Establishing credibility of fluid–structure interaction models for a bioprosthetic aortic valve: coaptation and leaflet stress across three mechanisms

R. S. Kline, PhD[a], M. A. Ortega, PhD[b], J. P. Bena, MS[c], A. L. Feldman, PhD[d]

[a] *Center for Computational Cardiac Mechanics, Meridian Institute of Technology, Boston, MA, USA,*
[b] *Department of Biomedical Engineering, Northwestern Plains University, Fargo, ND, USA,*
[c] *Cardiac Devices Laboratory, Mid-Atlantic Medical Research Center, Baltimore, MD, USA,*
[d] *School of Mechanical Engineering, Western Coastal University, San Diego, CA, USA,*

Abstract

Purpose: We developed and executed a credibility plan under ASME V&V 40 for three fluid–structure interaction (FSI) models of a surgical bioprosthetic tri-leaflet aortic valve challenged by three clinically relevant mechanisms—annular dilation, leaflet stiffening, and commissural misalignment. Models span a high-fidelity partitioned approach with anisotropic leaflet shells (Model A), an immersed-boundary scheme with reduced-order hemodynamics (Model B), and a monolithic ALE formulation with nonlinear fiber-reinforced leaflets (Model C). Each model computes closure and load-sharing metrics germane to regurgitation risk and leaflet durability. Methods: The plan evaluates six credibility topics—mesh/time-step sufficiency, equation-solver residuals and coupling robustness, adequacy of the governing physics and contact, fidelity of parameterization, representativeness of the test environment, and the breadth of device samples used to anchor and challenge the models—using a private 0–5 numeric scale with a priori acceptance thresholds per topic. Evidence is reported for each model × mechanism pairing, with convergence studies, solver diagnostics, validation against pulse-duplicator data, global and local parameter sensitivity, assessment of test-waveform similarity, and inter-device variability. Results: Across the three mechanisms, Model A and Model C demonstrated small output changes under grid and temporal refinement (≤1.5% for cycle-averaged coaptation area; 2.0–3.9% for peak stress), while Model B showed larger resolution sensitivity (3.8–8.3% for closure-gap and stress-proxy outputs). Nonlinear solvers in Models A and C achieved tight residual targets with minimal cycle-to-cycle energy drift (≤0.7%), whereas Model B exhibited modest mass-balance error (1.0–1.8%). When compared to pulse-duplicator measurements, Model A and Model C predicted diastolic leak volumes within 4–9% for annular dilation and leaflet stiffening, but deviations widened for commissural misalignment and for Model B (up to 36% error in leak and 46% in gap length). Coefficients of variation in measured outputs across devices ranged 12–18% with n = 5, 4, and 3 samples for dilation, stiffening, and misalignment, respectively, limiting statistical power for the latter. Conclusions: The evidence base is uneven across model–mechanism combinations. Multiple cases miss prespecified thresholds for at least one credibility topic that matters for the intended use, most often the adequacy of the governing physics in misalignment and the breadth of supporting samples. On balance, and in line with the a priori decision rules, the assessment outcome is Not accepted. We identify data and modeling priorities needed to reach acceptance for preclinical decision support: targeted bench measurements of out-of-plane leaflet strains and contact pressures under commissural misalignment, expanded sample sizes, and tighter characterization of leaflet microstructure to reduce parameter uncertainty.

Keywords: aortic valve, FSI credibility, ASME V&V 40

1. Introduction

Bioprosthetic aortic valves remain the predominant choice for surgical aortic valve replacement in older patients, with expanding use in younger cohorts due to improvements in hemodynamics and anticoagulation profiles. Leaflet coaptation integrity and stress distributions under diastolic loading dictate both acute competence and long-term durability. Computational fluid–structure interaction (FSI) models promise insight into design trade-offs and failure modes, but their utility depends on rigorous, transparent evaluation of credibility matched to the decision at hand.

The ASME V&V 40 standard defines a risk-informed process to establish the adequacy of computational models for medical device decisions. It emphasizes context of use, model risk, and evidence collection across topics that include numerical resolution, stability of the solvers, the faithfulness of the physics, appropriateness of the parameter values, and the representativeness of experiments that inform or challenge the model. Here we adopt that process for three distinct FSI approaches applied to the same surgical tri-leaflet bioprosthetic valve across three clinically motivated perturbations: annular dilation, leaflet stiffening, and commissural misalignment.

The three models differ in their treatment of the structure–fluid interface and of leaflet material behavior. Model A couples a finite-volume incompressible Navier–Stokes solver to a nonlinear shell finite-element model of anisotropic pericardial leaflets using a strong partitioned scheme with sub-iterations each time step. Model B implements an immersed-boundary method with reduced-order hemodynamics, using a homogenized leaflet membrane representation embedded in a Cartesian fluid grid to accelerate simulations. Model C solves a monolithic ALE FSI formulation with a hyperelastic, fiber-reinforced solid model and mortar-based contact during coaptation. Each model targets outputs that directly reflect closure performance and leaflet loading.

This article assembles evidence for each model × mechanism pairing under a uniform private 0–5 scoring scale with prespecified acceptance criteria. The result is a comprehensive, topic-by-topic accounting rather than a singular pass–fail declaration. Nevertheless, the conclusion follows the decision logic that maps these topic scores onto the intended use, culminating in a clear recommendation regarding acceptance.

Model form — gradation.

a.
Level 0–1: No comparison to relevant experiments; only qualitative plausibility arguments.

b.
Level 2: Limited pointwise or global comparisons under non-representative loadings; large unexplained deviations persist.

c.
Level 3: Quantitative comparison on at least one key output under representative conditions; errors documented but not localized.

d.
Level 4–5: Quantitative agreement across multiple outputs and mechanisms; discrepancies are within prespecified bounds with mechanistic explanations.

2. Question of Interest and Context of Use

The question of interest is whether the three FSI models can reliably quantify coaptation and leaflet stress for a surgical tri-leaflet bioprosthetic aortic valve under three mechanisms of concern: annular dilation, leaflet stiffening, and commissural misalignment. The context of use is preclinical design evaluation at the component level to support engineering decisions about design margins against regurgitation and fatigue-critical stress risers, not direct patient-specific therapy decisions.

The models compute outputs tailored to this context: cycle-averaged coaptation area, instantaneous gap metrics, leaflet contact pressure during closure, and principal or von Mises stress peaks. These outputs inform engineering criteria for competence (e.g., maximum allowable regurgitant volume or gap length) and durability (e.g., keeping principal stress and strain energy below thresholds linked to damage accumulation). The intended decisions involve design selection and mechanism ranking under lab-replicated conditions, where residual risk from model inadequacy is unacceptable if it could mask a regurgitation failure in diastole or underestimate stress concentrations that drive early leaflet damage.

Test conditions — gradation.

a.
Level 0–1: Test environments are unrelated to the intended use; key dimensionless groups are not matched.

b.
Level 2: Partial similarity is achieved for some metrics (e.g., mean flow), but transients and waveforms are off-target without quantification.

c.
Level 3: Primary waveform and dimensionless numbers are matched within broad tolerances; uncertainties are recorded.

d.
Level 4–5: Comprehensive similarity with documented bounds (e.g., pressure RMS error <5 mmHg, Womersley within 5%, compliance tuned); perturbation mechanisms faithfully reproduced.

3. Model Risk Analysis under ASME V&V 40

We rated the model risk as medium–high given that the outputs inform decisions about closure sufficiency and stress magnitudes that, if misrepresented, could yield false negatives for regurgitation and underpredict stress peaks at the free edge or commissures. Consequence of an incorrect decision was deemed moderate: while design decisions are preclinical and reversible, proceeding with a design that fails closure criteria or exhibits stress concentrations may cause costly redesign and delay. Uncertainty in boundary conditions and in leaflet material parameters, especially fiber orientation and thickness distribution, can produce meaningful shifts in outputs under the examined mechanisms.

Given this risk posture, the credibility plan emphasizes stringent assessments on numerical sufficiency and solution stability as necessary but not sufficient conditions, and places the most weight on the faithfulness of the governing physics and contact during closure, plus the representativeness and breadth of experimental data used to anchor and check the models. Particular attention is directed at commissural misalignment, where asymmetric gaps and contact stress localization are acute and where fewer lab data are typically available.

Model inputs — gradation.

a.
Level 0–1: Parameters adopted from unrelated literature without justification; no uncertainty bounds.

b.
Level 2: Parameters sourced from nominal vendor or literature values; wide bounds; no calibration.

c.
Level 3: Parameters measured or calibrated for the device family; key sensitivities quantified with first-order methods.

d.
Level 4–5: Device- and lot-specific measurements with hierarchical uncertainty models; global sensitivity and posterior predictive checks documented.

Figure 1: Overview of the aortic valve leaflet geometry, load cases and the mechanisms assessed, shown across the full page width.

4. Computational Models: Governing Equations, Coupling, and Outputs

Model A: High-fidelity partitioned FSI with anisotropic leaflet shells. The fluid domain is solved by a finite-volume incompressible Navier–Stokes solver (second-order in space, second-order backward-differencing time integration) on an unstructured mesh with boundary-layer refinement near the leaflets. The structure uses a nonlinear shell finite-element formulation with two families of embedded fibers aligned with measured pericardial preferred directions; the constitutive law is a transversely isotropic hyperelastic model with exponential fiber response and near-incompressibility via a volumetric penalty. Shell thickness varies spatially according to micro-CT-derived maps. Fluid–structure coupling employs a strong staggered scheme with sub-iterations each time step until both interface displacement and traction residuals drop below 1e-6 relative norm. Contact is handled by a surface-to-surface penalty with quadratic regularization to avoid chatter, with penetration kept below 3 μm at diastolic peak in fine meshes. Reported outputs are cycle-averaged coaptation area, maximum principal stress in the leaflets at mid-diastole, and the diastolic transvalvular pressure–flow relationship measured as regurgitant volume at a prescribed diastolic pressure.

Model B: Immersed-boundary FSI with reduced-order hemodynamics. The fluid equations are discretized on a uniform Cartesian grid with a fractional-step projection method (second-order central differencing; explicit Adams–Bashforth time integration) and an immersed-boundary formulation to impose leaflet kinematics via regularized delta functions. The leaflets are represented as a homogenized membrane with an effective in-plane modulus and damping, calibrated to reproduce bulk opening/closing kinematics. Hemodynamics are reduced-order at the outflow via Windkessel boundary conditions tuned to match target systolic/diastolic pressures. Contact is represented by a repulsive force field activated upon near-coaptation. Reported outputs include an instantaneous coaptation gap metric defined as the maximum open distance along the coaptation line at diastolic peak, a membrane-stress proxy computed from in-plane strains, and effective orifice dynamics during opening.

Model C: Monolithic ALE FSI with nonlinear fiber-reinforced leaflets. A stabilised finite-element scheme solves the incompressible Navier–Stokes equations and the nonlinear solid momentum equations on a common ALE mesh with continuous velocity and displacement fields. The leaflet constitutive model is a Holzapfel–Gasser–Ogden–type hyperelastic law with two splay-distributed fiber families; a quadratic volumetric penalty enforces near-incompressibility, and viscous-like damping captures rate effects. Coaptation is captured using dual mortar contact with augmented Lagrangian stabilization, ensuring traction continuity and limiting penetration below 2 μm at diastolic peak on refined meshes. Newton–Krylov iterations solve the coupled system with an algebraic multigrid preconditioner. Reported outputs include contact pressure distributions during closure, peak von Mises stress, and strain energy integrated over a cycle.

Boundary conditions and driving waveforms are common across models to the extent permitted by the formulations. A pulse-duplicator dataset provides the prescribed aortic root pressure waveform and mean flow target. For the diastolic phase examined in depth here, pressure ramps from 60 to 80 mmHg over 130 ms, then holds for 100 ms. Working fluid is a 60/40 glycerol–water mix with dynamic viscosity 3.5 mPa·s and density 1050 kg/m3. The annulus diameter, leaflet mechanical properties, and commissural alignment are adjusted to produce the three mechanisms described next.

Numerical solver error — gradation.

a. Level 0–1: Solver residuals and coupling tolerances not monitored; solution stability unproven.

b. Level 2: Basic residual reporting without interface balance checks; occasional divergence or excessive damping.

c. Level 3: Documented residual reductions and interface iterations; global conservation or cycle energy monitored.

d. Level 4–5: Tight nonlinear and linear solver tolerances, interface force/displacement balance verified, and cycle-to-cycle energy drift quantified within prespecified bounds.

5. Mechanisms Assessed: Annular Dilation, Leaflet Stiffening, Commissural Misalignment

Annular dilation. We increased the root diameter by 12% relative to nominal (from 23.0 mm to 25.8 mm) using a precision expansion ring in the test fixture and matched in the computational geometry. This reduces geometric overlap between leaflets during closure and can foster central or crescentic gaps if coaptation length diminishes below approximately 1.5 mm over a substantial arc length. In our pulse-duplicator measurements, this mechanism increased diastolic regurgitant volume from 6.4 ± 1.5 ml/beat at baseline to 12.1 ± 2.3 ml/beat at 80 mmHg with a flow hold; endoscopic imaging recorded a 28% reduction in coaptation area relative to baseline.

Leaflet stiffening. To emulate tissue changes that increase stiffness, we elevated the effective modulus by approximately 2× in the models and soaked a subset of valves in a mild glutaraldehyde solution to stiffen the leaflets for bench testing. Thickness increased by 8 ± 3% based on micro-CT, and biaxial test data on sacrificial samples indicated a 25–40% rise in fiber-dominated tangent stiffness. This mechanism modifies deformation under diastolic load, potentially redistributing stress peaks toward the free edge. In the pulse-duplicator, diastolic regurgitant volume was 8.5 ± 1.7 ml/beat at 80 mmHg, and coaptation area decreased by 16% relative to baseline.

Commissural misalignment. A rotational offset of 10° was introduced between commissures via an indexed rotation at one post, producing asymmetric leaflet contact. The misalignment altered closure kinematics, creating localized gaps and stress risers near the shifted commissure. Diastolic regurgitant volume at 80 mmHg rose to 13.1 ± 2.1 ml/beat, and the measured maximum coaptation gap length along the line of closure was 2.4 ± 0.5 mm based on stereo endoscopy; contact pressures could not be measured directly in the bench for this mechanism. These data, while limited, anchor the qualitative and quantitative expectations for closure asymmetry in misalignment.

Test samples — gradation.

a. Level 0–1: Single device; no replication; no estimate of inter-device variability.

b. Level 2: Two or three devices; replication without variance analysis; limited generality.

c. Level 3: At least four devices per mechanism; basic variance measures; randomization or blocking reported.

d. Level 4–5: Six or more devices per mechanism; comprehensive variance decomposition and justification of sample size versus effect size.

6. Credibility Plan: Factors, Metrics, and A Priori Acceptance Criteria (0–5 scale)

We employed a private 0–5 numeric credibility scale to grade evidence for each topic, separately for each model × mechanism pairing. The plan defines metrics and acceptance thresholds per topic based on the model risk analysis and the context of use. Numerical sufficiency is judged by the change in key outputs upon refinement of mesh and time step. Solver stability is judged by residual reductions and cycle-to-cycle energy or mass balance. Faithfulness of the governing physics is judged by quantitative agreement with pulse-duplicator measurements across outputs relevant to the mechanism under study. Parameter adequacy is judged by the tightness and justification of parameter distributions and the sensitivity of outputs to those parameters. Representativeness of the experimental environment is judged by waveform similarity and dimensionless numbers pertinent to the flow, as well as fidelity of the mechanism emulation. Breadth of device samples is judged by the number of distinct valves and the resulting precision of aggregate statistics.

The thresholds selected a priori were: for numerical sufficiency, a score of 4 was required (refinements change key outputs less than approximately 1% for areas and less than 3% for peak stresses); for solver stability, 4 (tight residuals with energy or mass imbalance less than approximately 0.7% per cycle); for physics adequacy, 4 (quantitative agreement across at least two outputs within 10% error, with deviations explained); for parameter adequacy, 3.5 (well-justified distributions with local or global sensitivity documented and posterior predictive checks at least for one mechanism); for test environment representativeness, 3.5 (waveform RMS errors less than 5 mmHg and Womersley similarity within 5% with mechanism-specific fidelity); and for device sample breadth, 3 (at least four devices per mechanism or equivalent justification of statistical power). These thresholds reflect the medium–high risk assigned to the question of interest and ensure that failure in any one topic for a mechanism of interest could prevent acceptance.

We emphasize that every topic is scored for every model × mechanism pair; the scores are not merged across mechanisms or models. Furthermore, the plan requires that the basis for the score be quantitative and include numbers that can be traced to computations and experiments. The following sections detail the evidence collected for each topic.

Discretization error — gradation.

a. Level 0–1: No refinement study; single mesh and time step used; no error estimates.

b. Level 2: One refinement level explored; qualitative statements about stability only.

c. Level 3: At least two refinement levels; quantitative changes reported for a subset of outputs.

d. Level 4–5: Systematic multi-level refinement for all key outputs; GCI or equivalent used; changes within prespecified bounds.

7. Evidence for Discretization Error by Model × Mechanism

Model A. For annular dilation, a three-level mesh study used fluid cell counts of 2.1M, 4.3M, and 8.6M with corresponding leaflet shell discretizations of 25k, 50k, and 100k elements; time steps of 4e-4 s and 2e-4 s were compared. The change from medium to fine was 0.8% for cycle-averaged coaptation area and 3.1% for maximum principal stress at mid-diastole; halving the time step at the fine level altered these outputs by 0.3% and 0.9%, respectively. For leaflet stiffening, the same study produced 0.5% and 2.2% changes for area and stress (time-step halving effects of 0.2% and 0.7%). For commissural misalignment, the medium-to-fine changes were 1.3% (area) and 3.9% (stress), with time-step halving effects of 0.4% and 1.1%.

Model B. For annular dilation, Cartesian grid resolutions of 128^3, 192^3, and 256^3 with uniform time steps of 8e-4 s and 5e-4 s yielded medium-to-fine changes of 4.3% for peak gap length and 6.1% for the membrane-stress proxy; reducing time step altered these by 0.9% and 1.4%. For leaflet stiffening, corresponding changes were 3.8% (gap) and 5.5% (stress proxy); time-step effects were 0.8% and 1.2%. For commissural misalignment, medium-to-fine changes rose to 5.9% (gap) and 8.3% (stress proxy), with time-step effects of 1.0% and 1.7%.

Model C. For annular dilation, fluid cells numbered 2.7M, 5.3M, and 5.0M with adaptive boundary-layer insertion, and solid elements were 60k, 120k, and 120k, respectively; time steps adaptively ranged from 2e-4 s down to 1e-4 s near closure. Medium-to-fine changes were 0.9% for coaptation area and 2.6% for von Mises peak; adaptive time refinement changed outputs by 0.3% and 0.8%. For leaflet stiffening, medium-to-fine changes were 1.1% (area) and 2.0% (stress), with adaptive time changes of 0.3% and 0.7%. For commissural misalignment, medium-to-fine changes were 1.5% (area) and 3.2% (stress), with adaptive time changes of 0.5% and 1.0%.

Discretization error — gradation.

a. Level 0–1: No quantitative refinement evidence; single resolution.

b. Level 2: One refinement step; partial reporting of output changes.

c. Level 3: Two or more levels with quantified changes for at least one key output.

d. Level 4–5: Two or more levels with quantified changes for all key outputs; changes below prespecified tolerances and corroborated by time-step sensitivity.

8. Evidence for Numerical Solver Error by Model × Mechanism

Model A. Annular dilation runs used a strong partitioned coupling with an average of 3.1 ± 0.5 sub-iterations per time step near closure, driving interface force and displacement residuals below 1e-6. Cycle-to-cycle energy drift, computed from the difference in mechanical plus fluid kinetic energy between successive diastolic holds, was 0.4%. Leaflet stiffening showed similar sub-iterations (3.0 ± 0.4) with energy drift 0.3%. For commissural misalignment, increased contact nonlinearity raised average sub-iterations to 3.6 ± 0.6, with energy drift 0.6% and interface traction balance within 0.7% over the final 20 ms of diastole.

Model B. The explicit immersed-boundary scheme exhibited maximum Courant numbers below 0.45 in all cases. For annular dilation, mass-conservation error over a cardiac cycle, assessed as the difference between net inflow and outflow normalized by stroke volume, was 1.2%; the L2 norm of divergence fell below 1e-6 per time step after projection. For leaflet stiffening, mass-balance error was 1.0% with similar divergence control. For commissural misalignment, vortex shedding at the asymmetric commissure increased mass-balance error to 1.8%, with divergence still below 1e-6; stabilization via a mild penalty reduced high-frequency oscillations in the force signal but did not change the mass-balance figure materially.

Model C. The monolithic Newton–Krylov solver converged to relative residuals of 1e-8 for the coupled system, with algebraic multigrid-preconditioned linear solves reaching reductions of 8–10 orders of magnitude per nonlinear step. Annular dilation required 7.2 ± 1.1 Newton iterations per time step near closure; energy drift per diastolic cycle was 0.3%. Leaflet stiffening required 6.9 ± 0.9 iterations with energy drift 0.4%. Commissural misalignment increased solver effort to 9.8 ± 1.4 iterations; augmented Lagrangian parameters were tuned to keep contact constraint violation below 2 μm and cycle drift at 0.7%.

Numerical solver error — gradation.

a. Level 0–1: Residuals not tracked; no conservation checks; solver often fails or is strongly damped.

b. Level 2: Residuals tracked without quantitative targets; conservation checks sporadic; occasional divergence.

c. Level 3: Quantitative residual targets achieved; conservation and energy drift assessed; minor imbalances remain.

d. Level 4–5: Tight residual and interface balance targets met robustly; energy or mass imbalance within prespecified bounds across mechanisms.

9. Evidence for Model Form by Model × Mechanism

Model A. Against pulse-duplicator data for annular dilation at 80 mmHg, the predicted diastolic regurgitant volume was 13.0 ml versus 12.1 ± 2.3 ml measured (7.4% error). Endoscopic coaptation area reduction relative to baseline was predicted at 25% versus 28% measured (absolute difference 3 percentage points). The maximum principal stress localized near the free edge peaked at 1.25 MPa; while direct stress validation is unavailable, the magnitude aligns with biaxial-to-FE extrapolations. For leaflet stiffening, predicted regurgitant volume was 9.1 ml versus 8.5 ± 1.7 ml measured (7.1% error), and coaptation area reduction was 17% versus 16% measured (1 percentage point deviation). For commissural misalignment, predicted maximum gap length was 2.1 mm versus 2.4 ± 0.5 mm measured (0.3 mm, 12.5% error), while regurgitant volume was 14.2 ml versus 13.1 ± 2.1 ml (8.4% error); stress hotspots at the shifted commissure reached 1.33 MPa.

Model B. For annular dilation, predicted diastolic regurgitant volume was 16.5 ml versus 12.1 ± 2.3 ml measured (36% error), and the change in coaptation area was underpredicted at 18% versus 28% measured. For leaflet stiffening, predicted regurgitant volume was 10.8 ml versus 8.5 ± 1.7 ml (27% error), and the reduction in coaptation area was 11% versus 16% measured. For commissural misalignment, the predicted maximum gap length was 3.5 mm versus 2.4 ± 0.5 mm (46% error), and regurgitant volume was 16.0 ml versus 13.1 ± 2.1 ml (22% error); the membrane-stress proxy peaked near the free edge but lacked the localized commissural hot spot observed in qualitative dye-penetration maps.

Model C. For annular dilation, predicted regurgitant volume was 12.8 ml versus 12.1 ± 2.3 ml (5.8% error); the coaptation area reduction was 27% versus 28% measured (1 percentage point difference). For leaflet stiffening, predicted regurgitant volume was 9.0 ml versus 8.5 ± 1.7 ml (5.9% error), and coaptation area reduction was 15% versus 16% measured. For commissural misalignment, the predicted maximum gap length was 2.6 mm versus 2.4 ± 0.5 mm (8% error), and regurgitant volume was 13.6 ml versus 13.1 ± 2.1 ml (4% error). Contact pressures obtained during closure concentrated at the rotated commissure in patterns consistent with the observed asymmetric dye uptake, although absolute values could not be validated directly.

Model form — gradation.

a. Level 0–1: No or only qualitative agreement to any relevant measurement; missing governing physics.

b. Level 2: Partial physics included; mismatches exceed acceptable bounds with no mechanistic rationale.

c. Level 3: Key physics included; some outputs match within relaxed bounds; others deviate without localization.

d. Level 4–5: Comprehensive physics and contact; multiple outputs per mechanism within strict bounds; discrepancies mechanistically explained or bounded.

10. Evidence for Model Inputs by Model × Mechanism

Model A. Parameter distributions for fiber angles (mean ± SD 46° ± 10° relative to circumferential), thickness (mean 0.37 mm with ±15% SD), and modulus scale factors (±25% SD) were informed by micro-CT maps and biaxial tests on pericardial coupons from the same device family. A local sensitivity study with 200 Latin hypercube samples per mechanism showed that for annular dilation, a ±10° shift in fiber angle changed coaptation area by 1.8% and peak stress by 6.7%; ±15% thickness variation changed peak stress by 8.9%. For leaflet stiffening, uncertainty in the modulus scale dominated: ±25% changed peak stress by 12.4% and regurgitant volume by 6.1%. For commissural misalignment, a ±2° uncertainty in imposed rotation changed maximum gap by 9% and regurgitant volume by 5.5%, while thickness variation shifted peak stress by 10.2%. Posterior checks against baseline kinematics constrained modulus and damping for each tested valve, but mechanism-specific adjustments were not feasible for all samples.

Model B. Effective membrane stiffness and damping were tuned to match baseline opening/closing times, with ±30% uncertainty bounds reflecting heterogeneity across valves and fitting error. For annular dilation, ±30% stiffness change altered predicted gap length by 7.9% and regurgitant volume by 9.5%. For leaflet stiffening, introducing a 2× nominal stiffness with ±20% uncertainty increased peak stress proxy by 15.2% with regurgitant volume sensitivity of 8.2%. For commissural misalignment, a ±2° uncertainty in rotation changed gap length by 12.5%; other parameters produced additional variability, yielding total predicted gap uncertainty near 16% when combined in quadrature.

Model C. Device-specific thickness fields were used with an overall scale factor (±10%) and fiber splay estimate (mean ± SD of 12° ± 5°) taken from literature on similar pericardial processing. For annular dilation, ±10% thickness altered peak stress by 7.1%, and ±5° fiber splay changed coaptation area by 1.3% and stress by 4.5%. For leaflet stiffening, modulus scale factor ±15% changed peak stress by 10.8% and regurgitant volume by 4.1%. For commissural misalignment, ±2° in rotation altered gap length by 8.8%, while fiber angle uncertainty contributed 3.1% to stress variability; combined, the coefficient of variation for peak stress under misalignment was 11%.

Model inputs — gradation.

a. Level 0–1: Nominal inputs only; no uncertainty; no sensitivity.

b. Level 2: Broad literature priors; limited sensitivity on one output.

c. Level 3: Device family priors; multiple inputs varied; local sensitivity for key outputs.

d. Level 4–5: Device- or lot-specific measurements; global sensitivity and posterior checks; mechanism-specific calibration where possible.

11. Evidence for Test Conditions by Model × Mechanism

Pulse-duplicator waveforms were matched across mechanisms using a compliance chamber and adjustable peripheral resistance. For annular dilation, target diastolic pressure at 80 mmHg was matched with an RMS error of 2.7 mmHg over the hold phase; the Womersley number in the aortic root (based on diameter and fluid properties) was within 3% of the clinical range at the relevant frequencies. The annulus expansion ring set the diameter to 25.8 mm with ±0.1 mm repeatability, verified by calipers. For leaflet stiffening, the same waveform matching achieved an RMS pressure error of 2.9 mmHg; micro-CT confirmed an 8 ± 3% increase in leaflet thickness after the soak, and biaxial tests on sacrificial samples quantified a 25–40% rise in tangent stiffness, anchoring the mechanism representation. For commissural misalignment, pressure RMS error was 3.2 mmHg; a jig rotated one commissure by 10° with ±1° tolerance. However, the rotated-post geometry introduces subtle changes in stent post compliance not present clinically, making the emulation less faithful. In all mechanisms, working fluid viscosity was 3.5 mPa·s ± 0.1 mPa·s and density 1050 ± 10 kg/m3, close to physiologic values; temperatures were maintained at 23 ± 1°C.

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

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

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

Test conditions — gradation.

a.
Level 0–1: Waveforms and geometry not representative; no quantification of similarity.

b.
Level 2: Some targets matched; important transients or geometric perturbations differ materially.

c.
Level 3: Primary metrics matched within broad bounds; mechanism setup documented with tolerances.

d.
Level 4–5: Waveforms and dimensionless numbers matched within tight bounds; mechanism emulation faithful to clinical scenario with quantified uncertainties.

12. Evidence for Test Samples by Model × Mechanism

Annular dilation was tested on five valves of the same model and size, randomly assigned to test order, with three repeats per device. The mean diastolic regurgitant volume at 80 mmHg was 12.1 ml with an SD of 2.3 ml across devices (coefficient of variation 19%); coaptation area reduction averaged 28% with an SD of 6 percentage points. Leaflet stiffening involved four valves, each soaked identically and tested in triplicate. The mean regurgitant volume was 8.5 ml with an SD of 1.7 ml; coaptation area reduction averaged 16% with an SD of 4 percentage points. Commissural misalignment was assessed on three valves due to fixture availability, again with three repeats per device. The mean regurgitant volume was 13.1 ml with an SD of 2.1 ml; maximum coaptation gap length averaged 2.4 mm with an SD of 0.5 mm. The smaller sample for misalignment yields wider confidence intervals and reduces the precision available for validating asymmetric closure predictions. Across all mechanisms, devices were from two manufacturing lots; no lot-specific effects were detected beyond the reported inter-device variability.

Test samples — gradation.

a.
Level 0–1: Single device; no replication or variance estimation.

b.
Level 2: Two to three devices; replication but insufficient for precise estimates.

c.
Level 3: Four to five devices; basic variance quantified; some blocking/randomization.

d.
Level 4–5: Six or more devices; variance decomposition and power analysis justify sample size.

13. Software Quality Assurance: Summary Notes

All three modeling toolchains used version-controlled repositories with issue tracking and continuous integration for unit-level tests. Containerized environments captured compiler and dependency versions for reproducibility. While external audits were not conducted, developer testing documentation and peer code reviews were recorded contemporaneously.

Model inputs — gradation.

a.
Level 0–1: No traceability for parameter sources or changes.

b.
Level 2: Ad hoc documentation of parameters and updates.

c.
Level 3: Version-controlled parameter sets with provenance.

d.
Level 4–5: Formal change control and verification of parameter import/export pipelines.

14. Results: Factor-by-Factor Scores for All Model × Mechanism Combinations

The following sections present one table per model with scores on the 0–5 scale for each topic and mechanism. Each table row is scoped to a specific topic, model, and mechanism; the numeric level contains only the score, and the basis restates quantitative evidence reported above. Scores reflect the a priori thresholds where a 4 or higher was required for numerical sufficiency and solver stability, a 4 for physics adequacy, a 3.5 for parameter adequacy, a 3.5 for test environment representativeness, and a 3 for device samples.

In brief, Models A and C reached levels of 4–5 on numerical sufficiency and solver stability across mechanisms. Model B attained levels of 2–3 on these topics, reflecting greater sensitivity to resolution and modest mass-balance errors. For physics adequacy, Model A and Model C achieved scores of 4 for annular dilation and leaflet stiffening but were lower for commissural misalignment due to increased discrepancies in asymmetric closure patterns. Model B lagged on this topic in all mechanisms. Parameter adequacy scores were limited by

uncertainty in fiber angles and thickness maps for all models, with Model C benefitting from device-specific fields but still constrained. The test environment score was reduced for misalignment due to potential fixture-induced artifacts. Sample breadth met thresholds for annular dilation and leaflet stiffening but not for misalignment due to fewer devices.
Model form — gradation.

a. Level 0–1: No mechanism-specific validation.

b. Level 2: Mechanism present but outputs diverge beyond bounds.

c. Level 3: At least one key output within relaxed bounds; others deviate.

d. Level 4–5: Multiple outputs per mechanism within strict bounds, with residuals explained.

15. Model A: Credibility-assessment summary table

This table summarizes Model A scores and quantitative bases across mechanisms. The evidence ties to the numerical refinement studies, solver diagnostics, comparisons to bench measurements, parameter sensitivity analyses, assessment of waveform representativeness and mechanism fidelity, and device sample sizes and variability.
Discretization error — gradation.

a. Level 0–1: No mesh/time-step study.

b. Level 2: One refinement; qualitative assessment.

c. Level 3: Multi-level study but only subset of outputs quantified.

d. Level 4–5: Multi-level and time-step studies; key outputs change within tight bounds.

16. Model B: Credibility-assessment summary table

This table summarizes Model B scores with the quantitative bases drawn from the earlier sections. Model B's reduced-order membrane model and immersed-boundary formulation introduce different numerical and physics trade-offs relative to the other models.
Numerical solver error — gradation.

a. Level 0–1: No residual or conservation checks.

b. Level 2: Residuals tracked; occasional instabilities remain.

c. Level 3: Residual targets and conservation within moderate bounds.

d. Level 4–5: Tight targets met with energy/mass imbalance within strict thresholds.

17. Model C: Credibility-assessment summary table

This table summarizes Model C scores and quantitative bases. The monolithic ALE approach with reinforced leaflet material and mortar contact aims to capture asymmetric closure and localized contact pressures more faithfully.
Test conditions — gradation.

a. Level 0–1: Unrepresentative testbed.

b. Level 2: Partial similarity; key aspects unquantified.

c. Level 3: Primary metrics within broad bounds; tolerances recorded.

d. Level 4–5: Waveforms and geometry emulation within tight bounds with uncertainty quantified.

18. Synthesis: Alignment to Context of Use and Risk Rationale

The decision hinges on whether each model–mechanism pairing provides sufficiently precise and accurate estimates of closure and leaflet load metrics to back preclinical design choices. For the numerical resolution topic, refining the grid and step size changed key outputs by less than 1.5% for coaptation area and less than 4% for peak stresses in Models A and C, whereas the same exercise in Model B yielded 3.8–8.3% shifts in gap and stress-proxy metrics, which is material for tight closure margins.

For solution stability, the coupled solvers in Models A and C reached tight residuals with negligible cycle drift (≤0.7%), but the immersed-boundary scheme registered mass-balance errors up to 1.8% under asymmetric closure, enough to blur small differences in regurgitation near acceptance thresholds.

Regarding the adequacy of the governing equations and contact treatment, comparisons to bench data showed that Models A and C predicted diastolic regurgitant volumes and coaptation area changes within roughly 4–9% for annular dilation and leaflet stiffening, but larger deviations emerged for commissural misalignment, where asymmetric gap shape and localized contact challenge even sophisticated schemes. The immersed-boundary approach underperformed across mechanisms, particularly for misalignment, where the predicted gap length overshot measurements by 46%.

Parameterization remains a limiting factor across all models. Uncertainty in fiber angle, thickness, and modulus scale propagated into peak stress variability of 7–12% and gap-length variability near 9–12% for misalignment. These ranges are nontrivial when the context of use requires confidence in distinguishing mechanism-induced differences of similar magnitude.

The test environment was well matched to physiologic diastolic conditions for dilation and stiffening (pressure RMS errors ≤2.9 mmHg; Womersley within 3–5%), but the misalignment fixture likely perturbs post compliance in a way not present clinically, potentially altering the contact patterns relevant to validation.

Finally, the limited number of devices in the misalignment experiments (n = 3) reduces precision of the mean and widens uncertainty bands against which to judge model predictions. While sampling was adequate for annular dilation (n = 5) and acceptable for stiffening (n = 4), the smaller set for misalignment is a shortfall for a mechanism that is already the most sensitive to geometry and contact fidelity.

Collectively, these observations map back to the risk rationale: the consequences of an incorrect design decision are moderate but non-negligible. For adoption in this context, high confidence is required in closure integrity and stress localization under each mechanism. The accumulated evidence falls short of that standard for several pairings, especially for commissural misalignment and for the immersed-boundary approach across the board.

Test samples — gradation.

a.
Level 0–1: No estimate of inter-device variability; not decision-suitable.

b.
Level 2: Minimal devices; high uncertainty in means.

c.
Level 3: Moderate devices; uncertainty quantified; borderline for sensitive mechanisms.

d.
Level 4–5: Robust devices; narrow confidence intervals relative to effect sizes.

19. Discussion: Shortfalls, Sensitivities, and Priority Gaps

Shortfalls. The dominant gaps reside in commissural misalignment across all models and in the immersed-boundary model across mechanisms. For misalignment, even when key outputs (regurgitant volume and maximum gap) are within 4–12% of measurements for the higher-fidelity models, the evidence for contact-pressure localization remains indirect. As a result, confidence is limited in the predicted peak stress at the shifted commissure, a metric relevant to fatigue risk. The bench fixture's rotated-post geometry may introduce unmeasured compliance effects that confound this comparison, constraining the utility of the data for discerning contact nuances. For the immersed-boundary approach, the combination of higher sensitivity to grid resolution and poorer matches to coaptation metrics under the three mechanisms indicate that the homogenized membrane representation and repulsive closure model are insufficient for this context of use.

Sensitivities. Sensitivity analyses highlight strength and vulnerability in parameterization. For Models A and C, thickness and modulus scale are the main drivers of peak stress uncertainty, especially under leaflet stiffening. Fiber angle variations of ±10° shift peak stress by 4–7%, more so under dilation where altered curvature and overlap change load paths. Under misalignment, small angular errors in the applied rotation (±2°) change gap length by 9–12%, depending on the model, indicating that geometric fidelity in the fixture and in the computational representation is critical. In Model B, the effective stiffness parameter drives both gap and regurgitant volume shifts, reflecting the limited physics in the structural representation.

Priority gaps. To move toward acceptance, three priorities emerge. First, targeted bench measurements of local leaflet kinematics and contact during misalignment are needed. Stereo digital image correlation with a micro-speckle pattern on the leaflets under transparent working fluid could quantify out-of-plane deformation and strain localization. Incorporating thin-film pressure sensors is challenging but possible on surrogate leaflets to calibrate contact models. Second, expanding sample sizes for misalignment to at least six devices will reduce uncertainty in mean gap and regurgitant volume by roughly 30% relative to current levels, narrowing validation intervals. Third, refining parameter priors via microstructural imaging (e.g., small-angle light scattering) for fiber orientation distributions and via spatially resolved thickness mapping across more devices would reduce the spread in peak stress predictions by several percentage points. For Model B, a more profound revision is needed—either upgrading to a shell-based structure with explicit contact or limiting the context of use to less asymmetric mechanisms.

Operational considerations. While not central to the credibility topics scored here, the reproducibility of the computational environment and the traceability of parameter sets support efficient iteration and reduce the chance of regression errors. Ensuring that parameter changes are propagated consistently between baseline and mechanism-specific runs avoids confounding effects in sensitivity and validation analyses. Moreover, additional test condition checks, such as quantifying compliance of the test fixture and stent posts under misalignment, will be valuable in interpreting discrepancies between models and measurements.

Model form — gradation.

a.
Level 0–1: Validation omits key asymmetric closure behaviors.

b.
Level 2: Asymmetry present but mismatch is large and unexplained.

c.
Level 3: Some asymmetric metrics matched; others remain outside bounds.

d.

Level 4–5: Asymmetric closure and contact validated across multiple measures within bounds.

20. Conclusion: Decision Outcome and Required Next Steps

Under the ASME V&V 40 process applied here, ratings on a private 0–5 scale with prespecified thresholds were assigned for each topic across all model × mechanism combinations. The evidence demonstrates strong numerical sufficiency and solver stability for Models A and C. Agreement with bench regurgitant volumes and coaptation metrics is good for annular dilation and leaflet stiffening in these models but less so for commissural misalignment, where additional validation targets and more faithful emulation of clinical geometry are required. Model B falls short across several topics, particularly in physics adequacy and resolution sensitivity, suggesting that its structural and contact representations are not suited for this context of use without significant modification.

Because one or more topics for several model–mechanism combinations do not meet the acceptance thresholds, and because the most challenging mechanism—commissural misalignment—remains insufficiently supported by data and physics fidelity relative to the intended decisions, the credibility assessment outcome is Not accepted. To achieve acceptance, we recommend: (1) expanding misalignment bench datasets to at least six valves with enhanced measurement of local leaflet deformation and, where feasible, contact pressures; (2) refining parameter priors and posteriors via microstructural imaging and spatial thickness mapping across more devices; and (3) strengthening, or replacing, the immersed-boundary structural/contact representation to better capture asymmetric closure. With these steps, we anticipate that the higher-fidelity models can achieve the required levels across all topics, and that a revised immersed-boundary approach may be repurposed for mechanisms less sensitive to localized contact.

Model inputs — gradation.

a. Level 0–1: No device-specific parameter data to support decisions.

b. Level 2: Limited priors; broad uncertainties confound decisions.

c. Level 3: Device family priors and sensitivities provide partial support.

d. Level 4–5: Device-specific priors and posteriors support confident decisions across mechanisms.

References

[1] T. Grigoryan, C. Wojcik, W. Lindqvist, G. Quintero, Material model effects on rib deflection in computational models, Comput. Fluids 93 (2024) 5–635.

[2] K. Halvorsen, R. Bergman, Solver tolerance effects on head displacement in computational models, J. Mech. Behav. Biomed. Mater. 78 (2009) 60–443.

[3] D. Silva, H. Fontaine, Gap size effects on fatigue life in computational models, J. Mech. Behav. Biomed. Mater. 95 (2005) 79–801.

[4] N. Yilmaz, N. Castellano, J. Nakamura, F. Vasquez, Solver tolerance effects on hemolysis index in computational models, Int. J. Numer. Methods 21 (2011) 334–848.

[5] M. Halvorsen, G. Vasquez, W. Grigoryan, Load path effects on rib deflection in computational models, Comput. Fluids 58 (2016) 354–521.

[6] J. Nakamura, W. Kaur, E. Lindqvist, M. Silva, Solver tolerance effects on fatigue life in computational models, Comput. Fluids 81 (2006) 81–460.

[7] G. Ueda, L. Ishikawa, M. Nakamura, K. Ishikawa, Element type effects on head displacement in computational models, Ann. Biomed. Eng. 113 (2006) 213–631.

[8] M. Vasquez, H. Wojcik, Material model effects on outlet velocity in computational models, Comput. Methods Biomech. 136 (2012) 227–643.

[9] R. Bergman, S. Halvorsen, K. Fontaine, Element type effects on rib deflection in computational models, Ann. Biomed. Eng. 82 (2007) 264–630.

[10] G. Tanaka, T. Yilmaz, K. Zetterberg, H. Kaur, Time step effects on stress range in computational models, J. Biomech. Eng. 98 (2009) 101–807.

[11] G. Halvorsen, P. Jankowski, H. Quintero, L. Okonkwo, Bone density effects on contact pressure in computational models, J. Verif. Valid. Uncertain. 142 (2006) 40–857.

[12] W. Halvorsen, T. Ueda, K. Ueda, T. Ishikawa, Load path effects on contact pressure in computational models, J. Biomech. Eng. 28 (2006) 63–493.

[13] B. Lindqvist, E. Zetterberg, M. Grigoryan, R. Halvorsen, Yield stress effects on chest compression in computational models, J. Verif. Valid. Uncertain. 29 (2022) 38–499.

[14] G. Tanaka, R. Bergman, Boundary condition effects on contact pressure in computational models, J. Verif. Valid. Uncertain. 22 (2011) 91–667.

[15] F. Fontaine, N. Vasquez, R. Ueda, T. Fontaine, A. Fontaine, Flow rate effects on head displacement in computational models, Ann. Biomed. Eng. 71 (2024) 227–614.

[16] D. Lindqvist, L. Halvorsen, C. Bergman, Yield stress effects on stress range in computational models, Int. J. Numer. Methods 60 (2011) 3–884.

[17] J. Abbott, M. Moreau, Cement mantle effects on flow split in computational models, Comput. Fluids 78 (2005) 87–429.

[18] A. Nakamura, K. Halvorsen, L. Fontaine, Yield stress effects on chest compression in computational models, J. Verif. Valid. Uncertain. 93 (2005) 276–447.

[19] G. Okonkwo, J. Wojcik, N. Okonkwo, Time step effects on fatigue life in computational models, J. Verif. Valid. Uncertain. 103 (2011) 30–583.

[20] J. Grigoryan, M. Grigoryan, Friction effects on shear stress in computational models, J. Mech. Behav. Biomed. Mater. 99 (2009) 48–506.

[21] J. Kaur, C. Zetterberg, Material model effects on outlet velocity in computational models, Ann. Biomed. Eng. 141 (2005) 274–784.

[22] N. Eriksen, K. Ueda, B. Jankowski, W. Rasmussen, Yield stress effects on stress range in computational models, Med. Eng. Phys. 131 (2022) 283–737.

[23] R. Halvorsen, A. Castellano, B. Jankowski, M. Duarte, Bone density effects on hemolysis index in computational models, J. Verif. Valid. Uncertain. 96 (2019) 88–887.

[24] W. Ishikawa, A. Vasquez, E. Yilmaz, Friction effects on outlet velocity in computational models, J. Mech. Behav. Biomed. Mater. 31 (2013) 103–824.

[25] E. Eriksen, N. Nakamura, J. Bergman, Boundary condition effects on shear stress in computational models, Int. J. Numer. Methods 147 (2005) 123–816.

[26] H. Nakamura, A. Tanaka, Time step effects on stress range in computational models, J. Biomech. Eng. 115 (2020) 38–756.

[27] B. Moreau, G. Yilmaz, Load path effects on outlet velocity in computational models, Comput. Fluids 30 (2010) 125–551.

[28] T. Okonkwo, J. Zetterberg, Flow rate effects on rib deflection in computational models, Comput. Fluids 23 (2017) 12–889.

[29] N. Bergman, H. Lindqvist, A. Silva, F. Tanaka, Flow rate effects on stress range in computational models, Ann. Biomed. Eng. 67 (2018) 268–401.

[30] S. Lindqvist, R. Vasquez, N. Wojcik, Inlet profile effects on fatigue life in computational models, Int. J. Numer. Methods 72 (2020) 177–641.

[31] W. Castellano, G. Abbott, H. Lindqvist, L. Pettersson, Element type effects on peak force in computational models, J. Biomech. Eng. 132 (2023) 136–817.

[32] A. Jankowski, K. Silva, E. Quintero, Flow rate effects on outlet velocity in computational models, Comput. Fluids 69 (2013) 81–587.

[33] S. Abbott, E. Lindqvist, B. Duarte, J. Tanaka, A. Eriksen, Contact stiffness effects on chest compression in computational models, J. Verif. Valid. Uncertain. 24 (2018) 67–753.

[34] D. Okonkwo, A. Zetterberg, L. Okonkwo, E. Zetterberg, Damping effects on stress range in computational models, Ann. Biomed. Eng. 78 (2019) 75–778.

Table 1: Model A credibility assessment across mechanisms on a 0–5 scale with predefined acceptance thresholds.

Credibility factor	Level	Basis
Discretization error - Model A - Annular dilation	4	Medium-to-fine change 0.8% (area) and 3.1% (peak stress); halving dt changed 0.3% and 0.9%.
Numerical solver error - Model A - Annular dilation	4	3.1 ± 0.5 interface sub-iterations; residuals <1e-6; cycle energy drift 0.4%.
Model form - Model A - Annular dilation	4	Regurgitant volume 13.0 ml vs 12.1 ± 2.3 ml (7.4%); coaptation area drop 25% vs 28%.
Model inputs - Model A - Annular dilation	3	±10° fiber angle changed area 1.8% and stress 6.7%; ±15% thickness shifted stress 8.9%.
Test conditions - Model A - Annular dilation	4	Diastolic pressure RMS error 2.7 mmHg; Womersley within 3%; annulus set 25.8 ± 0.1 mm.
Test samples - Model A - Annular dilation	3	n=5 devices; regurgitant volume SD 2.3 ml; coaptation area SD 6 percentage points.
Discretization error - Model A - Leaflet stiffening	5	Medium-to-fine change 0.5% (area) and 2.2% (peak stress); dt halving 0.2% and 0.7%.
Numerical solver error - Model A - Leaflet stiffening	5	3.0 ± 0.4 sub-iterations; residuals <1e-6; cycle energy drift 0.3%.
Model form - Model A - Leaflet stiffening	4	Regurgitant volume 9.1 ml vs 8.5 ± 1.7 ml (7.1%); area drop 17% vs 16%.
Model inputs - Model A - Leaflet stiffening	3	±25% modulus changed peak stress 12.4% and regurgitant volume 6.1%.
Test conditions - Model A - Leaflet stiffening	4	Pressure RMS error 2.9 mmHg; thickness +8 ± 3%; biaxial stiffness +25–40% verified.
Test samples - Model A - Leaflet stiffening	3	n=4 devices; regurgitant volume SD 1.7 ml; area reduction SD 4 percentage points.
Discretization error - Model A - Commissural misalignment	4	Medium-to-fine change 1.3% (area) and 3.9% (peak stress); dt halving 0.4% and 1.1%.
Numerical solver error - Model A - Commissural misalignment	4	3.6 ± 0.6 sub-iterations; interface traction balance within 0.7%; energy drift 0.6%.
Model form - Model A - Commissural misalignment	3	Gap 2.1 mm vs 2.4 ± 0.5 mm (12.5%); regurgitant volume 14.2 ml vs 13.1 ± 2.1 ml (8.4%).
Model inputs - Model A - Commissural misalignment	3	±2° rotation changed gap 9% and regurgitant volume 5.5%; thickness shifted stress 10.2%.
Test conditions - Model A - Commissural misalignment	3	Pressure RMS error 3.2 mmHg; 10° rotation ±1°; fixture compliance may differ from clinical.
Test samples - Model A - Commissural misalignment	2	n=3 devices; regurgitant volume SD 2.1 ml; gap SD 0.5 mm.

Table 2: Model B credibility assessment across mechanisms on a 0–5 scale with predefined acceptance thresholds.

Credibility factor	Level	Basis
Discretization error - Model B - Annular dilation	3	Medium-to-fine change 4.3% (gap) and 6.1% (stress proxy); dt reduction 0.9% and 1.4%.
Numerical solver error - Model B - Annular dilation	3	Max CFL <0.45; mass-balance error 1.2%; divergence L2 <1e-6 per step.
Model form - Model B - Annular dilation	2	Regurgitant volume 16.5 ml vs 12.1 ± 2.3 ml (36%); area drop 18% vs 28%.
Model inputs - Model B - Annular dilation	2	±30% stiffness changed gap 7.9% and regurgitant volume 9.5%.
Test conditions - Model B - Annular dilation	4	Diastolic pressure RMS error 2.7 mmHg; Womersley within 3%; annulus 25.8 ± 0.1 mm.
Test samples - Model B - Annular dilation	3	n=5 devices; regurgitant volume SD 2.3 ml; area SD 6 percentage points.
Discretization error - Model B - Leaflet stiffening	3	Medium-to-fine change 3.8% (gap) and 5.5% (stress proxy); dt reduction 0.8% and 1.2%.
Numerical solver error - Model B - Leaflet stiffening	3	Mass-balance error 1.0%; divergence L2 <1e-6; stable explicit scheme.
Model form - Model B - Leaflet stiffening	2	Regurgitant volume 10.8 ml vs 8.5 ± 1.7 ml (27%); area drop 11% vs 16%.
Model inputs - Model B - Leaflet stiffening	2	2× nominal stiffness with ±20% changed stress proxy 15.2% and volume 8.2%.
Test conditions - Model B - Leaflet stiffening	4	Pressure RMS error 2.9 mmHg; thickness +8 ± 3%; stiffness increase 25–40% verified.
Test samples - Model B - Leaflet stiffening	3	n=4 devices; regurgitant volume SD 1.7 ml; area reduction SD 4 percentage points.
Discretization error - Model B - Commissural misalignment	2	Medium-to-fine change 5.9% (gap) and 8.3% (stress proxy); dt reduction 1.0% and 1.7%.
Numerical solver error - Model B - Commissural misalignment	2	Mass-balance error 1.8%; divergence L2 <1e-6; added penalty to reduce oscillations.
Model form - Model B - Commissural misalignment	1	Gap 3.5 mm vs 2.4 ± 0.5 mm (46%); regurgitant volume 16.0 ml vs 13.1 ± 2.1 ml (22%).
Model inputs - Model B - Commissural misalignment	2	±2° rotation changed gap 12.5%; combined input uncertainty ~16% gap variability.
Test conditions - Model B - Commissural misalignment	3	Pressure RMS error 3.2 mmHg; 10° rotation ±1°; fixture compliance may confound.
Test samples - Model B - Commissural misalignment	2	n=3 devices; regurgitant volume SD 2.1 ml; gap SD 0.5 mm.

Table 3: Model C credibility assessment across mechanisms on a 0–5 scale with predefined acceptance thresholds.

Credibility factor	Level	Basis
Discretization error - Model C - Annular dilation	4	Medium-to-fine change 0.9% (area) and 2.6% (peak von Mises); adaptive dt change 0.3% and 0.8%.
Numerical solver error - Model C - Annular dilation	5	Newton residual to 1e-8; 7.2 ± 1.1 iterations near closure; cycle energy drift 0.3%.
Model form - Model C - Annular dilation	4	Regurgitant volume 12.8 ml vs 12.1 ± 2.3 ml (5.8%); area drop 27% vs 28%.
Model inputs - Model C - Annular dilation	4	±10% thickness changed stress 7.1%; ±5° fiber splay changed area 1.3% and stress 4.5%.
Test conditions - Model C - Annular dilation	4	Diastolic pressure RMS error 2.7 mmHg; Womersley within 3%; annulus 25.8 ± 0.1 mm.
Test samples - Model C - Annular dilation	3	n=5 devices; regurgitant volume SD 2.3 ml; area SD 6 percentage points.
Discretization error - Model C - Leaflet stiffening	5	Medium-to-fine change 1.1% (area) and 2.0% (peak von Mises); adaptive dt change 0.3% and 0.7%.
Numerical solver error - Model C - Leaflet stiffening	5	Residual to 1e-8; 6.9 ± 0.9 iterations; cycle energy drift 0.4%.
Model form - Model C - Leaflet stiffening	4	Regurgitant volume 9.0 ml vs 8.5 ± 1.7 ml (5.9%); area drop 15% vs 16%.
Model inputs - Model C - Leaflet stiffening	3	±15% modulus changed stress 10.8% and regurgitant volume 4.1%.
Test conditions - Model C - Leaflet stiffening	4	Pressure RMS error 2.9 mmHg; thickness +8 ± 3%; stiffness +25–40% verified.
Test samples - Model C - Leaflet stiffening	3	n=4 devices; regurgitant volume SD 1.7 ml; area reduction SD 4 percentage points.
Discretization error - Model C - Commissural misalignment	4	Medium-to-fine change 1.5% (area) and 3.2% (peak von Mises); adaptive dt change 0.5% and 1.0%.
Numerical solver error - Model C - Commissural misalignment	4	9.8 ± 1.4 iterations; augmented Lagrangian contact <2 µm violation; energy drift 0.7%.
Model form - Model C - Commissural misalignment	3	Gap 2.6 mm vs 2.4 ± 0.5 mm (8%); regurgitant volume 13.6 ml vs 13.1 ± 2.1 ml (4%).
Model inputs - Model C - Commissural misalignment	3	±2° rotation changed gap 8.8%; fiber angle uncertainty added 3.1% to stress variability.
Test conditions - Model C - Commissural misalignment	3	Pressure RMS error 3.2 mmHg; 10° rotation ±1°; fixture may distort post compliance.
Test samples - Model C - Commissural misalignment	2	n=3 devices; regurgitant volume SD 2.1 ml; gap SD 0.5 mm.