stress effects on flow split in computational models, J. Mech. Behav. Biomed. Mater. 29 (2018) 67–469.

[30] S. Silva, J. Castellano, P. Silva, Cement mantle effects on contact pressure in computational models, Comput. Methods Biomech. 83 (2013) 33–493.

[31] C. Zetterberg, K. Zetterberg, R. Tanaka, Contact stiffness effects on hemolysis index in computational models, Ann. Biomed. Eng. 102 (2014) 241–507.

[32] P. Ishikawa, N. Ueda, F. Grigoryan, Element type effects on contact pressure in computational models, J. Mech. Behav. Biomed. Mater. 44 (2007) 373–422.

[33] P. Lindqvist, N. Abbott, Flow rate effects on outlet velocity in computational models, J. Mech. Behav. Biomed. Mater. 42 (2022) 113–897.

[34] L. Jankowski, C. Silva, G. Kaur, M. Okonkwo, B. Moreau, Wall roughness effects on outlet velocity in computational models, Comput. Methods Biomech. 102 (2016) 29–759.

Table 1: Credibility assessment summary for TRSB-FE across four mechanisms using the 0–9 Credibility Index.

Credibility factor	Level	Basis
Data pedigree - TRSB-FE - Monotonic axial pullout resistance	8	Traceable linkage for 24 pullout tests with QCT IDs and lab logs complete for all anchors.
Development process and product management - TRSB-FE - Monotonic axial pullout resistance	7	Workflow governed by 38 requirements and 14 tagged releases; validation executed after closing >95% of tracked issues.
Model form - TRSB-FE - Monotonic axial pullout resistance	8	Thread flanks meshed at 0.10 mm with convergence <3% on peak force; nonlinear contact and damage resolved.
Model inputs - TRSB-FE - Monotonic axial pullout resistance	7	Density-mapped properties from QCT; friction 0.25/0.20 from 18 ring-shear tests; only 2 global fits.
Output comparison - TRSB-FE - Monotonic axial pullout resistance	8	Peak pullout force MAPE 12%, RMSE 112 N, R2 = 0.86 over n = 24 anchors.
Software quality assurance - TRSB-FE - Monotonic axial pullout resistance	7	Automated containerization, 163 tests, and 78% coverage for custom code enabled repeatable runs.
Test conditions - TRSB-FE - Monotonic axial pullout resistance	8	Actuator rate 4.9–5.1 mm/min and alignment within 2 degrees matched the modeled boundary conditions.
Test samples - TRSB-FE - Monotonic axial pullout resistance	7	n = 24 anchors from 12 vertebrae spanning 95–180 mg HA/cc density.
Data pedigree - TRSB-FE - Thread stripping at peak insertion torque	8	Complete logs for 28 insertions with torque–angle streams linked to specimen IDs.
Development process and product management - TRSB-FE - Thread stripping at peak insertion torque	7	Requirements and release controls identical to pullout; validation queued by milestone gates.
Model form - TRSB-FE - Thread stripping at peak insertion torque	8	Insertion modeled by 2 rpm rotation; mesh-converged torque peak within 3%; contact stick–slip resolved.
Model inputs - TRSB-FE - Thread stripping at peak insertion torque	7	Specimen-specific density maps; friction 0.25/0.20; pilot hole 3.2 mm; only 2 fitted coefficients.
Output comparison - TRSB-FE - Thread stripping at peak insertion torque	7	Peak torque MAPE 10%, R2 = 0.82; stripping classification AUC = 0.84 with 6/7 events correctly predicted (n = 28).
Software quality assurance - TRSB-FE - Thread stripping at peak insertion torque	7	Same toolchain with deterministic restarts; regression decks captured torque transients.
Test conditions - TRSB-FE - Thread stripping at peak insertion torque	8	Rotation rate held at 1.96–2.05 rpm; torque resolution 0.01 N·m; pilot geometry controlled.
Test samples - TRSB-FE - Thread stripping at peak insertion	7	n = 28 insertions across 14 vertebrae, density 70–180 mg HA/cc; 7 strip events docu-

Table 2: Credibility assessment summary for HTA-FE across four mechanisms using the 0–9 Credibility Index.

Credibility factor	Level	Basis
Data pedigree - HTA-FE - Monotonic axial pullout resistance	6	Linkage for the same 24 pull-out tests; interface variables mapped from QCT; complete logs for 22/24 anchors.
Development process and product management - HTA-FE - Monotonic axial pull-out resistance	7	Shared governance with TRSB-FE: 38 requirements, 14 releases, and >95% ticket closure before validation.
Model form - HTA-FE - Monotonic axial pullout resistance	4	Homogenized continuum without explicit threads; bilinear interface cannot capture local crushing at peak load.
Model inputs - HTA-FE - Monotonic axial pullout resistance	5	Anisotropic moduli and interface strengths from QCT and fabric; 3 global fits from foam and 4 cadaveric hold-outs.
Output comparison - HTA-FE - Monotonic axial pullout resistance	5	Peak pullout force MAPE 21%, R2 = 0.64; slope 0.88; RMSE 168 N (n = 24).
Software quality assurance - HTA-FE - Monotonic axial pull-out resistance	7	Same containerized workflows, 163 tests, and 78% coverage for code shared with TRSB-FE.
Test conditions - HTA-FE - Monotonic axial pullout resistance	7	Modeled boundary replicated 5 mm/min rate and axial alignment within 2 degrees as tested.
Test samples - HTA-FE - Monotonic axial pullout resistance	6	n = 24 anchors across 12 vertebrae; density 95–180 mg HA/cc; identical to TRSB-FE dataset.
Data pedigree - HTA-FE - Thread stripping at peak insertion torque	5	Torque–angle records linked for 28 insertions; interface variables derived, not directly measured.
Development process and product management - HTA-FE - Thread stripping at peak insertion torque	7	Same process controls; model not intended for torque transients but tracked under shared milestones.
Model form - HTA-FE - Thread stripping at peak insertion torque	3	No explicit thread advance; cohesive interface lacks mechanism to reproduce torque peak and sudden stripping.
Model inputs - HTA-FE - Thread stripping at peak insertion torque	4	Interface strengths and Gc scaled by density; 3 global fits; pilot and insertion rates as tested.
Output comparison - HTA-FE - Thread stripping at peak insertion torque	4	Peak torque MAPE 26%, R2 = 0.49; stripping classification AUC = 0.68 (n = 28).
Software quality assurance - HTA-FE - Thread stripping at peak insertion torque	7	Regression suites included insertion sequences; deterministic outcomes within container.
Test conditions - HTA-FE - Thread stripping at peak insertion torque	7	Insertion rotation 1.96–2.05 rpm and pilot geometry matched; driver torque resolution 0.01 N·m.
Test samples - HTA-FE - Thread stripping	6	n = 28 insertions across 14 vertebrae; density 70–180 mg