Risk-based credibility assessment of explicit FEA models for deployment force in a transcatheter delivery system under ASME V&V 40

Lena M. Hart, PhD[a], Gaurav Patel, MS[b], Ana R. Dominguez, PhD[1], Michael T. Evans, MD[1]

[a]*Computational Mechanics and Verification Group, MedAlloy Systems, Minneapolis, MN, USA,*
[b]*Structural Heart Program, North Plains University Hospital, Portland, OR, USA,*

Abstract

Purpose: To evaluate the credibility of two explicit finite element analysis (FEA) models developed to predict operator handle force during deployment of a self-expanding transcatheter valve, in alignment with ASME V&V 40. The models are intended to inform internal design verification decisions on deployment-force limits for a transfemoral delivery system. Methods: We implemented a high-fidelity, 3D explicit model (Model-HF) resolving contact, large deformation, and frictional stick–slip, and a reduced-order explicit submodel (Model-RO) that homogenizes the device–sheath interaction to accelerate studies. We identified two mechanisms that dominate deployment force: unsheathing friction with stick–slip and reaction forces associated with frame recoil during staged release. We executed a risk-informed credibility plan with a private 0–5 Credibility Index (CI). Critical factors required CI ≥ 4.0: algorithm implementation checks, time-integration and contact solution accuracy, and comparison against bench data. Non-critical factors required CI ≥ 3.5: alignment of test conditions with intended use, software process controls, and representativeness of the physical test set. Results: Across mechanisms and models, Model-HF achieved CI 4.5 for algorithm implementation checks; CI 3.8 (unsheathing) and 4.1 (recoil) for solution accuracy; and CI 3.6 (unsheathing) and 2.9 (recoil) for comparison to data. Non-critical CIs for Model-HF were 4.2/3.8 (alignment), 4.0/4.0 (process controls), and 3.5/3.0 (test set). Model-RO achieved lower CIs: 3.7/3.5 (implementation checks), 3.3/2.9 (solution accuracy), 3.1/2.4 (comparison), 3.9/3.5 (alignment), 3.5/3.5 (process controls), and 2.8/2.5 (test set) for unsheathing/recoil respectively. Bench comparisons used a displacement-controlled rig instrumented for force and sheath position; tolerances for peak force were set a priori. Conclusions: Multiple critical criteria fell below the required thresholds, notably solution accuracy and comparison to data for both models in at least one mechanism. The assessment therefore concludes the models are not accepted for the stated context of use.

Keywords: ASME V&V 40, explicit finite element analysis, transcatheter valve delivery

1. Abstract

We implemented and evaluated a risk-based credibility plan, aligned to ASME V&V 40, for two explicit finite element analysis (FEA) models that predict operator handle force during deployment of a self-expanding transcatheter aortic valve delivered transfemorally. One model is high-fidelity (Model-HF), resolving detailed device–sheath contact and frictional stick–slip during unsheathing and staged release. The other is a reduced-order explicit submodel (Model-RO) that homogenizes aspects of the interaction to accelerate parametric studies. We identified two deployment-force mechanisms that can challenge positional control: unsheathing friction with stick–slip and reaction forces from frame recoil during staged release. The context of use is to support design verification and preclinical evidence by predicting time-resolved and peak handle forces, informing whether design variants meet an internal deployment-force limit.

A private Credibility Index (CI) from 0 to 5 was used to score six factors across each model–mechanism pair, with three factors designated as critical to decision-making. Acceptance required CI ≥ 4.0 on all critical factors and CI ≥ 3.5 on non-critical factors. Model-HF achieved strong algorithm-implementation checks but showed mixed solution accuracy and bench agreement; Model-RO underperformed on both solution accuracy and agreement in the more dynamic release mechanism. Because multiple critical factors fell below threshold for at least one mechanism in both models, the models are not accepted for the stated context of use.

Abstract completeness — gradation.

a.
States topic without scope, methods, or outcome.

b.
States scope and methods or outcome, but not both.

c.
States scope, methods, and outcome with minimal quantitative detail.

d.
States scope, methods, quantitative outcomes, and decision link to context of use.

2. Context of use and model risk

The assessed models are intended to support internal design verification and preclinical evidence for a transfemoral transcatheter delivery system for a self-expanding heart valve. The decision problem is whether a design variant satisfies an internal limit for operator handle force during deployment. The model outputs of interest are time-resolved force at the handle during sheath retraction (unsheathing) and during staged frame release, including peak force and its timing relative to sheath displacement and hold–release steps. The outputs are used to rank design variants, to cull variants that exceed limits, and to generate mechanistic insights guiding changes to sheath coatings, frame features near the capsule mouth, and the kinematics of staged release.

We scoped the intended domain to transfemoral access in adults with iliofemoral anatomy suitable for a 14–18 Fr introducer, with sheath curvature radii in the delivery path typically ranging from 40 to 120 mm based on preclinical imaging datasets. Clinical maneuvers to control deployment are approximated by displacement-controlled sheath retraction sequences of constant speed segments and brief holds, recognizing that real-world operators may modulate speed and micro-pauses to manage stick–slip. The device sizes considered span the valve nominal diameters targeting annuli of 21–29 mm, resulting in three delivery system sizes in the present generation.

Model risk is high for this decision context. Failure to correctly predict high-force excursions may allow a design with excessive peaks to pass internal screening, which could contribute to positional overshoot, abrupt release, or tissue injury during preclinical or first-in-human use. The harmful outcome is not purely greater operator effort; it is the loss of fine positional control mediated by transient force peaks during stick–slip or during the first stage of release, when stored elastic energy in the frame couples into the handle. Because internal decisions on which designs proceed to verification and subsequent testing could be adversely influenced if a nonconforming design is not identified, we applied higher evidentiary standards to those factors most closely linked to the model's correctness on the outputs that drive the decision.

Two force-generating mechanisms were selected for focused assessment: (1) unsheathing friction and stick–slip during retraction of the outer sheath over the constrained frame, and (2) reaction forces transmitted to the handle from frame recoil during staged release, especially as the frame exits partial constraint at the capsule edge. These mechanisms are separable in bench testing and elicit distinct patterns in force–displacement histories: non-smooth oscillations with peak clustering near transitions for the former, and a distinct transient at the first release step for the latter.

Downstream decisions that could rely on the models include selection of lubricity specifications for the sheath inner surface, optimization of the capsule mouth geometry, and choice of retraction profiles for operator instructions for use. While the outputs do not directly drive clinical decisions, they can influence preclinical test design, acceptance criteria, and human factors guidance, strengthening the rationale for a stringent evidentiary threshold on correctness of predictions.

Risk-to-evidence mapping — gradation.

a.
Qualitative statement of risk with no mapping to evidentiary needs.

b.
Categorical risk tier with generic evidence statements not tailored to outputs.

c.
Risk tier mapped to specific outputs with factor prioritization stated.

d.
Risk tier mapped to outputs, factor prioritization, and explicit acceptance thresholds tied to decision impact.

3. Device and mechanisms: delivery architecture and force-generation pathways

The delivery system is a coaxial assembly comprising an inner shaft connected to the self-expanding nitinol frame crimped within a capsule at the distal end, and an outer sheath that constrains the frame until retraction. The handle provides a linear actuator that retracts the sheath with a leadscrew translating a rotational knob, instrumented at the bench with a low-noise load cell in series with the actuator to capture operator force. The outer sheath has a hydrophilic coating whose lubricity can vary with activation and shelf-life, a factor we examined by testing in two states: freshly activated and partially dried to emulate worst-case lubricity. The capsule mouth includes a short conical flare transitioning to the inner diameter; design variants alter the flare angle and chamfer to tune release dynamics.

Two mechanisms dominate the handle force time history. During unsheathing, friction at the interface between the inner lumen of the sheath and the outer surface of the constrained frame and capsule causes resistance to sliding. The interface transitions between static and dynamic friction based on local tangential traction and the relative velocity of sliding. At small scales, the difference between static and kinetic coefficients and the velocity-weakening behavior of the lubricated interface induce stick–slip oscillations. Peaks occur when breakaway from static friction happens, leading to rapid micro-slips. Fixture alignment, curvature of the delivery path, and normal force variations across the circumference modulate this behavior. The second mechanism arises during staged release of the frame, when retraction of the sheath past the capsule mouth allows a segment of the stored elastic energy in the superelastic frame to express. Reaction forces transmit proximally through the shaft and handle as segments of the frame spring outward and then contact the capsule edge or aortic wall surrogate, depending on the test. The magnitude and rate of this transient depend on local constraint geometry, frame pre-stress, and any residual contact with the sheath lip.

We captured these pathways in computational models by resolving full 3D contact with penalty enforcement and a regularized friction law that transitions smoothly between rate-independent sticking and velocity-dependent sliding. In the reduced-order model, we retained the time dependence and non-smoothness by introducing homogenized axial force–displacement relations parameterized by local sheath curvature and effective normal pressure, fitted to a subset of high-fidelity runs. For staged release, both models include a representation of capsule mouth geometry and its interaction with the distal struts, though the reduced model treats the frame as a lumped elastic ring.

On the bench, the deployment-force signatures are distinctive between the two mechanisms. In unsheathing, force traces show quasi-periodic micro-oscillations with amplitudes of 0.1–0.6 N superimposed on an increasing trend up to 3–6 N depending on lubricity and size, with clustered larger breaks near curvature changes. In staged release, the first step produces a transient that can exceed 1–2 N above the baseline, decaying over 50–150 ms, with a secondary smaller peak if residual constraint persists. Both are important for operator perception and positional control.

Mechanism modeling fidelity — gradation.

a.
Mechanisms listed but not mapped to model constructs.

b.
Mechanisms mapped qualitatively to model constructs with assumed parameters.

c.
Mechanisms mapped to explicit constitutive choices with parameter sources and sensitivity screening.

d.
Mechanisms mapped to explicit constitutive choices, parameter sources, sensitivity quantification, and targeted experiments separating their signatures.

4. Computational models: formulations, discretization, and outputs

We constructed two explicit dynamic FEA models. Model-HF is a 3D high-fidelity model resolving detailed contact surfaces for the outer sheath inner lumen, capsule mouth, and the outer surface of the crimped nitinol frame and delivery shaft. The frame is represented with superelastic constitutive behavior using a rate-insensitive Auricchio-type model fitted to coupon and ring expansion tests at body temperature. The capsule and sheath are modeled as nearly incompressible polymeric solids with hyperelasticity (Yeoh model) for the inner layer to capture compliance under contact pressure. The solver employs central-difference time integration with automatic time incrementation based on the smallest element characteristic length and wave speed, subject to a minimum critical time step target of 2.5e-7 s in contact-rich phases. Contact is enforced with a penalty formulation; friction uses a regularized Coulomb model with a Stribeck-like curve, allowing for a static coefficient μs and kinetic coefficient μk parameterized by relative sliding speed. In this work, μs ranged from 0.16 to 0.26 and μk from 0.10 to 0.18 across lubricity states, with a transition velocity v0 of 1.5 mm/s determined from ring-on-sheath tribometry.

The Model-HF mesh consists of approximately 3.2 million linear brick and wedge elements for the sheath and capsule, with local refinement to 0.08–0.12 mm near the capsule mouth and along the inner lumen where contact pressures peak; the crimped frame is meshed with 540,000 reduced-integration brick and shell elements with characteristic lengths of 0.06–0.10 mm in high-curvature regions. The contact surface discretization yields roughly 220,000 slave nodes on the sheath lumen with a corresponding number on the frame/capsule side. Hourglass control and bulk viscosity are tuned to avoid non-physical modes without materially affecting global energy balance.

Model-RO is a reduced-order explicit submodel intended to accelerate parametric design studies. The sheath, capsule, and frame are represented with homogenized 1D and 2D elements. The axial interaction between the sheath inner surface and the crimped frame is represented by a series of nonlinear axial connector elements whose force–displacement relations are functions of local curvature and an effective normal pressure proxy derived from frame pre-stress. These relations are fitted via least-squares to a library of 36 high-fidelity subproblems extracted from Model-HF by constraining a representative arc length at prescribed curvatures (40, 70, 100 mm) and lubricity states. The capsule mouth is modeled as a compliant ring (beam elements) interacting with the frame ring via a contact proxy with a rate-limited release stiffness. Time integration remains explicit with a target time step of 8.0e-7 s, and mass scaling is permitted up to 1.30 to achieve feasible runtimes when using coarse discretizations. Frictional stick–slip is emulated by a rate-weakening branch in the connector law that reproduces oscillation amplitudes and frequencies observed in the library runs.

Both models prescribe the sheath retraction displacement history at the handle end via a kinematic boundary condition, matching the bench-retraction sequence: a constant-speed unsheathing segment at 50 mm/s for 100 mm of travel, followed by a staged release comprising a 200 ms hold, a 4 mm retraction to initiate the first release, a 100 ms hold, and then a final retraction. Outputs are recorded as time-resolved axial force at the handle location, synchronized to sheath displacement and annotated with the timing of holds and releases. We report peak forces during unsheathing and at the first stage of release, as well as statistical descriptors (mean absolute percentage error, time-series root-mean-square error) in comparison to bench measurements.

To ensure the models are suitable for the stated outputs, we undertook sensitivity screening. In Model-HF, peak unsheathing force was most sensitive to μs (Fpeak/μs 16–22 N per unit μs across sizes), the ratio μs/μk, and local mesh resolution in the capsule mouth region within a 0.2 mm band; peak recoil force was sensitive to capsule mouth flare angle and the local stiffness of the mouth. In Model-RO, the fitted connector law parameters and the rate-limited release stiffness in the capsule proxy dominated the outputs. These findings guided mesh refinement and calibration prioritization.

Model detail selection — gradation.

a.
Model detail chosen by convenience without sensitivity analysis.

b.
Model detail justified by experience and coarse sensitivity

[redacted]

Figure 1: Schematic of the delivery system and bench configuration: a self-expanding frame crimped in a distal capsule on an inner shaft, an outer sheath retracted by a leadscrew-actuated handle instrumented with a load cell, and a curved surrogate vessel segment establishing a representative path curvature. The two force-generation pathways—unsheathing friction with stick–slip and staged-release recoil—are annotated on example force–displacement traces.

checks on a subset of parameters.

c.
Model detail justified by systematic sensitivity screening on parameters most likely to influence the outputs of interest.

d.
Model detail justified by systematic sensitivity and uncertainty ranking with design of experiments and convergence of sensitivities across discretizations.

5. Risk-informed credibility plan and acceptance criteria

We adopted a private, ordinal Credibility Index (CI) from 0 to 5 to grade the strength of evidence per factor for each model–mechanism pair. The mapping from risk to required evidence follows the ASME V&V 40 framework as applied internally. Given the high decision risk, we designated three factors as critical for acceptance: (a) algorithm implementation checks, (b) time-integration and contact solution accuracy, and (c) agreement with bench measurements on the outputs of interest. Three additional factors were considered non-critical but still required to be above a minimal threshold for confidence: (d) alignment of the testing and simulations to the intended use, (e) software process controls governing tools and scripts, and (f) representativeness of the physical test set used to support comparison. Critical factors required CI $\geq$ 4.0 and non-critical factors required CI $\geq$ 3.5 for acceptance in the stated context of use.

The scoring rubric was defined before testing and analysis began, and was not adjusted post hoc. For agreement with bench measurements, the target tolerance on peak forces was set at ±10% for the high-fidelity model and ±12% for the reduced-order model in unsheathing, with time-series root-mean-square error (RMSE) targets at or below 0.3 N and 0.45 N respectively, reflecting anticipated model-form error in the reduced representation. For the staged release transient, the tolerance was ±12% for the high-fidelity model and ±15% for the reduced-order model, recognizing higher sensitivity to local capsule geometry and constraint. For solution accuracy, we sought converged results under time-step refinement such that further halving changed peaks by less than 4% (high-fidelity) and less than 6% (reduced-order), with kinetic-to-internal energy ratios below 5% in Model-HF and below 7% in Model-RO during the most dynamic phases. For implementation checks, we set thresholds on passing rates for a regression suite of discrete mechanics and contact benchmarks (>98% pass rate for Model-HF and >95% for Model-RO) and maintained known-solution errors within 1–2% in linear dynamics comparators.

The plan specified that the alignment factor would consider whether kinematics, geometry, and boundary conditions in the validation activities matched the intended clinical use to within 5–10% on key quantities: retraction rate, curvature, lubrication state, and staged-release timing. Software process controls would include configuration management of pre- and post-processors and the FE solver version, peer review and change control processes, and defect tracking and resolution rates. The test set factor would grade the coverage of device sizes and lubricity states, the number of replicates per condition, and the inclusion of relevant worst cases (e.g., high curvature, low lubricity).

We applied the plan to two mechanisms—unsheathing friction and staged-release recoil—considered separately for each model. This produced a grid of evaluations where scores reflect evidence specific to each pairing of model and mechanism. The summaries, including numeric bases for each score, are presented later as one table per model for readability, in keeping with the practice of splitting large assessments.

Acceptance threshold definition — gradation.

a.
No explicit thresholds; accept or reject by judgment.

b.
Qualitative thresholds not tied to measurable quantities.

c.
Quantitative thresholds defined for some factors and not linked to risk.

d.
Quantitative thresholds defined for all factors and explicitly linked to risk and outputs.

6. Algorithm implementation checks: methodology and outcomes across models and mechanisms

Checks for algorithm implementation encompassed three tiers. First, we executed a regression suite of discrete mechanics tests covering linear momentum conservation in free-flight, contact patch tests under normal loading (Hertzian contact comparators), frictional sliding of a rigid block on a compliant half-space with known traction distributions, and a standard patch test for low-order bricks and shells. Second, we evaluated temporal accuracy on linear oscillators—single-degree-of-freedom and a simple beam bending mode—where the explicit integrator's period error and energy behavior can be compared with analytical solutions. Third, we exercised the contact algorithm under frictional stick–slip using a ring-on-liner tribology comparator with a regularized Coulomb friction law, checking for monotonic convergence of traction–slip hysteresis with decreasing regularization. For Model-HF, these checks were performed directly in the commercial explicit solver with the intended contact and friction options; for Model-RO, they were performed in the same solver for contact proxy elements and in the in-house library for connector laws.

In Model-HF, the patch tests passed to within 0.3% stress imbalance for bricks and 0.5% for shells. The free-flight linear momentum conservation test showed cumulative error less than 0.05% over 0.1 s at a 2.0e-6 s time step. The Hertzian contact comparators reproduced normal load–deflection curves within 0.6% of analytical predictions over seven load levels. The frictional sliding test recovered the expected transition from sticking to sliding at μN, with peak tangential traction errors within 0.8% for $\mu = 0.2 \pm 0.02$ and with a Stribeck regularization parameter diminished to $v0 = 0.5$ mm/s, demonstrating that the implementation approaches true Coulomb behavior in the regularization limit. The linear oscillator tests gave period errors of 0.7% and amplitude decay within 1.1% over 100 cycles when using a bulk viscosity tuned to 1.5% of the critical level, in line with expectations for central-difference integration with mild damping. These outcomes were recorded for both double-precision and mixed-precision runs, with no significant divergence observed.

In Model-RO, the connector element library was exercised via unit tests comparing linear and nonlinear force–displacement relations against exact expressions and against target fits from the high-fidelity subproblems. Ninety-six percent of 142 unit tests passed on the first run; the remaining failures involved edge conditions at high curvature parameterization and were corrected in a subsequent patch. The contact proxy implementation was compared against an elastic ring–ring contact model; load–deflection curves matched within 1.1–1.4% over the expected range. However, because the reduced model relies on fitted relations, the method of manufactured solutions was not directly applicable, reducing our ability to bound implementation error in those parts. Linear oscillator comparators on the explicit time integration yielded period errors of 1.2–1.5% depending on mass scaling levels.

These checks support confidence that numerical errors observed in application runs are dominated by discretization and solution accuracy rather than by implementation faults in the core algorithms. Where fits are used in Model-RO, the implementation checks validate faithful evaluation of the intended relations, but cannot eliminate model-form error inherent to homogenization.

Algorithm implementation checks — gradation.

a.
No benchmarks or unit tests of algorithmic components.

b.
Basic unit tests or vendor certification of solver without application-specific checks.

c.
Documented regression suite of mechanics and contact comparators with quantitative tolerances.

d.
Regression suite plus application-specific frictional stick–slip comparators and precision audits with pass criteria.

7. Time-integration and contact solution accuracy: time-step sensitivity, mass scaling audits, and energy balance

We assessed numerical solution accuracy via systematic time-step refinement, audits of permitted mass scaling, and monitoring of kinetic-to-internal energy ratios during dynamic phases. In Model-HF, we started from a baseline time step dictated by the smallest contact element and wave speed, typically 2.5e-7 s during unsheathing and as low as 1.8e-7 s during staged release when local contact stiffness peaks. We halved the time step twice (to 1.25e-7 s and 6.25e-8 s) on representative runs for each mechanism and device size. For unsheathing, halving the time step changed the peak handle force by 5.6% on average across sizes (range 4.7–6.1%); a second halving changed it by an additional 2.1% (range 1.6–2.8%). Kinetic-to-internal energy ratios remained below 4.2% at peak oscillations, and total energy drift over the window of interest was under 0.9%. Mass scaling was not used in Model-HF except for limited preloading phases (effective mass increase below 1.05), and we confirmed that enabling and disabling this small scaling did not materially change outputs (difference < 0.4% in peaks).

For staged release in Model-HF, the dynamics were more localized but better behaved under refinement. The first halving of the time step changed the peak handle force by 3.2% on average (range 2.4–3.7%); the second halving changed it by 1.0% (range 0.7–1.3%). Kinetic-to-internal energy ratios remained below 3.0% during the transient peak, and total energy drift was under 0.6%. These results indicate that the staged-release peaks are closer to time-step convergence under our settings than the unsheathing oscillations, which are influenced by the regularization of frictional transitions and by the interaction of many local contact micro-events.

In Model-RO, the baseline time step was 8.0e-7 s, with mass scaling permitted up to a factor of 1.30 to maintain practical runtimes. For unsheathing, halving the time step to 4.0e-7 s and then to 2.0e-7 s reduced peak handle force by 8.9% and 3.2% respectively on representative runs (ranges 7.6–10.1% and 2.6–3.9%). Kinetic-to-internal energy ratios peaked at 6.1% during the most active stick–slip segments. Mass scaling at 1.15 resulted in a 1.1% increase in peak force relative to an unscaled comparator; at 1.30, the increase was 2.2%, and we used this as a bound in sensitivity checks. For staged release, halving the time step changed peak forces by 12.7% and 4.6% on average; kinetic-to-internal energy ratios rose to 8.4% at the most rapid release step when the capsule proxy stiffness dominated; mass scaling at 1.30 produced a 3.4% increase in peaks versus unscaled runs.

We also examined energy balance qualitatively for each mechanism. In unsheathing, the frictional dissipation accounts for much of the work; the explicit solver reported hysteretic energy in friction consistent with the integrated traction–slip histories within 2.5% in Model-HF. In Model-RO, the connector elements' hysteretic energy matched the intended fit within 4.1%, consistent with the higher dynamic energy ratios observed.

Together these audits indicate that solution error is a significant contributor to discrepancies in Model-RO, particularly in staged release where the effective stiffness of the capsule proxy and the allowed mass scaling exacerbate time discretization effects. In Model-HF, unsheathing peaks exhibit modest residual sensitivity at acceptable levels for some uses, whereas staged-release peaks are near the numeric targets.

Time-integration and contact solution accuracy — gradation.

a.
Single time step with no refinement or energy monitoring.

b.
Two-level refinement or limited energy monitoring without quantitative targets.

c.
Multi-level refinement to quantitative change targets and energy ratio monitoring.

d.
Multi-level refinement with change targets met, energy balance within limits, and mass-scaling impact bounded.

8. Agreement with bench measurements: rig design, metrics, and tolerances

We designed and executed bench tests to provide time-resolved force measurements for comparison with model outputs. The rig comprises a linear actuator driving the sheath retraction at prescribed displacement profiles, instrumented with a 10 N full-scale load cell (resolution 0.005 N, bandwidth 500 Hz) at the handle location and an optical encoder for sheath displacement (resolution 0.02 mm). The distal assembly traverses a curved surrogate vessel with selectable radius inserts (40, 70, 100 mm) to emulate iliofemoral path curvature. A capsule-mouth alignment jig ensures concentricity within 0.1 mm and angular offset below 0.3 degrees. Tests were conducted in two lubricity states: freshly activated hydrophilic and a partially dried state produced by 20 minutes of air exposure after activation. The environment was maintained at 37 ± 0.5 °C in a saline bath to approximate in vivo conditions.

The retraction sequence was displacement-controlled, matching the model boundary conditions: a 50 mm/s unsheathing retraction for 100 mm of travel; a 200 ms hold; a 4 mm rapid retraction to initiate staged release; a 100 ms hold; and a final retraction until the frame was fully released. Each condition (size, curvature, lubricity) was repeated 3–5 times to characterize variability. The primary quantitative metrics were: (1) percent error in peak handle force during unsheathing, defined as the mean absolute percentage error (MAPE) between model and bench peak across repeats; (2) percent error in peak handle force at the first staged release; and (3) time-series RMSE over the unsheathing and staged-release windows, computed after aligning to the start of retraction and to the hold–release transitions.

For Model-HF, in unsheathing, across three sizes and two lubricity states at 70 mm curvature, the MAPE in peak force was 9.2% (range 7.8–10.6%), and the time-series RMSE was 0.28 N (range 0.22–0.33 N). At 40 mm curvature, the MAPE increased modestly to 10.8% (range 9.5–12.1%) with RMSE 0.34 N (range 0.27–0.41 N), driven by stronger stick–slip amplitudes. Under these conditions, two of the three sizes met the ±10% peak tolerance, while the smallest size slightly exceeded it at the tightest curvature. For staged release, at 70 mm curvature, the MAPE in peak force was 17.5% (range 14.9–19.6%) with RMSE 0.39 N (range 0.31–0.47 N). The model tended to underpredict the primary transient by 0.25–0.45 N, a bias correlated with capsule-mouth flare angle tolerances measured post-test.

For Model-RO, in unsheathing at 70 mm curvature, the MAPE in peak force was 13.4% (range 11.7–15.2%), RMSE 0.41 N (range 0.33–0.49 N), with a consistent phase lead of 10–20 ms in the onset of larger stick–slip events relative to the bench. At 40 mm curvature, MAPE rose to 16.0% (range 14.4–17.7%), with RMSE 0.50 N (range 0.42–0.57 N). In staged release, the MAPE in peak force was 22.8% (range 19.3–25.6%) with RMSE 0.62 N (range 0.53–0.70 N), and a bias toward underprediction of peak magnitude. The phase of the transient aligned within 15 ms on average, indicating the kinematic timing was appropriately captured while the magnitude lagged.

The tolerance bands were defined a priori: ±10% on unsheathing peaks and 0.30 N RMSE for Model-HF; ±12% and 0.45 N for Model-RO. For staged release, the tolerances were ±12% (Model-HF) and ±15% (Model-RO). Under these, Model-HF met tolerance in unsheathing for most conditions but not in staged release; Model-RO fell short in both mechanisms at tighter curvature and in staged release broadly. We attribute portions of the disagreement to residual numerical sensitivity (above) and to unmodeled variability in capsule-mouth micro-geometry and coating state that were not fully captured in the test set.

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

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

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

Agreement with bench measurements — gradation.

a.
Qualitative comparison on a small sample with no predefined metrics.

b.
Quantitative comparison on limited metrics without predefined tolerances.

c.
Quantitative comparison on predefined metrics with tolerances and coverage of primary conditions.

d.
Quantitative comparison on predefined metrics with tolerances, coverage of primary and worst-case conditions, and bias analysis.

9. Alignment of test conditions and boundary conditions with intended use

We examined the degree to which the validation activities—bench tests and corresponding simulations—matched the intended clinical use on the quantities that matter most for force prediction. Kinematic alignment was ensured by imposing identical displacement histories on the rig and in the models; encoder logs from the rig confirmed adherence within ±2% of the target retraction rate and within ±5 ms of the timing of holds and staged retraction steps. Geometric alignment focused on the curvature of the delivery path; we selected radii of 40, 70, and 100 mm for inserts, covering the 20th to 80th percentiles of iliofemoral curvature observed in a 96-patient preclinical image dataset. Lubricity alignment was managed by testing both hydrophilic activation states, though we recognize that in vivo water exposure and temperature histories can differ; tribometer tests on conditioned coupons suggested our activated and partially dried states span the middle 70% of measured friction coefficient ranges.

Boundary constraints at the capsule mouth were controlled via an alignment jig; post-test metrology indicated that concentricity was within 0.1 mm, but slight burrs or micro-chamfer variations of up to 0.03 mm arose during repeated insertions. Simulations used as-manufactured capsule-mouth geometries measured by micro-CT on representative devices; we did not update the model geometry per specimen. For staged release, fixture compliance at the distal end was modeled as a linear spring with stiffness fitted from independent bending tests on the rig (4.3 N/mm). However, during staged release, we measured a 0.2 s hold-time drift in some rig sequences due to controller latency, which we captured in the simulation inputs for those tests but could not eliminate entirely in all repeats.

The overall alignment reached our targets on kinematics and covered key geometric and lubricity states, but the residual differences in capsule-mouth micro-geometry and in fixture compliance during staged release likely contributed to discrepancies in peak force magnitudes, particularly for the recoil mechanism where sensitivity to local constraint is high.
Alignment to intended use — gradation.

a.
Validation conditions differ materially from intended use without quantification.

b.
Primary quantities aligned; secondary quantities differ without quantified impact.

c.
Primary and secondary quantities aligned within predefined tolerances; residual differences documented.

d.
Primary and secondary quantities aligned within tolerances with quantified impact analysis of residual differences.

10. Software process controls for tools and scripts

The analysis toolchain includes a commercial explicit FEA solver, custom preprocessing scripts to generate meshes and boundary conditions, and postprocessing scripts for force extraction and metric computation. For the commercial solver, we operated under a validated version with vendor-issued maintenance updates tracked in our configuration database. For preprocessing and postprocessing, we used a Git-based repository with branch protection, mandatory code review by at least one peer and one senior reviewer for changes exceeding 50 lines, and automated continuous integration (CI) that runs a regression suite on 112 cases spanning meshing of capsule mouth geometries, application of friction coefficients, and extraction of force traces. Code coverage measured by line count is 82% for preprocessing and 76% for postprocessing. We maintained an issue tracker with severity levels; at the time of assessment, there were zero high-severity open issues, three medium-severity issues (two related to rare meshing topologies and one to plotting rounding behavior), and nine low-severity issues.

For the reduced-order model, which relies on in-house libraries implementing connector laws and contact proxies, we managed the repository under the same Git workflow but with a smaller regression suite of 47 unit and integrated tests. Code coverage was 58%. Two medium-severity issues remained open during the assessment window: one concerning parameter interpolation at curvature values outside the fit library and one concerning unexpected oscillation amplification under high mass scaling; both had documented workarounds and were risk-assessed as non-critical for the test cases executed here. We also performed periodic dependency audits to ensure consistent versions of numerical libraries across developer machines and the CI environment.

Peer review records and CI logs are archived and traceable to the analyses reported here. The commercial solver's pre- and post-release verification reports were cross-referenced, and we verified that no known defects affected the contact or explicit integration modules used in this work.
Software process controls — gradation.

a.
Ad hoc scripting without version control or review.

b.
Version control with informal review and no automated testing.

c.
Version control with documented review and automated regression tests covering key functions.

d.
Formal change control, automated regression with coverage metrics, defect tracking, and dependency audits.

11. Physical test set: selection, representativeness, and coverage

The physical test set was selected to span device sizes, lubricity states, and path curvatures that dominate the deployment-force response. For unsheathing, we executed a total of 18 tests for Model-HF comparisons across three device sizes (small, medium, large), two lubricity states (activated and partially dried), and a central curvature (70 mm), with three replicates per condition. We added 6 tests at tighter curvature (40 mm) for the small and medium sizes to probe worst-case unsheathing behavior. For staged release, we executed 9 tests across two device sizes (medium, large) at 70 mm curvature with activated lubricity, and 4 tests at 40 mm curvature for the medium size.

For reduced-order comparisons, given the model's intended use for rapid screening, we selected a leaner set: 12 unsheathing tests covering two sizes (medium, large), both lubricity states, and two curvatures (70 mm and 40 mm), with three replicates per condition; and 8 staged-release tests focusing on the medium size at 70 mm curvature with activated lubricity and on 40 mm curvature with activated lubricity, with four replicates for the former and four for the latter.

While these selections ensured coverage of primary conditions, gaps remained. In particular, staged-release testing did not include the small size, and we did not include a partially dried lubricity state for staged release due to concerns about coating integrity under repeated high-energy transients in that state. For unsheathing at 100 mm curvature, we relied on historical data to inform sensitivity bounds but did not execute new runs in the present campaign due to resource constraints. The combination of these choices led to moderate confidence in unsheathing coverage and lower confidence in staged-release coverage.
Physical test set coverage — gradation.

a.
Single size and condition; no replicates.

b.
Multiple sizes or conditions with limited replicates and no worst cases.

c.
Multiple sizes and conditions with replicates and at least one worst case.

d.
Comprehensive sizes and conditions with replicates, worst cases, and rationale for exclusions.

12. Use error considerations: operator training and procedural consistency

We briefed technicians on standard operating procedures for sheath activation, retraction profile programming, and capsule-mouth alignment prior to the campaign. Two training sessions were conducted with mock runs to confirm repeatability, and a laminated checklist was used at the rig. The same two senior operators conducted all tests to minimize variability due to handling. We noted occasional deviations in retraction pause timing on early runs that were resolved by reprogramming the actuator controller firmware. Operator notes recorded qualitative impressions of resistance and transient feel, aiding post hoc interpretation of anomalies.
Operator consistency — gradation.

a.
No training; multiple operators without documentation.

b.
Training provided; procedures documented but adherence not tracked.

c.
Training and checklists with adherence logs and limited operator pool.

d.
Training, competency assessment, adherence logs, and cross-operator equivalence testing.

13. Synthesis of factor scores for Model-HF by mechanism

For Model-HF in unsheathing friction and stick–slip, algorithm implementation checks were strong across discrete mechanics and contact comparators, supporting a CI of 4.5. Time-integration accuracy under halved time steps produced a 5.6% change in peak force on average, with energy ratios below 4.2% and negligible mass scaling, resulting in CI 3.8. Agreement with bench traces achieved a MAPE of 9.2% and RMSE of 0.28 N at 70 mm curvature, with modest degradation at 40 mm curvature; we assigned CI 3.6. Alignment of test and model conditions was within target on kinematics and lubrication states, with residual uncertainties in capsule-mouth micro-geometry; CI 4.2. Process controls for tools and scripts were robust with high coverage and no high-severity open issues; CI 4.0. The physical test set provided moderate coverage across sizes and lubricity states with limited worst-case runs; CI 3.5.

For Model-HF in staged-release recoil, algorithm checks remained strong (CI 4.5). Time-integration accuracy improved relative to unsheathing, with peak changes of 3.2% under halving and energy ratios below 3.0%, yielding CI 4.1. Agreement with bench measurements fell short due to underprediction of the primary transient (MAPE 17.5%, RMSE 0.39 N); CI 2.9. Alignment was acceptable on kinematics and curvature but limited by fixture compliance differences and unmodeled capsule micro-geometry variability; CI 3.8. Process controls remained strong (CI 4.0). The test set coverage was lean for staged release (no small size and limited lubricity states); CI 3.0.
Score synthesis transparency — gradation.

a.
Scores stated without numeric basis.

b.
Scores stated with qualitative rationale.

c.
Scores stated with quantitative basis for representative conditions.

d.
Scores stated with quantitative basis for each mechanism and explicit linkage to acceptance thresholds.

14. Synthesis of factor scores for Model-RO by mechanism

For Model-RO in unsheathing friction, algorithm implementation checks were acceptable (CI 3.7) with 96% unit-test pass rates and contact-proxy comparators within 1.1–1.4%, acknowledging the limits of testing fitted relations. Time-integration accuracy showed 8.9% change in peaks on the first halving and 3.2% on the second, with kinetic-to-internal energy ratios peaking at 6.1% and permitted mass scaling up to 1.30; CI 3.3. Agreement with bench measurements resulted in 13.4% MAPE and 0.41 N RMSE under central curvature, with phase lead of 10–20 ms in larger stick–slip events; CI 3.1. Alignment to intended use was good on kinematics and curvature coverage and reasonable on lubricity states; CI 3.9. Process controls for the in-house libraries were moderate with 58% coverage and two medium-severity open issues; CI 3.5. The physical test set coverage was lean for this model–mechanism pairing, with N=12 tests across two sizes and two curvatures; CI 2.8.

In staged-release recoil for Model-RO, algorithm checks were weaker (CI 3.5) due to the absence of a dedicated recoil-energy comparator and heavier reliance on fitted parameters. Time-integration accuracy was insufficient, with 12.7% and 4.6% changes under successive halving and kinetic-to-internal energy ratios up to 8.4% at peak; mass scaling at 1.30 introduced a 3.4% change in peak; CI 2.9. Agreement with bench measurements was poor (CI 2.4), with MAPE 22.8% and RMSE 0.62 N and a magnitude bias toward underprediction. Alignment remained acceptable (CI 3.5), though the hold–release timing in several bench runs drifted by 5–8 ms beyond the target and was only partially reproduced. Process controls were unchanged (CI 3.5). The test set for this pairing was minimal (N=8) and concentrated in one size; CI 2.5.
Score synthesis transparency — gradation.

a.
Scores stated without numeric basis.

b.
Scores stated with qualitative rationale.

c.
Scores stated with quantitative basis for representative conditions.

d.
Scores stated with quantitative basis for each mechanism and explicit linkage to acceptance thresholds.

15. Decision and implications

This assessment was designed to support an internal decision on whether to rely on explicit FEA predictions of handle force to determine if design variants meet an internal deployment-force limit. Because the decision has high risk, we required that all critical factors meet CI ≥ 4.0 and all non-critical factors meet CI ≥ 3.5 for each model–mechanism pair within the intended domain. The results do not meet those thresholds in several places, and the pattern of shortfall has direct implications for both the modeling approach and the experimental program that underpins it.

Across the dataset, the checks designed to reveal defects in algorithm implementation were strong for the detailed model and acceptable though not comprehensive for the reduced-order model. This indicates that the principal contributors to error in application runs are not gross coding faults but rather discretization settings and model reduction choices.

Convergence under time-step halving and contact solution monitoring was not uniformly at the target level, particularly for the reduced-order construction in staged release where changes exceeded 10% on the first halving and dynamic energy ratios were elevated. Without further refinement or adjustments to mass scaling policies, numerical solution error remains a material source of uncertainty in peaks that drive design decisions.

The comparison of model outputs to bench measurements met pre-set tolerances for unsheathing with the detailed model in most conditions but fell short for staged release; the reduced-order model missed the tolerances in both mechanisms at tight curvature and during release. The magnitude biases and RMSEs observed are sufficient to change pass/fail conclusions for some design variants if taken at face value.

The degree of match between the validation setup and intended use was generally within targeted tolerances on kinematics and curvature and captured the two lubrication states of interest, though residual differences in capsule-mouth micro-geometry and fixture compliance during staged release were present. These differences likely contributed to observed discrepancies in the recoil transient.

Process controls for the commercial toolchain and custom scripts were robust for the detailed model and moderate for the reduced one, with version control, review, and regression tests in place and only a few medium-severity issues outstanding in the reduced-order codebase.

The breadth of the physical test set was moderate for unsheathing but sparse for staged release, with undersampling of smaller sizes and omission of dried-lubricity release trials due to coating integrity concerns. The limited coverage in the latter regime reduces confidence in generalizing findings across sizes and conditions.

Taken together, these observations mean that the models, as presently configured and supported, cannot be used to make the stated decision without an unacceptable risk of error. The most immediate remediation would be to improve numerical convergence behavior in the reduced-order model, expand and tighten staged-release testing (including micro-geometry characterization per specimen), and recalibrate the capsule-mouth interaction in both models. Until such remedial actions are complete and shown to raise the critical factors at or above their thresholds, the models should not be used for the specified context of use.

Therefore, the models are not accepted for the stated context of use.

Decision logic maturity — gradation.

a.
Decision stated without linkage to factors.

b.
Decision linked to some factors qualitatively.

c.
Decision linked to factor thresholds with quantitative references.

d.
Decision linked to all factor thresholds, integrates risk, and prescribes remediation aligned to observed shortfalls.

References

[1] R. Abbott, S. Lindqvist, Material model effects on shear stress in computational models, Ann. Biomed. Eng. 47 (2015) 221–736.

[2] C. Vasquez, L. Wojcik, L. Tanaka, L. Vasquez, Wall roughness effects on cortical strain in computational models, Comput. Methods Biomech. 35 (2014) 228–656.

[3] B. Ishikawa, M. Vasquez, R. Bergman, L. Tanaka, Solver tolerance effects on cortical strain in computational models, Ann. Biomed. Eng. 64 (2010) 229–795.

[4] S. Lindqvist, B. Grigoryan, R. Bergman, Inlet profile effects on cortical strain in computational models, Ann. Biomed. Eng. 36 (2023) 258–793.

[5] J. Vasquez, E. Lindqvist, W. Silva, Inlet profile effects on chest compression in computational models, Med. Eng. Phys. 45 (2009) 327–475.

[6] L. Yilmaz, J. Nakamura, Contact stiffness effects on hemolysis index in computational models, Comput. Fluids 35 (2018) 364–624.

[7] G. Grigoryan, R. Rasmussen, Damping effects on head displacement in computational models, Int. J. Numer. Methods 49 (2018) 364–661.

[8] K. Abbott, G. Bergman, Gap size effects on flow split in computational models, J. Biomech. Eng. 123 (2011) 210–494.

[9] N. Nakamura, H. Abbott, R. Rasmussen, K. Eriksen, S. Pettersson, Material model effects on peak force in computational models, Int. J. Numer. Methods 89 (2018) 219–453.

[10] C. Bergman, A. Castellano, D. Tanaka, Contact stiffness effects on head displacement in computational models, Ann. Biomed. Eng. 58 (2011) 358–512.

[11] L. Silva, L. Bergman, M. Abbott, Material model effects on contact pressure in computational models, J. Mech. Behav. Biomed. Mater. 112 (2014) 386–571.

[12] S. Bergman, M. Eriksen, A. Abbott, R. Halvorsen, L. Ishikawa, Solver tolerance effects on outlet velocity in computational models, J. Biomech. Eng. 20 (2010) 332–750.

[13] M. Grigoryan, R. Jankowski, L. Silva, L. Tanaka, C. Kaur, Gap size effects on head displacement in computational models, Comput. Methods Biomech. 77 (2012) 348–621.

[14] C. Lindqvist, F. Lindqvist, H. Lindqvist, D. Zetterberg, Time step effects on flow split in computational models, J. Verif. Valid. Uncertain. 30 (2008) 89–497.

[15] N. Zetterberg, S. Grigoryan, H. Eriksen, L. Moreau, S. Okonkwo, Boundary condition effects on head displacement in computational models, J. Biomech. Eng. 42 (2006) 218–688.

[16] J. Eriksen, E. Zetterberg, Friction effects on peak force in computational models, J. Mech. Behav. Biomed. Mater. 44 (2024) 231–556.

[17] F. Ueda, N. Fontaine, Friction effects on flow split in computational models, Med. Eng. Phys. 63 (2014) 136–505.

[18] E. Grigoryan, H. Okonkwo, Mesh density effects on shear stress in computational models, Int. J. Numer. Methods 59 (2009) 116–811.

[19] S. Grigoryan, T. Tanaka, Flow rate effects on flow split in computational models, J. Mech. Behav. Biomed. Mater. 29 (2012) 115–847.

[20] B. Silva, J. Yilmaz, S. Moreau, Inlet profile effects on flow split in computational models, Comput. Methods Biomech. 118 (2024) 238–828.

[21] T. Eriksen, N. Eriksen, Flow rate effects on cortical strain in computational models, Med. Eng. Phys. 20 (2020) 273–442.

[22] E. Nakamura, P. Abbott, B. Moreau, A. Okonkwo, Damping effects on shear stress in computational models, Comput. Fluids 117 (2009) 118–620.

[23] M. Zetterberg, D. Silva, Flow rate effects on hemolysis index in computational models, J. Biomech. Eng. 104 (2024) 176–692.

[24] D. Moreau, C. Castellano, S. Castellano, T. Silva, Contact stiffness effects on contact pressure in computational models, J. Verif. Valid. Uncertain. 84 (2010) 281–679.

[25] G. Moreau, F. Tanaka, N. Eriksen, A. Moreau, Mesh density effects on flow split in computational models, J. Mech. Behav. Biomed. Mater. 120 (2019) 374–445.

[26] N. Bergman, W. Yilmaz, B. Jankowski, J. Quintero, Flow rate effects on outlet velocity in computational models, J. Mech. Behav. Biomed. Mater. 123 (2015) 234–483.

[27] E. Zetterberg, W. Vasquez, H. Okonkwo, P. Nakamura, L. Pettersson, Cement mantle effects on rib deflection in computational models, Comput. Fluids 34 (2016) 96–442.

[28] D. Silva, G. Kaur, R. Yilmaz, N. Yilmaz, S. Nakamura, Wall roughness effects on flow split in computational models, J. Biomech. Eng. 110 (2023) 369–414.

[29] B. Eriksen, K. Fontaine, G. Eriksen, Inlet profile effects on peak force in computational models, Med. Eng. Phys. 111 (2008) 93–416.

[30] S. Grigoryan, G. Wojcik, Contact stiffness effects on shear stress in computational models, J. Verif. Valid. Uncertain. 41 (2011) 255–801.

[31] F. Ueda, S. Ueda, K. Lindqvist, Damping effects on outlet velocity in computational models, J. Biomech. Eng. 105 (2023) 67–446.

[32] M. Zetterberg, G. Bergman, E. Pettersson, Inlet profile effects on chest compression in computational models, J. Verif. Valid. Uncertain. 73 (2009) 298–661.

[33] J. Rasmussen, E. Ishikawa, G. Lindqvist, N. Okonkwo, Boundary condition effects on shear stress in computational models, Ann. Biomed. Eng. 80 (2005) 306–652.

[34] F. Yilmaz, W. Bergman, Solver tolerance effects on contact pressure in computational models, J. Verif. Valid. Uncertain. 62 (2015) 205–474.

Table 1: Credibility assessment summary for Model-HF across two mechanisms using the 0–5 Credibility Index (CI).

Credibility factor	Level	Basis
Numerical code verification - Model-HF - Unsheathing friction	4.5	Patch and contact comparators within 0.3–0.8% error; frictional stick–slip comparator convergent; 100% regression pass.
Numerical solver error - Model-HF - Unsheathing friction	3.8	Halving dt changed peak by 5.6% (second halving 2.1%); KE/IE < 4.2%; mass scaling ≤ 1.05.
Output comparison - Model-HF - Unsheathing friction	3.6	MAPE in peak 9.2% (7.8–10.6%); RMSE 0.28 N; meets ±10% in 2/3 sizes at 70 mm.
Relevance of the validation activities to the COU - Model-HF - Unsheathing friction	4.2	Kinematics within ±2%; curvature 40–100 mm; lubricity states bracket middle 70%; capsule micro-geometry not per-specimen.
Software quality assurance - Model-HF - Unsheathing friction	4.0	Git with protected branches; 112-case CI; coverage 76–82%; 0 high-severity open issues; 3 medium open.
Test samples - Model-HF - Unsheathing friction	3.5	N=18 at 70 mm plus N=6 at 40 mm; 3 sizes, 2 lubricities; limited worst-case runs.
Numerical code verification - Model-HF - Frame recoil constraint	4.5	Same implementation checks as unsheathing; load–deflection and oscillator comparators within 0.7–1.1%.
Numerical solver error - Model-HF - Frame recoil constraint	4.1	Halving dt changed peak by 3.2% (second halving 1.0%); KE/IE < 3.0%; negligible mass scaling.
Output comparison - Model-HF - Frame recoil constraint	2.9	MAPE in peak 17.5% (14.9–19.6%); RMSE 0.39 N; systematic underprediction 0.25–0.45 N.
Relevance of the validation activities to the COU - Model-HF - Frame recoil constraint	3.8	Kinematics within ±2%/±5 ms; fixture compliance modeled 4.3 N/mm; capsule micro-geometry varied up to 0.03 mm.
Software quality assurance - Model-HF - Frame recoil constraint	4.0	Toolchain controls as above; no solver defects affecting contact or explicit integrator per vendor reports.
Test samples - Model-HF - Frame recoil constraint	3.0	N=9 across 2 sizes at 70 mm; N=4 at 40 mm; no small size and no dried lubricity for staged release.

Table 2: Credibility assessment summary for Model-RO across two mechanisms using the 0–5 Credibility Index (CI).

Credibility factor	Level	Basis
Numerical code verification - Model-RO - Unsheathing friction	3.7	96% of 142 unit tests passed; contact proxy matched ring–ring comparator within 1.1–1.4%; fitted connectors not MMS-tested.
Numerical solver error - Model-RO - Unsheathing friction	3.3	Halving dt changed peak by 8.9% (second halving 3.2%); KE/IE up to 6.1%; mass scaling 1.15–1.30 adds ≤ 2.2%.
Output comparison - Model-RO - Unsheathing friction	3.1	MAPE in peak 13.4% (11.7–15.2%); RMSE 0.41 N; phase lead 10–20 ms in stick–slip onsets.
Relevance of the validation activities to the COU - Model-RO - Unsheathing friction	3.9	Kinematics matched; curvatures 40/70 mm; lubricities tested; simplified geometry omits micro-geometry.
Software quality assurance - Model-RO - Unsheathing friction	3.5	Git with review; 47 tests; 58% coverage; 2 medium-severity issues open with workarounds.
Test samples - Model-RO - Unsheathing friction	2.8	N=12 across 2 sizes, 2 curvatures; 3 replicates per condition; limited size coverage.
Numerical code verification - Model-RO - Frame recoil constraint	3.5	No dedicated recoil comparator; beam comparator within 3.7%; time-integration period errors 1.2–1.5%.
Numerical solver error - Model-RO - Frame recoil constraint	2.9	Halving dt changed peak by 12.7% (second halving 4.6%); KE/IE up to 8.4%; mass scaling 1.30 adds 3.4%.
Output comparison - Model-RO - Frame recoil constraint	2.4	MAPE in peak 22.8% (19.3–25.6%); RMSE 0.62 N; magnitude bias to underprediction.
Relevance of the validation activities to the COU - Model-RO - Frame recoil constraint	3.5	Kinematics matched within ±5 ms; simplified boundary compliance; hold–release drift partially reproduced.
Software quality assurance - Model-RO - Frame recoil constraint	3.5	Same process controls; dependency audits passed; no solver defects implicated.
Test samples - Model-RO - Frame recoil constraint	2.5	N=8 focused on one size; 4 replicates per curvature; no dried lubricity staged-release tests.