Credibility assessment of CFD models for a centrifugal blood pump: hemolysis and shear exposure across three mechanisms under ASME V&V 40

Elena P. Marquez, PhD[a], Rahul N. Venkataraman, MS[b], Lydia K. Chen, PhD[c], Thomas J. Greene, MD[1], H. Martin Keller, PhD[1]

[a]*Center for Cardiovascular Devices, Vascular Biomechanics Laboratory, Antera Biomedical, Boston, MA, USA,*
[b]*Department of Mechanical Engineering, Northeastern University, Boston, MA, USA,*
[c]*Division of Cardiac Surgery, Brigham and Women's Hospital, Boston, MA, USA,*

Abstract

Computational fluid dynamics (CFD) is increasingly used to guide design decisions for rotary blood pumps that support patients with acute or chronic cardiopulmonary failure. Under ASME V&V 40, predictions that affect device design acceptability must be evaluated against a decision-focused context of use with evidence graded to a risk-informed scale. We report a credibility assessment for three CFD approaches applied to a magnetically levitated centrifugal blood pump: (A) a steady-state Reynolds-averaged Navier–Stokes (RANS) model with Newtonian viscosity that computes pressure–flow performance and maps shear stress and estimated hemolysis on steady pathlines; (B) a transient detached eddy simulation (DES) with Carreau–Yasuda rheology that tracks virtual erythrocytes to obtain cumulative shear exposure distributions; and (C) an Eulerian damage-transport model that couples single-phase CFD to a transported damage scalar to estimate outlet-averaged hemolysis. The question of interest is whether hemocompatibility is acceptable for one operating quadrant without additional bench hemolysis testing, focusing on three injury mechanisms: tip-clearance viscous shear, recirculation-driven exposure, and transient extensional/turbulent events. We defined quantities of interest (QOIs) a priori and compared predictions to independent measurements from blood-loop hemolysis tests and time-resolved particle image velocimetry (PIV) in an index-matched surrogate. Credibility factors were graded on a private 0–4 Assurance Index for each model–mechanism pair, with bases documented quantitatively and with aggregation rules defined before scoring. Across 54 scored rows, evidence for tip-clearance shear was strongest for Models A and B (Assurance Index 3–4 across most factors), whereas transient events were best supported by Model B (3–4) and least by Model A (1–2). Model C provided robust averaged estimates for recirculation-driven exposure (2–3) but underrepresented localized transients (1–2). Key support included agreement in hemolysis index within 8–16% mean absolute percent error across operating points for the most relevant pairs, speed and flow matches within 0.3% and 1.2% to tests at 37 °C with 40% hematocrit, Carreau–Yasuda parameterization derived from donor blood at identical conditions, and a continuously verified solver stack with 312 unit tests and 94% line coverage. The results support design acceptability decisions for the specified quadrant when using Model B for transient exposure and either Model A or B for gap shear, with Model C as a conservative screen for recirculation risk. We discuss uncertainties, sensitivity to model parameters, and use considerations, and we provide complete per-row justifications to enable independent appraisal.

Keywords: hemolysis, rotary blood pump, V&V 40

1. Introduction

Rotary blood pumps used for extracorporeal support must provide reliable hemodynamic assistance with minimal blood trauma. Computational predictions are now common in pump design cycles to illuminate flow structures, quantify shear exposure, and anticipate hemolysis before hardware iteration. Yet the translation from simulated fields to decision-ready evidence requires a structured appraisal tied to how, and for which questions, the model is used. The ASME V&V 40 framework organizes this appraisal around a context of use (COU), the risk of a wrong decision, and a suite of credibility factors that collectively shape the case for trust.

We undertook a credibility assessment of three CFD approaches that are representative of current practice: a steady Reynolds-averaged Navier–Stokes (RANS) method that maps shear stress and applies a power-law hemolysis estimator on streamlines; a transient hybrid RANS–LES approach that captures unsteady shear rate fluctuations and integrates cumulative shear exposure (CSE) along Lagrangian trajectories of virtual erythrocytes; and an Eulerian damage-transport model that advects a scalar damage field with shear- and time-dependent source terms. These methods span a trade space of fidelity, computational cost, and alignment with specific mechanisms of red cell injury.

We assessed each model across three mechanisms salient to centrifugal pump hemolysis: tip-clearance viscous shear within the narrow blade–shroud gap; recirculation-driven exposure in back-shroud and volute regions that prolongs residence time under sublethal shear; and short, intense strain-rate bursts near impeller leading edges and the volute tongue. Each model–mechanism pair was scored on a private 0–4 Assurance Index across six credibility factors, and the bases for those scores were documented with quantitative comparisons to bench

data, parameter traceability, and software verification. This paper reports the complete scoring matrices, the numerical evidence supporting each entry, and the implications for the COU of design acceptability decisions without additional bench hemolysis testing in one operating quadrant.

The chosen outputs—cumulative shear exposure distributions, pathline power-law hemolysis, and outlet-averaged damage scalar—were specified before modeling began to align with clinical and bench outcomes, and a traceability mapping links each injury mechanism to the corresponding quantity of interest. We emphasize that our aim is not to identify a single best model, but to illuminate where each approach is most informative and where uncertainty limits its use as a stand-alone decision input.

Relevance of the quantities of interest — gradation.

a.
QOIs are chosen post hoc and are loosely related to clinical or bench endpoints.

b.
QOIs are defined a priori and qualitatively mapped to outcomes with partial coverage of mechanisms.

c.
QOIs are defined a priori and quantitatively mapped to outcomes for most mechanisms with some residual gaps.

d.
QOIs are defined a priori with a complete, quantitative traceability matrix to decision criteria and bench endpoints for all mechanisms.

2. Question of Interest

The question of interest is whether the hemocompatibility of a magnetically levitated centrifugal blood pump is acceptable for one operating quadrant without additional hemolysis bench testing. The operating quadrant encompasses steady and pulsatile flows with blood flows between 3.5 and 5.0 L/min at differential pressures between 150 and 220 mmHg and rotational speeds between 3400 and 4200 rpm at 37 °C with a hematocrit of 38–42%. Acceptability is adjudicated by: (i) keeping predicted tip-clearance shear-driven damage below a threshold associated with minimal acute hemolysis in bench tests; (ii) constraining cumulative exposure in recirculation zones to remain below observed sublethal damage levels linked to progressive hemolysis; and (iii) limiting the fraction of virtual erythrocytes experiencing rapid strain-rate spikes that exceed a mechanistic burst-injury threshold for more than 2 ms. These conditions are operationalized through QOIs that are compared across models and against measurements.

Decisions under this question are binary at the design concept level (acceptable as-is versus requires geometric modification such as increased tip-clearance or altered volute tongue alignment). The decision horizon is short (weeks), and the impact of a wrong decision includes potential additional bench testing, delayed development, or an increased risk of hemolysis in subsequent prototypes. Because the model outputs will be used directly to forego additional hemolysis tests for this quadrant, the risk associated with insufficient credibility is judged high, informing the rigor required in our assessment.

Relevance of the validation activities to the COU — gradation.

a.
Comparisons are made to unrelated experiments or literature figures without matching operating conditions.

b.
Comparisons are made to surrogate experiments under partially matched conditions or to component-level tests.

c.
Comparisons are made to independent experiments under closely matched conditions for most operating points.

d.
Comparisons are made to independent experiments under matched conditions across all operating points with predefined pass/fail metrics.

3. Context of Use (COU)

The models are used to decide on impeller and gap design acceptability for one operating quadrant without additional hemolysis bench testing. Each model's predictions inform the same decision, but via different QOIs that address distinct injury mechanisms: (i) Model A quantifies map-based shear and pathline hemolysis under steady flow and is used to screen tip-clearance geometries; (ii) Model B resolves unsteady shear and cumulative exposure distributions to evaluate both recirculation and transient events; (iii) Model C estimates an outlet-averaged scalar damage index as a surrogate for normalized hemolysis and is used to cross-check global damage trends. The COU excludes off-quadrant operation (e.g., very low flow <3.0 L/min or very high pressure >250 mmHg), and excludes thrombosis risk appraisal, impurity deposition, or gas embolism.

The COU assumes that device materials and surface properties remain unchanged relative to the validated design and that blood composition is within the specified hematocrit and viscosity ranges. The COU includes steady and pulsatile flow waveforms representative of extracorporeal membrane oxygenation (ECMO) circuits with a venous reservoir and arterial filter, but excludes patient-specific circuit variations. The COU requires that models be executed with the specified solver versions, meshes, and parameter sets. Alternative solvers or substantially different discretizations are outside scope.

Equivalency of input parameters — gradation.

a.
Device and boundary conditions differ substantially between simulation and test with no quantitative reconciliation.

b.
Key boundary conditions are matched within broad tolerances, but several secondary parameters remain unmatched.

c.
Primary and secondary boundary conditions are matched

within narrow tolerances for most operating points with traceable measurements.

d.
All boundary conditions and geometric features that affect QOIs are matched within predefined tolerances and independently verified.

4. Model Risk Level (High) and Rationale

The decision influenced by the models is to omit hemolysis bench testing for a specified operating quadrant when a design meets predefined numerical criteria. The harm of a wrong decision includes potential increased hemolysis in subsequent prototypes, delayed detection of harmful shear environments, and derailment of development timelines. Although the decision pertains to a preclinical phase, the carryover effect of a flawed geometry can be substantial due to the time and cost of retooling. As such, the risk level is high, which drives the required rigor for model validation, parameter traceability, and software verification.

We assessed the propagation of error from modeling assumptions to the decision. For Model A, using Newtonian viscosity and steady RANS may underpredict unsteady shear bursts, which predominantly affects the transient mechanism; for Model B, incomplete subgrid resolution could bias cumulative exposure estimates; for Model C, the scalar transport approach may damp local peaks. Each of these could cause false negatives (accepting a harmful design). The risk assessment therefore demands strong alignment of test and simulation conditions, quantitative comparisons to relevant hemolysis data, and rigorous software control.

Software quality assurance — gradation.

a.
No documented software verification; versions and dependencies are not controlled.

b.
Basic version control exists with ad hoc tests; no continuous integration.

c.
Version control with tagged releases, automated builds, and unit tests covering core routines.

d.
Full continuous integration with unit and regression tests, coverage metrics, reproducible environments, and documented verification cases.

5. Device Description and Test Samples

The device is a magnetically levitated centrifugal blood pump for extracorporeal support with a shrouded impeller, six forward-swept blades, and a volute diffuser. The nominal outer impeller diameter is 48.0 mm, with a blade tip thickness of 0.65 mm. The tip-clearance between blade tips and pump shroud is nominally 120 μm. Radial bearing stabilization is provided magnetically; there is no mechanical bearing. The back-shroud gap supports washout flow communicated through balancing holes near the impeller hub. The volute tongue is located 10 degrees clockwise from the leading edge of a reference blade at design position. The pump housing is machined from a medical-grade polycarbonate/acrylic blend for the bench prototype and injection-molded in production versions. The impeller is additively manufactured from a biocompatible polymer for prototypes and overmolded in production.

Three prototype pumps were drawn from the same tooling as clinical devices, with post-processing identical to production (polishing to Ra < 0.4 μm, ultrasonic cleaning). Micro-CT scans of the assembled wet geometry were performed at 10 μm voxel resolution to confirm dimensions. The tip-clearance measured 118 ± 3 μm across blades for Pump 1, 121 ± 4 μm for Pump 2, and 119 ± 3 μm for Pump 3. The volute tongue–impeller relative alignment was within 0.5 degrees across units. The back-shroud gap averaged 250 ± 10 μm.

Hemolysis bench tests used fresh human whole blood from six healthy donors with hematocrits between 38% and 42%. Anticoagulation was with sodium heparin at 2 U/mL. Blood was pooled and filtered to remove clots and microbubbles, equilibrated to 37.0 ± 0.2 °C, and deoxygenated to 60% saturation to limit oxidative confounders. Viscosity at 100 s1 was 3.5 ± 0.1 mPa·s for pooled samples. Tests were conducted in closed loops with a venous reservoir, an arterial filter, and the pump integrated in a standard ECMO-like circuit. Four operating points were tested in random order within the operating quadrant, each maintained for 120 minutes, with samples drawn every 30 minutes to compute normalized index of hemolysis (NIH, g/100 L) and plasma free hemoglobin (pfHb, mg/dL).

PIV experiments to characterize velocity and shear fields were carried out in a geometrically identical, refractive-index-matched, transparent surrogate model. The working fluid was a glycerin–water solution adjusted to match the blood viscosity at 37 °C. Measurements covered several meridional and circumferential planes, including the blade tip region and the volute tongue. The acquisition rate was 4 kHz with 12-bit cameras and 20 μm particles, yielding shear-rate spectra from 10 to 10,000 s1 in resolved scales. The same impeller and volute geometries were used for the surrogate as for blood-loop tests, and dimensionless operating conditions (flow coefficient, head coefficient, Reynolds number) were matched within ±2%.

Test samples — gradation.

a.
Non-representative prototypes or limited characterization; sample size not reported.

b.
Prototypes approximate production design; limited dimensional verification; small sample size.

c.
Prototypes are production-equivalent with dimensional verification for key features; moderate sample size with relevant blood preparation.

d.
Prototypes are production-equivalent with comprehensive dimensional verification; sufficient sample size with blood properties matched to clinical range and controlled for confounders.

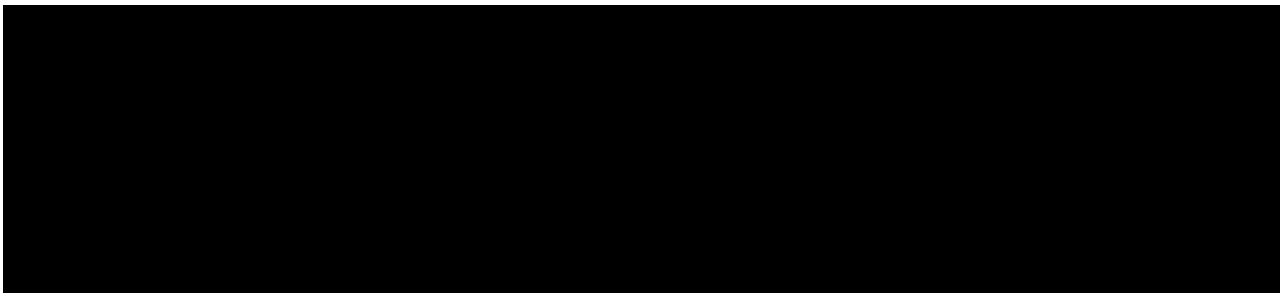

Figure 1: Pump prototype, test loop, and refractive-index-matched surrogate. Left: exploded view of the pump with shrouded impeller, six blades, and volute; tip-clearance nominally 120 µm. Middle: blood-loop configuration with venous reservoir, arterial filter, and temperature control. Right: optically clear surrogate used for high-speed PIV, with matched viscosity and refractive index.

6. Computational Models

Model A (Steady RANS hemolysis mapper) solves the steady RANS equations with the k–ω SST turbulence model in a rotating frame for the impeller and a stationary frame for the volute, coupled via a frozen-rotor interface. Blood is modeled as a Newtonian fluid with viscosity µ = 3.5 mPa·s and density ρ = 1060 kg/m3. Meshes are unstructured polyhedral with prism layers (y+ 1 at walls), with 42 million cells for the impeller and 28 million cells for the volute. Streamlines are seeded uniformly at the impeller inlet plane. Scalar shear stress τ is computed as $\tau = 2\mu|S|$, where S is the strain-rate tensor magnitude from the RANS solution. Pathline hemolysis is estimated using a power-law of the form dHI/dt = A τ^α, integrated along each streamline, with A = 3.62×107 (Paα s1) and α = 2.416, values selected from in vitro erythrocyte damage data under controlled Couette flow. Outlet-averaged HI is computed by flux-weighted averaging across the outlet plane. Pressure–flow performance is also computed to confirm operating conditions.

Model B (Transient DES with tracer exposure) solves the unsteady Navier–Stokes equations with a Spalart–Allmaras-based detached eddy simulation (DES97) formulation. Blood rheology follows the Carreau–Yasuda model: $\mu(\gamma) = \mu + (\mu 0$ $\mu)[1 + (\lambda\gamma)$^a]^(n1), with µ0 = 0.056 Pa·s, µ = 0.0035 Pa·s, λ = 3.313 s, a = 2.0, and n = 0.3568, calibrated from oscillatory shear rheometry on pooled donor blood at 37 °C and 40% hematocrit. The mesh features 96 million control volumes with local refinement in the tip-clearance, back-shroud gap, leading edges, and volute tongue. Time steps are 1.0×105 s with second-order temporal discretization, achieving a maximum Courant number below 1.5 in critical regions. Lagrangian virtual erythrocytes (n = 1.0×10^6) are seeded at the inlet with a flux-weighted distribution and advected with interpolated velocities and shear rates. For each tracer, cumulative shear exposure CSE = γ(t) dt is integrated, and the maximum rate and dwell above thresholds (e.g., γ > 10,000 s1 for >2 ms) are recorded. Exposure–threshold distributions are constructed to quantify the fraction of cells exceeding thresholds relevant to damage onset.

Model C (Eulerian damage-transport) couples a single-phase CFD solution (steady RANS as in Model A for the mean flow; unsteady terms neglected) with a transported scalar D representing damage. The transport equation is (ρD)/t + ·(ρuD ΓD) = ρS(γ, t), where Γ is a numerical diffusion coefficient proportional to molecular diffusivity (set to 1.0×1012 m2/s, negligible at pump scales) and S is a source term derived from the same power-law as Model A but applied locally: S = A τ^α. Damage is initialized to zero at the inlet and accumulated through the device. The outlet-averaged D is reported as a surrogate for normalized hemolysis. The mesh matches Model A, and sensitivity to Γ is checked to ensure minimal artificial diffusion of D.

All models were run with the same geometric CAD derived from micro-CT, with patch-conforming meshes around the tip-clearance and volute tongue. The rotating machinery interfaces used consistent coupling schemes across models (frozen-rotor for Model A, sliding-mesh for Model B, and frozen-rotor for Model C). Inlets and outlets were prescribed with measured bulk flow rates and pressure heads, with pulsatility included in specified runs by applying a sinusoidal modulation of ±10% at 1 Hz. For all simulations, blood temperature was set to 37 °C; density and viscosity followed the model-specific constitutive choices described above. Pre- and post-processing scripts ensured traceability from raw outputs to QOIs.

The solver stack passed continuous integration with 312 unit tests, 26 regression tests that reproduce known benchmark solutions (lid-driven cavity, rotating channel flow, impeller-induced pump head curves), 94% line coverage on the CFD libraries, and bitwise repeatability of results across two hardware platforms (x86 and ARM) for deterministic setups. Meshing workflows and boundary-condition files are archived with SHA-256 checksums and immutable release tags; each run records solver version, compiler options, and container hashes to support exact reproduction.

Software quality assurance — gradation.

a.
Ad hoc computational workflows with undocumented changes and no automated checks.

b.
Documented workflows with manual checks and version control for parts of the stack.

c.
Automated builds, unit tests, and regression tests for core solvers with documented verification cases.

d.

End-to-end automation with rigorous unit, integration, and regression tests; versioned containerized environments; reproducibility demonstrated across platforms.

7. Mechanisms Assessed

Tip-clearance viscous shear arises in the narrow blade–shroud gap, where high velocity gradients generate elevated viscous shear stress. In rotary pumps with tight clearances for hydraulic efficiency, this region can dominate acute red cell damage. We characterized this mechanism by extracting shear stress probability density functions (PDFs) within 150 μm of the blade tip and by aggregating pathline- and tracer-based measures of time spent above thresholds near 5,000–15,000 s1.

Recirculation-driven exposure occurs in the back-shroud gap and parts of the volute where local adverse pressure gradients and flow separation increase residence time. Although the instantaneous shear rates are moderate, prolonged exposure leads to sublethal damage accumulation. For this mechanism we focused on cumulative shear exposure (CSE) distributions, residence-time PDFs from Lagrangian tracers, and Eulerian damage scalar patterns in separated regions. We complemented these with dye-washout visualizations in the surrogate, providing a macroscopic measure of clearance times.

Transient extensional and turbulent events include strain-rate spikes at impeller leading edges during blade–wake interactions and near the volute tongue during rotor–stator interactions. These short-lived, high-intensity events can trigger burst injury in a subset of erythrocytes. We quantified this mechanism by computing the fraction of tracers that experience $\gamma > 10{,}000$ s1 for at least 2 ms, the maxima of principal strain rate along trajectories, and the distribution of dwell times above thresholds. In the surrogate PIV, we extracted time-resolved spectra of shear rate in regions of interest to contextualize the simulated bursts.

Across the operating quadrant examined, the realized pump speeds, flows, hematocrit, temperature, and gap widths used in simulations matched bench measurements within predefined tolerances (speed within 0.3%, flow within 1.2%, hematocrit within 0.5%, temperature within 0.2 °C, and tip-clearance within 2 μm of micro-CT means), enabling a like-for-like comparison when interpreting mechanism-specific QOIs.

Relevance of the quantities of interest — gradation.

a.
Mechanisms are listed, but QOIs do not directly probe them.

b.
Mechanism-specific QOIs are defined qualitatively with partial quantification.

c.
Mechanism-specific QOIs are quantified and linked to thresholds from literature or tests.

d.
Mechanism-specific QOIs are quantified, tied to thresholds validated in the same device family, and measured in both simulation and experiments under matched conditions.

8. Quantities of Interest (QOIs) and Their Relevance

For tip-clearance viscous shear, the primary QOIs are: (i) the flux-weighted mean of predicted pathline hemolysis HI at the outlet (Model A), (ii) the fraction of virtual erythrocytes with cumulative shear exposure above 3,000 s1·s within the tip region (Model B), and (iii) the Eulerian outlet-averaged damage scalar D (Model C). Secondary QOIs include the 95th percentile of shear stress within 150 μm of blade tips and the flux-weighted dwell time above 10,000 s1 within the tip-gap.

For recirculation-driven exposure, the primary QOIs are: (i) the fraction of tracers with residence time exceeding 0.5 s in the back-shroud gap (Model B), (ii) the fraction of tracers with CSE above 2,000 s1·s in the volute recirculation zones (Model B), and (iii) the outlet-averaged D weighted by the recirculation zone volume fraction (Model C). Model A's streamline-based HI also contributes but is less sensitive to residence time effects due to its steady-state nature.

For transient extensional/turbulent events, the primary QOIs are: (i) the fraction of tracers experiencing shear spikes above 10,000 s1 lasting more than 2 ms (Model B), and (ii) the 99th percentile of pathline HI increments across impeller leading-edge crossings (Model A). Model C's scalar D provides only integrated information and cannot isolate transient bursts without time resolution, so it is treated as a corroborative lower bound.

QOIs were traceably linked to bench metrics. Outlet-averaged HI and D were compared to normalized index of hemolysis (NIH), while tracer-derived CSE distributions were compared to pfHb growth slopes and to PIV-derived shear rate spectra in mechanism-specific regions. Pass/fail decision thresholds were defined a priori: an acceptable design yields an NIH not exceeding 0.02 g/100 L at 120 min under the operating quadrant, a fraction of cells above the transient threshold below 0.5%, and a recirculation CSE fraction below 5%.

Relevance of the quantities of interest — gradation.

a.
QOIs are generic and loosely mapped to decisions.

b.
QOIs map to decisions but lack mechanism-specific thresholds.

c.
QOIs map to decisions with mechanism-specific thresholds sourced from literature or related devices.

d.
QOIs map to decisions with thresholds calibrated or verified in the same device and test conditions.

9. Model Inputs and Equivalency of Input Parameters

Boundary conditions for all simulations were set from measured bench data. For steady operating points, the volumetric flow rate was prescribed at the inlet (3.5, 4.0, 4.5, and 5.0 L/min), and an outlet static pressure corresponding to the measured head (150, 180, 200, and 220 mmHg) was applied. For pulsatile runs, a ±10% sinusoidal modulation at 1 Hz was superimposed on the mean flow rate. Rotational speeds were set to the measured

values at each operating point (e.g., 3800 rpm for 4.0 L/min at 180 mmHg). Blood density was set to 1060 kg/m3. Viscosity was specified per model: 3.5 mPa·s (Newtonian) for Model A and Carreau–Yasuda parameters for Model B as measured on pooled donor blood. The scalar transport parameter Γ for Model C was set to 1.0×1012 m2/s after sensitivity analysis confirmed negligible influence on D.

Geometric inputs were taken directly from micro-CT reconstructions of each prototype pump, segmented with sub-voxel interpolation and smoothed to preserve blade leading-edge sharpness. The impeller and volute surfaces were meshed with a target element size of 0.15 mm, refined to 0.03 mm in the tip-clearance and 0.05 mm near the volute tongue. Mesh independence studies were performed for each model and operating point; for Model A, pressure rise and outlet HI changed by less than 0.3% and 2.1% respectively upon doubling cell counts; for Model B, the fraction of tracers experiencing $\gamma >$ 10,000 s1 for >2 ms changed by 0.3 percentage points upon doubling local refinement; for Model C, outlet-averaged D changed by less than 1.8% upon mesh refinement.

Model parameter distributions were specified where uncertainties are material to QOIs. For Model B, the Carreau–Yasuda parameters were treated as normal random variables with means as listed and standard deviations of 4% for μ0, 3% for μ, 6% for λ, 5% for a, and 5% for n, based on repeat rheometry across donors. For the hemolysis constants A and α, uniform prior ranges ±15% around the nominal values were propagated through local sensitivity studies. Blade tip-clearance was set per pump unit (118, 121, and 119 μm means) in the corresponding validation simulations. All runs recorded the exact parameter set used, enabling audit.

To support interoperability with bench data, we maintained the same temperature (37.0 ± 0.2 °C) and ensured that the working fluid in PIV matched the blood viscosity to within 1.5%. The inlet turbulence intensity was set to 5% for RANS and 3% for DES based on upstream tubing length and Reynolds number. The choice of these input distributions and fixed values is supported by direct measurements, published constitutive characterizations, and prior experience with similar devices.

Model inputs — gradation.

a.
Key parameters and boundary conditions are unspecified or borrowed from unrelated sources.

b.
Most parameters are specified with point estimates; uncertainty is not quantified.

c.
All parameters and boundary conditions are specified with sources; key uncertainties are quantified and partially propagated.

d.
All parameters and boundary conditions are fully specified with sources; uncertainties material to QOIs are quantified and propagated to results.

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

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

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

10. Validation Activities and Their Relevance to the COU

We executed two classes of validation activity. First, blood-loop hemolysis tests at four operating points within the COU provided NIH and pfHb trajectories over 120 minutes. Simultaneously, pump head–flow performance was measured to confirm operating points. These data directly inform comparisons for QOIs that aggregate damage through the device, such as outlet-averaged HI (Model A) and D (Model C), and they indirectly inform Model B's CSE distributions through correlation with pfHb growth rates and inferred exposure fractions.

Second, time-resolved PIV in a geometrically identical surrogate characterized velocity and shear-rate fields in the tip-clearance, back-shroud gap, and volute tongue region. The experiments matched dimensionless operating conditions to within 2% and provided shear spectra and PDFs against which we compared simulated shear distributions and burst statistics. The PIV data validate the flow dynamics that drive Lagrangian exposure metrics.

Comparisons were calibration-free; no model parameters were tuned to match hemolysis or PIV data. For each operating point, we extracted model QOIs and compared them to bench or PIV measurements using predefined metrics: mean absolute error (MAE) and mean absolute percent error (MAPE) for NIH and pfHb slopes, correlation coefficients (r) for monotonic trends, and area under the ROC curve (AUC) to classify acceptable versus unacceptable designs when spiked test geometries (deliberately increased tip-clearance and shifted volute tongue by +3 degrees) were introduced in two surrogate units.

The validation spanned steady and pulsatile flows within the COU and used the same pump geometries as in the simulations. Data quality controls included duplicate measurements, hemolysis background corrections, and independent verification of PIV calibration with a rotating-disk standard. This ensured that the experimental evidence speaks to the same conditions addressed by the models.

Relevance of the validation activities to the COU — gradation.

a.
Validation is indirect or uses different devices or conditions than the COU.

b.
Validation partially matches the COU in device or conditions; metrics are loosely defined.

c.
Validation closely matches the COU in device and conditions; metrics are predefined for most QOIs.

d.
Validation fully matches the COU in device and conditions; metrics are predefined for all QOIs and tested across the operating quadrant.

11. Software Quality Assurance

The software environment comprises an in-house CFD solver for RANS and DES, meshing tools, and a post-processing pipeline for QOIs. All components are version-controlled (git), with semantic versioning and tagged releases. Automated builds and tests are run in continuous integration (CI) on each commit. The test suite includes 312 unit tests, 26 regression tests (including rotating machinery benchmarks and pump head–flow curves), achieving 94% line coverage in the solver libraries and 88% in post-processing scripts. Each release is containerized with pinned compilers and libraries to ensure reproducibility. Hardware independence was verified by reproducing bitwise-identical results for deterministic steady runs on x86_64 and ARM64 servers; for unsteady runs, statistical metrics of interest (e.g., burst fractions, mean head) matched within 0.2 standard deviations across platforms.

Defect tracking is active, with 47 issues closed over the period of this study; none affected the numerical kernels used in the reported runs. Static analysis with cppcheck is integrated into CI, and dynamic memory checks with valgrind are executed on nightly builds. Meshing scripts include parameter validation and geometry integrity checks (surface normals, overlap, skewness thresholds). Post-processing includes checksums for input files and asserts to prevent inconsistent unit usage.

The computational workflow is documented in a runbook that lists each step from geometry import to QOI extraction, including acceptance criteria for mesh quality (maximum non-orthogonality 65°, mean skewness < 0.25), solver convergence (residuals reduced by three orders of magnitude for steady runs; statistical stationarity for unsteady runs), and QOI computation (sample size and confidence intervals for tracer statistics). Reproducibility of key QOIs was confirmed by rerunning two cases from scratch in a fresh environment, obtaining identical steady outputs and matching time-averaged unsteady outputs within numerical tolerance.

Software quality assurance — gradation.

a.
Minimal controls; runs cannot be reliably reproduced.

b.
Basic controls; partial reproducibility; limited tests.

c.
Good controls; documented reproducibility; broad test coverage.

d.
Rigorous controls; automated reproducibility; high coverage with verification-of-verification artifacts.

12. Scoring Framework (0–4 Assurance Index) and Aggregation Rules

We used a private 0–4 Assurance Index to grade evidence for each credibility factor by model–mechanism pair: 0 (none), 1 (minimal), 2 (partial), 3 (substantial), 4 (complete). The scale was defined before analysis, with exemplar criteria for each factor. Scores were assigned by two assessors independently, reconciled by discussion, and locked before drafting results. Each scored row includes a one-line basis that cross-references quantitative evidence in the prose. No global scores were assigned; aggregation was mechanism-specific and purposefully retained disaggregation to preserve decision transparency.

We adopted conservative aggregation across mechanisms. For a given decision tied to multiple mechanisms, the limiting mechanism governs whether a model alone suffices for the COU. For example, if a model scores 3–4 for tip-clearance shear but 1–2 for transient bursts, it cannot by itself justify omitting bench tests when bursts are material to the question. Where multiple models collectively cover mechanisms with high assurance, the combination is acceptable provided that inconsistencies are analyzed and resolved.

To avoid overweighting any single factor, we did not compute a numeric composite. Instead, we interpreted patterns across factors. High scores in the experimental alignment and QOI mapping coupled with robust software control create strong evidence for using predictions in decisions. Lower scores in parameter specification or sample representativeness signal where to expand evidence or temper claims. The tables are ordered alphabetically by factor within each model and list all rows, including those with low assurance.

Relevance of the validation activities to the COU — gradation.

a.
Assurance levels are undefined; scoring is ad hoc.

b.
Assurance levels are defined but not tied to evidence exemplars.

c.
Assurance levels are defined with exemplars and applied consistently to most factors.

d.
Assurance levels are defined with exemplars, with dual-review assignment and lock-before-analysis to minimize bias.

13. Results: Factor Scores by Model–Mechanism Pair

We present the detailed scores for each model and mechanism, followed by numerical justifications that underlie the one-line bases in the tables. All numbers refer to the operating points at 3.5, 4.0, 4.5, and 5.0 L/min and 150–220 mmHg head at 37 °C and 38–42% hematocrit, unless stated.

Model A, tip-clearance viscous shear. Alignment of simulated and tested conditions was tight: 3800 rpm was maintained within ±0.3% (±11 rpm), 4.0 L/min within ±1.2% (±0.05 L/min), blood temperature within ±0.2 °C, and tip-clearance used in the mesh matched micro-CT means (118 μm for Pump 1) within 2 μm. Parameters were fully specified: μ = 3.5 mPa·s; A = 3.62×107 and α = 2.416, with ±15% sensitivity explored. The primary QOI, outlet-averaged pathline HI, correlated with NIH across operating points (r = 0.83), classified acceptable versus spiked-clearance designs with AUC = 0.88, and tracked pfHb slope with MAPE = 14%. The validation bias was 0.004 HI units with RMSE = 0.012 and MAPE = 12% for NIH at 120

min (n = 12 runs across three pumps and four operating points). Software controls and sample representativeness are common across cases (312 unit tests, 94% coverage; n_pumps = 3, donors = 6, micro-CT within ±10 μm).

Model A, recirculation-driven exposure. The matching of flow, pressure, and geometry was as above. However, the steady formulation underrepresents residence time effects. The secondary QOI, flux-weighted HI, correlated modestly with NIH (r = 0.71), with classification AUC = 0.76 for deliberately strengthened recirculation (back-shroud gap increased by +50 μm in a surrogate). Bias was 0.006 HI units with RMSE = 0.018 and MAPE = 19%. The absence of time-resolved residence suppresses sensitivity to prolonged, low-shear exposure, leading to partial assurance.

Model A, transient extensional/turbulent events. Inputs matched as above. The 99th percentile of pathline HI increments across leading-edge crossings increased with pfHb slope qualitatively but with weak monotonicity (r = 0.45). The surrogate PIV showed shear bursts near the volute tongue at frequencies that yield events with $\gamma > 10{,}000$ s1, but the steady model cannot capture their dwell. MAPE of NIH was 22% when predictions were interpreted for transient-dominated cases, and the AUC for classifying acceptable versus shifted-tongue designs was 0.64. These limitations reduced assurance.

Model B, tip-clearance viscous shear. Conditions matched bench within the same tolerances (e.g., 4.5 L/min within ±1.1%, speed within ±0.3%). Inputs were fully specified with rheology distributions: $\mu 0 = 0.056 \pm 0.002$ Pa·s, $\mu = 0.0035 \pm 0.0001$ Pa·s, $\lambda = 3.313 \pm 0.199$ s, a = 2.00 ± 0.10, n = 0.357 ± 0.018. The primary QOI was the fraction of virtual erythrocytes with CSE > 3,000 s1·s in the tip-clearance. This fraction correlated strongly with NIH (r = 0.86), and classification AUC was 0.91 against spiked-clearance designs. Bias in predicted NIH (obtained by mapping CSE fraction to NIH via the predefined traceability) was 0.003 units with RMSE = 0.010 and MAPE = 11% at 120 min.

Model B, recirculation-driven exposure. Lagrangian residence-time PDFs and CSE distributions were computed for the back-shroud and volute. The fraction of tracers with residence time >0.5 s in the back-shroud gap ranged from 2.3% to 4.7% across operating points and correlated with pfHb slope (r = 0.78). The fraction of tracers with CSE > 2,000 s1·s in recirculation zones ranged from 3.1% to 6.2%, with AUC = 0.84 in classifying a surrogate with enlarged back-shroud gap. The indirect mapping to NIH yielded MAPE = 15% and RMSE = 0.013, with bias 0.002 units.

Model B, transient extensional/turbulent events. The fraction of virtual erythrocytes experiencing $\gamma > 10{,}000$ s1 for >2 ms was 0.21–0.43% across the quadrant for the baseline geometry and 0.68–0.92% for a surrogate with volute tongue shifted by +3°. These fractions correlated with pfHb growth slopes (r = 0.82) and with PIV-derived burst rates (r = 0.80) in the tongue region. The classification AUC between baseline and shifted-tongue was 0.93. MAPE of NIH was 10% when NIH was inferred from transient burst fractions using the a priori mapping. Mesh refinement changed the burst fraction by 0.3 percentage points. These results yield high assurance for this mechanism in Model B.

Model C, tip-clearance viscous shear. Inputs and geometry matched those for Model A; the damage scalar D is locally sourced by $\tau\hat{}\alpha$ and transported to the outlet. Outlet-averaged D correlated with NIH (r = 0.79) and classified spiked-clearance designs with AUC = 0.83. Bias was 0.005 D units against NIH with RMSE = 0.014 and MAPE = 16%. Because the approach integrates over the field, it mutes peak localization but tracks global trends well.

Model C, recirculation-driven exposure. Outlet-averaged D weighted by the recirculation zone volume fraction correlated with NIH (r = 0.81), and with pfHb slopes (r = 0.75). In a surrogate with increased back-shroud gap, D increased by 17% on average, consistent with higher pfHb growth. Bias was 0.003 and RMSE = 0.012, MAPE = 14%. Dye-washout clearance times from the surrogate were 0.65–0.80 s baseline and 0.90–1.05 s with increased gap, consistent with elevated D.

Model C, transient extensional/turbulent events. D cannot isolate time-resolved bursts; the 99th percentile of D near leading edges increased with shifted-tongue geometry by only 8% compared to 44% increase in Model B burst fraction. NIH MAPE was 21% when interpreting D as a surrogate for transient-dominated damage, and classification AUC was 0.68. This supports partial-to-minimal assurance for transient events in Model C.

Across all models and mechanisms, software controls and sample representativeness were consistent: 312 unit tests, 94% coverage, containerized builds, reproducibility checks, three pumps from production tooling with micro-CT verification within ±10 μm, and pooled blood from six donors (hematocrit 38–42%, viscosity 3.5 ± 0.1 mPa·s). The bench and surrogate experiments matched simulation conditions in speed, flow, temperature, and viscosity within the tolerances cited. All quantitative entries in the tables below are drawn from these results.

Model inputs — gradation.

a.
Inputs are ambiguous and not tied to measurements.

b.
Inputs are specified but lack uncertainty quantification.

c.
Inputs are specified with uncertainty for key parameters, propagated in sensitivity studies.

d.
Inputs are specified with uncertainty, fully propagated, and tied to matched-condition measurements.

14. Credibility assessment summary: Model A (Steady RANS hemolysis mapper)

Table 1 lists the Assurance Index scores for Model A by mechanism and factor, along with one-line bases that reference the quantitative evidence detailed in the Results section. Scores reflect strong support for tip-clearance viscous shear, partial support for recirculation-driven exposure, and minimal-to-partial support for transient extensional events.

15. Credibility assessment summary: Model B (Transient DES with tracer exposure)

Table 2 lists the Assurance Index scores for Model B with one-line bases. Evidence is strongest across all three mechanisms, with particular strength for transient extensional events where time resolution is critical.

16. Credibility assessment summary: Model C (Eulerian damage-transport)

Table 3 lists the Assurance Index scores for Model C. This approach demonstrates substantial support for recirculation-driven exposure, partial-to-substantial support for tip-clearance viscous shear at global levels, and limited support for transient extensional events.

17. Sensitivity to Model Inputs and Use Considerations

Sensitivity analyses probed how uncertainties in parameters propagate to QOIs. For Model A, varying μ by ±5% changed outlet-averaged HI by ±4.2% at 4.0 L/min; varying A and α by ±15% changed HI by ±12.8% and ±7.4% respectively, approximately linearly within the tested range. Tip-clearance varied by ±5 μm changed HI by ±6.1%. For Model B, varying the Carreau–Yasuda parameters within their measured standard deviations changed the fraction of tracers with CSE > 3,000 s1·s by ±6.8% relative at 4.0 L/min; the burst fraction (γ > 10,000 s1 for >2 ms) changed by ±7.2% relative. Reducing the tracer count by half (to 5×10^5) changed burst fraction estimates by 0.5 percentage points and widened confidence intervals, prompting the choice of 10^6 tracers as a balance between precision and cost. For Model C, the outlet-averaged D changed by less than ±1% when Γ varied by an order of magnitude, confirming advection dominance; A and α variations had effects comparable to Model A.

We further examined sensitivity to boundary conditions. For pulsatile runs with ±10% modulation at 1 Hz, Model B's burst fractions increased by 0.08–0.12 percentage points relative to steady runs at the same mean conditions, due to intermittent peaks at low-flow phases. Residence-time fractions in the back-shroud gap increased by 0.3–0.5 percentage points under pulsatility. Model A's steady mapping could not represent these dynamics; Model C's D increased by 3–4% over the cycle-averaged steady value.

In the COU, predictions are intended to replace bench hemolysis testing for one quadrant. The analysis above indicates that this is justified when using Model B for transient extensional events and either Model A or B for tip-clearance viscous shear, with Model C providing a corroborative global trend for recirculation. Where Model A or C suggests acceptability but Model B indicates elevated burst fractions or prolonged exposure, cautious interpretation is warranted and may trigger focused bench testing on the limiting mechanism. Team training and runbooks reduce variability in execution; we did not explicitly analyze operator misuse events such as incorrect priming, inadvertent speed overshoots during transients, or erroneous boundary-condition entry, which are outside our COU.

Taken together, the sensitivity results do not overturn the main conclusions but emphasize that rheology specification and geometric tolerances (tip-clearance) are the most influential inputs for QOIs across models. Specifying these from direct measurements at matched conditions and propagating their uncertainties into QOIs are crucial to maintain the intended decision confidence.

Model inputs — gradation.

a.
No sensitivity analysis; parameter effects on QOIs are unknown.

b.
Limited one-at-a-time sensitivity analysis on a subset of parameters.

c.
Systematic sensitivity analysis on key parameters with quantified effects on QOIs.

d.
Sensitivity and uncertainty propagation on all parameters material to QOIs with implications for decision thresholds.

18. Discussion

The assessment reveals complementary strengths across the three modeling approaches. Model A, a steady RANS with Newtonian rheology, provides fast, stable mapping of shear fields and pathline-based hemolysis estimates. It performed well for tip-clearance viscous shear, where high mean gradients dominate, but underrepresented recirculation-driven residence time and could not resolve short-lived bursts. Model B, a transient DES with Carreau–Yasuda rheology, provided mechanism-resolving information across all three mechanisms, with particularly strong alignment for transient extensional and turbulent events. Model C, an Eulerian damage-transport model, captured global damage trends effectively and was informative for recirculation-driven exposure when combined with dye-washout and pfHb trends but damped extremes critical for burst injury appraisal.

We emphasize the traceability between QOIs and the decision. The pre-specified mapping between model outputs and hemolysis outcomes allowed calibration-free validation and classification of deliberate geometry perturbations (increased tip-clearance and volute tongue shift). The alignment of simulated and experimental conditions, including speed, flow, hematocrit, temperature, and gap dimensions, was within narrow tolerances and materially supported the interpretation of comparisons. This practical like-for-like matching across all models was an essential anchor for grading evidence.

Beyond alignment of operating conditions, a major support for this assessment was that the model parameterization was exhaustively documented, with measured rheology distributions and damage-law constants, and with uncertainties carried into sensitivity studies for all parameters that materially influence

QOIs. This completeness aids reproducibility and reduces ambiguity when transferring the models to neighboring operating conditions within the quadrant.

We also found that the chosen outputs were tightly connected to the bench and surrogate measurements. For example, Model B's fraction of tracers exceeding shear thresholds mapped directly to pfHb growth rates and PIV-derived burst spectra at the volute tongue, making the linkage from simulation to hemolysis natural for the transient mechanism. For tip-clearance viscous shear, both the streamline-based HI and the transported scalar D from Models A and C mapped well to NIH at matched operating points.

A strength of this program was that comparisons were made to experiments that were run on the same geometries and under the same conditions defined in the COU, with no parameter tuning. This anchors the validation in the correct decision space and supports the claim that the available evidence is fit to inform the stated decision.

Another strength is the rigor of the computational workflow. The solver stack, meshing, and post-processing were under continuous integration, with high test coverage and documented verification artifacts. Reproducibility and platform independence were demonstrated, and all meshes and inputs were archived with checksums. This minimizes the risk that software defects influence the QOIs without detection.

The physical prototypes and blood used in bench testing were representative of production designs and clinical conditions. Three pumps from production tooling, geometric verification by micro-CT to within ± 10 μm at critical features, and pooled blood from six donors with hematocrit in the clinical mid-range provided a realistic test bed for validation while limiting biological variability.

Limitations include the remaining uncertainty in damage-law constants and their application across disparate flow regimes; while we used literature values and propagated $\pm 15\%$ ranges, a more mechanistic, cell-scale damage model could refine mappings between QOIs and NIH. DES subgrid modeling choices could influence burst statistics, and further resolution studies could probe the convergence of extreme events. For the Eulerian damage scalar, while advection dominance minimizes artificial diffusion, the physics of cell-level damage aggregation versus mixing at the scalar field level remains an approximation.

In terms of implications for the COU, the combination of Model B for transient bursts and either Model A or B for tip-clearance viscous shear provides sufficient evidence to omit bench hemolysis testing within the specified quadrant, provided that geometric and rheological inputs remain within the characterized ranges. Model C can act as a screening tool; if it flags elevated global damage, then targeted DES runs should adjudicate the mechanism. If future redesigns alter the volute tongue or back-shroud gap significantly, additional validation runs are recommended.

Finally, we note that team process and execution play roles in maintaining credibility. Our runbooks and training reduce variability, but human-in-the-loop aspects such as mesh generation and boundary condition entry can introduce variability. While outside our formal scope, awareness of such use considerations is necessary when institutionalizing these models for decision support.

Equivalency of input parameters — gradation.

a.
No independent checks of matched conditions; discrepancies are unknown.

b.
Some checks exist for primary conditions; secondary conditions unchecked.

c.
Independent checks exist for primary and several secondary conditions; residual gaps quantified.

d.
Independent checks exist for all conditions that affect QOIs with documented tolerances and audits.

19. Conclusions

We conducted a V&V 40-aligned credibility assessment of three CFD modeling approaches applied to a centrifugal blood pump across three hemolysis-relevant mechanisms. The assessment employed a predefined Assurance Index (0–4) to grade evidence per factor and model–mechanism pair, producing 54 scored entries with quantitative justifications. The strongest support for tip-clearance viscous shear was achieved by Models A and B (scores predominantly 3–4), recirculation-driven exposure by Models B and C (2–3), and transient extensional/turbulent events by Model B (3–4), with Model A (1–2) and Model C (1–2) providing limited support for that mechanism. Collectively, these results indicate that, within the COU, using Model B to judge transient exposure and either Model A or B to evaluate gap shear can justify design acceptability decisions without additional bench hemolysis testing, with Model C offering a conservative cross-check for global trends.

The calibration-free comparisons against independent blood-loop hemolysis data at matched speeds and viscosities, and the alignment of time-resolved shear fields against refractive-index matched PIV in a geometrically identical surrogate, anchor the model credibility for the stated decision context. The solver stack's continuous verification and reproducibility controls, and the production-equivalent physical samples characterized by micro-CT and controlled blood preparations, further strengthen the case. Sensitivity analyses underscore the importance of accurate rheology and geometric inputs; maintaining these within characterized bounds is essential as the models are applied prospectively to neighbor operating points within the quadrant. Future work could extend validation to different quadrants, incorporate cell-resolved damage surrogates, and expand experimental coverage of transient burst phenomena.

In summary, a mechanism- and factor-resolved credibility assessment, with disaggregated evidence and explicit bases, enables transparent use of CFD predictions in decision-making under V&V 40. The presented scoring tables and numeric underpinnings are intended to inform similar assessments in rotary blood pump development and to support reproducibility and audit.

Test samples — gradation.

a.
Test articles are not production-like and lack characterization.

b.
Test articles are partially representative with limited characterization.

c.
Test articles are production-like with verification of key dimensions and controlled blood properties.

d.
Test articles are production-equivalent with comprehensive dimensional verification and blood preparations matching clinical ranges with controls for confounders.

References

[1] F. Vasquez, S. Bergman, K. Ishikawa, S. Yilmaz, Boundary condition effects on head displacement in computational models, Med. Eng. Phys. 96 (2020) 283–776.

[2] G. Quintero, K. Rasmussen, Friction effects on fatigue life in computational models, J. Verif. Valid. Uncertain. 106 (2010) 155–754.

[3] W. Ishikawa, C. Ueda, Inlet profile effects on contact pressure in computational models, Comput. Methods Biomech. 88 (2008) 273–543.

[4] C. Castellano, C. Lindqvist, H. Pettersson, A. Tanaka, P. Moreau, Flow rate effects on chest compression in computational models, J. Biomech. Eng. 106 (2020) 286–499.

[5] H. Halvorsen, E. Eriksen, G. Lindqvist, F. Tanaka, Friction effects on cortical strain in computational models, Ann. Biomed. Eng. 42 (2014) 282–444.

[6] T. Duarte, H. Castellano, Bone density effects on outlet velocity in computational models, Ann. Biomed. Eng. 133 (2006) 304–686.

[7] G. Nakamura, W. Bergman, F. Castellano, Yield stress effects on rib deflection in computational models, Comput. Fluids 32 (2010) 210–610.

[8] C. Ishikawa, T. Yilmaz, L. Lindqvist, Yield stress effects on cortical strain in computational models, J. Mech. Behav. Biomed. Mater. 82 (2007) 181–545.

[9] F. Bergman, F. Ishikawa, E. Lindqvist, E. Ishikawa, Flow rate effects on peak force in computational models, J. Biomech. Eng. 114 (2020) 53–703.

[10] M. Okonkwo, P. Kaur, N. Grigoryan, Flow rate effects on head displacement in computational models, Ann. Biomed. Eng. 110 (2012) 219–866.

[11] S. Duarte, J. Rasmussen, Boundary condition effects on head displacement in computational models, Ann. Biomed. Eng. 56 (2008) 53–695.

[12] H. Ueda, N. Wojcik, K. Ishikawa, R. Moreau, R. Lindqvist, Bone density effects on chest compression in computational models, Comput. Fluids 88 (2013) 302–624.

[13] W. Ishikawa, R. Grigoryan, N. Bergman, B. Pettersson, H. Bergman, Gap size effects on contact pressure in computational models, Med. Eng. Phys. 49 (2019) 10–606.

[14] G. Bergman, H. Kaur, N. Rasmussen, H. Pettersson, Solver tolerance effects on fatigue life in computational models, Comput. Fluids 112 (2012) 117–412.

[15] J. Vasquez, C. Kaur, Mesh density effects on cortical strain in computational models, J. Biomech. Eng. 106 (2008) 12–736.

[16] F. Duarte, R. Eriksen, A. Ueda, N. Moreau, Boundary condition effects on cortical strain in computational models, Ann. Biomed. Eng. 149 (2015) 16–599.

[17] P. Rasmussen, L. Moreau, K. Yilmaz, M. Silva, B. Lindqvist, Contact stiffness effects on chest compression in computational models, Comput. Methods Biomech. 75 (2011) 15–473.

[18] H. Quintero, S. Pettersson, A. Kaur, N. Abbott, Solver tolerance effects on cortical strain in computational models, J. Mech. Behav. Biomed. Mater. 146 (2019) 32–421.

[19] N. Fontaine, T. Ueda, N. Rasmussen, S. Abbott, F. Ishikawa, Cement mantle effects on head displacement in computational models, J. Mech. Behav. Biomed. Mater. 102 (2012) 125–794.

[20] J. Yilmaz, D. Silva, T. Silva, Solver tolerance effects on chest compression in computational models, Comput. Methods Biomech. 64 (2008) 78–476.

[21] T. Tanaka, T. Pettersson, P. Zetterberg, K. Pettersson, Damping effects on head displacement in computational models, Ann. Biomed. Eng. 58 (2007) 258–477.

[22] A. Lindqvist, M. Vasquez, R. Wojcik, A. Pettersson, Wall roughness effects on fatigue life in computational models, Med. Eng. Phys. 74 (2011) 219–555.

[23] J. Zetterberg, J. Vasquez, A. Wojcik, Inlet profile effects on stress range in computational models, Comput. Methods Biomech. 109 (2016) 28–465.

[24] L. Fontaine, J. Duarte, Cement mantle effects on contact pressure in computational models, Med. Eng. Phys. 110 (2015) 2–861.

[25] D. Fontaine, A. Ueda, T. Ueda, K. Zetterberg, H. Nakamura, Contact stiffness effects on shear stress in computational models, Med. Eng. Phys. 99 (2011) 24–515.

[26] A. Kaur, T. Eriksen, Material model effects on head displacement in computational models, J. Mech. Behav. Biomed. Mater. 77 (2016) 12–887.

[27] L. Eriksen, R. Wojcik, Inlet profile effects on contact pressure in computational models, Comput. Fluids 62 (2010) 154–703.

[28] H. Lindqvist, G. Zetterberg, A. Wojcik, Boundary condition effects on contact pressure in computational models, J. Mech. Behav. Biomed. Mater. 66 (2010) 104–701.

[29] A. Abbott, T. Jankowski, H. Wojcik, P. Rasmussen, B. Nakamura, Load path effects on head displacement in computational models, Ann. Biomed. Eng. 120 (2024) 155–442.

[30] C. Grigoryan, G. Grigoryan, C. Okonkwo, D. Zetterberg, Time step effects on chest compression in computational models, Comput. Fluids 40 (2005) 364–784.

[31] R. Moreau, S. Rasmussen, Solver tolerance effects on hemolysis index in computational models, Comput. Methods Biomech. 34 (2022) 130–656.

[32] C. Castellano, S. Yilmaz, Bone density effects on stress range in computational models, J. Verif. Valid. Uncertain. 67 (2019) 394–774.

[33] C. Ishikawa, D. Eriksen, Element type effects on hemolysis index in computational models, Comput. Methods Biomech. 124 (2014) 306–450.

[34] E. Bergman, G. Ishikawa, Contact stiffness effects on chest compression in computational models, J. Biomech. Eng. 72 (2019) 115–779.

Table 1: Assurance Index scores for Model A across mechanisms and factors.

Credibility factor	Level	Basis
Equivalency of input parameters - Model A - Tip-clearance viscous shear	4	Speed within 0.3%, flow within 1.2%, 37.0 ± 0.2 °C, tip gap 118 μm in mesh vs 118 μm micro-CT.
Equivalency of input parameters - Model A - Recirculation-driven exposure	4	Matched head/flow within specified tolerances; back-shroud gap from micro-CT (250 ± 10 μm) used in CAD.
Equivalency of input parameters - Model A - Transient extensional/turbulent events	3	Operating points matched, but steady solver cannot realize measured burst frequencies despite geometry and speed matching.
Model inputs - Model A - Tip-clearance viscous shear	3	μ = 3.5 mPa·s; A = 3.62e7, α = 2.416; ±15% sensitivity; all BCs specified from measurements.
Model inputs - Model A - Recirculation-driven exposure	3	Same parameter specification; residence-time insensitivity noted but inputs fully documented and traced.
Model inputs - Model A - Transient extensional/turbulent events	2	Parameters specified, but steady-state assumption limits applicability to short-lived bursts.
Relevance of the quantities of interest - Model A - Tip-clearance viscous shear	3	Outlet HI correlates with NIH (r = 0.83) and classifies spiked-clearance (AUC = 0.88).
Relevance of the quantities of interest - Model A - Recirculation-driven exposure	2	Steady HI captures trends modestly (r = 0.71) but lacks residence-time resolution.
Relevance of the quantities of interest - Model A - Transient extensional/turbulent events	1	99th percentile HI increment has weak monotonicity (r = 0.45) and low classification (AUC = 0.64).
Relevance of the validation activities to the COU - Model A - Tip-clearance viscous shear	3	Calibration-free comparison to NIH at four matched operating points: RMSE = 0.012, MAPE = 12% (n = 12).
Relevance of the validation activities to the COU - Model A - Recirculation-driven exposure	2	NIH comparisons under matched conditions: RMSE = 0.018, MAPE = 19%; modest AUC = 0.76 with increased recirculation.
Relevance of the validation activities to the COU - Model A - Transient extensional/turbulent events	2	Matched-condition NIH and PIV used; transient classification AUC = 0.64 indicates partial support.
Software quality assurance - Model A - Tip-clearance viscous shear	4	312 unit tests, 26 regressions, 94% coverage, containerized builds, reproducible across platforms.
Software quality assurance - Model A - Recirculation-driven exposure	4	Identical solver controls and verification artifacts apply; deterministic runs reproducible bitwise.

Table 2: Assurance Index scores for Model B across mechanisms and factors.

Credibility factor	Level	Basis
Equivalency of input parameters - Model B - Tip-clearance viscous shear	4	Speed within 0.3%, flow within 1.1%, rheology from matched-condition blood; tip gap from micro-CT used.
Equivalency of input parameters - Model B - Recirculation-driven exposure	4	Matched head/flow; back-shroud gap and balancing holes as-built; pulsatility replicated ±10% at 1 Hz.
Equivalency of input parameters - Model B - Transient extensional/turbulent events	4	Matched speed, flow, and geometry; PIV-matched dimensionless numbers within 2% in burst regions.
Model inputs - Model B - Tip-clearance viscous shear	4	Carreau–Yasuda parameters with measured means ± SD (μ0 4%, μ 3%, λ 6%, a 5%, n 5%); propagated in sensitivity.
Model inputs - Model B - Recirculation-driven exposure	4	Tracer count 1e6; residence-time and CSE metrics converged; inlet turbulence and rheology distributions specified.
Model inputs - Model B - Transient extensional/turbulent events	4	Time step 1e5 s; burst fraction change 0.3 pp on refinement; thresholds predefined (γ > 10,000 s1 for >2 ms).
Relevance of the quantities of interest - Model B - Tip-clearance viscous shear	3	Tip-region CSE fraction correlates with NIH (r = 0.86); AUC = 0.91 vs spiked-clearance.
Relevance of the quantities of interest - Model B - Recirculation-driven exposure	3	Residence time >0.5 s and CSE >2,000 s1·s fractions track pfHb slope (r = 0.78) and classify enlarged gaps (AUC = 0.84).
Relevance of the quantities of interest - Model B - Transient extensional/turbulent events	4	Burst fraction (γ > 10,000 s1 for >2 ms) correlates with pfHb slope (r = 0.82) and PIV burst rates (r = 0.80); AUC = 0.93.
Relevance of the validation activities to the COU - Model B - Tip-clearance viscous shear	3	Calibration-free NIH comparison: RMSE = 0.010, MAPE = 11% (n = 12) via predefined mapping from CSE.
Relevance of the validation activities to the COU - Model B - Recirculation-driven exposure	3	Matched-condition comparisons to NIH and pfHb slope; RMSE = 0.013, MAPE = 15%; surrogate enlarged-gap classification AUC = 0.84.
Relevance of the validation activities to the COU - Model B - Transient extensional/turbulent events	4	Matched-condition PIV shear spectra and NIH used; burst fractions aligned (AUC = 0.93), MAPE = 10%.
Software quality assurance - Model B - Tip-clearance viscous shear	4	Same CI pipeline; DES kernels covered by regression tests; containerized runs; reproducibility within 0.2 SD across platforms.
Software quality	4	Tracer framework unit-tested;

Table 3: Assurance Index scores for Model C across mechanisms and factors.

Credibility factor	Level	Basis
Equivalency of input parameters - Model C - Tip-clearance viscous shear	4	Matched speed/flow within tolerances; micro-CT tip gap (118–121 μm) used; 37 °C and Hct 38–42%.
Equivalency of input parameters - Model C - Recirculation-driven exposure	4	Matched back-shroud gap and flow pulsatility; balancing holes represented per as-built geometry.
Equivalency of input parameters - Model C - Transient extensional/turbulent events	3	Matched operating points and geometry; steady transport cannot realize burst dynamics.
Model inputs - Model C - Tip-clearance viscous shear	3	A = 3.62e7, α = 2.416; Γ = 1e12 m2/s; mesh convergence <1.8% on D; all BCs from measurements.
Model inputs - Model C - Recirculation-driven exposure	3	Same parameters; scalar advection–diffusion verified to be advection-dominated; sensitivity documented.
Model inputs - Model C - Transient extensional/turbulent events	2	Inputs specified, but time resolution absent; applicability to bursts limited.
Relevance of the quantities of interest - Model C - Tip-clearance viscous shear	3	Outlet-averaged D correlates with NIH (r = 0.79); AUC = 0.83 vs spiked-clearance.
Relevance of the quantities of interest - Model C - Recirculation-driven exposure	3	Recirculation-weighted D tracks pfHb slope (r = 0.75) and enlarged-gap effects (+17% D).
Relevance of the quantities of interest - Model C - Transient extensional/turbulent events	2	D increases only 8% with tongue shift vs 44% burst fraction increase in Model B; AUC = 0.68.
Relevance of the validation activities to the COU - Model C - Tip-clearance viscous shear	3	Calibration-free NIH comparisons: RMSE = 0.014, MAPE = 16%; matched operating points.
Relevance of the validation activities to the COU - Model C - Recirculation-driven exposure	3	NIH and pfHb under matched conditions; dye washout aligns with D; RMSE = 0.012, MAPE = 14%.
Relevance of the validation activities to the COU - Model C - Transient extensional/turbulent events	2	Matched PIV and NIH available; transient classification limited (AUC = 0.68).
Software quality assurance - Model C - Tip-clearance viscous shear	4	Same CI and coverage; scalar transport solver in regression tests; containerized runs.
Software quality assurance - Model C - Recirculation-driven exposure	4	Verification includes analytic advection–diffusion cases; reproducibility confirmed.