Credibility Assessment of Implicit FEA Models for Bone Screw Thread Insertion Torque Under Four Mechanisms

Elena M. Vargas, PhD[a], Navid Rahman, PhD[1], Priya Deshmukh, MS[1], Harold C. Lin, PhD[1]

[a]*Center for Orthopedic Simulation and Analytics, Department of Mechanical and Biomedical Engineering, Northbridge University, Northbridge, USA,*

Abstract

We report a structured credibility assessment under 7009A of three implicit finite element analysis (FEA) models that estimate insertion torque versus depth for a self-tapping cortical bone screw. The evaluated models are: (1) an Axisymmetric Equivalent-Thread Model (AETM) that homogenizes thread–bone contact with Coulomb friction and ploughing resistance; (2) a 3D Half-Turn Helical Segment (HTHS) that resolves detailed flank geometry and represents chip formation via a viscoplastic cutting surrogate; and (3) a Multi-Scale Cortical–Cancellous Coupled Model (MS-CCC) that couples a macroscale screw–bone system to an orthotropic damage representative volume element (RVE). Four physical mechanisms are examined: frictional and ploughing interface; pilot hole and cortex tolerance; self-tapping and chip formation; and bone microstructure and damage evolution. Evidence encompasses geometry metrology, tribological tests on cortical bone, microstructure-informed properties, torque–depth records for 24 human diaphyseal insertions and 18 synthetic blocks, and numerical verification of implicit contact and nonlinear material routines. We define a private 0–7 Credibility Units scale with anchors for each factor and grade the eight clear factors per model–mechanism pair, yielding 96 scored entries. Median absolute percentage error (MAPE) against mechanism-specific benchmarks was 12%, 10%, 38%, and 16% for AETM; 11%, 9%, 28%, and 14% for HTHS; and 14%, 12%, 41%, and 13% for MS-CCC across the four mechanisms respectively. Convergence metrics met residual norms below 1e6 (force) and 1e8 (displacement) for all steady-contact cases, with mesh refinement changing peak torque by less than 4% except in self-tapping, where sensitivity rose to 12%. While steady frictional drag and pilot tolerance effects were captured within pre-set comparison targets, transient cutting peaks were not consistently reproduced: HTHS overshot peak torque by 22% with 0.7 mm timing delay, and AETM/MS-CCC missed spike morphology entirely. Provenance of friction and damage inputs was moderate, with coefficients from 42 pin-on-bone trials (μcortex = 0.30 ± 0.05) and damage softening calibrated to 14 donors, but limited inter-laboratory replication. Because the self-tapping mechanism failed to meet comparison criteria and input traceability for friction and damage was insufficient to support the reported scores, the model suite is not accepted for predicting insertion torque for this screw under the declared scope. We identify additions required: a verified chip segmentation formulation, expanded multi-site cutting evidence with quantified uncertainty, and higher-grade tribological and damage data linked to donor phenotypes.

Keywords: bone screw torque, implicit finite element analysis, model credibility

1. Introduction

Insertion torque of cortical bone screws governs fixation security, risk of intraoperative stripping, and postoperative stability. For a self-tapping thread, torque–depth profiles are shaped by interfacial friction and near-surface plasticity, interference set by pilot and cortex geometry, transient chip formation during cutting, and microstructure-driven softening as the tool advances. Within this landscape, implicit finite element analysis (FEA) is widely used to plan designs and predict torque envelopes. Yet model-based decisions depend on whether the models are sufficiently credible for the stated question and context. The present study applies a structured framework consistent with 7009A to three implicit FEA models of a 3.5 mm self-tapping cortical screw, each embodying different abstractions of the screw–bone interaction.

We examine four mechanisms that materially change the torque response: (1) steady interfacial shear and subsurface ploughing that dominate in fully engaged cortices; (2) geometric tolerances of pilot diameter and cortical thickness that alter interference; (3) cutting and chip evacuation that produce transient torque spikes and peak loads; and (4) damage evolution in anisotropic bone that modulates escalation and relaxation with depth. For each mechanism, we bring traceable input sources, describe numerical solution and checks, and compare outputs to mechanism-specific benchmarks. Factors assessed include provenance and traceability of data used to set model constants, how consistently inputs are represented across models and tests, structural adequacy of the equations and discretization, appropriateness of numerical values fed to analyses, checks that implementation and contact algorithms recover known solutions, numerical error control, representativeness of the physical tests, and quantitative agreement to outputs. The study reports all scores on a private 0–7 Credibility Units scale with anchors calibrated to the decision context.

A set of numerical checks was executed prior to comparisons: return-mapping unit tests for elastoplastic routines reproduced manufactured solutions to within 1.2% in stress update norms, frictional contact linearization recovered analytic responses of an elastic cylinder on a rigid plane with errors under 0.7%, and patch tests for axisymmetric and helical meshes retained compatibility to within 0.5% energy error. These results indicate that, within the scope of steady-state contact and small-strain plasticity considered here, the implemented procedures track known responses closely enough to proceed with confidence to validation steps.

The paper is structured to first define the device and measured quantity, then the models, the mechanisms, and the evidence. Each factor in the assessment is argued from primary sources and bench measurements where possible. The outcome is a not-accepted decision for the suite as a predictor of insertion torque because the mechanism involving chip formation does not meet comparison targets and the provenance of certain high-influence inputs is not strong enough to sustain the claimed levels. We also delineate additions required to close those specific gaps.

Implementation checks prior to external comparison — gradation.

a.
Minimal internal checks; no quantitative bounds on deviations from analytic cases.

b.
Unit tests on isolated routines with qualitative pass/fail criteria; limited use of manufactured solutions.

c.
Structured suite including manufactured solutions and contact linearization with stated error tolerances tied to decision needs.

d.
Independent replication of verification results and stress updates, with uncertainty budgets allocated to code-level approximations.

2. Device and Measured Quantity: Self-Tapping Bone Screw Insertion Torque

The device is a 3.5 mm diameter, self-tapping cortical bone screw with a 1.25 mm thread pitch, a 30-degree flank angle, and a single-start helical form. The cutting tip includes two flutes with a rake-like wedge angle of approximately 20 degrees. Dimensional metrology by optical scanning on three manufacturing lots (n = 30 screws total) yielded a mean outer diameter of 3.52 mm (standard deviation 0.02 mm), core diameter of 2.70 mm (0.03 mm), and mean flank roughness Ra of 0.9 μm (0.2 μm). The driver interface is a standard 2.5 mm hex socket. The pilot hole nominal for cortical engagement is 2.50 mm with a manufacturing tolerance ±0.05 mm. Measured cortical thickness in harvested mid-diaphyseal femoral sections used in testing ranged from 2.0 to 3.2 mm, determined by micro-CT (voxel 20 μm) and verified with optical sections post-test.

Insertion torque was recorded with a reaction torque transducer sampled at 2 kHz and calibrated to an expanded uncertainty of 0.2 N·m (k = 2). Torque–depth curves were time-synchronized to optical tracking of the driver axial motion (frame rate 500 fps) and converted to depth by integrating driver rotation with a pitch-corrected helix lead. The test program included human cortical specimens from 14 donors (age 47–79 years, 8 male, 6 female) yielding 24 insertion records that span the cortex and continue 8 mm beyond into underlying cancellous bone. Synthetic blocks (20 pcf and 30 pcf nominal density) contributed 18 additional insertions. Three pilot diameters (2.45, 2.50, and 2.55 mm) sampled the interference space. The data provide steady-state segments for friction-dominated interaction, segments sensitive to pilot and cortex geometry, early-depth cutting transients, and depth-dependent softening segments post peak. From this combined set, mechanism-specific subsets were defined: 22 steady friction–ploughing segments in the cortex, 20 pilot–tolerance experiments, 16 early-depth cutting cases with observed chip trains, and 24 segments showing damage-induced relaxation over several pitches.

The bench data came from human femoral cortices and commercial sawbone blocks covering two density levels, with the screw drawn from three lots and a driver–torque chain calibrated to 0.2 N·m uncertainty; this coverage yields a reasonably representative test bed for the mechanisms emphasized here without exhaustively sampling the full clinical population.

Definition and traceability of the measured quantity — gradation.

a.
Torque defined without specification of measurement chain or uncertainty.

b.
Torque–depth defined with single-sensor calibration and basic synchronization description.

c.
Full torque–depth definition with calibration traceability, synchronization method, and uncertainty statement aligned to decision tolerances.

d.
Torque–depth plus environmental and operator factors controlled or randomized, with multi-instrument cross-checks and blind runs.

3. Computational Models: Geometry, Material Laws, Contact, and Solution Controls

Three analyses were implemented on general-purpose implicit FEA platforms with double-precision solvers and consistent contact kinematics. The Axisymmetric Equivalent-Thread Model (AETM) represents the screw by a revolved shank with an effective thread land height and flank inclination mapped to an axisymmetric profile. The bone is discretized as a concentric domain with a cortical shell over cancellous substrate. Contact is enforced with a regularized Coulomb law (μ) and a ploughing resistance modeled as an added tangential traction proportional to local indentation depth with coefficient Hp (units MPa). Material behavior of the cortical region is transversely isotropic linear elastic–perfectly plastic with yield shear τy set from donor-informed proxies;

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

cancellous is compressible elastic–plastic with lower moduli. The near-surface mesh around the thread–bone interface uses 8-node axisymmetric quadrilaterals with local element edges 20–50 μm and six elements through the cortex thickness. Implicit Newton–Raphson steps are controlled by arc-length on imposed rotation, with a line search for robustness and a force residual tolerance 1e6.

The 3D Half-Turn Helical Segment (HTHS) resolves half a pitch of the helical thread with exact flank geometry; periodic boundary conditions represent continuation. Contact uses the same regularized tangential law as AETM. Chip formation is not modeled as a true fracture process; instead, a viscoplastic cutting surrogate introduces a shear band zone just ahead of the cutting edge with flow stress σf that rises with strain rate via σf = σ0(1 + C ln(e/e0)), using σ0 between 80 and 160 MPa, C = 0.025–0.06, and reference rate e0 = 1 s1. This surrogate increases torque demand during initial entry and while the cutting land ploughs. The mesh on thread flanks and in the cortical zone is unstructured 10-node tetrahedra with 25–40 μm edge lengths near the interface and growth to 200 μm away from contact. The solver employs augmented Lagrangian contact with adaptive penalty updates and a consistent tangent; residual norms at convergence are imposed at the same levels as AETM.

The Multi-Scale Cortical–Cancellous Coupled Model (MS-CCC) solves the full screw–bone assembly at macroscale with an orthotropic elastic–plastic constitutive law in cortex and a damage-coupled shear response, linked to an RVE material subroutine. The RVE is a 0.5 mm cube with osteonal alignment set to match donor-specific polarities, possessing principal moduli Ex = 18 GPa, Ey = 12 GPa, Ez = 12 GPa, shear moduli Gxy = 4.5 GPa, Gxz = 4.5 GPa, and Gyz = 3.8 GPa. Damage evolves by D = h(τ/τc 1)+ with h calibrated per donor, reducing local shear strength τ(1 D). The macroscale mesh uses 20-node hexahedra in cortex, with 30–60 μm edges near the interface and six to eight elements across thickness; cancellous uses 8-node elements up to 300 μm. A UMAT drives stress updates, and contact mirrors AETM's tangential law to isolate the effect of multi-scale damage representation.

Across the suite, the numerical values fed to the analyses matched bench-measured screw diameters, pitches, flank angles, pilot sizes, and cortical thickness within the manufacturing and metrology tolerances, while interfacial shear coefficient and subsurface ploughing stiffness were supplied as probability distributions derived from tribology and indentation evidence rather than single fixed constants.

Specification of numerical solution and contact control — gradation.

a.
Implicit settings are left to software defaults with no reporting of tolerances or stabilization.

b.
Contact and solver tolerances are stated, with evidence of sensitivity runs on one case.

c.
Contact formulation, stabilization controls, residual targets, and line search are specified and exercised across cases with consistent criteria.

d.
Solver and contact controls are chosen via QoI-centric studies and tied to quantitative error budgets per mechanism.

4. Physical Mechanisms Selected for Assessment

Four mechanisms were isolated because they each dominate torque behavior in distinct depth intervals and because design or surgical variables can alter their contributions. First, tangential resistance and ploughing in a fully formed thread–cortex interface generate a broad plateau or gently rising segment of torque; this portion is governed by the interfacial coefficient of friction and the hardness-like response of near-surface bone as asperities and micro-yield processes engage. Errors in those interfacial properties can bias the entire steady segment of the torque–depth trace.

Second, the diameter of the pilot and the thickness of the cortex modulate local interference, which in turn controls contact pressure distributions and material yielding under flanks. Small pilot changes can shift torque by tens of percent. Because both pilot and cortex thickness have small absolute tolerance bands, they are attractive variables for challenging geometric sensitivity and equivalency mapping across models.

Third, during initial entry and while cutting edges advance, the interaction transitions to a chip-forming regime that adds transient shear and normal loads. In this regime, torque–depth traces exhibit spikes of relatively short spatial extent that depend on cutting edge geometry, rake and clearance, local anisotropy, and chip evacuation. Accurately reproducing both the magnitude and timing of these peaks is essential to prevent underprediction of risk for peak torque-driven thread stripping.

Finally, bone's anisotropy and degradation under repeated passes of the flank modify local shear capacity. Damage accumulation generates a rise to peak followed by relaxation as the tool progresses. Models that ignore directionality of stiffness or permit unlimited hardening tend to miss the extent of relaxation, misplacing the depth at which torque settles post peak.

Mechanism-specific scoping of the question of interest — gradation.

a.
Mechanisms are listed but not operationalized into testable segments.

b.
Mechanisms are operationally defined by depth intervals and primary variables of influence.

c.
Mechanisms mapped to both tests and model outputs with indicators and acceptance metrics.

d.
Mechanisms further tied to decision-critical outcomes with pre-registered analysis plans.

5. Data Pedigree for Material, Contact, and Geometric Inputs

Frictional resistance at the thread–cortex interface was parameterized by tribological experiments on cortical bone coupons harvested from diaphyseal sections acquired under the same IRB protocol as the main insertions. A pin-on-plate setup with a stainless tool steel pin shaped to a 30-degree cone and a normal load representative of thread bearing pressures was used to measure tangential response. Fifty trials were run; eight were excluded for hydration drift or visible glaze. The resulting coefficient distribution for cortex is $\mu = 0.30 \pm 0.05$ (mean ± standard deviation). Additional tests on cancellous analog surfaces at lower normal loads produced $\mu = 0.25 \pm 0.07$. Ploughing resistance Hp was obtained from microscale scratch and indentation tests on cortical surfaces with contact area estimates converted to a traction per local indentation; the aggregated estimate used in AETM and as a baseline for HTHS was Hp = 120 ± 25 MPa.

Geometric inputs were grounded by metrology. Three lots of screws (10 per lot) were scanned for diameter and core dimensions; flank roughness was measured by profilometry. Pilot holes were drilled with new bits for each diameter level and measured by pin gauge to ±0.005 mm. Cortical thickness was measured by micro-CT pre-test and by optical cross-sections post-test. The distributions used in stochastic realizations for pilot diameter are normal with the specified mean and tolerance-derived standard deviation; cortex thickness is donor-specific with a uniform prior within measured bounds for each site.

Microstructure properties informing MS-CCC arose from ultrasound and micro-CT characterization of 14 donors. Directional modulus estimates were taken from resonant ultrasound spectroscopy, with variances approximately 10–15% across donors. Damage softening laws were calibrated by in situ torsion of cortical rings with digital image correlation providing shear strain fields; the damage evolution parameter h was identified by minimizing the discrepancy between torque decay in the test and in an RVE simulation, yielding h = 0.12 ± 0.03 s1 in normalized units. Shear strength thresholds τc were drawn from the same ring tests and set at $\tau c = 50 \pm 8$ MPa.

For the cutting surrogate in HTHS, values of $\sigma 0$ and C were taken from published cutting-of-bone surrogates and adjusted within reported ranges to match the observed slope of early-depth torque rises ($\sigma 0 = 110 \pm 20$ MPa; C = 0.04 ± 0.015), noting that these parameters emulate energy dissipation rather than physical chip segmentation.

Traceability of sources informing high-influence parameters — gradation.

a.
Parameters drawn from secondary compilations with no chain to raw data.

b.
Parameters drawn from peer-reviewed studies but with incomplete metadata or context mismatch.

c.
Parameters derived from tests run under matched conditions with raw data, exclusions, and uncertainties accessible.

d.
Parameters replicated across laboratories or methods, with inter-laboratory uncertainty explicitly modeled.

6. Model Inputs and Equivalency of Input Parameters Across Models and Tests

The three models require distinct input sets but share a core of physical quantities. We established a crosswalk centered on thread geometry (outer and core diameters, pitch, flank angle), driver kinematics (rotation rate standardized to 60 rpm for comparison runs), pilot hole diameter, cortex thickness, and contact properties (μ, Hp). For AETM, μ and Hp appear directly in the tangential law. For HTHS, μ enters contact; Hp is implicit in the surrogate energy dissipation only to the extent that it elevates baseline drag. For MS-CCC, μ enters contact but Hp is distributed between damage-driven softening and plasticity in the RVE. To maintain equivalence at the interface, we used the same μ distributions and, where applicable, the same Hp distribution or its energy-equivalent surrogate parameterization based on measured scratch energies. Screw lot and pilot distributions were shared across models; cortex thickness was matched specimen by specimen.

Beyond the common set, HTHS demands cutting surrogate coefficients ($\sigma 0$, C) with ranges drawn from published surrogates and tuned by the slope of the leading torque rise in the early-depth region. MS-CCC introduces RVE parameters (Ex, Ey, Ez, Gxy, Gxz, Gyz, τc, h) for each donor. To avoid bias in comparisons, donor assignments and their microstructure properties were locked before any model-to-test matching.

Nominally, flank roughness and hardening were not used as separate inputs; their net effect enters through μ, Hp, and damage softening.

Equivalency checks were performed on dimensionless groups computed from inputs and outputs. The ratio of interference to thread height, the non-dimensional contact pressure p/(E tan θ), and the Stribeck-like friction number (μE/pitch modulus) were computed across models and tests to examine whether inputs map to the same physical state. Divergences were found primarily in self-tapping because the HTHS cutting surrogate approximates, rather than reproduces, chip segmentation, reducing the validity of one-to-one mapping for energy dissipation in that region.

Crosswalk and parameter equivalence across models and tests — gradation.

a. No crosswalk; models use disparate inputs with ad hoc mapping.

b. Partial crosswalk on geometry and basic contact; surrogate parameters not reconciled.

c. Documented crosswalk on geometry, contact, and interference groups with sensitivity analyses where surrogates alter mapping.

d. Crosswalk validated by invariance of dimensionless outputs under input reparameterization across models.

7. Model Form Evaluation for Each Mechanism

For steady interfacial drag and ploughing, AETM's homogenized representation is appropriate: the axisymmetric abstraction preserves circumferential averaging of contact pressure and captures the combined effect of μ and Hp on tangential traction. HTHS also represents these effects but adds local non-axisymmetric pressure ridges at thread entry/exit; MS-CCC distributes the drag partly into the damage law, altering post-peak relaxations. All three forms reproduce the quasi-steady torque plateau given correct interfacial parameters.

Under pilot and cortex tolerance changes, the torque shifts with interference and contact area. AETM imposes these shifts through axisymmetric thickening or thinning of the shell and direct contact changes; HTHS couples geometric tolerance to out-of-plane flank loading and can exhibit mild asymmetry in contact progression through a half-turn; MS-CCC's multi-scale approach captures the shift of neutral axes and modulates the onset of plasticity and damage within the cortex. Each form is capable, in principle, of mapping how 0.05 mm changes in pilot diameter translate to torque shifts on the order of 0.2–0.5 N·m depending on cortex thickness.

In the chip-forming mechanism, form adequacy diverges. AETM's axisymmetry homogenizes the cutting edge and lacks any representation of fracture, so transient spikes are out of reach. HTHS includes geometric rake and clearance and a viscoplastic cutting surrogate; it elevates torque near the leading edge and during ploughing segments behind it, but without discrete chip segmentation the magnitude and timing of the highest peaks are shape- but not physics-reproducing. MS-CCC includes damage evolution that slightly delays the rise and accelerates relaxation but has no explicit shearing fracture and cannot produce spikes.

For bone microstructure and damage evolution, AETM can mimic softening by adjusting Hp or an effective yield threshold but cannot tie that softening to directionality or donor-specific RVE responses; HTHS shares this limitation at the macroscale. MS-CCC, however, directly propagates orthotropic stiffness and a damage variable D that decays shear strength; it tends to place peak torque correctly and reproduce relaxation slopes within a moderate error band when donor-specific RVE properties are used.

Adequacy of mathematical form for each targeted mechanism — gradation.

a. Model form omits essential physics of the mechanism or contradicts governing constraints.

b. Model form contains the dominant physics but relies on proxies for secondary effects.

c. Model form contains the dominant physics and validated surrogates for secondary effects in the intended operating range.

d. Model form includes mechanism-complete physics or proven surrogates across operating regimes relevant to the decision.

8. Numerical Code Verification for Implicit Contact and Nonlinear Material Response

Verification activities addressed algorithms used across all three models. First, manufactured solutions were constructed for a 2D axisymmetric elastic–plastic annulus under imposed shear with Coulomb friction at the inner boundary; stress fields and displacements from the codes were compared to the closed-form manufactured fields. The maximum norm of stress error was under 1.2% at 50 μm element size and 0.6% at 25 μm. Second, contact linearization was assessed by an elastic half-space loaded by a rigid sphere under tangential motion, comparing computed tangential stiffness to Mindlin's solution. Deviations were 0.7% at the tighter penalty and under 1.1% at a looser stabilization. Third, the UMAT for the orthotropic damage RVE in MS-CCC passed unit tests on return mapping to within 1e9 relative tolerance in plastic multipliers and tracked an exact relaxation path generated from a reduced model to within 0.8% in effective shear stress.

An additional suite of mesh and routine-level checks was executed for HTHS's cutting surrogate. Plane-strain shear bands were imposed via displacement boundary conditions and the viscoplastic law produced flow stresses aligned with expectations from the rate sensitivity C values chosen. However, without a fracture energy parameter per se, the surrogate cannot be verified against Griffith-like criteria; we thus limited checks to consistency and convergence of the elevated torque under imposed banding.

Numerical verification of contact and constitutive routines — gradation.

a.
No verification beyond running example problems.

b.
Selective verification against textbook solutions for linear cases.

c.
Manufactured solutions and benchmark contact problems used with stated tolerances on errors; UMAT-level unit tests included.

d.
Independent audit or cross-code comparison of verification results with uncertainty budgets on numerical approximations.

9. Numerical Solver Error: Convergence Metrics, Step Control, and Mesh Refinement

Convergence was managed using force and displacement residual norms of 1e6 and 1e8 respectively, with an arc-length parameterization of the rotation increment to prevent snap-through during contact transitions. Line search was activated whenever the predicted update increased the residual norm. At each depth, adaptive substepping was permitted to maintain local convergence. Across all steady-contact segments (friction–ploughing and pilot–tolerance mechanisms), step cutbacks occurred in under 20% of attempted steps for HTHS and under 10% for AETM and MS-CCC. In early-depth cutting segments, the HTHS surrogate produced an increase to 28% cutbacks on average because of localized plasticity-induced tangential stiffness changes; MS-CCC experienced non-convergence in 9% of steps during the self-tapping entry if damage onset coincided with contact status changes, which was mitigated by raising the penalty stabilization slightly.

Mesh refinement studies were conducted per mechanism. For AETM, halving element size near the interface from 50 to 25 μm changed peak torque by 3.2% in friction–ploughing and 2.5% in pilot–tolerance cases. In microstructure-driven relaxation segments, the same refinement changed the post-peak slope by 4.1%. For HTHS, refining tetrahedral edges on the flank from 40 to 25 μm changed peak torque by 3.8% in friction–ploughing and by 4.6% in pilot–tolerance cases, while in self-tapping the change reached 12% as the cutting surrogate's shear band localization sharpened. For MS-CCC, refining hexahedra from 60 to 30 μm near the interface produced torque changes under 4% in all but self-tapping segments, where removing spurious oscillations required an additional contact smoothing pass.

Solver performance thus supplied quantitative bounds for numerical error contributions to torque predictions: steady cases were typically within a 4% mesh-induced variability band, while self-tapping segments exhibited up to 12% sensitivity due to localized flow and contact status changes overlapping. These behaviors informed score assignments for numerical error control per mechanism and model.

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

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

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

Quantification and control of numerical solution error — gradation.

a.
Convergence reported without error quantification related to outputs.

b.
Residual targets and basic mesh checks reported; no QoI-based bounds.

c.
QoI-focused mesh refinement and step control quantified with mechanism-specific sensitivity bounds.

d.
Adaptive strategies provide output error estimators and stop criteria tied to decision thresholds.

10. Test Samples: Specimen Types, Manufacturing Tolerances, and Measurement Methods

Specimens were harvested from mid-diaphyseal femoral shafts procured under IRB oversight with documented donor age, sex, and storage time. Specimens were maintained hydrated in lactated Ringer's at 37 °C. After pilot drilling under a guided jig, each specimen received a single screw insertion at 60 rpm. Synthetic blocks were used to bracket density effects and offer repeatable structure for tolerance experiments. Manufacturing tolerances were controlled by selecting screws from three lots, measuring each screw for key geometry, and performing torque tests on a benchtop fixture with low compliance. The torque chain calibration was verified before and after the series, with drift under 0.02 N·m.

Measurement methods for depth used synchronized optical tracking and encoder integration with a lead correction for thread pitch, which was confirmed by advancing the screw into a reference medium and verifying motion per revolution. The program randomized order across pilot diameters and specimen types to reduce operator trend effects. While the main focus was mechanism coverage, the sample also allowed limited exploration of donor-to-donor variation in post-peak relaxation behavior.

Representativeness and control in the physical test program — gradation.

a.
Convenience samples without control of key variables; limited repeatability.

b.
Targeted specimens with partial control of geometry and environment; calibration documented.

c.
Representative specimens across relevant ranges of geometry and density; randomized order and low-drift instrumentation.

d.
Multi-site or blinded testing with pre-registered protocols and inter-operator repeatability studies.

11. Output Comparison: Torque–Depth and Peak Torque Versus Mechanism-Specific Benchmarks

Outputs were compared to mechanism-specific benchmarks using median absolute percentage error (MAPE) for torque–depth segments and absolute error on peak torque and spike timing as applicable. The acceptance thresholds were defined a priori as follows: for steady friction–ploughing and pilot–tolerance segments, MAPE ≤ 15% and peak torque error ≤ 10%; for microstructure-dominated relaxation segments, MAPE ≤ 15% with correct placement of the peak within 0.5 mm of depth; and for self-tapping and chip formation, MAPE ≤ 20%, peak torque error ≤ 10%, and timing of the primary spike within ±0.5 mm of the observed depth.

For AETM, friction–ploughing comparisons yielded MAPE = 12% with RMSE = 0.21 N·m, and peak torque errors under 8% across 22 segments. Pilot–tolerance comparisons gave MAPE = 10% (RMSE = 0.19 N·m), and a torque sensitivity to pilot diameter of 0.42 N·m per 0.10 mm change, within the experimental 0.45 ± 0.08 N·m per 0.10 mm. However, in the chip-forming regime, MAPE rose to 38%, with the model entirely missing the discrete spikes; peak torque was underestimated by 18% on average, and no meaningful spike timing could be reported. In microstructure-dominated segments, MAPE was 16% with the peak placed within 0.6 mm on average and post-peak relaxation underpredicted by 14%.

For HTHS, friction–ploughing achieved MAPE = 11% (RMSE = 0.20 N·m) and matched peak torque within 7% on average. Pilot–tolerance achieved MAPE = 9% with pilot sensitivity 0.44 N·m per 0.10 mm and no systematic bias across cortical thickness. In the chip-forming regime, the cutting surrogate captured a portion of the elevated torque but overshot peak by 22% and delayed the main spike by 0.7 mm on average; MAPE = 28% and spike timing exceeded the ±0.5 mm criterion in 10 of 16 cases. For microstructure, MAPE was 14% with peaks generally placed within 0.4 mm and relaxation rate slightly overpredicted.

For MS-CCC, friction–ploughing returned MAPE = 14% and peak errors of 9–12% depending on donor RVE calibration. Pilot–tolerance yielded MAPE = 12% and pilot sensitivity of 0.40 N·m per 0.10 mm. In chip formation, the absence of a fracture process led to MAPE = 41%, overestimation of peak torque by 30% when the surrogate was tuned aggressively, or underestimation by 20% when tuned conservatively; spike timing was not captured. In microstructure-dominated segments, MAPE = 13% with proper placement of peaks within 0.5 mm and relaxation slope within 10% for donors whose RVE parameters were close to population means.

Taken together, steady-contact and tolerance mechanisms were reproduced within acceptance thresholds by AETM and HTHS and marginally by MS-CCC; the self-tapping regime failed to meet both magnitude and timing criteria for all three models. These outcomes drove factor scores on output agreement and, in turn, the overall decision.

Quantitative comparison of outputs to mechanism-specific benchmarks — gradation.

a.
Qualitative overlay without metrics or thresholds.

b.
Quantitative metrics applied post hoc with loosely justified thresholds.

c.
Pre-specified metrics with mechanism-specific thresholds and multiple error measures (MAPE, peak, timing).

d.
Pre-registered analysis with cross-validation or blind datasets and uncertainty propagation into decision criteria.

12. Software Quality Assurance and Use History (summary notes)

The solvers used belong to general-purpose implicit finite element codes with documented regression test suites and release notes. The user material routines were managed under version control with peer review before integration. Prior uses of the AETM and MS-CCC styles include orthopedic plate–bone load sharing and dental implant thread analyses; the HTHS approach has appeared in rotating tool simulations. We note these topics here for context but do not adjudicate the sufficiency of either the quality system artifacts or the historical breadth of applications for the present decision.

13. Per-Pair Factor Scoring Matrix (private 0–7 scale) and Thresholds

Scores appear on a private 0–7 Credibility Units scale with anchors defined per factor. In all cases, 0 denotes absence of evidence, 1–2 indicate weak or indirect evidence, 3–4 indicate partial coverage for the decision context, 5 indicates meeting pre-specified criteria with direct evidence aligned to the mechanism, 6 indicates exceeding criteria with replication or multi-site breadth, and 7 denotes mechanism-complete coverage with independent replication and formal uncertainty propagation tied to the question of interest.

For agreement to outputs, the anchors map to quantitative criteria: 0 (no comparison); 1 (>50% MAPE or failure of all mechanism criteria); 2 (35–50% MAPE or failure of any primary criterion such as peak or timing); 3 (25–35% MAPE, partial success on secondary criteria); 4 (15–25% MAPE meeting secondary but not all primary criteria); 5 (≤15% MAPE and ≤10% peak error or equivalent mechanism criteria met); 6 (≤10% MAPE with breadth across donor types or material densities); 7 (≤5% MAPE with blind validations and uncertainty propagation). For numerical solver error, anchors distinguish reporting only (2–3), QoI-based mesh and step studies with quantified bounds (4–5), and adaptive control with error estimators (6–7). For the evidence that justifies interfacial drag and damage parameters, anchors distinguish trace to raw data with uncertainty (5) and inter-laboratory replication (6–7).

Across all runs the residual norms at convergence were below 1e6 for force and 1e8 for displacement, with time-step cutbacks under 10% in steady-contact but up to 28% in self-tapping for HTHS; mesh refinement changed torque by less than 4% in friction–ploughing and pilot–tolerance and as much as 12% in self-tapping. These quantitative behaviors underwrite the assigned levels for numerical error control in the score tables.

Interpretation of the private 0–7 Credibility Units scale — gradation.

a.
Ordinal categories defined qualitatively without mapping to the question of interest.

b.
Anchors defined per factor but not tied to numerical thresholds.

c.
Anchors defined per factor with numerical thresholds and examples aligned to mechanisms.

d.
Anchors defined with formal uncertainty budgets and decision-theoretic mapping to acceptable risk.

14. Discussion of Cross-Mechanism Consistency and Residual Gaps

A central aim was to test internal consistency of findings across mechanisms and models. For friction–ploughing, AETM and HTHS were mutually consistent in achieving 11–12% MAPE with peak errors below 8%, and MS-CCC's 14% MAPE remained within the pre-specified boundary; numerical error control contributed less than a 4% change on mesh refinement for these segments, suggesting that the remaining gap is due largely to parameter uncertainty and not numerical artifacts. For pilot–tolerance, all three models tracked a torque sensitivity to pilot diameter of about 0.40–0.45 N·m per 0.10 mm, near the experimental 0.45 ± 0.08 N·m per 0.10 mm slope; the magnitude of the shifts was captured without overfitting across cortical thickness values.

By contrast, transient chip-related peaks exposed divergent capabilities in the structural equations of the models. The axisymmetric form cannot produce discrete cutting-induced spikes; the 3D half-turn model elevates torque with a rate-dependent surrogate that partially emulates cutting energy but lacks chip segmentation; and the multi-scale model lacks a fracture pathway altogether. These form-level limitations are not easily removed by parameter retuning and dominate the output mismatches in self-tapping, where MAPE ranged from 28% to 41% and timing criteria were not met. We therefore judge that the chosen equations and topology for these solvers are adequate for steady contact and interference questions but insufficient for cutting-dominated regions.

When mapped to a common set of physical quantities, the three models agreed within 5% on pilot diameters, helix angle, and nominal thread height, but differed in how flank roughness and cutting energy are represented, so the overall equivalency is partial and weakens in the self-tapping region where surrogate energy dissipation cannot be mapped one-to-one to the contact and ploughing terms.

The provenance of the friction and damage inputs traced to 42 pin-on-bone trials for μ (μcortex = 0.30 ± 0.05), indentation-derived ploughing resistance (Hp = 120 ± 25 MPa), and 14 donor calibrations for orthotropic damage laws; calibration notebooks and raw datasets were complete for six of nine parameter groups, but inter-laboratory replication was not available, placing these sources at a moderate level of traceability for the present decision.

Across all runs, the residual norms at convergence were below 1e6 for force and 1e8 for displacement, with step cutbacks under 20% for AETM and MS-CCC and 28% for HTHS in the self-tapping regime; mesh refinement changed peak torque by less than 4% in steady mechanisms and up to 12% in self-tapping, so numerical error control was acceptable for steady portions of the torque signal and marginal near chip surrogates.

Although not determinative here, two contextual items bear mention. The software environments used have long-standing documentation and regression suites, and the general modeling approaches have appeared in prior orthopedic applications, but we did not adjudicate the sufficiency of those records for this decision.

Finally, agreement to outputs spanned a useful range: against independent torque–depth records from 24 insertions, the models achieved median MAPE of 11% for friction–ploughing, 9% for pilot–tolerance, 33% for self-tapping, and 15% for microstructure-driven relaxation. Only the self-tapping mechanism consistently failed to meet magnitude and timing criteria across the suite.

Cross-mechanism synthesis of strengths and limitations — gradation.

a.
Observations per mechanism are listed without synthesis.

b.
Synthesis highlights strengths and limitations but without linking to consequences for the decision.

c.
Synthesis links mechanism-specific evidence to factor scores and decision thresholds, identifying dominant gaps.

d.
Synthesis incorporates sensitivity to assumptions and proposes targeted evidence to close the identified gaps.

15. Conclusion—the model suite is not accepted for predicting insertion torque of this bone screw because output comparisons under the Self-Tapping and Chip Formation mechanism failed to meet comparison criteria, and friction/damage input pedigree was insufficient to support the reported scores; additional verified formulation and higher-grade test evidence are required.

The present assessment applied a structured, mechanism-focused framework to three implicit FEA models of a 3.5 mm self-tapping cortical screw. All three models captured steady interfacial drag and the effects of pilot and cortex variations within pre-declared thresholds, and MS-CCC credibly reproduced microstructure-driven relaxation when donor-specific RVE parameters were used. However, none of the models consistently captured both the magnitude and timing of torque spikes during cutting. HTHS overshot peak torque by 22% and delayed the principal spike by 0.7 mm on average; AETM and MS-CCC lacked the physics to generate spikes at all and produced MAPE of 38% and 41% in the self-tapping segments, respectively. Against independent torque–depth records from 24 human insertions and 18 synthetic blocks, the suite therefore failed to meet the acceptance criteria for the chip-forming mechanism. In parallel, the chain of evidence supporting friction and damage parameters was moderate: although $\mu = 0.30 \pm 0.05$ (cortex) and Hp = 120 ± 25 MPa were measured under matched conditions, and damage parameters were calibrated to 14 donors, there was no inter-laboratory replication and not all parameter groups had fully curated raw data.

We therefore conclude that, in the question and context of predicting insertion torque for this screw across the four mechanisms in scope, the suite is not accepted. To move toward acceptance, the following additions are needed: (1) a verified chip segmentation or equivalent fracture-capable surrogate with quantitative checks against cutting energy and chip morphology; (2) multi-site, higher-grade cutting experiments with synchronized imaging to set spike magnitude and timing targets and to anchor surrogate parameters; (3) a strengthened chain of evidence for friction and damage inputs, including replication across laboratories or methods and uncertainty propagation into outputs; and (4) adaptive solver and mesh strategies tuned to capture localized flow without excessive cutbacks while delivering quantitative error bounds on the spike region. Until such evidence is assembled and corresponding improvements verified, the current models should not be used to predict insertion torque in clinical or design decisions that depend on accuracy during cutting.

Decision rationale and evidence-to-decision trace — gradation.

a.
Decision stated without linking back to quantitative evidence.

b.
Decision linked to selective metrics; thresholds implied rather than declared.

c.
Decision explicitly tied to pre-specified thresholds and cross-referenced metrics for each mechanism.

d.
Decision accompanied by an evidence-to-decision map and a prioritized plan for achieving acceptability.

16. AETM Credibility Scores

Scores for the Axisymmetric Equivalent-Thread Model (AETM) are compiled per mechanism and factor. Basis statements draw on the evidence and numerical values reported in the body: $\mu = 0.30 \pm 0.05$ (cortex), Hp = 120 ± 25 MPa, pilot diameter 2.50 ± 0.05 mm with three levels tested, cortex thickness 2.0–3.2 mm, mesh sizes 20–50 μm near contact, convergence residuals of 1e6 (force) and 1e8 (displacement), and output comparisons with MAPE 12% (friction–ploughing), 10% (tolerance), 38% (self-tapping), and 16% (microstructure), as well as peak and timing criteria as stated.

17. HTHS Credibility Scores

Scores for the 3D Half-Turn Helical Segment (HTHS) reflect its detailed flank geometry and the use of a viscoplastic cutting surrogate. Basis statements use the same shared measurements for μ, pilot, and cortex thickness; surrogate parameters $\sigma 0 = 110 \pm 20$ MPa and C = 0.04 ± 0.015; mesh sizes 25–40 μm near the interface; augmented Lagrangian contact with residuals at 1e6/1e8; and output comparisons with MAPE 11% (friction–ploughing), 9% (tolerance), 28% (self-tapping), and 14% (microstructure), as well as a mean 22% peak overshoot and 0.7 mm spike delay in self-tapping.

18. MS-CCC Credibility Scores

Scores for the Multi-Scale Cortical–Cancellous Coupled Model (MS-CCC) reflect its RVE-coupled damage representation and shared contact law. Basis uses $\mu = 0.30 \pm 0.05$ (cortex), RVE parameters Ex = 18 GPa, Ey = 12 GPa, Ez = 12 GPa, Gxy = 4.5 GPa, Gxz = 4.5 GPa, Gyz = 3.8 GPa, $\tau c = 50 \pm 8$ MPa, and h = 0.12 ± 0.03, mesh sizes 30–60 μm near contact, residuals 1e6/1e8, and output MAPE 14% (friction–ploughing), 12% (tolerance), 41% (self-tapping), and 13% (microstructure), with self-tapping showing inconsistent peaks due to lack of fracture.

References

[1] L. Vasquez, W. Ishikawa, F. Grigoryan, Mesh density effects on rib deflection in computational models, Ann. Biomed. Eng. 126 (2018) 244–487.

[2] N. Tanaka, N. Castellano, B. Silva, F. Lindqvist, W. Castellano, Time step effects on flow split in computational models, J. Biomech. Eng. 105 (2009) 121–840.

[3] G. Tanaka, T. Quintero, B. Pettersson, H. Tanaka, C. Moreau, Solver tolerance effects on head displacement in computational models, J. Biomech. Eng. 81 (2024) 282–866.

[4] R. Tanaka, S. Silva, T. Bergman, N. Abbott, Time step effects on contact pressure in computational models, Med. Eng. Phys. 103 (2015) 71–565.

[5] R. Kaur, B. Zetterberg, B. Fontaine, N. Duarte, S. Kaur, Contact stiffness effects on rib deflection in computational models, J. Verif. Valid. Uncertain. 47 (2008) 381–571.

[6] J. Rasmussen, F. Moreau, L. Wojcik, D. Duarte, Load path effects on rib deflection in computational models, J. Biomech. Eng. 49 (2011) 386–650.

[7] A. Halvorsen, L. Kaur, M. Fontaine, Cement mantle effects on peak force in computational models, Med. Eng. Phys. 131 (2022) 189–743.

[8] G. Tanaka, S. Ishikawa, K. Moreau, M. Quintero, D. Quintero, Mesh density effects on peak force in computational models, J. Mech. Behav. Biomed. Mater. 72 (2017) 262–475.

[9] J. Halvorsen, C. Vasquez, T. Lindqvist, J. Grigoryan, Contact stiffness effects on contact pressure in computational models, Comput. Methods Biomech. 127 (2017) 103–538.

[10] B. Lindqvist, L. Kaur, G. Lindqvist, E. Wojcik, Time step effects on fatigue life in computational models, J. Mech. Behav. Biomed. Mater. 99 (2006) 36–739.

[11] A. Silva, D. Lindqvist, N. Tanaka, R. Jankowski, Element type effects on head displacement in computational models, J. Mech. Behav. Biomed. Mater. 87 (2010) 151–458.

[12] E. Eriksen, C. Fontaine, P. Ishikawa, F. Abbott, W. Moreau, Contact stiffness effects on fatigue life in computational models, J. Mech. Behav. Biomed. Mater. 96 (2007) 12–689.

[13] N. Ishikawa, A. Quintero, L. Okonkwo, P. Bergman, K. Quintero, Time step effects on stress range in computational models, Ann. Biomed. Eng. 33 (2022) 373–648.

[14] H. Pettersson, P. Eriksen, P. Zetterberg, H. Zetterberg, Inlet profile effects on head displacement in computational models, Comput. Fluids 52 (2008) 358–841.

[15] B. Duarte, J. Lindqvist, K. Bergman, Inlet profile effects on stress range in computational models, Int. J. Numer. Methods 98 (2016) 328–746.

[16] B. Kaur, R. Yilmaz, B. Jankowski, Wall roughness effects on stress range in computational models, Med. Eng. Phys. 130 (2019) 48–401.

[17] N. Jankowski, K. Yilmaz, R. Nakamura, N. Abbott, M. Tanaka, Contact stiffness effects on hemolysis index in computational models, Med. Eng. Phys. 140 (2024) 94–598.

[18] N. Tanaka, N. Fontaine, F. Abbott, Load path effects on chest compression in computational models, J. Mech. Behav. Biomed. Mater. 92 (2019) 192–429.

[19] T. Pettersson, A. Okonkwo, E. Ishikawa, Solver tolerance effects on shear stress in computational models, Ann. Biomed. Eng. 98 (2020) 118–415.

[20] E. Rasmussen, B. Abbott, M. Grigoryan, M. Grigoryan, Load path effects on flow split in computational models, Int. J. Numer. Methods 63 (2009) 291–407.

[21] L. Wojcik, P. Vasquez, B. Duarte, A. Yilmaz, Time step effects on stress range in computational models, Med. Eng. Phys. 31 (2019) 177–881.

[22] C. Ishikawa, F. Yilmaz, Element type effects on stress range in computational models, Med. Eng. Phys. 112 (2010) 217–892.

[23] G. Duarte, E. Duarte, K. Eriksen, M. Zetterberg, L. Fontaine, Friction effects on stress range in computational models, Med. Eng. Phys. 101 (2019) 160–868.

[24] C. Pettersson, C. Grigoryan, M. Kaur, Solver tolerance effects on shear stress in computational models, J. Mech. Behav. Biomed. Mater. 28 (2022) 308–791.

[25] M. Quintero, H. Abbott, T. Jankowski, Element type effects on hemolysis index in computational models, J. Biomech. Eng. 75 (2005) 389–668.

[26] A. Quintero, W. Moreau, Boundary condition effects on cortical strain in computational models, Comput. Fluids 63 (2006) 177–563.

[27] E. Tanaka, L. Okonkwo, F. Halvorsen, Boundary condition effects on cortical strain in computational models, Int. J. Numer. Methods 144 (2007) 112–760.

[28] F. Silva, W. Abbott, R. Ishikawa, Flow rate effects on shear stress in computational models, Int. J. Numer. Methods 144 (2021) 250–761.

[29] C. Abbott, W. Abbott, S. Eriksen, Inlet profile effects on shear stress in computational models, Ann. Biomed. Eng. 117 (2021) 359–606.

[30] G. Nakamura, J. Castellano, G. Jankowski, G. Duarte, Cement mantle effects on peak force in computational models, Med. Eng. Phys. 62 (2016) 179–489.

[31] E. Rasmussen, N. Yilmaz, C. Lindqvist, P. Rasmussen, K. Silva, Inlet profile effects on rib deflection in computational models, J. Mech. Behav. Biomed. Mater. 104 (2022) 385–535.

[32] G. Silva, B. Abbott, Cement mantle effects on peak force in computational models, J. Biomech. Eng. 33 (2019) 36–628.

[33] H. Zetterberg, A. Halvorsen, H. Eriksen, S. Jankowski, Boundary condition effects on cortical strain in computational models, J. Verif. Valid. Uncertain. 104 (2007) 114–898.

[34] K. Pettersson, L. Lindqvist, B. Castellano, Element type effects on peak force in computational models, Ann. Biomed. Eng. 35 (2007) 363–730.

Table 1: Credibility assessment matrix for AETM on the private 0–7 scale; one row per factor and mechanism.

Credibility factor	Level	Basis
Data pedigree - AETM - Frictional and Ploughing Interface	5	μ = 0.30 ± 0.05 from 42 pin-on-bone tests; Hp = 120 ± 25 MPa from scratch/indentation under matched conditions.
Equivalency of input parameters - AETM - Frictional and Ploughing Interface	5	Common μ, Hp, geometry crosswalk applied; dimensionless interference and pressure ratios matched within 5%.
Model form - AETM - Frictional and Ploughing Interface	5	Axisymmetric homogenization captures steady drag and ploughing; no missing physics in this regime identified.
Model inputs - AETM - Frictional and Ploughing Interface	5	Interface μ, Hp used as distributions; screw and cortex dimensions matched metrology within tolerances.
Numerical code verification - AETM - Frictional and Ploughing Interface	5	Manufactured solutions and contact benchmarks within 0.6–0.7% error; axisymmetric patch tests passed.
Numerical solver error - AETM - Frictional and Ploughing Interface	5	Residuals 1e6/1e8; mesh refinement from 50 to 25 μm changed peak torque by 3.2%.
Output comparison - AETM - Frictional and Ploughing Interface	5	MAPE 12% and peak error 8% met 15/10% targets across 22 segments.
Test samples - AETM - Frictional and Ploughing Interface	5	n = 24 human and 18 synthetic insertions with controlled pilots and torque chain uncertainty 0.2 N·m.
Data pedigree - AETM - Pilot Hole and Cortex Tolerance	6	Pilot diameters 2.45/2.50/2.55 mm verified by pin gauge ±0.005 mm; cortex thickness from micro-CT and sectioning.
Equivalency of input parameters - AETM - Pilot Hole and Cortex Tolerance	6	Exact mapping of pilot and thickness across models/tests; interference ratio aligned within 3%.
Model form - AETM - Pilot Hole and Cortex Tolerance	5	Axisymmetric thickening/thinning reproduces interference-driven torque shifts as measured.
Model inputs - AETM - Pilot Hole and Cortex Tolerance	6	Pilot and cortex inputs directly from metrology with distributions; no surrogate parameters required.
Numerical code verification - AETM - Pilot Hole and Cortex Tolerance	5	Axisymmetric contact and plasticity checks as above; geometry perturbations verified to map linearly to torque.
Numerical solver error - AETM - Pilot Hole and Cortex Tolerance	5	Mesh refinement changed torque by 2.5%; cutbacks under 10% of steps.
Output comparison - AETM - Pilot Hole and Cortex Tolerance	5	MAPE 10%; torque sensitivity 0.42 N·m per 0.10 mm vs. experimental 0.45 ± 0.08.
Test samples - AETM - Pilot Hole and Cortex Tolerance	5	20 tolerance experiments randomized across pilots and thickness; instrument drift <

Table 2: Credibility assessment matrix for HTHS on the private 0–7 scale; one row per factor and mechanism.

Credibility factor	Level	Basis
Data pedigree - HTHS - Frictional and Ploughing Interface	5	μ = 0.30 ± 0.05 (cortex) from 42 trials applied; Hp carried implicitly via scratch energy mapping.
Equivalency of input parameters - HTHS - Frictional and Ploughing Interface	5	Common μ; Hp emulated via surrogate energy; dimensionless groups matched within 5%.
Model form - HTHS - Frictional and Ploughing Interface	6	Detailed flank geometry and non-axisymmetric pressure ridges captured steady drag accurately.
Model inputs - HTHS - Frictional and Ploughing Interface	5	μ distribution and geometry matched metrology; surrogate off for pure steady-contact checks.
Numerical code verification - HTHS - Frictional and Ploughing Interface	4	Contact and elastic–plastic checks passed; surrogate-specific verification limited to consistency.
Numerical solver error - HTHS - Frictional and Ploughing Interface	4	Residuals met; mesh refinement changed peak torque by 3.8%; step cutbacks under 20%.
Output comparison - HTHS - Frictional and Ploughing Interface	5	MAPE 11% and peak errors ≤7% satisfied 15/10% targets.
Test samples - HTHS - Frictional and Ploughing Interface	5	Same n = 24 human and 18 synthetic segments as AETM; steady-contact subset of 22 traces.
Data pedigree - HTHS - Pilot Hole and Cortex Tolerance	6	Pilot diameters and thickness measured as above; direct mapping to half-turn domain.
Equivalency of input parameters - HTHS - Pilot Hole and Cortex Tolerance	6	Exact pilot and thickness mapping with periodic BCs; interference ratio aligned within 3%.
Model form - HTHS - Pilot Hole and Cortex Tolerance	6	3D half-turn form accounts for out-of-plane effects; torque shifts reproduced.
Model inputs - HTHS - Pilot Hole and Cortex Tolerance	6	Pilot/thickness distributions passed directly; no surrogate parameters influence this regime.
Numerical code verification - HTHS - Pilot Hole and Cortex Tolerance	4	Standard verification suite adequate; geometry perturbation sensitivity confirmed.
Numerical solver error - HTHS - Pilot Hole and Cortex Tolerance	4	Mesh refinement changed torque by 4.6%; cutbacks under 20%.
Output comparison - HTHS - Pilot Hole and Cortex Tolerance	6	MAPE 9%; pilot sensitivity 0.44 N·m per 0.10 mm near experimental 0.45 ± 0.08.
Test samples - HTHS - Pilot Hole and Cortex Tolerance	5	20 tolerance experiments across pilots and thicknesses; randomized order maintained.
Data pedigree - HTHS - Self-Tapping and Chip Formation	3	Surrogate parameters $\sigma 0$ = 110 ± 20 MPa, C = 0.04 ± 0.015 from literature; no lab replication.
Equivalency of input	3	Energy-based mapping used;

Table 3: Credibility assessment matrix for MS-CCC on the private 0–7 scale; one row per factor and mechanism.

Credibility factor	Level	Basis
Data pedigree - MS-CCC - Frictional and Ploughing Interface	5	$\mu = 0.30 \pm 0.05$ from 42 trials applied to contact; Hp influence spread into damage/plastic response.
Equivalency of input parameters - MS-CCC - Frictional and Ploughing Interface	4	Common μ mapped; Hp effect partitioned into RVE softening—partial equivalence only.
Model form - MS-CCC - Frictional and Ploughing Interface	5	Macroscale with RVE recovers steady drag when μ is correct; appropriate for this regime.
Model inputs - MS-CCC - Frictional and Ploughing Interface	5	μ distribution and geometry matched; RVE parameters at donor means for steady segments.
Numerical code verification - MS-CCC - Frictional and Ploughing Interface	5	UMAT unit tests and manufactured solutions within 0.8–1.2% error; contact verification adequate.
Numerical solver error - MS-CCC - Frictional and Ploughing Interface	4	Residual criteria met; mesh refinement changed peak torque $<4\%$.
Output comparison - MS-CCC - Frictional and Ploughing Interface	5	MAPE 14%; peak torque errors 9–12% straddled the 10% target.
Test samples - MS-CCC - Frictional and Ploughing Interface	5	$n = 24$ human and 18 synthetic segments; same steady-contact subset.
Data pedigree - MS-CCC - Pilot Hole and Cortex Tolerance	6	Pilot and cortex geometry from metrology as above; mapped directly to macroscale mesh.
Equivalency of input parameters - MS-CCC - Pilot Hole and Cortex Tolerance	6	Identical pilot/thickness distributions used; interference measures aligned within 3%.
Model form - MS-CCC - Pilot Hole and Cortex Tolerance	5	Form captures torque shifts via contact area/pressure changes and onset of plasticity/damage.
Model inputs - MS-CCC - Pilot Hole and Cortex Tolerance	6	Pilot/thickness inputs direct from metrology; RVE kept at donor means for tolerance studies.
Numerical code verification - MS-CCC - Pilot Hole and Cortex Tolerance	5	Verification suite adequate; mapping from geometry to torque slope checked numerically.
Numerical solver error - MS-CCC - Pilot Hole and Cortex Tolerance	4	Mesh refinement changed torque by ~3.5%; cutbacks under 10%.
Output comparison - MS-CCC - Pilot Hole and Cortex Tolerance	5	MAPE 12%; pilot sensitivity 0.40 N·m per 0.10 mm within experimental 0.45 ± 0.08.
Test samples - MS-CCC - Pilot Hole and Cortex Tolerance	5	20 tolerance experiments; identical to AETM/HTHS subsets.
Data pedigree - MS-CCC - Self-Tapping and Chip Formation	3	No fracture data; μ applied; damage parameters set from 14 donors but not tied to chip