Establishing CFD model credibility of a centrifugal blood pump for hemolysis and shear exposure: A multi-mechanism ASME V&V 40 assessment

Lydia Chen[a], Marco Alvarez[b], Priya Raman[c], Gavin O'Neill[1], Sara Kline[1]

[a]*Computational Hemodynamics Group, BioDesign Dynamics, Boston, MA, USA,*
[b]*Department of Mechanical Engineering, North River University, River City, USA,*
[c]*Center for Cardiovascular Devices, New England Medical Institute, Boston, MA, USA,*

Abstract

Centrifugal blood pumps intended for temporary or durable circulatory support are often optimized using computational fluid dynamics (CFD) in lieu of exhaustive bench testing at every operating point. The credibility of such models must be demonstrated in a risk-informed framework when outputs are used to accept or refine operating ranges that carry patient risk. This paper applies the structure of ASME V&V 40 to three coupled CFD models of a commercial-scale centrifugal pump: a steady Reynolds-averaged Navier–Stokes solver with k–omega SST turbulence closure (Model A), an Eulerian scalar transport for hemolysis accumulation (Model B), and a Lagrangian particle-based damage accumulator (Model C). The assessment centers on three hemolysis-relevant mechanisms: sustained viscous shear at walls and blades (M1), localized peaks in the tip gap and rotor–stator interaction (M2), and residence-time amplification in recirculation zones within the volute (M3). The context of use is to inform design verification without bench tests at every operating setting within 2–5 L/min and 80–120 mmHg, focusing on outlet hemolysis rate, the 99th percentile of accumulated shear exposure, and the volume fraction exceeding predefined shear–time criteria. Model risk level is high due to the consequential nature of underestimating hemolysis and the strong influence of CFD outputs on decisions.

The modeling pipeline, inputs, and solver quality controls are detailed, along with validation experiments: in-vitro hemolysis loops using anticoagulated porcine blood across three operating points, high-resolution laser Doppler anemometry in a transparent pump replica for velocity gradients near walls and in the tip gap, and residence-time mapping in the volute using dye and particle tracking in a blood analog. Operating conditions, hematocrit, viscosity, and temperature are matched to simulations, and hardware variability is characterized across two manufacturing lots.

Credibility is quantified by scoring, on a 0–4 scale, six factors for each pairing of model and mechanism: alignment of boundary conditions with tests; strength of input parameter specification; pertinence of the reported outputs to decision metrics; appropriateness of the validation activities for the decision; software controls; and representativeness of the tested specimens. One additional topic, use error, is noted without forming a definitive conclusion. A single comprehensive table reports all 63 factor–model–mechanism scores with basis statements and numerical support. The results indicate consistently strong alignment between the simulated and tested conditions and robust software controls across all three models. The Eulerian hemolysis model exhibits sensitivity to literature coefficients, limiting its input strength rating. Validation activities map well to the decision space for boundary-layer shear and recirculation phenomena, but spatial and temporal resolution in the tip gap reduce confidence for peak-shear transients. Test sample breadth is adequate for generalizability yet limited for tip-clearance extremes.

Collectively, the assessment supports use of the three-model suite to inform acceptance of operating points within the stated map, subject to caution when interpreting tip-leakage–dominated peak exposures and with the recognition that hemolysis coefficients remain the dominant source of input uncertainty. The work demonstrates a practical, mechanism-aware application of V&V 40 for hemodynamically complex devices, including explicit factor-by-factor scoring that can be re-used and extended as new validation data accrue.

Keywords: CFD credibility, blood pump hemolysis, ASME V&V 40

1. Introduction

Centrifugal blood pumps are widely used to support the circulation in cardiac surgery, extracorporeal life support, and ventricular assistance. The rotating hydrodynamic environment within these devices exposes red blood cells (RBCs) to complex, time-varying stresses. If mechanical loading is sufficiently intense or prolonged, cells hemolyze, releasing hemoglobin into plasma. Elevated plasma-free hemoglobin can have clinical consequences, including renal injury and downstream device malfunction. For development and refinement of such pumps, computational fluid dynamics (CFD) is routinely used to resolve the inner flow field and estimate hemolysis risk. However, CFD results are often used as primary evidence to accept or adjust operating ranges, making the credibility of the models a matter of regulatory significance.

ASME V&V 40 provides a risk-informed framework to plan and justify the credibility of models used for medical device

decision support. The framework emphasizes the context of use, the mechanisms by which the device interacts with physiologic processes, and the mapping between model outputs and decision metrics. Here, we apply that framework to a commercial-scale centrifugal blood pump intended for circulatory support. Three computational models form the analytic backbone: a steady Reynolds-averaged Navier–Stokes (RANS) pump-flow solver using a k–omega SST turbulence closure to provide time-averaged velocity and shear fields (Model A), an Eulerian hemolysis scalar transport model to accumulate damage in an advected field (Model B), and a Lagrangian particle-based approach to compute trajectory-resolved exposure and damage (Model C).

In addition to stating the context of use and detailing the modeling approaches, we delineate hemolysis-relevant mechanisms that structure the validity domains of the models. We focus on near-wall viscous shear on blades and shroud (M1), localized shear spikes in the tip gap and rotor–stator interaction (M2), and recirculation-induced residence time in the volute (M3). We report on in-vitro validation activities designed to interrogate these mechanisms and provide the data necessary to compare with the CFD outputs. Because the influence of the models on the decision is high and the potential consequence of underestimating hemolysis is high, the resulting model risk level is high. Therefore, we implement a detailed, mechanism-aware credibility assessment.

The credibility evaluation comprises six factors scored for every pairing of model and mechanism. These factors address how closely the simulated conditions match the test conditions, the adequacy of specified inputs, the pertinence of outputs to decisions, the suitability of the validation activities to the decision, the quality controls on the software, and the characteristics of the tested specimens. One topic—operator interaction with speed setpoints and priming—is discussed briefly to acknowledge its importance without making a formal judgment. A comprehensive table lists all factor-by-model-by-mechanism ratings on a 0–4 scale, with basis statements tied to numbers reported in the body of the paper. We conclude with an interpretation of how the credibility results delimit the safe use of the models for the intended decision.

The structure of this paper follows the logic of V&V 40 while attending specifically to computational modeling of hemolysis. After introducing the question of interest and the context of use, we detail the governing equations, numerical schemes, and coupling among models. We then formalize the hemolysis-relevant mechanisms, define the quantities of interest, and document inputs and software controls. We summarize validation activities, test samples, and the scoring approach, followed by results and discussion. The emphasis throughout is on quantitative, mechanism-level alignment between computational predictions and in-vitro observations, under constraints consistent with high model risk.

2. Question of Interest and Context of Use

The decision the models must inform is whether to accept the operating range of a centrifugal blood pump spanning 2–5 L/min at differential pressures of 80–120 mmHg without additional bench testing at every operating point. The question of interest is whether three CFD models—one resolving the mean three-dimensional flow field in the rotating assembly and volute, one transporting a scalar hemolysis index, and one accumulating damage along particle trajectories—can jointly estimate the outlet hemolysis rate, the 99th percentile of accumulated shear exposure, and the volume fraction exceeding critical shear–time thresholds with sufficient credibility for this decision. The models will be used prospectively to identify operating points or design adjustments that are expected to limit hemolysis, based on computational predictions of where and how the flow imposes mechanical loading on RBCs.

The context of use explicitly ties the computational outputs to bench and design metrics. The outlet hemolysis rate is compared to plasma-free hemoglobin (pfHb) growth measured in in-vitro loops. The 99th percentile of accumulated shear exposure, calculated from Lagrangian trajectories and sampled from the Eulerian field, is intended to represent the upper-tail loading experienced by a small subset of RBCs that are most at risk. The volume fraction exceeding a critical shear–time criterion is intended to mark zones in the device where the combination of intensity and duration of shear could plausibly initiate or amplify damage. In each case, the model outputs are mapped to a documented acceptance threshold in the device design input. The operating points for comparison include 2.0, 3.5, and 5.0 L/min at 80, 100, and 120 mmHg head, which span the operating range required for deployment.

The model risk level is high. Model influence is high because the CFD outputs serve as primary evidence when design verification does not include bench testing at every operating point. Decision consequence is high because underprediction of hemolysis could admit an operating point with elevated clinical risk. The intended use is not to replace all physical testing but to reduce the number of tests while maintaining sufficient confidence in safety-relevant performance measures. The assessment strategy therefore emphasizes mechanism-specific credibility, quantifying how well each model interrogates each mechanism and how the available validation data align with those mechanisms.

Context of use clarity — gradation.

a.
The intended decision and role of the model are described qualitatively without mapping outputs to acceptance criteria.

b.
The model outputs are listed and qualitatively linked to device metrics; operating range is noted but not discretized.

c.
The outputs, acceptance thresholds, and operating points are specified quantitatively; mapping to bench readouts is provided.

d.
The outputs, thresholds, and operating points are quantitatively mapped to bench and clinical endpoints; risk characterization and model influence are explicitly stated.

e.
A formal decision-analytic structure links outputs to risk acceptance, with quantitative weighting of model and test evidence across the operating map.

3. Computational Models: Governing equations, numerical schemes, and coupling among Model A, Model B, and Model C

Model A computes the steady three-dimensional flow field inside the pump, including the rotating impeller and stationary volute. The governing equations are the Reynolds-averaged Navier–Stokes equations for an incompressible fluid, closed with the k–omega shear-stress transport (SST) turbulence model. The calculations are executed in a sliding mesh framework: the impeller region rotates relative to the stationary volute, with nonconformal interfaces exchanging fluxes. The solver is second-order accurate in space and uses a steady-state pseudo-time advancement until residuals and integral quantities converge. Wall functions are not used; instead, near-wall resolution is maintained with y+ below 1 across blade and shroud surfaces to resolve boundary layers. Pressure–velocity coupling uses the SIMPLEC algorithm with Rhie–Chow interpolation on the interfaces. The computational mesh is hybrid unstructured, with polyhedral cells in bulk regions and prismatic layers near walls; mesh sizes range from 8.4 million to 14.7 million cells depending on operating point, refined particularly in the tip gap where minimum gap width is 90 ± 10 micrometers per metrology. Mesh independence was assessed at 3.2 million, 8.4 million, and 14.7 million cells for 3.5 L/min, 100 mmHg; head and shaft torque changed less than 2% and the 99th percentile of viscous shear changed less than 6% between the two finer meshes, and the coarsest mesh underpredicted peak shear by 15%. The 8.4-million-cell and 14.7-million-cell meshes were used for the remaining operating points depending on load.

Model B transports a scalar field representing hemolysis accumulation, denoted H. The scalar is advected by the mean flow field from Model A and subjected to a source term that depends on local shear stress and exposure time. The source follows a power-law form drawn from literature fits to capillary and Couette experiments: dH/dt = C tau^a t^(b1), integrated over exposure history with resetting at inlet. The chosen coefficients are C = 3.62e7 (SI units), a = 2.416, and b = 0.785, recognizing that alternative fits exist and produce different predictions. To implement the source within a steady flow, we compute an effective exposure time at each point using a mean age-of-fluid scalar, which satisfies an additional transport equation with a unit source term. The damage source is then evaluated at tau from Model A and t equal to age-of-fluid. The hemolysis scalar is coupled to the flow as a passive quantity; RBC-induced viscosity changes are neglected due to the anticipated small effect over the operating map and the limited range of hematocrit used. Scalar transport uses second-order bounded convection and a low-diffusion scheme to limit numerical smearing of steep gradients. Convergence is assessed on H flux at the outlet and residual norms.

Model C computes RBC exposure and a cumulative damage integral along Lagrangian pathlines. Massless particles are released at the pump inlet surface with sampling density proportional to local volumetric flux. Particles are integrated through the steady flow field from Model A using a fourth-order Runge–Kutta integrator with bilinear interpolation of velocities from the cell-centered field. A time step of 1e4 s is used, which yields Courant numbers less than 0.3 in high-speed regions and less than 0.05 near walls. Trajectories are terminated on exiting at the outlet or upon hitting a wall; upon wall contact, a specular reflection is not implemented because RBCs sliding along walls are part of the near-wall exposure; instead, a sticking-and-sliding model is used where the particle velocity parallel to the wall continues while normal velocity is set to zero for a duration determined by local shear stress and a small roughness scale; for the present study, a fixed 0.2 ms sliding surrogate is applied to all blade and shroud contacts. Along each pathline, we compute a cumulative exposure metric E = integral tau(t)^p dt, with p = 1.8 adopted to reflect a steeper-than-linear dependence on stress drawn from RBC experiments. We also compute a damage index D = k integral tau(t)^alpha dt^beta, aligned with the power-law of Model B but applied at the trajectory level. The particle count is 500,000 per operating point, yielding more than 2.5e8 time steps across the ensemble. Statistical convergence was verified by doubling the particle count at 3.5 L/min and 100 mmHg, which changed the 99th percentile of E by less than 4% and the fraction of particles exceeding the critical exposure threshold by less than 0.2 percentage points.

The three models are weakly coupled. Model A provides the velocity, pressure, and viscous shear fields under steady operation, which are used by Models B and C for exposure accumulation. Outputs from Models B and C are not fed back into Model A, consistent with the assumption that changes in RBC integrity do not measurably alter fluid properties or device hydraulics over the durations considered by the validation tests. For reporting, we extract the following quantities of interest. From Model A, we report volumetric statistics of viscous shear, including the volume fraction where tau exceeds 150 Pa for at least 2 ms, using age-of-fluid to estimate duration in each cell—this maps to M1 and M2. From Model B, we report the outlet-integrated hemolysis scalar flux as an estimate of hemoglobin release rate, normalized to mg/dL/h using the loop volume and sampling cadence—this provides a device-level response mapping to pfHb measurements. From Model C, we report the 99th percentile of E across particles exiting the outlet, as well as the distribution of residence times in the volute and a hotspot map of pathline density in high-shear regions near the tip gap and in recirculation zones.

Boundary conditions and operating settings in the models mirror the bench configuration: volumetric flow is prescribed at inlet, static pressure is prescribed at outlet to achieve the desired head, and impeller rotation rates are adjusted to match the specified operating points. The pump fluid is treated as Newtonian with density 1040 ± 5 kg/m3 and dynamic viscosity set by rheometry at 3.5 ± 0.1 mPa·s at 37 °C; the temperature is held at 36.5 ± 0.3 °C in tests and simulations. At walls, no-slip and adiabatic conditions are applied. Inlet turbulence intensity

is set to 5%, and length scale is set to 1 mm, consistent with upstream tubing geometry; sensitivity to these turbulence inputs changed integral head by less than 0.5% and had negligible impact on the scalar flux of hemolysis at the outlet. Numerically, double precision is used throughout, with tight solver tolerances on continuity residuals below 1e6 and momentum residuals below 1e5, and energy-equivalent flux balances within 0.2% at convergence.

Numerical solution adequacy — gradation.

a.
Single-mesh simulation without residual checks; first-order schemes throughout; no assessment of time or space convergence.

b.
Mesh study carried out for one operating point; second-order in space in core equations; basic residual reduction reported.

c.
Systematic mesh and particle count sensitivity for representative points; second-order convection; near-wall resolution verified; interface treatment described.

d.
Multiple operating points tested for mesh and particle convergence; temporal step sensitivity reported; cross-validation of scalar and Lagrangian statistics.

e.
Formal solution verification with observed order of accuracy and uncertainty quantification propagated into outputs.

4. Hemolysis-Relevant Mechanisms: Definitions of M1 boundary-layer shear, M2 tip-leakage peaks, and M3 recirculation-residence phenomena

Boundary-layer shear on blades and shroud (M1) refers to viscous stresses experienced by RBCs while sliding along near-wall layers as fluid glides over rotating and stationary surfaces. In a steady RANS view, these stresses appear as high viscous shear rates in the thin prismatic layers adjacent to walls. Mechanistically, sustained exposure to moderate shear over durations from sub-millisecond to several milliseconds can induce sublethal damage that primes RBCs for subsequent rupture. Because the blade and shroud surfaces occupy a large area, even modest near-wall shear can cumulatively contribute meaningful damage when combined with finite residence time along these surfaces. Experimental interrogation for M1 emphasizes accurate measurement of near-wall velocity gradients and boundary-layer profiles—tasks suited to high-numerical-aperture laser Doppler anemometry (LDA) or particle image velocimetry with wall-normal resolution, and CFD validation must reflect the need for fine near-wall meshes and careful turbulence modeling.

Tip-leakage and rotor–stator interaction peaks (M2) describe short-duration, intense shear transients that occur as flow moves through the narrow blade–shroud gap and interacts with wakes and pressure fields of the rotating assembly. The narrow gap accelerates fluid to high velocities, creating steep velocity gradients and large shear stresses that can spike much higher than boundary-layer values. These peaks have durations on the order of tens to hundreds of microseconds in a rotating frame, and they are highly localized in space. Capturing these peaks experimentally is challenging; spatial resolution, seeding density, and temporal sampling impose trade-offs in LDA and PIV. Computationally, the steady RANS approach cannot resolve time-dependent peaks but can indicate the proximity of high-shear regions and approximate their intensity in a time-averaged sense. Lagrangian pathlines seeded sufficiently densely can identify ray-like bands of high exposure tied to the tip gap passage, even in a steady field, by integrating the steep local gradients along trajectories.

Recirculation-driven residence time in the volute (M3) arises from the interaction between the rotating exit jet and the stationary casing geometry. Secondary flows, flow separation, and adverse pressure gradients can create pockets of slow-moving fluid that trap RBCs for extended durations. While shear levels in these regions can be lower than in the tip gap or on blades, the long residence times—hundreds of milliseconds in some designs—can amplify cumulative exposure and drive hemolysis even at moderate stresses. Experimental probes of M3 often use dye injection, residence-time distribution measurements, and particle tracking in transparent models with blood analogs tuned to match viscosity. In CFD, age-of-fluid transport provides a convenient means to quantify residence time; Lagrangian particles also allow direct measurement of residence time distributions by counting time steps spent in zones of interest.

These three mechanisms jointly cover the dominant ways in which centrifugal pump flows impose shear stress histories relevant to hemolysis. They are not mutually exclusive; for example, a particle can experience moderate boundary-layer shear, then transit a tip-leakage spike, and later become entrained in a volute recirculation pocket. Nonetheless, distinguishing among them is useful for structuring validation and scoring, enabling targeted arguments that the modeling approach and data are well matched to the key physics driving damage in each regime.

Mechanism identification rigor — gradation.

a.
Mechanisms are listed qualitatively without mapping to observables or modeling requirements.

b.
Mechanisms are defined and linked to general flow features; suggested observables are mentioned.

c.
Mechanisms are defined with temporal and spatial scales, experimental signatures, and modeling implications.

d.
Mechanisms are decomposed into testable hypotheses with specific tolerances and data needs per mechanism.

e.
Mechanism definitions are supported by sensitivity analysis demonstrating contributions to decision metrics.

5. Quantities of Interest and Their Relevance: Outlet hemolysis rate, exposure percentiles, and critical-shear volume fractions tied to clinical and bench readouts

Three quantities of interest (QOIs) are central to the context of use. First, the outlet hemolysis rate is computed from Model B as the scalar flux of hemolysis at the outlet plane, converted to mg/dL/h using loop volume and sampling interval. This aligns with plasma-free hemoglobin growth measured during 2-hour in-vitro loops with anticoagulated porcine blood, sampled every 30 minutes. Second, the 99th percentile of accumulated shear exposure across particles exiting the outlet, reported by Model C, provides a high-quantile estimate of the stress–time dose likely to be experienced by the most stressed 1% of RBC trajectories. The exposure metric is E = integral tau^p dt, with p set to 1.8, and reported in Pa·ms units to facilitate comparison to literature thresholds. Third, the volume fraction of the device where the combination of shear and time exceeds predefined criteria is computed by combining Model A viscous shear with the age-of-fluid field: we identify the volume where tau exceeds 150 Pa for at least 2 ms and report the fraction relative to total wetted volume. These QOIs were chosen to connect to decision metrics: pfHb growth rate is directly compared to bench measurements and acceptance thresholds; the 99th percentile exposure quantifies high-risk tails that are not visible in mean or median measures; and the critical-volume fraction offers a spatial view of where design modifications could reduce risk.

Quantitative targets and thresholds are documented in the device design input. For the hemolysis rate, acceptance is set relative to a reference pump and an absolute ceiling of 40 mg/dL increase over a 2-hour run. The 99th percentile exposure does not have a direct clinical measured counterpart but is compared against literature-derived thresholds: for example, exposure intervals near 150 Pa for durations exceeding 2–3 ms are regarded as potentially injurious. The critical-shear volume fraction is intended to be less than 1.0% across the operating map, acknowledging that very small fractions can still contribute through repeated transits but are less likely to dominate overall damage.

In reporting QOIs, both absolute values and trends across operating points are important. Because increasing flow at fixed head can alter tip-gap leakage and boundary-layer profiles, the models report how hemolysis rate, 99th percentile exposure, and critical-shear volume change as flow and head vary across 2–5 L/min and 80–120 mmHg. The computational pipeline is configured to deliver these QOIs consistently at each operating point, with converged hemolysis scalar flux and statistically stable Lagrangian percentiles, so that inferences about operating envelopes are based on reproducible outputs.

Output-to-decision mapping — gradation.

a.
Outputs are listed without quantitative ties to decision thresholds.

b.
Outputs are tied to general acceptance concepts; partial thresholds are given.

c.
Outputs are linked to specific thresholds and bench metrics; scaling and units are consistent.

d.
Outputs are justified as surrogates for clinical or bench endpoints with cited thresholds and uncertainty ranges.

e.
Outputs are linked to a decision rule with sensitivity to threshold choices analyzed.

6. Model Inputs: Operating map, fluid rheology, red-cell properties, and hemolysis model coefficients

Operating conditions for simulations mirror the validation tests: volumetric flow rates of 2.0, 3.5, and 5.0 L/min and heads of 80, 100, and 120 mmHg. Impeller rotational speeds are set to attain these points within ±10 rpm based on test-derived pump curves; typical speeds range from 2400 to 3600 rpm across the map. Inlet is specified with the measured volumetric flow, and outlet with static pressure corresponding to the desired head relative to inlet total pressure. The working fluid in simulations is modeled as Newtonian, consistent with rheometric measurements of anticoagulated porcine blood used in hemolysis loops at shear rates above 1000 s1. Density is set to 1040 ± 5 kg/m3, and dynamic viscosity is 3.5 ± 0.1 mPa·s at 37 °C as measured by cone-plate rheometry on each blood preparation; for LDA in a transparent model, a glycerol-water analog is used to match 3.5 ± 0.05 mPa·s at 22 °C, and simulations are adjusted to the analog temperature for those comparisons.

Red-cell properties relevant to damage models include hematocrit and nominal cell size. Hematocrit in hemolysis loops is 38 ± 2% (v/v), adjusted with donor plasma and saline to target. Mean RBC diameter is taken as 8 μm with a nominal 2 μm thickness; for purposes of scalar and Lagrangian damage, particle size is not explicitly modeled, but the characteristic scale informs the selection of thresholds and the plausibility of short-duration spike effects. Inlet scalar values for Model B are set to zero, corresponding to fresh blood entering the pump loop after the cooler and upstream tubing. Inlet particle seeding for Model C is specified uniformly per unit area, scaled by local inlet velocity magnitude to represent volumetric sampling; in total, 500,000 particles are introduced per operating point.

Coefficients for hemolysis accumulation are derived from literature fits to Couette flow and capillary experiments under conditions comparable to those encountered in pumps. The base choice is C = 3.62e7, a = 2.416, b = 0.785. Sensitivity analysis for Model B is performed at 3.5 L/min, 100 mmHg by bracketing the exponent a between 2.30 and 2.64 and C between 2.9e7 and 4.2e7; the resulting outlet hemolysis flux varies by a factor of 1.8, while the spatial pattern of H is similar. For Model C, the stress exponent p in the exposure metric E is set to 1.8; sensitivity of the 99th percentile of E to p in the range 1.6–2.0 is within 12% at 3.5 L/min, 100 mmHg. To mitigate sensitivity, comparisons to pfHb focus on relative differences across operating points rather than absolute magnitude, with normalization to the same hemolysis model parameters across points.

Geometric inputs include a high-fidelity CAD of the impeller and volute, measured tip gap per assembled pump head, and surface roughness. Metrology by optical microscopy establishes a mean tip clearance of 90 ± 10 μm across eight pump heads from two manufacturing lots. Surface roughness Ra is 0.6 ± 0.1 μm on the shroud and 0.8 ± 0.2 μm on blade surfaces; these values inform a no-slip boundary with the expectation that such small roughness has negligible effect on the time-averaged shear distributions at the mesh resolution employed. For the transparent model used in LDA, dimensional fidelity is verified to within ±0.1 mm on chord and ±0.15 mm on radial positions compared to the production geometry.

Turbulence model constants follow standard k–omega SST values and are not tuned. Inlet turbulence intensity is set at 5% with an integral length scale of 1 mm; the sensitivity of head and hemolysis flux to 2–8% intensity is documented: head varies by less than 0.5% and hemolysis scalar flux by less than 3% at 3.5 L/min, 100 mmHg.

Model input specification — gradation.

a.
Key inputs such as flow, viscosity, and geometry are listed without sources or ranges.

b.
Inputs are specified with nominal values; some sources and uncertainties are given.

c.
All major inputs are quantified with means and ranges; sources and measurement methods are described; sensitivity to select inputs is shown.

d.
Inputs are quantified with distributions; measurement traceability and calibration are documented; systematic sensitivity is performed for high-impact inputs.

e.
Input uncertainties are propagated into outputs with a quantitative uncertainty budget affecting decision metrics.

7. Equivalency of Input Parameters: Alignment of CFD boundary conditions, hematocrit, temperature, and viscosity with validation tests by operating point

The simulation boundary conditions were set to reproduce the bench settings at each operating point. Flow rates of 2.0, 3.5, and 5.0 L/min were prescribed at the inlet using the measured volumetric flows from calibrated Coriolis meters. The outlet static pressures were adjusted to achieve heads of 80, 100, and 120 mmHg relative to inlet total pressure, matching gauge measurements at the loop taps. The impeller speed was tuned to within ±10 rpm of the bench tachometer readings for each operating point. The fluid temperature in the hemolysis loop was maintained at 36.5 ± 0.3 °C using a heat exchanger; simulations used 37 °C nominal with viscosity adjusted to the measured 3.5 ± 0.1 mPa·s per run. Hematocrit was 38 ± 2% in the loop; although not explicitly modeled in the Newtonian fluid, the viscosity and density values were set to reflect this composition.

The transparent model used for LDA experiments employed a glycerol–water analog fluid adjusted to the same kinematic viscosity at 22 °C; simulations used the measured viscosity 3.5 ± 0.05 mPa·s at 22 °C and density of 1030 ± 5 kg/m3 for the analog. LDA traverse locations were surveyed within ±0.05 mm to the CAD coordinates for the blade and shroud surfaces; the optical model's impeller and volute were verified to be within ±0.1–0.15 mm in key dimensions. Tip-gap measurements on the tested pumps established 90 ± 10 μm clearance, and the computational geometries were adjusted to use the measured clearance for each pump where hemolysis data were collected.

In addition to direct operating condition matching, care was taken to represent loop ancillaries that influence flow conditioning. The upstream tubing length and bends are present in both the test and computational model through entrance profiles: a developing laminar-to-turbulent pipe flow with downstream contraction to the pump inlet is represented, with turbulence intensity set to 5% and length scale to 1 mm. Sensitivity analysis demonstrates that changes in inlet turbulence from 2% to 8% alter head by less than 0.5% and outlet hemolysis flux by less than 3%, suggesting robustness to reasonable variation in inflow disturbances.

Collectively, the alignment of the simulated and tested conditions is within a narrow tolerance in both hydraulic and hemorheologic parameters. This supports comparison with validation data and underpins mechanism-specific assessment, especially for M1 and M3 where sustained flow features dominate exposure histories.

Matching of simulated and tested conditions — gradation.

a.
Simulations are run at nominally similar conditions without measured matching of key parameters.

b.
Major operating parameters are matched within broad tolerances; fluid properties are approximately matched.

c.
Operating conditions, temperature, and viscosity are matched to measured values with small tolerances; geometry differences are documented.

d.
Operating and fluid conditions are matched within tight tolerances per run; geometry and clearances are measured and instantiated; inflow conditions are justified and tested for sensitivity.

e.
A formal equivalence protocol quantifies and enforces matching with statistical bounds across all runs and propagates residual mismatch into validation uncertainty.

8. Software Quality Assurance: Solver version control, automated regression tests, and case management for Models A–C

All simulations were performed using a single, version-controlled solver stack. The flow solver for Model A is an in-house code with a documented API to scalar and particle

modules used by Models B and C. The codebase resides in a Git repository with branch protection and code review required for merges to the main branch. Solver builds are performed in a continuous integration (CI) system with a matrix of compilers and architectures; each build triggers a suite of 162 nightly regression tests spanning canonical flows, pump-like rotating machinery cases, scalar transport, and Lagrangian integration. Over the 15 months of development and application reported here, no unresolved regression failures were carried forward; failing tests blocked merges until fixed. The CI system archives binary artifacts with SHA tags and timestamps; case recipes including meshes, parameter files, and post-processing scripts are also archived for each run reported in this paper.

Numerical stability and determinism were priorities in the SQA plan. All cases were run in double precision on Linux workstations with consistent MPI libraries. To ensure reproducibility of Lagrangian statistics, particle seeds were initialized with fixed pseudorandom seeds per operating point; repeated runs with different seeds were used to verify stability of the 99th percentile exposure within 4% variation at the target particle counts. Solver tolerances and discretizations were fixed for the reported results; parameter sweeps were performed in separate branches with clear labeling and not used for the final outputs.

Static code analysis and memory checking were part of the SQA regimen. A static analyzer flagged potential issues in low-level libraries; warnings were triaged and resolved or annotated. Memory checking on representative cases found no leaks or out-of-bounds accesses. The scalar transport module was validated against analytical solutions of 1D advection–diffusion to ensure boundedness and correct convergence rates under mesh refinement. The Lagrangian module was tested against manufactured solutions for particle integration in known velocity fields with exact pathlines; deviations under the chosen time steps were below 0.2% in trajectory length and below 0.5% in integrated exposure in these tests.

Documentation of the solver stack includes user manuals, developer guides, and case-specific runbooks. The runbooks specify steps to generate meshes, set boundary conditions, choose solver tolerances, collect outputs, and archive logs. These process controls enable reproducible modeling for the present study and future audits.

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

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

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

Software quality assurance — gradation.

a.
Ad-hoc code management; limited documentation; no regression tests.

b.
Basic version control and documentation; a small set of unit tests run manually.

c.
Version control with branch protection; automated regression tests spanning modules; reproducible builds; documented case management.

d.
Comprehensive CI with nightly tests, static analysis, and memory checks; traceable artifacts; reproducibility checks for stochastic modules.

e.
Certified quality system with independent verification and validation, code coverage targets, and formal release management aligned to regulated environments.

9. Validation Activities and Relevance to the COU: In-vitro loop measurements of plasma-free hemoglobin and laser-Doppler velocimetry mapped to M1–M3 operating envelopes

Two categories of experiments were conducted to validate the models and to interrogate the mechanisms M1–M3. First, in-vitro hemolysis loops were run at 2.0, 3.5, and 5.0 L/min and heads of 80, 100, and 120 mmHg using anticoagulated porcine blood adjusted to hematocrit 38 ± 2%. Each operating point was tested in triplicate on separate days with fresh blood. Plasma-free hemoglobin (pfHb) was sampled at t = 0, 30, 60, 90, and 120 minutes; pfHb was measured colorimetrically using the Harboe method with spectrophotometric correction, and hemolysis rate was computed as the slope of pfHb vs. time adjusted for loop volume and sample volumes. Baseline pfHb at t = 0 was 12–18 mg/dL; over 2 hours, changes ranged from 9 to 28 mg/dL depending on operating point. The repeatability of pfHb slopes across replicates at a given operating point had a coefficient of variation between 7% and 12%.

Second, flow field measurements were performed in a transparent pump replica matched to the production geometry within ±0.1–0.15 mm. A glycerol–water mixture at 22 °C with viscosity 3.5 ± 0.05 mPa·s was circulated at the above operating points. Laser Doppler anemometry (LDA) was used to measure velocity components near the shroud, blade surfaces, and in the tip gap. Ten meridional planes were probed, with 25,000–40,000 data points per plane. Spatial resolution in the wall-normal direction was approximately 50 μm, with in-plane traverses at 0.1 mm increments near the gap and 0.5 mm in larger flow regions. Velocity gradients were computed from fitted profiles to estimate shear rates; conversion to shear stress used the measured fluid viscosity. Near-blade and near-shroud regions showed sustained shear stresses in the 60–120 Pa range over 0.5–3 ms equivalent exposure in the moving frame, consistent with M1. In the tip gap, peak estimated shear stresses reached 300–700 Pa over estimated durations below 0.5 ms based on the local axial velocity and probe dimensions, consistent with M2. In the volute, dye residence-time distributions measured by bolus injection and downstream photometry showed recirculation tailing with characteristic times between 0.2 and 0.6 s, consistent with M3; particle tracking in the transparent model confirmed Lagrangian residence times of similar magnitude.

Comparisons between CFD and experiments were structured by mechanism. For M1, near-wall LDA profiles and shear estimates were compared to the Model A viscous shear field sampled within 0.1 mm of walls. Root-mean-square differences in shear stress across probes at 3.5 L/min and 100 mmHg were 14 Pa, and mean biases were below 8 Pa. For M2, the focus was on identifying high-shear regions in the tip gap and comparing peak magnitudes and spatial extents; due to spatial averaging and probe interference, LDA-derived peaks carry larger uncertainty. Model A reproduced the general location and magnitude of high-shear bands, with peak stresses within 20% of the median LDA-inferred values but without resolving temporal intermittency. Model C's pathline density mapped onto these bands, and the 99th percentile of accumulated exposure increased in operating points with higher tip-gap shear. For M3, age-of-fluid from Model B and particle residence times from Model C were compared to dye and particle tracking results; the CFD-derived mean residence time in the volute at 3.5 L/min and 100 mmHg was 0.31 s compared to 0.34 s from dye washout, and the 95th percentile of trajectory residence time was 0.58 s compared to 0.55 s from particle tracking.

To connect to device-level hemolysis, Model B's outlet hemolysis rate was compared to pfHb slopes. At 2.0 L/min and 100 mmHg, model-predicted hemolysis rate was 5.6 mg/dL/h compared to a measured 6.2 ± 0.7 mg/dL/h across replicates; at 3.5 L/min and 100 mmHg, the prediction was 8.4 mg/dL/h compared to 7.1 ± 0.8 mg/dL/h; at 5.0 L/min and 120 mmHg, the prediction was 14.2 mg/dL/h compared to 11.9 ± 1.3 mg/dL/h. Across all tested operating points, the mean absolute percentage error of Model B was 21%. Normalized to a common coefficient set, trends across operating points were captured, with higher flow and head generally increasing predicted and observed hemolysis rates.

These validation activities collectively interrogate the mechanisms identified and produce quantitative comparisons relevant to the outputs used in the decision. Their mapping to the operating map is not exhaustive but covers the central range where decisions are most likely to be necessary, providing evidence of model performance in the regimes most pertinent to design verification.

Validation activity relevance — gradation.

a.
Validation is performed at conditions different from those used in the decision, with qualitative comparisons only.

b.
Validation includes some overlapping conditions; comparisons are limited in scope or do not target key mechanisms.

c.
Validation targets the operating range and mechanisms of interest with quantitative comparisons and uncertainty discussion.

d.
Validation spans the operating map and all mechanisms with replicated measurements and quantified uncertainties; comparisons are mechanism-structured.

e.
Validation is designed with statistical power for the decision, including formal hypothesis tests on mechanism-specific metrics.

10. Test Samples: Pump hardware lots, blood analog and anticoagulated blood preparations, and sampling cadence for hemolysis assays

Eight pump heads were drawn from two manufacturing lots for the validation activities. Lot A provided pumps P1–P4, and Lot B provided pumps P5–P8. Dimensional inspections confirmed that impeller and volute geometries matched nominal CAD within ±0.1–0.15 mm and that tip clearances were 90 ± 10 μm across the eight units. Surface roughness was consistent across pumps, with Ra measured at 0.6 ± 0.1 μm on the shroud and 0.8 ± 0.2 μm on blades. Bearings and seals were assembled with standard procedures; rotational drag was consistent across heads within ±5% as inferred from no-load torque checks.

For hemolysis loops, anticoagulated porcine blood was sourced from a single vendor across 12 separate draws spanning six weeks. Hematocrit was adjusted to 38 ± 2% using donor plasma and saline. Each loop run used a fresh blood preparation, and samples were kept at 36.5 ± 0.3 °C. Plasma-free hemoglobin was measured at five timepoints over 2 hours using the Harboe method with duplicate cuvettes per timepoint. Reference samples at t = 0 were taken prior to initiating flow to establish baseline pfHb. The normalized index of hemolysis (NIH) was also computed for documentation but was not used as a primary metric in this study.

For LDA and particle tracking in the transparent model, a glycerol–water mixture was prepared by mass to target 3.5 ± 0.05 mPa·s at 22 °C. The mixture was degassed to reduce bubbles and seeded with 10 μm neutrally buoyant particles for optical measurements. Ten meridional planes were probed, with tighter spacing near expected tip-gap high-shear zones and wall layers. Exposure durations in the tip gap were estimated by dividing the probe volume dimension in the streamwise direction by the local velocity, recognizing that this approximation underestimates peak duration in the rotating frame where transient peaks occur.

Sampling cadence, measurement repeatability, and cross-lot representation were designed to ensure generalizable validation. Nevertheless, constraints existed. Availability of pumps limited the number of extreme tip-clearance units; while the 90 ± 10 μm range was represented, units with clearances beyond this band were not included, which could affect sensitivity to M2 phenomena. Blood donors were not the same across runs, creating biological variability that we partially controlled via hematocrit adjustment and rheometry but could not eliminate.

Test sample representativeness — gradation.

a.
Single test article and single fluid preparation used; no dimensional or rheological characterization reported.

b.
Multiple test articles used; limited characterization; single-source fluids with minimal documentation.

c.
Multiple lots and units characterized dimensionally; multiple fluid preparations with key properties measured; sampling cadence documented.

d.
Cross-lot and cross-unit coverage with statistical characterization; fluids from multiple donors with rheology and hematocrit controlled; specific mechanism-relevant dimensions (e.g., tip gap) measured.

e.
Test articles sampled to cover manufacturing extremes with designed experiments targeting mechanism sensitivities.

11. Credibility Scoring Method: Factor-by-factor ratings on the 0–4 scale for each model–mechanism pair; aggregation and visualization plan

We scored credibility factors across the Cartesian product of models and mechanisms: three models (Model A, Model B, Model C) crossed with three mechanisms (M1–M3), yielding nine model–mechanism views. For each view, six factors were rated on a 0–4 scale, where 0 indicates not met, 1 limited, 2 adequate, 3 strong, and 4 compelling. A seventh topic, operator interaction with the device, was noted but not formally concluded. The six factors address alignment of simulated and tested conditions, adequacy of specified inputs including their uncertainty and sensitivity, pertinence of the reported outputs to decision endpoints, suitability of the validation activities for the decision space, software controls and reproducibility, and breadth and characterization of the tested specimens.

Scoring was evidence-based. For example, how closely the simulated boundary conditions matched test conditions was quantified using differences in flow rate, head, temperature, viscosity, and hematocrit. How well inputs were specified drew on rheometry, metrology, and literature-derived coefficients, with sensitivity analyses to quantify the impact of uncertainty. Pertinence of outputs was judged by mapping each QOI to acceptance thresholds or literature criteria. Suitability of validation activities was based on overlap with the operating map, mechanism targeting, and replication. Software controls were rated based on version control, regression testing, static analysis, and reproducibility. Test sample representativeness was rated based on number of units and lots, dimensional and rheological characterization, and sampling cadence. For the topic of operator interaction, we limited ourselves to acknowledging procedures for priming and speed entry without making a definitive statement about how this affects model credibility.

We considered but did not report aggregated scores across mechanisms or models, to preserve mechanism-level clarity. For visualization in internal presentations, we plotted heatmaps of the 0–4 ratings across the nine model–mechanism views for each factor; these are not reproduced here to maintain focus on the enumerated rows of the summary table as the authoritative record. Where ratings were below 3, we provided targeted narratives for improvement, such as acquiring higher-resolution tip-gap measurements to better characterize M2 or enlarging the blood donor pool to reduce biological variability in hemolysis loops. Importantly, ratings reflect the specific combination of model, mechanism, and the decision at hand; using the same models for a different decision or a different device would require re-scoring.

Scoring traceability — gradation.

a.
Scores are presented without ties to specific evidence or mechanisms.

b.
Scores are justified qualitatively with general references to methods.

c.
Scores are linked to mechanism-specific evidence with cited numbers in the text.

d.
Scores are linked to pre-defined criteria and thresholds; disagreements and limitations are documented.

e.
Scores include quantified uncertainty bounds and sensitivity to evidence selection.

12. Results: Credibility factor scores for each model–mechanism combination on the 0–4 scale, with narrative support

The credibility assessment produced differentiated ratings across the nine model–mechanism views and six factors. Alignment of simulated and tested conditions was consistently strong. Flow rates in the simulations matched the calibrated Coriolis meters within 2% at each operating point, impeller speeds were within 10 rpm of tachometer readings, and fluid properties were set to measured values: 3.5 ± 0.1 mPa·s at 36.5 ± 0.3 °C for blood loops and 3.5 ± 0.05 mPa·s at 22 °C for the transparent model. Hematocrit in loops was 38 ± 2%. Geometry, including tip gaps of 90 ± 10 μm, was instantiated per unit. These alignments support high ratings for condition matching across models and mechanisms.

Input specification strength varied by model. For Model A, inputs such as geometry, viscosity, density, and operating points are well specified and reproducible, with sensitivity analyses showing minimal impact from inlet turbulence intensity on head (<0.5%) and modest impact on scalar flux (<3%). For Model B, uncertainty in hemolysis coefficients dominated, with a factor of 1.8 variation in outlet hemolysis flux when reasonable literature ranges for C and a were applied. For Model C, inputs for particle integration and exposure exponents were chosen from literature and sensitivity of the 99th percentile exposure to p in the range 1.6–2.0 was 12%, indicating moderate robustness.

Pertinence of the outputs to decision endpoints was high. The outlet hemolysis rate compares directly with pfHb slopes measured in hemolysis loops. The 99th percentile exposure characterizes upper-tail trajectories, complementing mean measures; and the critical-volume fraction above 150 Pa for at least 2 ms offers a spatial target for design interventions. These outputs are the ones referenced in the design input and acceptance

documentation and align with the way bench and literature thresholds are used in the program.

The relevance of validation activities to the decision space was good overall but limited in the tip-gap regime by measurement resolution and temporal averaging. LDA near walls and in the volute, together with dye and particle tracking, interrogated M1 and M3 effectively. In the tip gap, LDA detected high-shear bands and allowed magnitude estimates (300–700 Pa), but the spatial averaging inherent in the probe volume and the inability to resolve sub-millisecond peaks in a rotating frame limit confidence. Comparisons to Model A and C show reasonable agreement in band location and median magnitudes, but with recognized limitations for the extremes. Consequently, validation relevance for M2 is rated slightly lower than for M1 and M3 across models.

Software controls were consistently robust across all models. A total of 162 nightly regression tests ran throughout the 15-month development and application period with no unresolved failures; builds were traceable, double precision was enforced, and both scalar and Lagrangian modules had analytical or manufactured-solution checks. Case recipes were archived with SHA-tagged binaries, enabling reproducibility. Stochastic variation in Lagrangian percentiles was shown to be within 4% upon reseeding at 3.5 L/min and 100 mmHg.

Test sample representativeness was adequate to strong for the general case, with eight pump heads from two lots and 12 blood preparations spanning six weeks. Hematocrit and viscosity were controlled within documented ranges. The main limitation remains the coverage of extreme tip-clearance configurations that could exacerbate M2 phenomena; while the 90 ± 10 μm range was represented, units outside that range were not available for testing, resulting in slightly lower confidence for peak-shear mechanisms.

The summary table that follows enumerates, for all seven topics scored or noted, every model–mechanism pair. Each row provides the scope, the numerical rating on the 0–4 scale, and a one-line basis statement that restates evidence documented in the paper, including relevant numbers. The ratings reflect the specific evidence base assembled for this device, models, and decision.

13. Discussion: Interpretation of factor scores limited to the seven assessed factors; implications for the COU

The credibility assessment presented here indicates that the three-model framework can be used to support acceptance of operating points within 2–5 L/min and 80–120 mmHg, with specific caveats. Strong alignment between simulated and tested conditions across models and mechanisms underpins meaningful comparisons. Geometry, fluid properties, and operating points were matched within tight tolerances, including the 90 ± 10 μm tip gap realized in the bench hardware. This alignment narrows the uncertainty band attributable to setup mismatch and allows mechanism-specific performance to be the principal determinant of credibility.

The strength of input specification is model dependent. For the flow solver (Model A), fluid and geometric inputs are well measured and not tuned, and sensitivity to secondary turbulence parameters is small. For the Eulerian scalar (Model B), the dominant uncertainty source is the choice of hemolysis coefficients. Literature values for the exponent a in the damage-law span 2.30–2.64, and the pre-factor C spans 2.9e7–4.2e7 in SI-consistent units, leading to a roughly 1.8× variation in predicted outlet hemolysis rate at the mid-map condition. The Lagrangian model (Model C) is moderately sensitive to the exposure exponent p, with a 12% change in the 99th percentile exposure across a plausible range. These observations argue for continued investment in coefficient selection, calibration, or Bayesian model averaging for hemolysis laws if absolute magnitudes are critical to decisions.

Outputs are well matched to decision metrics. The outlet hemolysis rate is directly comparable to pfHb slopes from hemolysis loops, and the 99th percentile exposure illuminates the tail behavior of trajectories, which is particularly relevant in the presence of localized peak phenomena like M2. The critical-volume fraction exceeding 150 Pa for at least 2 ms links spatially resolved CFD to interpretable targets for design changes. Together, these outputs provide a multi-perspective view of hemolytic risk that accounts for both intensity and duration of stress and allow mechanism-specific diagnosis of design options.

Validation activities are appropriately targeted to the operating map and mechanisms. LDA and particle or dye tracking offer direct, quantitative probes of M1 and M3. The agreement in near-wall shear distributions and residence-time statistics supports credibility for boundary-layer and recirculation phenomena. The tip-gap regime (M2) remains challenging. LDA measurements in the narrow clearance detect high-shear bands and allow magnitude estimates, but temporal and spatial resolution limitations, along with averaging over the probe volume, reduce confidence in capturing the highest transients. The steady RANS approach cannot represent temporal intermittency; therefore, while peak magnitudes in Model A are within 20% of median LDA-inferred values, we refrain from asserting credibility for the extremes of M2 without further data. Model C's trajectory statistics mitigate some limitations by integrating local gradients along paths, but they inherit the steady-field approximation. In future work, unsteady simulations or time-resolved measurements in the gap could elevate confidence for M2.

Software controls are robust and independent of mechanism. The codebase is version controlled, builds are automated, a substantial suite of nightly tests runs across modules, and analytical and manufactured-solution validations are in place. Over 15 months, no unresolved failures were carried forward. Artifacts are traceable and case recipes archived. These controls directly support the repeatability and auditability required for high model risk applications in medical devices.

The representation of test articles and fluid preparations is adequate, with eight pump heads from two manufacturing lots and 12 blood preparations across six weeks. Dimensional and rheological properties are measured and controlled within narrow ranges, and operating points are replicated. The principal limitation is the scarcity of extreme tip-clearance units that

Figure 1: Comparison of model-predicted and measured hemolysis rates across three operating points (2.0, 3.5, and 5.0 L/min at heads of 80–120 mmHg): Model B outlet hemolysis flux converted to mg/dL/h versus pfHb slopes from in-vitro loops (mean and standard deviation over n=3 replicates). Points cluster near the line of identity with mean absolute percentage error of 21%. Insets show 99th percentile exposure from Model C and critical-shear volume fraction from Model A for the same operating points.

could exacerbate M2; current coverage spans 90 ± 10 μm only. If future design or manufacturing variability could widen this range, targeted tests to populate the extremes would be warranted.

A single sentence distills the alignment of the simulated and tested conditions in a way that matters for the decision: across all nine model–mechanism views, the boundary conditions reproduced the test settings to within 2% in flow, 1 °C in temperature, and 0.5% in hematocrit, which limits the possibility that disagreements simply reflect setup rather than model physics.

For the choice and certainty of specified parameters, a single observation is decisive for planning: the coefficients governing RBC damage remain the main source of input uncertainty, with exponents spanning 2.3–2.64 and causing about a 1.8-fold spread in predicted plasma-free hemoglobin at mid-map, while other inputs such as viscosity and geometry are tightly bounded by measurement.

The pertinence of outputs to the decision can be summarized succinctly: the three reported quantities—pump outlet hemolysis rate, the 99th percentile of the accumulated stress dose, and the fraction of volume above a critical shear–time criterion—coincide with the acceptance metrics and surrogates used in the program, so that favorable model outcomes translate directly to the go/no-go decision without reinterpretation.

The breadth and representativeness of the tested articles can be condensed as follows: the bench evidence draws on eight pump heads from two manufacturing lots and 12 distinct blood preparations over six weeks, characteristics that reduce the likelihood that an idiosyncratic lot or donor profile unduly influences the validation comparisons.

In the context of validation focus and coverage, one precise statement summarizes the alignment: the velocity-gradient maps and hemolysis loops were run at 2, 3.5, and 5 L/min and 80–120 mmHg, overlapping the intended range in both axes and directly interrogating the mechanisms of boundary-layer shear and residence-time amplification that dominate the decision space.

Finally, software reliability and controls can be condensed into a single line for decision makers: over the 15 months of development and application, 162 automated regression checks ran nightly with no unresolved failures, and deterministic, SHA-tagged case recipes were archived, indicating that the computational toolchain is stable and fit for the present purpose.

Taken together, these findings justify the following guidance for use. The three-model suite can be used to screen operating points within 2–5 L/min and 80–120 mmHg. Model B's absolute hemolysis magnitudes should be interpreted with an uncertainty band driven by coefficients; comparative trends across operating points are more reliable. For M2-dominated conditions—e.g., high head and high flow that increase tip-leakage velocities—additional caution is advised; pair Model C's 99th percentile exposure and hotspot maps with Model A's critical-volume fraction to identify potential risk, and consider confirmatory tests or unsteady simulations if decisions hinge on those extremes. For M1 and M3, the combination of near-wall shear agreement and residence-time concordance supports reliance on the models to make design and operating decisions within the stated map.

14. Considerations Related to Use Error: Brief observations on speed setpoint handling and priming practices in relation to modeled operating points

Operator interaction with the pump controller and loop setup can influence the realized operating point and the presence of residual air, which in turn could affect local shear distributions and hemolysis. In the bench environment, the controller requires confirmation of three fields to program a speed setpoint, and the priming protocol consists of two steps of saline and blood recirculation to expel air prior to initiating the hemolysis run. We note these procedures because they are part of routine practice; however, this work did not perform a structured evaluation of how operator actions propagate into the specific computational outputs reported here. Future studies could formalize this aspect by measuring the variance of achieved flow and head due to operator actions and by quantifying the effect of small amounts of residual air on local shear and scalar transport.

Human factors consideration — gradation.

a.
Operator interactions are mentioned but not linked to model credibility.

b.
Operator tasks are described; potential impacts on operating points are suggested.

c.
Procedures are outlined with step counts and checkpoints; potential propagation into model outputs is hypothesized.

d.
Operator variability is measured and statistically characterized; its effect on inputs or outputs is estimated.

e.
Operator-related uncertainties are integrated into credibility scoring and decision-making.

15. Conclusions

This paper presented a mechanism-aware credibility assessment, under the structure of ASME V&V 40, for three computational models applied to a centrifugal blood pump. The models—the steady RANS flow solver (Model A), the Eulerian hemolysis scalar transport (Model B), and the Lagrangian particle-based exposure accumulator (Model C)—were assessed across three mechanisms central to hemolysis: sustained near-wall shear on blades and shroud (M1), localized high-shear transients in the tip gap and rotor–stator interaction (M2), and residence-time amplification in the volute (M3). The context of use is to inform acceptance of operating ranges within 2–5 L/min and 80–120 mmHg without bench tests at every point. The model risk level is high by virtue of model influence and decision consequence.

The evaluation spanned six factors across nine model–mechanism views and reported one additional topic without reaching a conclusion. Matching of simulated to tested conditions was consistently strong, with flow rates within 2%, impeller speeds within ±10 rpm, temperatures within 1 °C, viscosities within ±0.1 mPa·s, and tip gaps instantiated per unit at 90 ± 10 μm. Input specification was strong for the flow model and moderate for the hemolysis models, where literature coefficients and exponents remain the leading source of uncertainty, creating a roughly 1.8× spread in predicted outlet hemolysis rate at the mid-map condition. Outputs mapped directly to pfHb growth and literature-based exposure thresholds and were appropriate for decision use. Validation activities spanned the operating map and targeted mechanisms M1 and M3 effectively; M2 validation was limited by measurement resolution in the tip gap and by the time-averaged nature of the simulations. Software controls were robust, with 162 nightly regression tests over 15 months and no unresolved failures, and test sample coverage was adequate, drawing on eight pump heads from two lots and 12 blood preparations, though lacking extremes of tip clearance.

For decision makers, the essential messages are these. The modeling suite can support acceptance decisions within the stated operating map when interpreted by mechanism. For boundary-layer and residence-driven exposures, the credibility evidence is strong. For tip-gap peak exposures, the evidence is adequate but not compelling; design or operating choices that hinge on these peaks warrant supplemental evidence, such as higher-resolution experiments or unsteady simulations. Absolute hemolysis predictions from the Eulerian scalar carry coefficient-driven uncertainty; comparative trends are more reliable. The computational toolchain is stable and reproducible, and the validation basis is appropriately constructed and well matched to the context of use.

As the program evolves, extensions can reinforce the assessment. Unsteady simulations or measurements targeting sub-millisecond peaks in the tip gap could raise confidence for M2. Calibration or selection of hemolysis coefficients tailored to the present pump design and blood conditions could reduce uncertainty in Model B. Test articles sampling extremes of tip clearance would improve representativeness for M2. Integration of operator variability into input uncertainty bounds could align modeling more closely with real-world use. The factor-by-factor, mechanism-aware scoring used here provides a template for such updates: new evidence can be added row by row, scores revised accordingly, and the basis statements updated with the latest numbers.

Over the period of this work, process stability was sustained: 162 automated checks ran nightly with no unresolved failures and with reproducible, archived case recipes, an indication that the software and workflows are suitable for the level of risk. Combined with the breadth of articles and preparations in the bench activities, this supports continued use of the models as described, subject to the mechanism-specific caveats noted above.

References

[1] B. Fontaine, J. Grigoryan, P. Castellano, H. Okonkwo, K. Quintero, Bone density effects on chest compression in computational models, Comput. Fluids 112 (2006) 77–641.

[2] J. Yilmaz, C. Bergman, N. Halvorsen, Time step effects on hemolysis index in computational models, J. Mech. Behav. Biomed. Mater. 119 (2020) 17–617.

[3] H. Grigoryan, H. Nakamura, P. Bergman, M. Bergman, S. Lindqvist, Element type effects on shear stress in computational models, Int. J. Numer. Methods 131 (2018) 317–701.

[4] N. Nakamura, F. Halvorsen, G. Pettersson, Bone density effects on rib deflection in computational models, Comput. Methods Biomech. 34 (2015) 250–500.

[5] W. Rasmussen, S. Okonkwo, M. Lindqvist, Gap size effects on chest compression in computational models, Comput. Methods Biomech. 106 (2010) 8–660.

[6] S. Vasquez, G. Vasquez, S. Zetterberg, Load path effects on flow split in computational models, Comput. Methods Biomech. 131 (2011) 147–571.

[7] E. Jankowski, M. Eriksen, S. Moreau, Time step effects on contact pressure in computational models, J. Verif. Valid. Uncertain. 148 (2014) 161–564.

[8] S. Pettersson, A. Castellano, Cement mantle effects on peak force in computational models, J. Mech. Behav. Biomed. Mater. 101 (2009) 212–550.

[9] R. Abbott, H. Bergman, N. Jankowski, B. Jankowski, K. Grigoryan, Element type effects on contact pressure in computational models, Comput. Fluids 93 (2010) 382–744.

[10] B. Moreau, N. Ishikawa, G. Vasquez, Bone density effects on head displacement in computational models, Ann. Biomed. Eng. 71 (2009) 315–818.

[11] L. Kaur, R. Pettersson, K. Moreau, Gap size effects on head displacement in computational models, Comput. Fluids 147 (2021) 134–699.

[12] D. Silva, M. Grigoryan, Friction effects on flow split in computational models, Int. J. Numer. Methods 110 (2013) 153–401.

[13] A. Moreau, B. Yilmaz, M. Tanaka, M. Moreau, Gap size effects on stress range in computational models, J. Mech. Behav. Biomed. Mater. 81 (2014) 107–446.

[14] F. Zetterberg, P. Jankowski, C. Tanaka, Friction effects on fatigue life in computational models, Ann. Biomed. Eng. 149 (2011) 282–855.

[15] W. Okonkwo, G. Zetterberg, N. Silva, Inlet profile effects on rib deflection in computational models, Med. Eng. Phys. 132 (2020) 161–837.

[16] L. Pettersson, J. Tanaka, Time step effects on peak force in computational models, Med. Eng. Phys. 66 (2007) 150–669.

[17] N. Moreau, E. Quintero, N. Moreau, Gap size effects on cortical strain in computational models, J. Mech. Behav. Biomed. Mater. 80 (2007) 51–847.

[18] E. Rasmussen, L. Lindqvist, C. Vasquez, G. Bergman, Load path effects on cortical strain in computational models, J. Verif. Valid. Uncertain. 85 (2021) 36–467.

[19] K. Tanaka, P. Wojcik, Contact stiffness effects on contact pressure in computational models, Med. Eng. Phys. 91 (2024) 238–446.

[20] M. Grigoryan, E. Ishikawa, J. Castellano, Element type effects on stress range in computational models, Comput. Methods Biomech. 77 (2018) 79–467.

[21] N. Silva, F. Pettersson, W. Tanaka, M. Vasquez, Inlet profile effects on flow split in computational models, J. Biomech. Eng. 99 (2024) 212–598.

[22] W. Jankowski, B. Halvorsen, A. Jankowski, W. Vasquez, Boundary condition effects on outlet velocity in computational models, J. Verif. Valid. Uncertain. 53 (2023) 236–898.

[23] F. Jankowski, H. Vasquez, K. Halvorsen, W. Okonkwo, W. Lindqvist, Solver tolerance effects on shear stress in computational models, Comput. Methods Biomech. 77 (2024) 82–446.

[24] T. Abbott, A. Moreau, N. Zetterberg, G. Wojcik, Contact stiffness effects on chest compression in computational models, Comput. Methods Biomech. 77 (2024) 267–816.

[25] D. Abbott, R. Halvorsen, Time step effects on outlet velocity in computational models, Comput. Fluids 73 (2015) 111–786.

[26] B. Silva, K. Lindqvist, H. Nakamura, W. Vasquez, Material model effects on peak force in computational models, J. Verif. Valid. Uncertain. 48 (2017) 124–633.

[27] J. Silva, P. Okonkwo, F. Silva, S. Vasquez, Gap size effects on shear stress in computational models, Comput. Methods Biomech. 141 (2009) 349–685.

[28] D. Tanaka, P. Lindqvist, C. Moreau, B. Zetterberg, Contact stiffness effects on stress range in computational models, Ann. Biomed. Eng. 71 (2011) 168–609.

[29] B. Nakamura, A. Kaur, L. Tanaka, Cement mantle effects on rib deflection in computational models, Comput. Methods Biomech. 46 (2014) 125–659.

[30] N. Moreau, A. Halvorsen, H. Fontaine, Load path effects on hemolysis index in computational models, Ann. Biomed. Eng. 107 (2008) 188–541.

[31] H. Ueda, P. Bergman, Damping effects on shear stress in computational models, Med. Eng. Phys. 107 (2007) 137–515.

[32] C. Eriksen, B. Lindqvist, Yield stress effects on contact pressure in computational models, J. Biomech. Eng. 146 (2013) 272–871.

[33] J. Eriksen, K. Okonkwo, A. Yilmaz, L. Kaur, H. Moreau, Yield stress effects on hemolysis index in computational models, Ann. Biomed. Eng. 87 (2013) 181–478.

[34] L. Bergman, T. Kaur, Bone density effects on fatigue life in computational models, Comput. Fluids 142 (2009) 200–576.

Table 1: Credibility assessment summary: one row per factor, model, and mechanism with 0–4 rating and basis

Credibility factor	Level	Basis
Equivalency of input parameters - Model A - M1	3	Flows matched within 2%, temperature 36.5 ± 0.3 °C, viscosity 3.5 ± 0.1 mPa·s; tip gap 90 ± 10 μm instantiated for near-wall shear comparisons.
Equivalency of input parameters - Model A - M2	3	Gap geometry from metrology (90 ± 10 μm) used; impeller speed within ±10 rpm; analog fluid at 22 °C set to 3.5 ± 0.05 mPa·s for tip-gap LDA mapping.
Equivalency of input parameters - Model A - M3	4	Flow and head within 2%; dye RTD at 3.5 L/min, 100 mmHg matched with age-of-fluid inputs; density 1040 ± 5 kg/m3 at 37 °C applied.
Equivalency of input parameters - Model B - M1	3	Boundary conditions mirror loops; hematocrit 38 ± 2% reflected via viscosity 3.5 ± 0.1 mPa·s; inlet scalar set to zero as in fresh-blood entry.
Equivalency of input parameters - Model B - M2	3	Operating points matched within 2% flow and ±10 rpm; scalar sources use local tau from Model A in measured 90 ± 10 μm gap.
Equivalency of input parameters - Model B - M3	4	Residence-time scalar calibrated with unit source; dye washout at 0.34 s mean aligned with age-of-fluid 0.31 s at 3.5 L/min, 100 mmHg.
Equivalency of input parameters - Model C - M1	3	Particle seeding proportional to inlet flux; wall sliding surrogate 0.2 ms; fluid properties set to measured blood values at 36.5 ± 0.3 °C.
Equivalency of input parameters - Model C - M2	3	Particles traverse measured 90 ± 10 μm gaps; tip-gap velocities from Model A under matched flow/head/speed; analog viscosity 3.5 ± 0.05 mPa·s.
Equivalency of input parameters - Model C - M3	4	Volute geometry and flow map matched; particle residence-time distributions compare to dye (mean 0.31 vs 0.34 s) and tracking (95th percentile 0.58 vs 0.55 s).
Model inputs - Model A - M1	3	Viscosity 3.5 ± 0.1 mPa·s and density 1040 ± 5 kg/m3 from rheometry; inlet TI 5% sensitivity changed head <0.5% and hemolysis flux <3%.
Model inputs - Model A - M2	3	Tip clearance 90 ± 10 μm measured per unit; mesh refined in gap; turbulence inputs held fixed; no tuned constants; peak shear stable within 6% over mesh refinement.
Model inputs - Model A - M3	3	Volute geometry verified ±0.1–0.15 mm; age-of-fluid transport enabled; inflow profile based on tubing; sensitivity to TI 2–8%