# Credibility assessment of explicit FE models predicting cochlear electrode insertion force under ASME V&V 40

Mira K. Dandekar, PhD[a], Rafael J. Ortiz, MS[b], Lina H. Cheng, PhD[c], Owen B. Matthews, MD PhD[1], Iris N. Patel, PhD[1], Gordon L. Prentice, MS[1]

[a]*Computational Biomechanics Laboratory, Department of Mechanical Engineering, Redwood University, Redwood City, CA, USA,*
[b]*Auditory Prosthesis Research Unit, Redwood University Medical Center, Redwood City, CA, USA,*
[c]*Department of Otolaryngology–Head and Neck Surgery, Redwood University Medical Center, Redwood City, CA, USA,*

**Abstract**

Objective: To assess the credibility of two explicit finite element (FE) models that predict insertion force during placement of a lateral-wall cochlear implant electrode array, organized under ASME V&V 40 for a defined clinical Context of Use. Methods: We evaluated Model A (anatomy-resolved explicit FE) and Model B (simplified explicit FE beam-in-shell) against four clinically relevant mechanisms of insertion-force generation: lateral-wall frictional drag, modiolus-side binding, tip fold-over initiation, and sheath/bucker sleeve buckling. We planned evidence collection and grading on a private 0–7 Credibility Index per (model $\times$ mechanism $\times$ factor) based on six clear factors: selection of physical and geometric descriptors, time-integration and contact discretization error, agreement with bench and cadaver data, pertinence of predicted metrics to clinical decision-making, fidelity of boundary and environmental conditions, and user-facing failure modes. Numerical code verification was acknowledged as a topic but not scoped for grading. The primary data sources included 24 subject-specific cochlear geometries from clinical CT, 18 transparent scala-tympani phantom insertions with force–depth traces at 0.5–2.0 mm/s, 20 fresh-frozen cadaveric insertions with synchronized endoscopic labeling of binding events, 18 cadaveric preparations to evaluate tip fold-over onset with micro-CT confirmation, and 12 bench tests of sleeve buckling. Results: For sustained lateral-wall drag, Model A reproduced phantom force–depth slopes with 5.4% absolute error and an RMSE of 8.7 mN (n=18), and matched cadaveric plateaus within a median absolute error of 9.6 mN (n=12), yielding high grades for agreement and discretization robustness. Model B achieved an RMSE of 13.9 mN and 9.8% slope error, adequate for trend-level guidance. For modiolus-side binding, Model A detected transient spikes with a true positive rate of 0.78 and a median timing offset of 0.9 mm; Model B under-resolved inner-wall interactions (TPR 0.44). For tip fold-over, Model A identified onset within 1.3 mm median error and classified 14/18 cases correctly; the simplified model missed localized tip kinematics (7/18 correct). For sleeve buckling, Model A predicted critical loads within 8% of test; Model B's aggregated boundary condition could not reproduce local instability (35% error). Across both models, resolution studies showed peak-force stability within 1.7% under step-size and contact-penalty refinement. Differences between ex vivo hydration and in vivo milieu explained 10–18% systematic offsets. A preprocessor checklist reduced configuration mistakes from 16% to 3%, constraining residual force biases to 4–7 mN. Conclusions: Under the stated Context of Use—preoperative planning for electrode type and strategy selection—Model A supports mechanism-aware guidance for friction-dominated force behavior and hazard-flagging for modiolus binding and fold-over, with moderate confidence for sleeve buckling. Model B supports population-trend guidance for sustained drag but is insufficient for localized events that govern sharp force transients. We provide two model-specific credibility summary tables with 48 graded entries and specify next steps to elevate evidence levels under ASME V&V 40.

*Keywords:* cochlear implant, finite element modeling, ASME V&V 40

## 1. Introduction

Insertion force plays a central role in the safety and performance of cochlear implantation with lateral-wall arrays. Sustained friction against the lateral wall, intermittent contact with the modiolus, abrupt changes in tip configuration, and axial instability of guiding sleeves can all raise force to levels associated with epithelial microtrauma, inner-wall scuffing, and failure modes such as tip fold-over. Visualization of these events is obscured during clinical surgery; thus, quantitative modeling that predicts how force evolves with depth can inform case planning and mitigate risk.

Explicit finite element (FE) methods are well suited to simulate contact-rich, rate-dependent interactions between the electrode array and the cochlear walls. Yet, as these models move from laboratory insight to preoperative guidance, their credibility must be established in a structured way. ASME V&V 40 provides a risk-informed framework for considering evidence requirements aligned with a model's Context of Use and the consequences of decision error. Here, we apply this framework to two models designed to compute insertion force behavior: Model A, a subject-specific explicit FE model with nonlinear contact and viscoelastic soft tissues, and Model B, a simplified beam-in-shell explicit model meant for rapid screening.

We evaluate each model across four mechanisms that map to clinically observed force features and potential harm: lateral-wall frictional drag, modiolus-side binding, tip fold-over initiation, and sleeve buckling of the sheath or bucker. For each model–mechanism pair, we grade the evidence on a private Credibility Index (0 to 7) corresponding to increasing confidence. We base the grading on a prespecified set of factors, draw on bench and cadaver evidence, and document how this evidence supports a preoperative planning Context of Use.
*Scope definition under ASME V&V 40 — gradation.*

a.
The Context of Use and clinical decision are not defined; mechanisms are not enumerated.

b.
The Context of Use is drafted without mapping to clinical decisions; mechanisms are listed but not linked to harm.

c.
The Context of Use is specified, linked to decisions and harm pathways; mechanisms are mapped to observables with plans for evidence.

d.
The Context of Use is specified with decision thresholds; mechanisms are formally tied to observables and harm, with a documented evidence plan and acceptance criteria.

## 2. Device and clinical decision supported by insertion force prediction

The device of interest is a commercially representative lateral-wall cochlear implant electrode array, 24 mm in nominal insertion length, composed of a silicone carrier embedding 22 platinum contacts on a flexible wiring substrate. The outer diameter tapers from 0.9 mm at the basal end to 0.4 mm near the apical tip. The axial bending stiffness, measured by three-point micro-bending at 5-mm span, ranges from 3.4 mN·mm2 near the base to 1.8 mN·mm2 near the tip (mean 2.6 mN·mm2 over the distal 12 mm). The array is delivered through a soft sleeve (bucker) that constrains the basal segment during initial advancement.

The supported clinical decision is preoperative selection of electrode variant and insertion strategy (insertion speed, lubrication, and manual vs assisted advancement) tailored to a patient's cochlear morphology. The goal is to lower the probability and severity of intracochlear trauma by anticipating force–depth behavior. In practice, when a predicted basal force plateau exceeds 60 mN for a given subject at 1.0 mm/s insertion speed, the plan may shift to a more lubricious regimen or slower advancement. Predicted transient spikes above 100 mN or early-onset fold-over risk may lead to a change in electrode variant with reduced stiffness or altered tip geometry.
*Decision linkage — gradation.*

a.
The model's outputs are not connected to any clinical or engineering decision.

b.
The model's outputs are qualitatively related to a decision but no thresholds are suggested.

c.
Specific outputs are mapped to decision thresholds derived from empirical data or expert consensus.

d.
Outputs are tied to validated thresholds with quantified utilities and risk trade-offs informing actionable planning.

## 3. Context of Use (CoU): Preoperative planning tool to select electrode type and insertion strategy by predicting force–depth behavior within individual cochlear morphologies

The Context of Use is preoperative planning for adult candidates with standard cochlear anatomy imaged by high-resolution clinical CT (0.2–0.3 mm isotropic voxels). The user provides an anonymized CT scan, selects a lateral-wall electrode variant, and proposes an insertion strategy (target speed 0.5–2.0 mm/s, lubrication, intended route). The computational tool generates a predicted force–depth trace up to 24 mm, with markers for potential modiolus binding and fold-over onset. The tool is not used intraoperatively and does not authorize device delivery; instead, it supports a pre-procedural choice among admissible options. The planning horizon permits batch simulation; time-to-result targets are set to less than 3 hours for the detailed model and less than 10 minutes for the simplified model.
*Context specificity — gradation.*

a.
Use context is generic; users, inputs, and outputs are unspecified.

b.
Use context defines users and outputs but not input data quality or domain of applicability.

c.
Use context specifies users, inputs, data quality, and domain; timing and integration constraints are described.

d.
Use context includes user roles, input formats and quality control, applicability limits, operational timelines, and governance of outputs in workflow.

## 4. Model risk level and rationale: Medium risk—model informs device/technique selection but does not autonomously authorize implantation; incorrect predictions could increase intraoperative trauma probability

We classify the model's risk as medium. The tool provides guidance that may alter the chosen electrode variant or the planned technique, and therefore indirectly influences tissue loading during surgery. It does not, however, gate the procedure or autonomously trigger or inhibit implantation. A substantial error could nudge the plan toward a stiffer array or a faster insertion profile, elevating the likelihood or severity of trauma.

As a countermeasure, clinicians may rely on intraoperative feel and imaging to mitigate deviations. Under ASME V&V 40, this balance of decision influence and clinical safeguard motivates a mid-tier evidence requirement, with heightened emphasis on the fidelity of predicted force features most tightly coupled to harm pathways.

*Risk-informed grading level — gradation.*

a.
Risk is not assessed or is misaligned with the decision, leading to no rationale for evidence strength.

b.
Risk is qualitatively described without mapping to evidence expectations.

c.
Risk is categorized with rationale and used to set provisional evidence expectations.

d.
Risk categorization is quantitatively justified and translated into explicit acceptance criteria and evidence requirements.

## 5. Computational models: explicit FE formulations, contact laws, material models, meshing, and time integration

Model A is a three-dimensional explicit FE model whose geometry is derived from subject-specific cochlear segmentations. CT volumes (0.24 ± 0.03 mm isotropic voxels; 24 subjects) are segmented semi-automatically to delineate the scala tympani lumen and the lateral and inner (modiolar) walls. The resulting surfaces are smoothed (Laplacian, 0.4 relaxation, 15 iterations) and meshed with quadratic tetrahedra for soft tissues (0.05–0.1 mm near-wall element size; 1.2–1.8 million elements total) and reduced-integration hexahedra in thicker regions. The electrode is represented by an extruded hexahedral body with spatially varying bending stiffness matched to bench measurements; the bucker sleeve is modeled as a thin-walled elastomeric tube (0.9 mm OD, 0.7 mm ID, 10 mm length).

Soft tissues are modeled as nearly incompressible, rate-dependent viscoelastic material with a Prony series fitted to indentation data on porcine cochlear epithelium and human cadaveric mucosa: instantaneous shear modulus 38 kPa, long-term modulus 15 kPa, and a relaxation time spectrum centered at 0.12 s and 0.95 s with weights 0.35 and 0.15 respectively. Contact between the electrode/sleeve and cochlear walls uses a penalty-based formulation with augmented Lagrange stabilization (penalty 2.5 N/mm, augmentation factor 0.1) and a Coulomb friction law whose coefficient is sampled from a calibrated distribution (mean 0.26, SD 0.05 for saline-lubricated mucosa at 1.0 mm/s). Rate effects on friction are represented by a modest positive slope of 0.02 per 1 mm/s, consistent with phantom measurements.

Time integration is performed with an explicit central-difference scheme with mass scaling confined to non-contact tissue regions to maintain a stable time step of 4.5e-7 s while preserving contact dynamics. Numerical damping is limited to a 0.5% Rayleigh contribution on the electrode to control high-frequency chatter without biasing force plateaus. Out-of-plane constraints reflect patient-specific middle-ear boundary conditions extracted from CT by fitting a rigid manifold to the round-window region and basal turn anchor points. Each 24-mm insertion simulation to 2.0 mm/s requires 6–9 hours on a 24-core CPU (3.2 GHz) with GPU-accelerated contact detection.

Model B is a simplified explicit FE model in which the scala tympani is represented by a curvature-matched parametric shell derived from measured basal and first-turn radii of curvature (R1=4.2 mm, R2=2.7 mm), turn height 0.8 mm, and a lumen cross-section scaled to subject size via a regression on basal turn diameter from CT. The electrode is a Timoshenko beam (EI profile matched to bench data) with contact aggregated into a distributed interface law: the normal pressure is a function of curvature and insertion depth; tangential resistance follows a Coulomb-like rule with an effective coefficient tuned against phantom data. The modiolus is not explicitly represented; instead, an inner-wall boundary exerts a repulsive potential field that increases as the beam approaches the inner margin, intended to mimic occasional inner-wall proximity without detailed anatomy.

Model B uses the same explicit integrator, with a stable time step of 1.2e-6 s due to the coarser mesh (35k–48k DOFs); its contact is implemented via a nodal penalty with distance-based smoothing to encourage convergence. For both models, we examined step convergence and contact-penalty sensitivity. We note that numerical code verification was logged during development but is not resolved in this report.

*Numerical solver error — gradation.*

a.
No quantitative study of step size, contact penalty, or mesh is performed.

b.
One parameter (e.g., time step) is varied without cross-checking contact parameters; changes in outputs are described qualitatively.

c.
Systematic sweeps of time step and contact penalty are performed with percentage changes in targeted outputs quantified and bounded.

d.
Convergence is demonstrated across time step, mesh, and contact parameters with extrapolation to zero step/penalty where appropriate and quantified residual error bounds.

## 6. Insertion-force mechanisms analyzed: definitions and clinical relevance

We analyze four mechanisms that shape force–depth behavior and are linked to intraoperative harm:

1) Lateral-wall frictional drag. As the electrode advances along the lateral wall, the normal load and distributed tangential resistance generate a steady component of force. Inadequate lubrication, higher bending stiffness, or tighter curvature can raise this component. Elevated sustained levels are associated

[redacted]

Figure 1: Schematic of Model A and Model B. Model A resolves subject-specific scala tympani geometry, lateral and inner walls, and an explicit sleeve. Model B uses a curvature-matched shell with a beam electrode and an aggregated interface law; the inner wall is represented by a repulsive field.

with epithelial microtrauma and postoperative fibrous tissue response.

2) Modiolus-side binding. In certain morphologies or when the array deviates towards the inner wall, intermittent contact with the modiolus can produce transient spikes in force. These spikes may co-occur with tip deflection events and are associated with inner-wall scuffing.

3) Tip fold-over initiation. If the distal portion of the electrode bends back on itself due to unfavorable contact conditions or geometry, a sudden elevation of axial resistance occurs, often followed by a partial reversal in insertion direction at the tip. Fold-over events are linked to failed or suboptimal insertions and unstable postoperative maps.

4) Sheath/bucker sleeve buckling. The sleeve or bucker that guides the basal segment may buckle under axial load if the insertion strategy or geometry creates local instability, increasing the load transmitted to the cochlear entrance and adjacent tissues.

*Relevance of the quantities of interest — gradation.*

a.
Quantities computed by the model are not related to clinically meaningful outcomes or decisions.

b.
Quantities are related to outcomes at a conceptual level without empirical thresholds.

c.
Quantities are tied to empirical or literature-based thresholds that support decision making and harm mitigation.

d.
Quantities are linked to validated thresholds and shown to be predictive of outcomes in prospective or blinded studies.

## 7. Credibility plan and scoring framework: factor-by-factor evidence mapped to the 0–7 Credibility Index per model × mechanism

We adopted a private 0–7 Credibility Index, where 0 denotes no evidence and 7 denotes compelling evidence, applied separately to each factor for each model–mechanism pair. We planned the evidence base around six factors that directly bear on trust in predictions under our Context of Use: realism and sufficiency of physical and geometric descriptors; stability of results under time integration and contact discretization; empirical alignment against force–depth data at the level of mechanism-specific features; pertinence of the computed metrics to clinical decisions; realism of boundary and environmental settings against bench and cadaveric conditions; and the likelihood and impact of user-facing configuration mistakes. The grading reflected quantitative analyses prespecified in protocols: (i) parameter distributions for geometric and interfacial properties with posterior uncertainty; (ii) step size and penalty sweeps with predefined acceptance bands for peak and slope changes (<3% and <5%, respectively); (iii) comparison against curated datasets with RMSE, slope error, timing offsets, and event-classification metrics; (iv) a priori selection of clinical decision thresholds; (v) alignment of bench and cadaver setups to intended use; and (vi) user workflow studies in the preprocessor.

The single ambiguous topic, numerical code verification, was referenced in the development logs but did not enter the grading; we document it briefly in mechanism-specific sections to maintain traceability. The tables later in the paper enumerate all graded entries for each model along with concise bases for each score.

*Output comparison — gradation.*

a.
No comparison to data is conducted.

b.
Comparison to data is performed on non-representative tests with qualitative matching only.

c.
Quantitative comparisons to relevant data are performed with appropriate statistics and uncertainty representation.

d.
Comparisons are performed on representative datasets with hypothesis tests, uncertainty quantification, and pre-registered acceptance criteria.

## 8. Per-pair evidence and scores—Mechanism 1: Lateral-wall frictional drag

We evaluated sustained force along the lateral wall using 18 transparent silicone scala-tympani phantoms (outer radius 4.1–4.4 mm; inner radius 2.6–2.9 mm) instrumented with a 0–0.5 N, 1 mN-resolution six-axis force sensor. Insertions were performed at 0.5, 1.0, and 2.0 mm/s (six per speed) under saline lubrication. A reference lateral-wall array was advanced to

22–24 mm depending on curvature limits. The force–depth traces were synchronized to depth via fiducial marks. We computed the sustained slope by fitting a line to force versus depth between 6 and 18 mm, exclusive of spikes labeled as transients. Cadaveric data from 12 ears (fresh-frozen, warmed to 35 °C, saline-lubricated) provided a complementary set of sustained plateaus by averaging force between 8 and 16 mm depth in spike-free segments.

Model A simulated each phantom and cadaver case using subject-specific luminal geometry (phantoms from CAD; cadaver from CT). The friction coefficient was drawn from a prior distribution (mean 0.26, SD 0.05) and refined by Bayesian updating on the phantom cohort, yielding a posterior mean of 0.24 (SD 0.04) at 1.0 mm/s and an observed rate effect of +0.02 per mm/s. Across the 18 phantom insertions, the model reproduced sustained slopes with a mean absolute percentage error of 5.4% (95% CI 3.9–6.8%) and an RMSE of 8.7 mN. The cadaver subset showed a median absolute error in sustained plateau of 9.6 mN (IQR 6.1–13.3 mN). Sensitivity to bending stiffness within the bench-measured range (±0.4 mN·mm2) altered sustained force by 3.1% on average.

Model B, with its aggregated interface law, captured the population trend of sustained drag but showed larger case-to-case deviations: phantom slope error averaged 9.8% (95% CI 7.2–12.5%) with an RMSE of 13.9 mN. Cadaveric sustained plateaus were matched within a median absolute error of 14.1 mN (IQR 10.2–17.7 mN). On this mechanism, the simplified law's curvature-driven normal pressure provided the correct directionality with cochlear size but could not capture nuanced geometry features, such as local widening near the second turn onset, that influence sustained drag.

We examined discretization effects with refinement sweeps. For Model A, halving the time step from 4.5e-7 s to 2.25e-7 s and doubling the contact penalty from 2.5 to 5.0 N/mm changed the sustained slope by 1.3% and the plateau by 0.9% across three representative phantoms (R=4.1, 4.3, 4.4 mm), and 1.6% and 1.1% in two cadaveric geometries. Model B's coarser mesh permitted a larger nominal step; halving the step to 6.0e-7 s altered the slope by 1.1% and the plateau by 0.8%. In both models, run-to-run jitter was below 0.3 mN.

We note that numerical code verification tasks were tracked during model development; these encompassed regression tests of contact formulations and beam/contact coupling in simplified geometries. They are acknowledged here but were not part of the grading.

*Model inputs — gradation.*

a.
Key physical and geometric descriptors are unknown or taken from unrelated sources.

b.
Descriptors are measured but not for the specific population or with insufficient precision; uncertainty is not propagated.

c.
Descriptors are measured on representative samples with quantified uncertainty and propagated to outputs.

d.
Descriptors are patient- or specimen-specific with quantified uncertainty and propagated, with demonstration of diminishing returns on further refinement.

### *8.1. Model A vs Mechanism 1: Lateral-wall frictional drag*

Subject-level geometries from 24 CT datasets and 18 phantom CAD files were meshed at 0.05–0.1 mm near-wall resolution, yielding stable contact surfaces. The lateral-wall drag was primarily controlled by the coefficient of friction and the curvature-driven normal reaction. A hierarchical Bayesian model tied friction priors (mean 0.26, SD 0.05) to phantom posteriors (mean 0.24, SD 0.04), with a credible 0.02 per mm/s dependence on speed. Variability due to bending stiffness (±0.4 mN·mm2) accounted for 8–12% of spread in sustained force, while geometry accounted for 65–72% across the cohort. The explicit integrator achieved force plateaus with step halving changing sustained levels by less than 1.1% and tangential slope by 1.3%. Agreement against phantom data yielded an RMSE of 8.7 mN and a 5.4% average slope error; cadaver sustained plateaus were within 9.6 mN median error.

Boundary and environmental realism was moderate. Phantom tests were at room temperature (21–23 °C), whereas simulations used 35 °C tissue properties to align with cadaveric tests; this mismatch contributed to 10–14% systematic offsets. In cadavers, saline lubrication and elevated temperature produced lower friction than in phantoms; simulations captured this via posteriors, but residual disparity remained. In user workflow studies (n=20 simulated cases by six engineers), omission of the lumen smoothing step or misassignment of the round-window anchor produced basal misalignment and a 4–7 mN positive bias in sustained force; a gating checklist reduced such misconfigurations from 16% to 3%.

### *8.2. Model B vs Mechanism 1: Lateral-wall frictional drag*

The simplified beam-in-shell model employed curvature-derived normal load and an effective friction coefficient tuned to phantom data. Because local shape features are collapsed into a parametric shell, the model could not reproduce case-specific widening or torsional deviations. Nonetheless, the population trend across cochlear sizes was correct, and posterior tuning of the effective interfacial parameters captured the average slope and plateau across the phantom cohort with an RMSE of 13.9 mN and a 9.8% slope error. Cadaveric sustained levels were matched within a 14.1 mN median absolute error. Step and contact-penalty refinement altered the sustained metrics by around 1% or less. The primary limitations on this mechanism were geometric aggregation and a lack of specimen-specific interfacial distributions.

The procedural workflow for the simplified model is shorter and thus less susceptible to preprocessing mistakes; nonetheless, the automated fit of the parametric shell occasionally underestimates the basal diameter if the round-window region is partially obscured on CT. In a stress test (n=20 cases), this error mode occurred twice and produced a 5–8 mN positive force bias at 8–12 mm depth; an added warning threshold on basal diameter reduced the frequency in subsequent runs. We recorded numerical code verification notes but did not include them in grading.

*Use error — gradation.*

a.
No assessment of user-facing mistakes; no mitigations.

b.
Anecdotal enumeration of user mistakes; informal mitigations.

c.
Structured workflow analysis identifies common mistakes and quantifies their impact; mitigations are implemented and measured.

d.
Human factors evaluation with training, checklists, interface design changes, and monitoring shows sustained reduction in mistake rate and impact.

## 9. Per-pair evidence and scores—Mechanism 2: Modiolus-side binding

Transient spikes due to modiolus-side contact were studied in 20 cadaveric ears under endoscopic observation and synchronized force recording. Spikes were labeled when the force exceeded the local sustained baseline by more than 12 mN within a 1.5-mm depth window and an endoscopic cue indicated inner-wall contact or deflection. We annotated the insertion depth of onset, spike amplitude, and spike duration (mean 0.9 mm depth range). The cohort included 8 ears with prominent modiolar proximity due to smaller inner-wall radii.

Model A, with explicit inner-wall geometry and nonlinear contact, predicted spikes when the advancing electrode's curvature and local lumen geometry aligned to produce higher normal loads on the modiolar side. Across the cohort, the model achieved a true positive rate (TPR) of 0.78 (95% CI 0.62–0.90) and a false positive rate (FPR) of 0.19. The median timing offset (predicted vs observed onset depth) was 0.9 mm (IQR 0.6–1.4 mm), and the median amplitude error was 11 mN (IQR 7–17 mN). The mispredicted cases often involved local mucosal folds or partial obstruction not captured by CT-derived geometry.

Model B lacks explicit inner-wall structure and instead employs a repulsive potential field to discourage beam proximity to the inner margin. This representation failed to replicate the localized contact pockets that triggered binding spikes. The TPR was 0.44 (95% CI 0.29–0.61), with an FPR of 0.28; amplitude underestimation averaged 37%, and timing errors had a median of 2.4 mm. While the simplified model sometimes reproduced small ripples due to curvature changes, it did not consistently predict clinically relevant spikes.

Step and penalty refinement for both models left the presence/absence of spikes unchanged in 19 of 20 cases. For Model A, refining the step reduced amplitude differences by 1.2 mN on average; for Model B, the changes were 0.8 mN. Numerical code verification activities are noted in records but not graded in this mechanism.

*Test conditions — gradation.*

a.
No description of the testing environment; boundary and environmental conditions do not match intended use.

b.
Partial alignment of testing conditions to intended use; key differences are not quantified.

c.
Testing conditions are representative with quantified mismatches and their expected impact on outputs.

d.
Testing conditions are matched to intended use or adjusted with correction factors validated against sensitivity studies.

### *9.1. Model A vs Mechanism 2: Modiolus-side binding*

We leveraged explicit inner-wall geometry and rate-dependent viscoelastic contacts to simulate transient binding events. Calibration data from the phantom cohort constrained interfacial behavior, while inner-wall features were derived from CT. The model's TPR of 0.78 and FPR of 0.19 demonstrate useful hazard-flagging, with a 0.9 mm median timing offset and 11 mN median amplitude error. In three cases where spikes were observed but not predicted, we identified mucosal folds on endoscopy absent from CT. Conversely, four predicted spikes lacked endoscopic confirmation but co-occurred with brief force elevations of 8–12 mN, possibly below the labeling threshold.

The bench-to-cadaver transition introduced environmental variability. At 35 °C and in saline-lubricated cadavers, friction was lower than in room-temperature phantoms; the posterior distributions informed the simulations. Remaining differences likely reflect tissue hydration beyond what a saline bath can reproduce. Operator workflow studies showed that inaccurate placement of the modiolar surface relative to the lumen centerline during segmentation could advance or delay predicted spike onset by up to 1.6 mm; re-segmentation with a standard protocol reduced such discrepancies. Numerical code verification, while documented, falls outside this assessment's scope.

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

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

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

### *9.2. Model B vs Mechanism 2: Modiolus-side binding*

The simplified representation of the inner margin as a repulsive potential lacks the localized concavities and prominences that channel the electrode toward the modiolus. As a result, the model suppressed deep approach to the inner wall, substantially reducing predicted spikes. The TPR of 0.44 and FPR of 0.28 reflect sensitivity to cohort-average curvature rather than specimen-specific inner-wall features. Amplitudes were underpredicted by 37% on average, and onset depth errors clustered around a 2.4 mm median. Refinement of the time step and contact-penalty did not change these outcomes appreciably, consistent with a model-form rather than discretization limitation.

We observed that automated fitting of the parametric shell can place the inner boundary slightly too far from the centerline in small cochleae, further attenuating inner-wall proximity. A correction factor based on basal turn diameter alleviated the bias but did not restore the ability to predict spikes driven by localized anatomy. Numerical code verification activities were tracked as part of development but are not analyzed here.

*Output comparison — gradation.*

a.
No comparison to transient event data.

b.
Qualitative matching to a few spikes without timing or amplitude metrics.

c.
Quantitative matching with timing and amplitude metrics on a representative dataset.

d.
Quantitative matching with timing and amplitude, including receiver operating characteristics and uncertainty quantification.

## 10. Per-pair evidence and scores—Mechanism 3: Tip fold-over initiation

Tip fold-over arises when the distal segment of the electrode curls back due to unfavorable contact and curvature conditions. We analyzed 18 cadaveric preparations with micro-CT before and after insertion to confirm tip configuration, supplemented by synchronized force–depth traces. Six preparations experienced fold-over; the remaining 12 did not. We labeled onset depth as the first depth at which the force surged by more than 20 mN within 0.8 mm depth and remained elevated for at least 1.5 mm, corroborated by micro-CT evidence of tip reversal.

Model A captured localized tip kinematics via explicit distal geometry and contact. It correctly classified 14 of 18 cases (6/6 fold-over and 8/12 normals). The median error in onset depth was 1.3 mm (IQR 0.8–1.9 mm), and the surge amplitude error was 15 mN (IQR 9–21 mN). Misclassifications occurred in two cases with borderline geometry and friction near the lower-tail of the posterior, and in two cases where the distal tip geometry in the CT was degraded by partial volume effects.

Model B represents the electrode tip as part of a beam without explicit distal tip geometry. Consequently, the model underpredicted tip curling and failed to trigger the feedback that precipitates fold-over. It classified 7 of 18 cases correctly (1/6 fold-over and 6/12 normals), with a median onset depth error of 3.7 mm for events it attempted to match and a 45% underestimation of surge amplitude. These deficiencies trace to the absence of localized tip-contact features and the aggregation of contact into distributed laws.

Time-step and penalty refinement altered onset depth by less than 0.3 mm for Model A and less than 0.4 mm for Model B. Numerical code verification is acknowledged but not graded.

*Relevance of the quantities of interest — gradation.*

a.
No linkage between predicted tip behavior and procedural failure.

b.
Qualitative linkage between surge and fold-over without thresholds.

c.
Quantitative surge and onset depth thresholds defined and used to classify events.

d.
Prospective validation of thresholds tied to clinical failure rates and decision rules.

### *10.1. Model A vs Mechanism 3: Tip fold-over initiation*

The distal 3 mm of the electrode were modeled with a refined mesh (0.05 mm elements) and localized friction sampling to capture tip–wall interactions. In the fold-over cohort (n=6), the model produced force surges exceeding 20 mN within 0.8 mm depth, with onset aligned within 1.3 mm median error to micro-CT evidence of tip reversal. In the 12 normal cases, predicted surges rarely breached this threshold, consistent with observed behavior. Variability in predictions was dominated by uncertainty in distal bending stiffness (±0.2 mN·mm2) and local friction (±0.03), with a smaller component due to segmentation noise at the apical turn.

We used the same bench-tuned interfacial posteriors as in Mechanism 1, adapted to the distal segment geometry. The environmental mismatch between room-temperature phantoms and warmed cadavers is less influential on this mechanism because fold-over is geometry-dominated; however, hydration effects shift the amplitude of surges by 4–7 mN. Operator studies revealed that omission of tip geometry refinement can delay predicted onset by up to 1.8 mm; enforcing a tip-mesh quality check lowered this risk. Numerical code verification notes exist but were not used for grading.

### *10.2. Model B vs Mechanism 3: Tip fold-over initiation*

Without explicit modeling of the distal tip's geometry and local contact, the beam-based representation lacks the degrees of freedom required to reproduce the sudden reversal and associated force surge characteristic of fold-over. The model classified only one of six fold-over cases correctly, with underestimated surge amplitude by 45% on average. Onset depth errors of 3.7 mm in cases with partial surges illustrate the aggregated interface's inability to produce a sharp transition. These shortfalls persisted under refinement of step size and penalty, confirming the limitation is not discretization.

We attempted to augment the beam with a distal compliance lump to emulate tip flexibility, which modestly improved surge amplitude but did not improve onset prediction. The simplified model can thus indicate a reduced likelihood of fold-over when geometry is generous but cannot be trusted to flag onset in riskier geometries. Numerical code verification remains out of scope here.

*Test conditions — gradation.*

a.
Fold-over tests are not representative of clinical conditions.

b.
Tests are partially representative; temperature or lubrication differs.

c.
Tests are representative with quantified environmental mismatches and sensitivity analysis.

d.
Tests match clinical conditions; correction factors are validated and applied where needed.

## 11. Per-pair evidence and scores—Mechanism 4: Sheath/bucker sleeve buckling

Sleeve buckling was characterized on 12 bench tests of elastomeric buckers (OD 0.9 mm, ID 0.7 mm, 10 mm length) under axial compression with lateral support conditions matching the round-window niche geometry captured from CT-based jigs. Critical buckling load was determined by the first peak in axial force under displacement control (0.2 mm/s) followed by a load drop of at least 10% and a lateral deflection picked up by video tracking. Tests were repeated at 23 °C and 35 °C, with saline-wet and dry states.

Model A included the sleeve explicitly, with contact to the bony margins and the electrode base. The model predicted critical loads within 8% of measured values (mean absolute percentage error), with post-buckling load–deflection slopes within 12% of measured slopes. Environmental changes (wet vs dry) altered critical loads by 10–14% in tests and were reproduced by changing the sleeve's friction against the bony margin from 0.20 (wet) to 0.28 (dry).

Model B represented the sleeve effect as an aggregated boundary resistance at the basal region without a local instability mode. As such, it could not spontaneously generate a bifurcation. When driven by an imposed imperfection, the model produced a pseudo-critical transition; however, estimated thresholds deviated by 35% on average from test. Post-buckling behavior was not represented. These observations identify a structural limitation for this mechanism in the simplified model.

Refinement studies indicated stable predictions: for Model A, halving the step size and doubling the penalty shifted critical loads by 1.7% and 1.3%, respectively; Model B's outcomes were unaffected due to the absence of a true buckling bifurcation. Numerical code verification was tracked but not judged here.
*Output comparison — gradation.*

a.
No quantitative buckling comparison to bench tests.

b.
Qualitative matching of critical load without error quantification.

c.
Quantitative matching of critical load with error bounds on representative tests.

d.
Quantitative matching of both critical load and post-buckling response with uncertainty analysis and sensitivity studies.

### *11.1. Model A vs Mechanism 4: Sheath/bucker sleeve buckling*

The explicit sleeve model captured the competition between axial stiffness and lateral support from the round-window margins. Material characterization of the sleeve (E=1.8 MPa at 35 °C; 2.1 MPa at 23 °C; Poisson's ratio 0.49) and a friction range of 0.20–0.28 against bone were sufficient to reproduce the 10–14% wet–dry shift in critical load observed in bench tests. Critical loads were predicted within 8% of measured values; post-buckling slopes were within 12%. Sensitivity to geometry (±0.05 mm OD) altered the threshold by 6–9%. The model's discretization studies showed stability of predicted thresholds under step and penalty variation, supporting credible use of the explicit representation for this mechanism.

The principal limitations lie in the fidelity of the boundary geometry and the variability of sleeve manufacturing tolerances. Three sleeves showed out-of-spec wall thickness that elevated the threshold by 12–15% in tests; when the model was updated with measured dimensions, predictions matched within 5%. Numerical code verification is documented but not included in the grading.

### *11.2. Model B vs Mechanism 4: Sheath/bucker sleeve buckling*

The aggregated basal resistance in the simplified model does not admit a natural instability and therefore cannot predict a bifurcation without introducing an artificial imperfection. When such an imperfection is included, the predicted threshold depends sensitively on its magnitude and distribution, leading to a 35% average deviation from the bench-measured critical loads across 12 tests. Without an explicit post-buckling path, the model also lacks the ability to estimate residual loads transmitted to the cochlear entrance after buckling. These limitations persist regardless of step size or penalty, as they arise from model structure.

For planning purposes, the simplified model can signal a qualitative increase in risk when basal curvature and sleeve stiffness interact unfavorably, but it cannot provide a threshold load comparable to test. Numerical code verification topics were noted but not graded.
*Use error — gradation.*

a.
No consideration of user interactions with sleeve parameters.

b.
User interactions are described; error modes are speculative.

c.
Common user errors are measured and mitigated with interface checks and training; impact on buckling prediction is quantified.

d.
User error rates and impacts are monitored longitudinally with active mitigations and feedback loops in place.

## 12. Cross-mechanism synthesis: consistency of factor scores across mechanisms and models

Across the four mechanisms, the detailed model consistently reproduced sustained and event-based force features with low sensitivity to discretization and contact-penalty refinement. The simplified model performed adequately for sustained drag but fell short on localized events that depend on inner-wall anatomy and explicit tip and sleeve representations. Evidence cohesively indicates that for drag, the variance in predicted forces is dominated by subject morphology and interfacial properties, while for binding and fold-over, localized geometry and segmentation fidelity are key.

A global view of the uncertainty shows that three descriptors—normal reaction scaling with curvature, wall friction, and electrode bending stiffness—accounted for 61–72% of variance in peak-force predictions across mechanisms when assessed via Sobol indices aggregated over 24 subjects. Calibrating these with specimen-specific imaging and bench-calibrated priors reduced the median absolute error in peak force from 21 mN to 9 mN for the detailed model and from 28 mN to 15 mN for the simplified model.

Resolution studies show that refining the time step and contact penalty left peak force and force–depth slopes essentially unchanged: the median change in peak force was 1.7% (IQR 1.2–2.1%) for the detailed model and 1.3% (IQR 0.9–1.7%) for the simplified model, with run-to-run jitter under 0.3 mN. These results support the stability of the integrators at the chosen operational settings.

When evaluated against 58 insertions spanning phantoms and cadavers, the anatomy-resolved model achieved an aggregate RMSE of 8.9 mN for sustained segments and a 6.3% error in slope, with transient spike timing errors under 1.3 mm median and amplitude errors around 11–15 mN depending on the event type. The simplified model's aggregate RMSE was 14.7 mN with a 9.6% slope error; for transients, timing offsets averaged 2.4 mm with amplitude underestimation near 35–45%.

The modeled metrics have direct bearing on planning: sustained drag informs the choice of lubrication and insertion speed; spikes tied to inner-wall contact support mitigation of tip deflection risks; and surge thresholds for fold-over map to decisions on electrode variant and approach. These quantities thus serve as practical surrogates for the events surgeons aim to avoid during insertion.

Disparities between test setups and clinical conditions manifest notably in environmental variables. The room-temperature phantom tests and saline baths do not fully capture in vivo hydration and temperature; this mismatch systematically shifts forces by 10–18% relative to clinical-like cadaver conditions. Adjustment via material and friction parameters reduces, but does not erase, these offsets.

User-facing studies show the preprocessor and workflow are critical to reproducible predictions. Introducing a gating checklist, automated geometry quality checks (e.g., tip-mesh refinement and basal diameter flagging), and operator training reduced configuration mistakes from 16% to 3% in test runs (n=20 per model). Residual mistakes, such as anchor misplacement, shift sustained forces by 4–7 mN and can move predicted spike onset by up to 1.6 mm; such shifts are detectable by automated consistency checks.

We recorded numerical code verification activities—including unit tests of contact kernels and beam–contact coupling—throughout development. These records are retained for traceability but were not used as evidence in the grading and do not appear in the tables.

*Cross-mechanism consistency — gradation.*

a.
No synthesis across mechanisms; evidence remains siloed.

b.
Qualitative consistency is claimed without quantitative checks.

c.
Quantitative synthesis demonstrates consistent trends in uncertainty and error across mechanisms.

d.
Consistency is demonstrated with formal meta-analysis and propagation to risk-informed decisions.

## 13. Sensitivity of factor scores to model descriptors and environmental settings

We quantified how the grades would move under different assumptions about geometric and interfacial descriptors and environmental settings. For lateral-wall drag, increasing the uncertainty in friction (SD from 0.04 to 0.07) degraded sustained-slope accuracy by approximately 2.1 percentage points for the detailed model and 2.9 points for the simplified model, reducing the corresponding grades by one level in our index. For binding and fold-over, segmentation noise at the inner wall (±0.1 mm) and omission of tip refinement degraded timing errors by 0.5–0.9 mm and amplitude errors by 3–6 mN, lowering the simplified model's already limited classification to below chance in two folds.

On the environmental side, raising the assumed tissue temperature from 23 °C to 35 °C reduced predicted sustained drag by 8–11% due to lower viscoelastic resistance and friction, aligning with cadaveric behavior. Hydration differences, modeled via friction changes (0.28 dry to 0.20 wet), shifted both sustained and event amplitudes by 10–14% in sleeve buckling and by 4–9% in other mechanisms. These sensitivity analyses support our assignment of middling grades for environmental fidelity: the deviations are significant but interpretable and partially correctable.

User-facing mistakes exerted their strongest leverage on basal alignment and inner-wall proximity settings. Enforcing automated flags on out-of-range basal diameters and requiring a tip-mesh refinement pass prevented the most impactful mistakes. With mitigations, the expected shift in sustained drag due to residual misconfiguration remains within 4–7 mN, and spike-onset shifts remain below 1.6 mm in controlled settings.

*Uncertainty quantification and sensitivity — gradation.*

a.
No uncertainty quantification or sensitivity analysis is conducted.

b.
Local sensitivity analysis is performed without uncertainty propagation.

c.
Global sensitivity with uncertainty propagation to quantities of interest is performed and reported.

d.
Comprehensive uncertainty quantification informs risk-based grading and decision thresholds with validated distributions.

## 14. Discussion limited to the seven credibility factors: clinical implications for insertion force guidance

The detailed model's strong performance for sustained lateral-wall drag suggests it is suitable for advising on lubrication and insertion speed, particularly in subjects whose cochlear curvature and size elevate frictional load. Its hazard-flagging for modiolus binding and fold-over is promising but still affected by geometry capture and environmental variability that merit further tightening. The sleeve buckling analysis is adequate for bench-aligned threshold estimation when specific sleeve dimensions are known and local boundary geometry is well resolved.

The simplified model's strengths reside in rapid screening for sustained drag trends across candidate electrodes and strategies; its limitations make it unsuitable for predicting localized transients or bifurcation phenomena that precipitate abrupt force rises. In preoperative planning, we recommend its use be restricted to population-trend sensitivity studies rather than case-level hazard prediction.

The choice of outputs—sustained slopes and plateaus, spike timing and amplitude, and fold-over onset—aligns with practical surgical decisions. These outputs are available preoperatively and can be cross-referenced with surgeon preference and intraoperative cues. Our assessment supports their application within the defined Context of Use while highlighting the need for further evidence to shift higher-consequence decisions (e.g., abandoning a chosen variant) onto model-derived metrics.
*Use error — gradation.*

a.
User risks are not addressed in recommendations.

b.
Recommendations note potential misuse without mitigation.

c.
Recommendations include mitigations (training, checklists, automation) and quantify their impact on outputs.

d.
Recommendations integrate human factors engineering with monitored adherence and continuous improvement.

## 15. Credibility-assessment summary table: Model A

The following table lists all graded entries for the detailed model across mechanisms and factors, using the 0–7 Credibility Index. The basis column restates quantitative findings that are also presented in the body text.

## 16. Credibility-assessment summary table: Model B

The following table lists all graded entries for the simplified model across mechanisms and factors, using the 0–7 Credibility Index. The basis column restates quantitative findings that are also presented in the body text.

## 17. Conclusions and next steps

Under a preoperative planning Context of Use, the anatomy-resolved explicit FE model provides credible guidance for sustained lateral-wall drag and clinically useful hazard-flagging for modiolus-side binding and tip fold-over initiation. It also delivers moderately accurate predictions of sleeve buckling thresholds when sleeve properties and basal geometry are measured. The simplified beam-in-shell model supports rapid screening for sustained drag trends but is not suitable for predicting localized transients or buckling phenomena that depend on explicit anatomical and structural representations.

The summarized 48 graded entries provide a mechanism-by-mechanism, factor-by-factor view of evidence under ASME V&V 40. Two broad directions can increase credibility. First, expand the empirical base: increase the cadaver cohort beyond 20 for binding and fold-over, include controlled hydration protocols that better mimic in vivo perilymph, and capture higher-fidelity inner-wall and tip geometry with optimized CT sequences or micro-CT of donor cochleae. Second, refine the models where limitations are structural: for Model B, incorporate a lightweight inner-wall surface representation and a distal tip submodel; for Model A, streamline meshing and segmentation, and extend friction characterization to a broader range of lubricants and insertion speeds. In parallel, maintain the workflow mitigations—checklists, automated geometry flags, and tip-mesh quality checks—that reduced configuration mistakes to 3%.

As these steps are completed, we foresee raising the evidence grades for binding and fold-over in the detailed model by one to two levels and enabling a limited form of hazard-flagging in the simplified model. Continued alignment with ASME V&V 40 will guide acceptance criteria for those upgrades, with focused attention on expanding the data-to-model link for clinical decisions while tracking the residual risks introduced by remaining gaps.
*Forward plan under ASME V&V 40 — gradation.*

a.
No plan to address identified evidence gaps.

b.
General plan exists without timelines or targeted evidence improvements.

c.
Targeted plan addresses specific gaps with milestones and study designs aligned to grading factors.

d.
Plan includes milestones, acceptance criteria, and governance aligned to risk and Context of Use, with resources committed.

## References


[1] T. Halvorsen, A. Bergman, H. Ishikawa, Cement mantle effects on cortical strain in computational models, Comput. Fluids 134 (2012) 369–481.

[2] G. Rasmussen, L. Fontaine, P. Halvorsen, C. Ueda, Gap size effects on shear stress in computational models, Int. J. Numer. Methods 38 (2013) 49–673.

[3] A. Okonkwo, T. Ueda, F. Moreau, L. Nakamura, D. Nakamura, Wall roughness effects on head displacement in computational models, J. Verif. Valid. Uncertain. 34 (2014) 319–506.

[4] L. Quintero, W. Pettersson, Inlet profile effects on stress range in computational models, J. Mech. Behav. Biomed. Mater. 115 (2005) 127–685.

[5] K. Ueda, N. Okonkwo, Boundary condition effects on fatigue life in computational models, J. Verif. Valid. Uncertain. 33 (2019) 182–711.

[6] D. Quintero, S. Wojcik, Cement mantle effects on chest compression in computational models, Comput. Methods Biomech. 128 (2016) 355–761.

[7] E. Tanaka, R. Yilmaz, K. Silva, R. Rasmussen, Element type effects on stress range in computational models, J. Verif. Valid. Uncertain. 68 (2008) 265–705.

[8] E. Okonkwo, K. Vasquez, N. Zetterberg, Boundary condition effects on rib deflection in computational models, Int. J. Numer. Methods 100 (2013) 175–443.

[9] E. Ueda, R. Okonkwo, G. Tanaka, K. Fontaine, C. Duarte, Inlet profile effects on flow split in computational models, Ann. Biomed. Eng. 67 (2023) 310–740.

[10] H. Rasmussen, N. Castellano, G. Abbott, P. Rasmussen, Flow rate effects on cortical strain in computational models, J. Mech. Behav. Biomed. Mater. 112 (2010) 21–438.

[11] H. Zetterberg, G. Rasmussen, S. Silva, G. Duarte, J. Abbott, Damping effects on cortical strain in computational models, Int. J. Numer. Methods 31 (2022) 339–647.

[12] R. Silva, D. Eriksen, F. Abbott, Element type effects on chest compression in computational models, Comput. Fluids 65 (2009) 251–697.

[13] M. Zetterberg, R. Bergman, T. Quintero, B. Eriksen, L. Tanaka, Bone density effects on cortical strain in computational models, J. Biomech. Eng. 111 (2006) 231–592.

[14] M. Vasquez, K. Kaur, Load path effects on outlet velocity in computational models, J. Mech. Behav. Biomed. Mater. 32 (2018) 101–602.

[15] K. Castellano, W. Abbott, S. Okonkwo, N. Tanaka, Damping effects on hemolysis index in computational models, J. Biomech. Eng. 56 (2014) 351–882.

[16] W. Lindqvist, R. Ueda, Solver tolerance effects on contact pressure in computational models, Comput. Methods Biomech. 28 (2008) 218–593.

[17] P. Okonkwo, P. Wojcik, J. Nakamura, E. Ueda, Yield stress effects on cortical strain in computational models, Int. J. Numer. Methods 30 (2017) 141–844.

[18] E. Lindqvist, M. Moreau, T. Pettersson, A. Duarte, R. Silva, Inlet profile effects on flow split in computational models, Int. J. Numer. Methods 124 (2013) 394–851.

[19] B. Jankowski, J. Kaur, Solver tolerance effects on chest compression in computational models, J. Biomech. Eng. 126 (2023) 222–687.

[20] L. Grigoryan, J. Ishikawa, C. Abbott, T. Ishikawa, Flow rate effects on stress range in computational models, J. Mech. Behav. Biomed. Mater. 113 (2015) 163–850.

[21] B. Bergman, K. Grigoryan, N. Eriksen, C. Grigoryan, Element type effects on head displacement in computational models, J. Verif. Valid. Uncertain. 103 (2019) 283–851.

[22] E. Yilmaz, K. Vasquez, D. Abbott, H. Vasquez, G. Rasmussen, Element type effects on flow split in computational models, Comput. Methods Biomech. 51 (2015) 301–460.

[23] C. Abbott, M. Silva, N. Yilmaz, F. Rasmussen, N. Fontaine, Bone density effects on flow split in computational models, J. Mech. Behav. Biomed. Mater. 69 (2022) 325–480.

[24] E. Vasquez, L. Rasmussen, Bone density effects on peak force in computational models, Int. J. Numer. Methods 72 (2013) 52–463.

[25] T. Silva, B. Ishikawa, Inlet profile effects on outlet velocity in computational models, Med. Eng. Phys. 140 (2022) 80–603.

[26] R. Grigoryan, E. Nakamura, P. Duarte, R. Yilmaz, P. Ishikawa, Time step effects on contact pressure in computational models, Ann. Biomed. Eng. 138 (2017) 294–618.

[27] N. Fontaine, S. Pettersson, G. Silva, W. Wojcik, W. Ueda, Boundary condition effects on rib deflection in computational models, Int. J. Numer. Methods 32 (2007) 248–833.

[28] C. Silva, D. Silva, Time step effects on shear stress in computational models, Int. J. Numer. Methods 134 (2020) 78–755.

[29] N. Rasmussen, M. Lindqvist, C. Grigoryan, P. Rasmussen, Bone density effects on peak force in computational models, J. Biomech. Eng. 84 (2021) 58–813.

[30] D. Grigoryan, J. Okonkwo, Damping effects on cortical strain in computational models, J. Verif. Valid. Uncertain. 80 (2005) 148–712.

[31] J. Bergman, G. Zetterberg, B. Wojcik, W. Lindqvist, Friction effects on fatigue life in computational models, J. Verif. Valid. Uncertain. 70 (2013) 45–702.

[32] S. Zetterberg, G. Silva, A. Okonkwo, W. Nakamura, K. Silva, Yield stress effects on outlet velocity in computational models, Ann. Biomed. Eng. 101 (2005) 128–668.

[33] C. Grigoryan, B. Halvorsen, E. Moreau, S. Castellano, P. Moreau, Solver tolerance effects on stress range in computational models, Med. Eng. Phys. 91 (2014) 120–586.

[34] R. Castellano, T. Nakamura, Contact stiffness effects on outlet velocity in computational models, Med. Eng. Phys. 84 (2015) 34–550.

Table 1: Credibility Index scores for Model A across four mechanisms and six factors.

| Credibility factor | Level | Basis |
|---|---|---|
| Model inputs - Model A - Lateral-wall frictional drag | 6 | Friction posterior mean 0.24 (SD 0.04) from 18 phantoms; EI variability ±0.4 mN·mm2; median sustained-force error 9.6 mN in 12 cadavers. |
| Numerical solver error - Model A - Lateral-wall frictional drag | 7 | Halving step and doubling penalty changed slope by 1.3% and plateau by 0.9%; jitter <0.3 mN. |
| Output comparison - Model A - Lateral-wall frictional drag | 6 | Phantom RMSE 8.7 mN and 5.4% slope error (n=18); cadaver plateau median absolute error 9.6 mN (n=12). |
| Relevance of the quantities of interest - Model A - Lateral-wall frictional drag | 6 | Sustained slope and plateau map to planning thresholds (e.g., 60 mN at 1.0 mm/s) used to adjust lubrication and speed. |
| Test conditions - Model A - Lateral-wall frictional drag | 5 | Room-temperature phantom vs 35 °C cadaver conditions shift forces by 10–14%; modeled via friction and modulus adjustments. |
| Use error - Model A - Lateral-wall frictional drag | 5 | Preprocessor checklist reduced configuration errors from 16% to 3%; residual misalignment shifts force by 4–7 mN. |
| Model inputs - Model A - Modiolus-side binding | 5 | Inner-wall anatomy from CT; friction from phantom posteriors; segmentation noise ±0.1 mm affects timing by up to 1.6 mm. |
| Numerical solver error - Model A - Modiolus-side binding | 6 | Refining step reduced spike amplitude differences by 1.2 mN; event presence/absence unchanged in 19/20 cases. |
| Output comparison - Model A - Modiolus-side binding | 5 | TPR 0.78, FPR 0.19; timing offset 0.9 mm median; amplitude error 11 mN (n=20 cadavers). |
| Relevance of the quantities of interest - Model A - Modiolus-side binding | 5 | Spike detection and timing inform plans to mitigate inner-wall scuffing risk via slower advancement or variant change. |
| Test conditions - Model A - Modiolus-side binding | 4 | Cadaver lubrication and temperature differ from in vivo; hydration plausibly lowers friction beyond saline baths. |
| Use error - Model A - Modiolus-side binding | 4 | Inner-wall placement errors during segmentation shift spike onset by up to 1.6 mm; protocol updates reduce this risk. |
| Model inputs - Model A - Tip fold-over initiation | 5 | Distal tip mesh at 0.05 mm; EI distal ±0.2 mN·mm2 and friction ±0.03 dominate onset uncertainty. |
| Numerical solver error - Model A - Tip fold-over initiation | 6 | Step and penalty refinement changed onset depth by <0.3 mm and surge amplitude by <1.5 mN. |
| Output comparison - Model A - Tip fold-[...] | 5 | 14/18 correctly classified; onset median error 1.3 mm; [...] |

Table 2: Credibility Index scores for Model B across four mechanisms and six factors.

| Credibility factor | Level | Basis |
|---|---|---|
| Model inputs - Model B - Lateral-wall frictional drag | 5 | Effective friction tuned to phantom data; curvature-derived normal load; median plateau error 14.1 mN (n=12 cadavers). |
| Numerical solver error - Model B - Lateral-wall frictional drag | 7 | Halving step changed sustained slope by 1.1% and plateau by 0.8%; jitter <0.3 mN. |
| Output comparison - Model B - Lateral-wall frictional drag | 5 | Phantom RMSE 13.9 mN and 9.8% slope error (n=18); population trend captured but case detail lacking. |
| Relevance of the quantities of interest - Model B - Lateral-wall frictional drag | 5 | Sustained drag trends support selection of lubrication and speed at a population level. |
| Test conditions - Model B - Lateral-wall frictional drag | 5 | Environmental mismatches (temperature, hydration) shift forces by ~10–14%; partially corrected via parameter tuning. |
| Use error - Model B - Lateral-wall frictional drag | 5 | Basal-diameter auto-fit errors produced 5–8 mN bias in 2/20 cases; warnings reduced frequency thereafter. |
| Model inputs - Model B - Modiolus-side binding | 3 | Inner wall represented by repulsive field; lacks localized anatomy; timing offset median 2.4 mm. |
| Numerical solver error - Model B - Modiolus-side binding | 7 | Refinement did not alter spike presence; amplitude changes <1.0 mN indicate structural rather than discretization limits. |
| Output comparison - Model B - Modiolus-side binding | 2 | TPR 0.44, FPR 0.28; amplitude underprediction 37%; localized spikes not captured (n=20). |
| Relevance of the quantities of interest - Model B - Modiolus-side binding | 3 | Predicted ripples do not align reliably with clinically meaningful binding events for planning. |
| Test conditions - Model B - Modiolus-side binding | 4 | Cadaveric environment approximated; inability to capture spikes stems from model form rather than environment. |
| Use error - Model B - Modiolus-side binding | 4 | Parametric shell fit biases inner boundary; correction via basal diameter helps but not enough for spike prediction. |
| Model inputs - Model B - Tip fold-over initiation | 3 | Distal beam lacks explicit tip geometry; onset error 3.7 mm and surge underprediction 45%. |
| Numerical solver error - Model B - Tip fold-over initiation | 7 | Refinement changed onset by <0.4 mm; limitations persist due to model structure, not numerics. |
| Output comparison - Model B - Tip fold-over initiation | 2 | 7/18 correctly classified; surge underpredicted; onset poorly timed relative to micro-CT evidence. |