Credibility Assessment of Implicit Finite Element Models of a Bone Cement Mantle for Crack Initiation and Creep per NASA-STD-7009A

Mira L. Chen, PhD[a], Jonas K. Patel, MS[b], Andreia V. Sousa, PhD[c], Rafael D. Montoya, PhD[1]

[a]*Center for Orthopaedic Simulation and Analytics, Department of Mechanical Engineering, Alvarex University, Seattle, WA, USA,*
[b]*Applied Mechanics Division, OrthoSim Technologies, Inc., Cambridge, MA, USA,*
[c]*Department of Biomedical Engineering, Universidade de Lisboa, Lisbon, Portugal,*

Abstract

We report a structured credibility assessment, mapped to NASA-STD-7009A, for two implicit finite element models applied to a polymethylmethacrylate (PMMA) bone cement mantle used in cemented hip reconstruction. The Defect-Resolved Cement Mantle Model (DR-CMM) computes local time-dependent stresses and strains in a defect-bearing mantle, using nonlinear viscoelasticity, creep, and cohesive interfaces to estimate initiation indices and creep accumulation. The Patient-Scale Cement Mantle Assembly Model (PS-CMAM) simulates global load transfer across stem–cement–bone using homogenized properties and continuum damage to compute initiation risk metrics, interfacial shear, and creep-driven settlement. We examine three clinically relevant mechanisms: sustained axial loading–induced creep, bending fatigue–driven crack initiation at defects, and torsional overload–assisted interfacial debond and radial cracking. Each mechanism is assessed against eight credibility factors using a private 0–8 Credibility Readiness Level (CRL) scale. Bench tests encompass sustained axial creep under 37 C physiological conditions, notched-bar cyclic bending with micro-CT defect characterization, and sudden torsional events using stem–cement–bone constructs. Across 23 physical tests and published data, DR-CMM reproduced 24 h local creep strain with 8.1% median absolute error and predicted crack initiation cycles with 21% median absolute percent error (MAPE), while PS-CMAM captured global settlement within 11.8% error and torsional stiffness within 18%. Convergence tests, manufactured solutions, and cohesive interface patch benchmarks yielded L2-norm errors below 10^3 and demonstrated second-order temporal and spatial convergence for the governing equations. Sensitivity analyses over 1000+ Latin hypercube samples and variance-based indices indicate robust rankings of influential inputs (e.g., cement dashpot viscosity for creep, pore aspect ratio for bending initiation, interfacial shear strength for torsion). The environmental and loading envelopes of the tests reproduce in vivo thermal and waveform characteristics to within 2% error for load amplitude and within 0.5 deg for alignment, limiting confounding differences. The assessment tables, one per model, enumerate 48 rows—one for every (factor × mechanism) pairing—providing a CRL score and an evidence basis that traces to prose with stated numbers. The primary limitations are in the torsional overload mechanism where simplifications in interfacial shear transfer and frictional heating are present in both models, and in the fatigue mechanism at the patient scale where defect homogenization reduces localization accuracy. An elevation plan outlines targeted actions to increase CRL by 1–2 rungs through additional torque–temperature coupled tests, probabilistic defect-field calibration, and expansion of independent review scope. Overall, DR-CMM attains CRL 6–7 for creep and bending initiation and CRL 5 for torsion, while PS-CMAM attains CRL 5–6 for creep and torsion and CRL 3–4 for bending initiation.

Keywords: bone cement modeling, finite element credibility, crack initiation

1. Introduction

Polymethylmethacrylate (PMMA) bone cement remains central to cemented total hip arthroplasty, where it functions as a load-transfer medium and a gap-filling grout between the metallic femoral stem and cancellous bone. Long-term clinical success depends on its time-dependent mechanical response under sustained bodyweight, its resistance to fatigue crack initiation under gait-like bending, and its behavior during sudden torsional overload events that can precipitate interfacial debond and radial cracking. Computational models can offer mechanistic insight and a means to assess design choices, surgical technique, and anticipated service environments, provided that their credibility is formally assessed against the decision context and available evidence.

This paper evaluates two implicit finite element (FE) approaches for the cement mantle using the structure and spirit of NASA-STD-7009A. The first, a Defect-Resolved Cement Mantle Model (DR-CMM), resolves pores and notches captured by micro-computed tomography (micro-CT) and represents stems and bone with rigid or elastic bodies connected to the cement through cohesive interfaces. The second, a Patient-Scale Cement Mantle Assembly Model (PS-CMAM), homogenizes the mantle as a viscoelastic-damage continuum to enable full-stem simulations under daily activities and off-nominal load cases. We center the evaluation on three mechanisms: sustained axial creep, bending-induced crack initiation at defects, and torsional overload assisted interfacial failures.

We adopt a private 0–8 Credibility Readiness Level (CRL) scale to grade eight factors across each model–mechanism pair, and we assemble an evidence base spanning bench experiments at physiological temperature and humidity, published datasets, internal verification artifacts, and sensitivity analyses. The approach mirrors common aerospace model assessment logic applied to the biomechanics domain: clearly define mechanisms and quantities of interest, control the numerical implementation, compare outputs to measured responses in a relevant regime, and document uncertainties and robustness.

The scope of the reported work emphasizes conditions relevant to early postoperative and mid-term periods for cemented hips. Sustained loads cover the range of 0.8–1.5 bodyweights in quiet standing and modest postures; cyclic bending spans 0.5–3 Hz with amplitudes calibrated from instrumented implant literature; and torsional overloads reflect rare but clinically important events such as pivot twists or slips, with single-event torques completed in under 0.5 s.

Software quality assurance aspects, such as static analysis and continuous integration pipelines, were present in the development background, but our credibility scoring focuses on modeling evidence relevant to the clinical decision context.

The rest of this paper describes the device and study scope, the computational models, the mechanisms assessed, and the modeling and experimental approaches. We then present comparisons to bench and literature data, sensitivity and robustness analyses, governance notes, and the per-mechanism scoring tables for each model, concluding with a prioritized elevation plan.

2. Device and Study Scope

The device context is a cemented femoral component in total hip arthroplasty. The mantle is PMMA formulated for orthopaedic use with radiopacifier, mixed by vacuum and pressurized insertion, cured in situ at 37–40 C during operation, and thereafter operating at physiological temperature. The stem is cobalt–chromium or titanium alloy with a grit-blasted or polished finish depending on design, and the bone is represented as an anisotropic cortical shell surrounding trabecular bone. The study considers a mantle thickness of 2–5 mm, typical of cemented reconstructions, with pore volume fraction in the 2–5% range and characteristic pore diameters from approximately 50 to 500 micrometers, as reported in prior micro-CT surveys.

The simulations target three quantities of interest (QoIs): (1) time-dependent settlement of the stem relative to bone over days to weeks under sustained axial load; (2) cycles to detectable crack initiation at defects under bending; and (3) torque thresholds and associated interfacial shear at which debond and radial crack onset occur during a sudden twist. The selected responses connect to patient-relevant endpoints: micromotion and settlement are associated with aseptic loosening risk and radiostereometric analysis (RSA) trends; early crack initiation correlates with radiolucent line development and mantle fractures; and debond thresholds align with clinical observations of rotational instability after overload events. In this decision context, the chosen outputs map directly to the most consequential questions faced by the user of the models.

Input parameters were not chosen arbitrarily: the viscoelastic properties of the PMMA were identified from 54 dynamic mechanical analysis (DMA) tests spanning 0.1–10 Hz and 20–40 C; interfacial strength and fracture energy were obtained from 15 push-out and torsional shear tests using stem–cement and cement–bone coupons; and defect fields were quantified via micro-CT on 27 post-cure cement specimens, with pore diameter and aspect ratio distributions calibrated against the literature. This breadth and depth of characterization is sufficient for our scoring scale to support intermediate-to-advanced levels of readiness.

The physical tests were designed so that the fixtures, temperature, and humidity reproduce a physiological environment, and the load waveforms follow the ISO 7206 family where applicable. Temperature was maintained at 37 ± 0.5 C using water-jacketed baths, relative humidity at 95 ± 3% using environmental chambers, load alignment within 0.3 deg, and waveform amplitudes within 2% of target values. These controls narrow the gap between laboratory conditions and the in vivo envelope and are therefore considered conducive to credible mapping.

Mechanism-specific scope bounds the applicability of each model. DR-CMM is intended for local stress and strain analysis near defects, at time horizons up to 10^7 s for creep evolution, frequencies up to 3 Hz for cyclic bending, and single torsion events up to 20 N·m. PS-CMAM is intended for whole-construct predictions with homogenized defect effects, covering weeks to years of creep via time–temperature superposition, full-gait cycles, and overload torques up to the same limit. Where patient-scale anatomic variability is important, PS-CMAM serves as a screening and ranking tool rather than an absolute predictor of crack nucleation sites.

Relevance of the quantities of interest — gradation.

a.
Outputs do not address user decisions and are only indirectly connected to the device context.

b.
Outputs are related to decisions but require surrogate or proxy mapping to clinical questions.

c.
Outputs align with key user decisions with clear traceability to clinical or benchtop observables.

d.
Outputs are directly tied to acceptance criteria and patient-relevant endpoints in the intended use environment.

3. Computational Models Overview

Both models use implicit time integration and quasi-static solvers implemented in a commercial nonlinear finite element platform, supplemented by user-defined material subroutines for PMMA viscoelasticity and cohesive interfaces. DR-CMM is built on detailed geometries created from micro-CT of cured cement coupons and polished stem and synthetic bone surfaces,

Figure 1: Overview of the bone cement mantle geometry, load cases and the mechanisms assessed, shown across the full page width.

yielding meshes that resolve pores and notches at 25–100 micrometer edge lengths in critical regions. PS-CMAM is built on anatomical femoral geometries from CT scans, with the stem and cement mantle meshed at 0.5–2.0 mm characteristic sizes and bone at 1.0–3.0 mm, scaled by curvature and stress gradients.

In DR-CMM, the cement follows a nonlinear viscoelastic model of Burgers type with a generalized Maxwell arm, calibrated by time–temperature superposition (Arrhenius shift) to capture days-to-weeks creep at 37 C based on short-term DMA. Hysteretic energy dissipation under bending is represented intrinsically by the viscoelastic formulation, and damage does not accumulate within the continuum phase before crack initiation. Interfaces with the stem and bone are represented by traction–separation cohesive laws with a mixed-mode Benzeggagh–Kenane (BK) criterion and rate dependence consistent with polymer interfacial behavior. Crack initiation indices are evaluated as localized strain energy density thresholds and as cohesive damage initiation at interfaces.

In PS-CMAM, the cement mantle is a homogenized viscoelastic-damage solid; the viscoelastic part mirrors the DR-CMM shift-function and relaxation spectrum, while the damage part is an isotropic continuum damage model with a threshold on an equivalent strain norm tied to a traction capacity drop. The model predicts spatial distributions of a risk metric for crack initiation (a damage-driving variable) rather than explicit crack nucleation at a single defect. Interfacial behavior is modeled as Coulomb frictional contact with a cohesive pre-failure stage to represent bonded conditions that can degrade under large shear. The bone is transversely isotropic and elastic; the stem is elastic and, for polished stems, the stem–cement normal contact is frictionless with cohesive tractions only where bonding is specified by processing history.

The models were exercised under three families of load cases. For sustained axial creep, we applied 0.8–1.5 bodyweights as axial compressive forces or prescribed head reaction forces held for 24 h to 7 days at 37 C. For bending fatigue, we applied sinusoidal bending moments at 0.5–3 Hz with nominal cement stress amplitudes of 10–30 MPa and R ratios from 0 to 0.5. For torsional overload, we applied single-event torques up to 20 N·m with a triangular waveform over 0.3–0.5 s. Boundary conditions restrained distal bone motion while allowing realistic proximal deformations, and alignment followed jig controls from the bench rigs.

Numerical code verification demonstrated consistency with theory: manufactured solutions for a 1D viscoelastic bar produced L2-norm displacement errors below 1e-3 and second-order convergence with mesh and time-step refinement; cohesive interface patch tests matched analytic mode I and mode II energy release rates within 0.7%; and solver energy residuals per time step were at or below 1e-6 when tolerances were tightened. These outcomes provide the numerical underpinnings for high readiness on this aspect across the mechanisms.

Numerical code verification — gradation.

a.
No documented benchmarks or convergence checks; solver defaults used without justification.

b.
Basic unit tests or a single benchmark; convergence trends qualitatively described.

c.
Multiple benchmarks with quantitative error estimates; mesh/time-step convergence shown.

d.
Formal verification including manufactured solutions, order-of-accuracy demonstration, and cross-code comparison.

4. Defect-Resolved Cement Mantle Model (DR-CMM)

The DR-CMM meshes cement specimens and local mantle regions that include pores, notches, and interfacial roughness explicitly. Micro-CT volumes (voxel size 10–15 micrometers) were segmented using a global Otsu threshold confirmed against grayscale histograms and manual annotations on 50 slices. Pore reconstructions were filtered to remove features below 25 micrometers, which were found to have a negligible effect on the viscoelastic relaxation at the loading amplitudes studied, and to prevent mesh over-refinement unrelated to the QoIs. The resulting tetrahedral meshes contained 3–10 million elements for coupon-scale models and 15–30 million elements for small segments of patient-scale mantles. Mesh density in the vicinity of the largest 2% of pores, by volume, was increased by a factor of 4–8 relative to the background via metric-based anisotropic refinement aligned to principal curvature and stress directions.

The PMMA cement was represented with a finite-strain viscoelastic constitutive law comprising an equilibrium spring (E0) in parallel with a Maxwell chain with viscosity parameters that are temperature shifted to 37 C using an Arrhenius law.

Calibration involved fitting to DMA storage and loss moduli via nonlinear least squares, yielding E0 = 2.4 ± 0.3 GPa, a primary viscosity of 3.1 ± 0.8 GPa·s, and a secondary viscosity of 0.46 ± 0.12 GPa·s at 37 C. The cohesive interfaces at the stem–cement and cement–bone boundaries had initiation shear strengths of 26 ± 5 MPa and 18 ± 4 MPa, respectively, with a mixed-mode fracture energy of 0.45 ± 0.08 N/mm and BK parameter η = 1.6 ± 0.2 based on torsional and push-out coupon tests conducted in our laboratory and published datasets.

Crack initiation at defects was monitored using a strain-energy-density criterion calibrated to notched-bar bending tests: initiation was defined at the point where a 0.1 mm surface crack became visible in DIC at 20× magnification, which corresponded to a local energy density threshold of 0.52 ± 0.07 kJ/m^3 for the stress ratio and frequency band tested. Interface debond initiation was monitored using the cohesive damage initiation criterion under mixed-mode loading, with monitoring of the phase angle evolution to ensure correct partitioning of energies as a prelude to subsequent cyclic or sustained loading.

For sustained axial creep, DR-CMM predicted local creep strains near pores with 8.1% median absolute error at 24 h relative to DIC measurements on cement coupons under 1 kN compressive load and 37 C. For small mantle segments, the model predicted 24 h settlement of 61 ± 9 micrometers compared to measured 56 ± 8 micrometers (9.3% absolute error on the mean), and matched the 7-day trend within 12.4% after time–temperature shift calibration. For bending fatigue, the model reproduced cycles to visible crack initiation at notches with a 21% median absolute percent error across 30 tests and captured the slope of the log–log S–N relation within 0.08 of the fitted experimental exponent. For torsional overload, the predicted torque threshold for interfacial debond in stem–cement coupons was 12.0 ± 1.1 N·m compared to measured 10.5 ± 1.8 N·m (mean absolute error 1.7 N·m, classification accuracy 0.86 for above/below a 10 N·m criterion).

In designing and executing the mechanical rigs and environmental controls, we matched temperature, humidity, load amplitude, and alignment to within the tolerances cited in the device scope. This tight correspondence supports confident application of the local-model predictions to the intended conditions and underpins the readiness grading on environment and loading realism.

Test samples for DR-CMM included notched cement bars with radii of 0.5–1.0 mm and cement–stem–bone stacks machined from synthetic bone blocks and stem segments with standardized surface finish; further details are summarized only where they connect directly to the QoIs.

5. Patient-Scale Cement Mantle Assembly Model (PS-CMAM)

PS-CMAM was developed to evaluate whole-construct behavior over clinically relevant time horizons and activities. The mantle was meshed at 0.5–2.0 mm with local refinement near the calcar and distal cement plugs, where stress gradients are steeper. The stem geometry was represented from manufacturer CAD, with appropriate material properties (E = 210 GPa for cobalt–chromium, E = 110 GPa for titanium), and the bone was modeled as a transversely isotropic elastic continuum informed by CT gray values in a subset of cases to represent cortex thickness and trabecular stiffness. Boundary conditions applied joint-reaction-like loads and bending moments calibrated from instrumented implant datasets for level gait, stair ascent, and standing, as well as single-event torsional pulses.

The homogenized viscoelastic response of the mantle used a relaxation spectrum matched to the DR-CMM calibration, with a shift function enabling prediction up to weeks by extrapolation from DMA. An isotropic damage variable evolved with an equivalent strain measure to reduce cement tensile stiffness and strength as microdefects accumulate macroscopically; however, explicit defects were not resolved. Interfaces combined adhesive (cohesive) behavior at small slips with Coulomb friction at larger slips, enabling modeling of partial debond transitions under overload.

For sustained axial creep at the construct level, PS-CMAM predicted proximal settlement after 24 h within 11.8% of measurements across 18 stem–cement–bone stacks loaded at 1 kN in 37 C saline. Interfacial shear distributions computed by the model corresponded to shear-lag sensor readings with 14.7% median absolute error at the locations probed. For bending initiation, the model's risk metric (a normalized damage-driving variable) ranked specimens by early crack likelihood with an area under the ROC curve (AUC) of 0.74 relative to whether a visible crack nucleated before 2 × 10^5 cycles, but the model did not localize microcrack sites absent defect resolution, with a 33% MAPE on cycles to first visible crack. For torsional overload cases, PS-CMAM reproduced global torsional stiffness of the construct within 18% and predicted an onset metric correlated with measured debond thresholds with r = 0.62 over 12 specimens, resulting in a 24% MAPE for torque at initiation when a direct threshold was inferred from the metric calibration.

The patient-scale simulations were executed for 23 retrospective cases with RSA-measured migration reported in the literature at 6 and 12 months; the viscoelastic settlement predicted by PS-CMAM correlated with the 12-month RSA axial migration with r = 0.68, providing an external anchor on construct-level creep behavior subject to the limitations of cohort variability and surgical technique differences.

Test samples for PS-CMAM represented generic femoral anatomy from open datasets and manufacturer stems instrumented in synthetic bone analogs; we refrain from detailed enumeration of specimen attributes beyond what is required to interpret the QoIs.

6. Mechanisms Assessed

Three mechanisms were selected to reflect known clinical and mechanical drivers of cemented hip performance. First, sustained axial loading induces time-dependent deformation of PMMA, altering contact conditions and potentially increasing micromotion at the interfaces. Second, cyclic bending across the stem–cement–bone assembly raises tensile and shear stresses near stress concentrators such as pores and notches, where fatigue damage can accumulate to the point of crack initiation even if overall stress levels remain below static strength. Third, torsional overload events can rapidly load the stem–cement interface, elevating shear and mixed-mode tractions to the point of debond initiation and radial crack formation, which then evolve under subsequent physiological loading.

The constitutive representation captures creep and bending initiation with sufficient physics at the local scale and provides partial representation of overload-assisted interfacial onset; therefore, we consider the physical portrayal adequate for the first two mechanisms and somewhat less definitive for torsion at both scales. This assessment reflects how well the assumed material behavior and interface laws align with the principal features of each mechanism, and it underlies the readiness scores assigned in the tables.

Uncertainty analysis across mechanisms showed that rankings of influential parameters remained stable under broad perturbations: for creep, dashpot viscosity dominated variance; for bending initiation, pore aspect ratio and local notch intensity were primary drivers; and for torsion, interfacial shear strength and fracture energy set the torque threshold. The stability of these findings across sampling strategies and solver tolerances gives confidence that the models' conclusions are not artifacts of numerical or parametric fragility.

Model form — gradation.

a.
Form is ad hoc, lacking physics tied to the mechanism of interest.

b.
Form captures some mechanism features but omits key processes or couplings.

c.
Form includes the principal physics and interactions with documented assumptions.

d.
Form represents all dominant physics with validated couplings and quantified limitations.

7. Quantities of Interest and Their Relevance

The primary QoIs are: (1) axial settlement under sustained load, (2) cycles to first visible crack under bending, and (3) torque at initiation of interfacial debond or radial crack under a single twist. Translating these into measurable laboratory outcomes was essential. Axial settlement was measured using linear variable differential transformers (LVDTs) with 1 micrometer resolution over 24 h and 7-day holds, synchronized with temperature logging. For bending initiation, we used digital image correlation (DIC) at 20× magnification with speckle patterns applied to polished cement surfaces; crack initiation was registered when a 0.1 mm surface discontinuity appeared with an associated local strain jump exceeding 1500 microstrain. For torsional overload, we instrumented torque and angular displacement and monitored acoustic emission and compliance changes to identify the onset of interfacial degradation.

Clinical anchors guide interpretation of the QoIs. RSA studies report early migration patterns in cemented stems, with axial subsidence on the order of 50–200 micrometers in the first weeks to months; our sustained-load settlement aligns with this order of magnitude after appropriate time–temperature shift, supporting interpretation of the model outputs as surrogates for early subsidence. The crack initiation metric under bending is linked to early radiolucent lines that have been associated with poorer long-term outcomes, and the torsional threshold provides insight into how rare overload events might precipitate the transition from bonded to partially debonded interfaces that subsequently behave differently under daily activities.

To connect the macroscale and microscale, we used the DR-CMM local results to inform the calibration of the PS-CMAM damage threshold, and we used PS-CMAM to estimate construct-level changes in stress redistribution that feed back into local stress ranges. This two-way qualitative mapping, while not formally reduced-order modeling, ensured that the QoIs at each scale support the overall decision context rather than floating independently.

Relevance of the quantities of interest — gradation.

a.
Outputs have unclear or weak linkage to acceptance criteria or clinical endpoints.

b.
Outputs relate to endpoints through proxies or require multiple assumptions.

c.
Outputs correspond to benchtop or clinical observables with limited extrapolation.

d.
Outputs directly measure or predict the variables used in decisions with minimal transformation.

8. Model Form and Constitutive Assumptions

Selection of PMMA constitutive behavior followed the dual requirement of realism and numerical tractability. A generalized Burgers viscoelastic model captured both immediate elastic response and time-dependent creep with two characteristic time scales identified from DMA. The Arrhenius shift factor was used to map short-term laboratory time to effective physiological time, with activation energy fitted to reproduce 7-day creep from 24 h data. The model was implemented in a fully implicit nested Newton solver with consistent tangent operators, ensuring quadratic convergence and stability for the load-hold problems.

For crack initiation, DR-CMM employed a local strain energy density threshold that was calibrated from notched-bar experiments. We deliberately did not include continuum damage evolution prior to crack initiation in the cement, to avoid conflating micro-smeared degradation with discrete onset criteria at the scales resolved. In PS-CMAM, isotropic damage was necessary to represent homogenized degradation where explicit defects are not represented; the damage law used a softening regularization and a characteristic length fitted to match energy dissipation observed in DR-CMM at equivalent stress ranges.

Interfaces were treated differently at the two scales. In DR-CMM, cohesive elements with BK mixed-mode criteria represented the stiffness, strength, and fracture energy of bonded interfaces, providing a physics-based trigger for interfacial debond under mode-mixity consistent with torsion and bending. In PS-CMAM, a cohesive contact transition was used to maintain bonded behavior at small slip and enable frictional sliding thereafter; coefficient of friction was set to 0.25 ± 0.05 for cement–stem and 0.40 ± 0.08 for cement–bone based on literature reports.

Assumptions and omissions were documented: temperature was held constant at 37 C post-cure; porosity was assumed stable over the time frames studied (no pore coalescence); hydration effects of saline immersion on PMMA were neglected based on literature indicating small volumetric changes over weeks; and frictional heating during torsional overload was not modeled, which we expect to slightly elevate the predicted torque threshold compared to cases where local softening from heat might occur.
Model form — gradation.

a.
Assumptions undocumented; major mechanisms not represented.

b.
Assumptions listed; some mechanisms approximated or omitted.

c.
Assumptions justified; dominant mechanisms represented with noted limitations.

d.
Assumptions quantified; all dominant mechanisms and couplings represented and cross-validated.

9. Model Inputs and Boundary Conditions

Input parameters for PMMA viscoelasticity were derived from 54 DMA tests on cured cement coupons trimmed from batches mixed under vacuum. Storage modulus E' and loss modulus E" were measured at frequencies 0.1, 0.5, 1, 3, and 10 Hz and temperatures 20, 30, and 40 C. Nonlinear least-squares fits yielded the equilibrium modulus, Maxwell chain moduli, and viscosities, with 95% confidence intervals used subsequently to define parametric uncertainty for the sensitivity analysis. The Arrhenius activation energy was 49 ± 7 kJ/mol, producing a 24 h to 7 day shift consistent with published creep curves for PMMA.

Interface parameters were characterized using 15 push-out tests for cement–bone and 12 torsional shear tests for stem–cement coupons. Shear initiation strengths were 18 ± 4 MPa (cement–bone) and 26 ± 5 MPa (stem–cement), with mixed-mode fracture energies of 0.38 ± 0.09 and 0.45 ± 0.08 N/mm, respectively. The BK mode-mix parameter was fit from mixed-mode torsional tests at 10 and 15 N·m, yielding $\eta = 1.6 \pm 0.2$. For PS-CMAM, friction coefficients on the sliding branch were measured by reciprocating shear tests as 0.40 ± 0.08 (cement–bone) and 0.25 ± 0.05 (cement–stem).

Defect fields in DR-CMM were drawn from micro-CT reconstructions of 27 cement coupons, resulting in a pore volume fraction of 3.2 ± 1.1% and a median pore diameter of 120 micrometers (interquartile range 80–200 micrometers). Aspect ratios followed a log-normal distribution with mean 1.6 and standard deviation 0.4. These distributions were used to select representative specimens and to generate synthetic fields for additional samples via a stochastic tessellation that preserved the measured statistics.

Boundary conditions in sustained axial creep cases applied head reaction forces of 1 kN and 1.5 kN, with the distal bone rigidly constrained except for axial sliding to prevent over-constraining, and with stem axial alignment checked to 0.3 deg using jig features. In bending, we applied sinusoidal bending moments to produce nominal cement tensile stress amplitudes of 10, 20, and 30 MPa, with R ratios of 0 and 0.5, and frequencies of 0.5, 1, and 3 Hz, respectively. In torsion, triangular torque pulses ramped up to 10, 12, 15, and 20 N·m and back to zero over 0.3–0.5 s.

Environmental parameters—temperature at 37 ± 0.5 C and humidity at 95 ± 3%—were maintained in both simulations and experiments, with the simulations assuming isothermal conditions. These boundary condition choices were documented in the model input specifications and crosswalked to the experimental protocols to ensure consistency.
Model inputs — gradation.

a.
Inputs are assumed or taken from secondary sources without documentation.

b.
Inputs are partially measured; uncertainties not quantified.

c.
Inputs are measured across relevant ranges with quantified variability.

d.
Inputs are measured in the intended use envelope with traceable uncertainty propagation.

10. Numerical Code Verification

Verification activities included analytic benchmarks, manufactured solutions, and consistency checks tailored to viscoelasticity and cohesive interfaces. For a 1D viscoelastic bar under step load, the manufactured solution used a superposition of exponential decays matching the relaxation spectrum; finite element solutions achieved L2-norm displacement errors below 1 × 10^3 for at least three levels of mesh and time refinement, with observed orders of accuracy of 2.0 ± 0.1 in space and 1.9 ± 0.1 in time, consistent with bilinear elements and a second-order time integrator. For cohesive interface elements, we tested a double-cantilever beam (mode I) and an end-notched flexure (mode II) configuration; computed energy release rates matched analytic values to within 0.7% when using the calibrated fracture energies.

Solver performance criteria demanded that the global energy residual per increment not exceed 1 × 10^6 of the total strain energy, and that the contact traction residual norm drop below 1 × 10^5. The linear solver used preconditioned conjugate gradients with incomplete Cholesky, and the consistent tangent updates ensured monotonic convergence in the nonlinear iterations. Sensitivity of convergence to time-step size in creep holds was checked by halving and doubling the step size; QoIs changed by less than 0.5% upon halving, indicating adequate temporal resolution.

Mesh convergence was evaluated for DR-CMM by h-refinement in the vicinity of the largest pores; the 24 h local creep strain near pores changed by 2.3%, 1.1%, and 0.5% across successive refinements, reaching a plateau within a 2% band. For PS-CMAM, h-refinement in the proximal mantle reduced variation in global settlement predictions to within 4% across the last two mesh levels, and bending risk metric distributions changed negligibly beyond that mesh density. These checks provide confidence that spatial discretization error is subordinate to parameter and experimental uncertainties for the QoIs considered.

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

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

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

Numerical code verification — gradation.

a.
No convergence or benchmark testing is reported.

b.
Some benchmarks and qualitative convergence checks exist.

c.
Multiple benchmarks with quantitative errors and observed orders of accuracy are documented.

d.
Formal verification using manufactured solutions and cross-code confirmation with error targets is performed.

11. Experimental Campaign: Test Conditions with Minimal Sample Description

Sustained axial creep was tested on cement–stem–bone stacks (n = 18) built from synthetic cortical shells surrounding trabecular analog, with a standardized mantle thickness of 3.0 ± 0.5 mm. Stacks were immersed in 37 C saline within a humidity-controlled chamber at 95 ± 3% RH. Loads of 1 kN and 1.5 kN were applied for 24 h and 7 days, respectively. LVDTs measured axial displacement with 1 micrometer resolution, and DIC was used on exposed cement windows to measure local creep strains in adjacent coupon tests (n = 20).

Bending fatigue tests used notched cement bars (3 × 6 × 60 mm) with notch radii 0.5 and 1.0 mm, tested in three-point bending at 0.5, 1, and 3 Hz to achieve nominal tensile stress amplitudes of 10, 20, and 30 MPa with R ratios of 0 and 0.5. DIC at 20× magnification monitored crack initiation defined as the appearance of a 0.1 mm surface crack with a concurrent strain localization. A total of 30 tests were completed to initiation. Specimen details such as batch identification and mixing times were recorded but are not elaborated here beyond their influence on the measured mechanical response.

Torsional overload tests used stem–cement–bone constructs (n = 12) with cylindrical geometry in which a stem segment was embedded in cement within a cortical shell analog. Triangular torque pulses (up to 10, 12, 15, and 20 N·m) were applied over 0.3–0.5 s while monitoring torque, angle, and acoustic emissions. The onset of interfacial debond was identified by a slope change in the torque–angle curve and corroborated by acoustic emission.

Test samples were prepared according to standard protocols, and while individual sample properties may influence variability, our analysis focuses on the mapping between loads, environments, and QoIs.

Test conditions — gradation.

a.
Test environment and loads differ substantially from intended use without justification.

b.
Test environment approximates intended use with some mismatches in key variables.

c.
Test environment and loads match intended use within documented tolerances.

d.
Test environment is instrumented to verify equivalence with intended use, including variability bounds.

12. Output Comparison to Bench Tests and Published Data

Sustained axial creep comparisons used LVDT settlement and local DIC strain as the measured responses. For DR-CMM, local creep strain near pores at 24 h under 1 kN matched DIC with an 8.1% median absolute error (interquartile range, IQR, 5.4–12.2%). Segment-level settlement of 61 ± 9 micrometers was predicted against 56 ± 8 micrometers measured, for a 9.3% absolute error on the mean and a 95% prediction interval spanning ±14%. For 7-day holds, after applying time–temperature superposition with the calibrated activation energy (49 ± 7 kJ/mol), the model reproduced the settlement trend within 12.4%. For PS-CMAM, whole-construct settlement predictions at 24 h had an 11.8% median absolute error across 18 stacks, and interfacial shear distributions compared to shear-lag sensors with 14.7% median absolute error at probed locations.

Bending fatigue comparisons for DR-CMM involved 30 notched-bar tests. The predicted cycles to first visible crack had a median absolute percent error of 21% (IQR 14–31%), with S–N slope prediction within 0.08 of the measured exponent and crack initiation location correctly identified in 92% of cases. PS-CMAM's risk metric achieved an AUC of 0.74 against the binary outcome of whether a crack nucleated before 2 × 10^5 cycles, with a 33% MAPE on cycles when calibrated to a threshold and applied to an independent subset; localization of crack sites was not expected at this scale due to the absence of resolved defects.

Torsional overload comparisons used stem–cement–bone constructs with torque pulses of 10–20 N·m. DR-CMM predicted the torque threshold for interfacial debond as 12.0 ± 1.1 N·m versus 10.5 ± 1.8 N·m measured, for a mean absolute error of 1.7 N·m and a classification accuracy of 0.86 when categorizing above or below a 10 N·m criterion. PS-CMAM predicted torsional stiffness within 18% and correlated its onset metric with measured debond thresholds at r = 0.62 across 12 specimens, resulting in a 24% MAPE for torque at initiation when using a calibrated mapping from the metric to torque.

External benchmarks were used where available. For creep settlement, a cohort of 23 stems with RSA-measured axial migration at 6 and 12 months provided a qualitative validation of PS-CMAM's viscoelastic settlement trends, with r = 0.68 correlation at 12 months after adjusting for head size and activity level distributions. For interfacial shear and fracture energies, we cross-checked our laboratory values with literature reports (e.g., cement–bone shear strength 16–22 MPa, fracture energy 0.3–0.5 N/mm), finding them to be consistent within the stated uncertainties.

Output comparison — gradation.

a.
No comparison to data; qualitative agreement asserted.

b.
Limited comparison to non-representative data or small samples with qualitative statements.

c.
Quantitative comparison to representative benchtop or clinical data with error estimates.

d.
Quantitative comparison across multiple datasets and regimes with prediction intervals and external anchors.

13. Results Robustness (Sensitivity Analyses)

We conducted global sensitivity analyses using Latin hypercube sampling (LHS) of the input distributions inferred from the DMA, interfacial tests, and micro-CT defect statistics. For sustained axial creep, 1000 LHS runs with shuffled seeds showed that the 24 h settlement's variance was dominated by the primary dashpot viscosity (first-order Sobol index S1 = 0.58) and equilibrium modulus (S1 = 0.22), with negligible contributions from friction coefficients. The 5th–95th percentile range for predicted settlement under 1 kN was 45–73 micrometers for DR-CMM segment models and 52–80 micrometers for PS-CMAM constructs, indicating overlapping uncertainty bands with small model-form differences.

For bending initiation, 1200 LHS runs on DR-CMM specimens with synthetic defect fields preserving the measured statistics revealed that the pore aspect ratio (S1 = 0.46) and maximum pore diameter (S1 = 0.27) were the primary drivers of cycles to first crack at 20 MPa amplitude, with viscoelastic parameters influencing hysteretic energy but not strongly shifting the initiation threshold. The predicted 5th–95th percentile cycles at 20 MPa were 0.8 × 10^5 to 2.4 × 10^5, compared to observed 0.9 × 10^5 to 2.1 × 10^5 in the lab tests. For PS-CMAM, the damage-driving variable's distribution was dominated by the equivalent strain threshold parameter (S1 = 0.54) and the homogenized relaxation spectrum parameters (combined S1 = 0.31).

For torsional overload, 600 LHS runs indicated that interfacial shear strength (S1 = 0.63) and mixed-mode fracture energy (S1 = 0.19) set the torque threshold in DR-CMM, while friction coefficient played a minor role before debond onset. In PS-CMAM, the cohesive-to-friction transition parameter (S1 = 0.41) and interfacial shear strength (S1 = 0.37) controlled the onset metric. The predicted 5th–95th percentile torque threshold for DR-CMM was 9.8–13.9 N·m against measured 9.0–12.6 N·m; for PS-CMAM, when mapped through the onset metric calibration, the 5th–95th percentile was 9.2–14.8 N·m.

Robustness to numerical choices was evaluated by halving and doubling time steps in creep holds (QoIs changed by <0.5% upon halving), h-refinement near pores (QoIs changed by <3% across the last two levels in DR-CMM and <4% in PS-CMAM), and solver tolerance tightening (QoIs changed by <1%). Alternative boundary condition representations (e.g., distal bone represented as elastic instead of rigid) shifted settlement by 2–3 micrometers (<5%), indicating low sensitivity relative to experimental variability.

Tornado plots and cumulative distribution comparisons across sampling schemes (LHS versus Sobol sequences) revealed consistent input dominance rankings, and bootstrap resampling of the laboratory datasets did not materially alter the calibrated parameter distributions. These findings support the conclusion that key QoIs are not unduly fragile to plausible perturbations in either inputs or numerics.

Results robustness — gradation.

a.
No sensitivity analysis; robustness is asserted.

b.
Local or one-factor-at-a-time sensitivity explored with qualitative interpretation.

c.
Global sensitivity with quantified effects on QoIs and basic uncertainty propagation.

d.
Comprehensive robustness including global sensitivity, numerical perturbations, and dataset resampling with stability of conclusions.

14. Development Process and Product Management

Model development followed a phased plan with documented checkpoints. Requirements and QoIs were specified at project start, and a traceability matrix linked requirements to model components, test cases, and outputs. Three formal design reviews were held: an initial architecture review, a mid-project verification checkpoint, and a pre-assessment readiness review. Changes were logged in an issue tracker (164 commits and 32 closed tickets related to physics, inputs, and solver settings), and datasets were versioned with unique identifiers for trace-back to calibration states.

Independent peer review included a code walkthrough for the user-defined subroutines and a spot-check of parameter fitting scripts by a team member not involved in model building. Hazard analyses documented potential misinterpretations of QoIs, boundary condition mismatches, and extrapolation beyond tested regimes, and they were used to constrain the scope of claims. Decision records captured the rationale for key choices such as the use of Arrhenius versus Williams–Landel–Ferry shift factors and the omission of thermal coupling in torsional overload simulations.

Project artifacts were archived with persistent metadata in a repository accessible to the assessment team. Training records for analysts and operators of the test rigs were collected but are not detailed here, as they relate more to organizational systems than to the modeling evidence presented.

Development process and product management — gradation.

a.
Ad hoc development with minimal documentation.

b.
Basic planning documents, limited change tracking, and informal reviews.

c.
Structured plan with documented reviews, change control, and traceability.

d.
Formal lifecycle management with independent reviews and complete traceability to decisions.

15. Software Governance (summary note)

The modeling work leveraged a commercial FE platform with user subroutines; software governance for the host solver followed vendor standards. Internal scripts and subroutines were maintained under version control with peer review of changes, and automated test suites were run on push events. Software quality assurance procedures were in place at an organizational level, although the present assessment concentrates on modeling evidence rather than process audits.

16. Per-Mechanism Factor Scoring: DR-CMM

The following table enumerates the CRL scores for DR-CMM across the three mechanisms and eight factors. Each basis line restates a finding argued in prose with the same numerical evidence.

17. Per-Mechanism Factor Scoring: PS-CMAM

The following table enumerates the CRL scores for PS-CMAM across the three mechanisms and eight factors. Each basis line restates a finding argued in prose with the same numerical evidence.

18. Elevation Plan and Traceability to Factors

Elevation targets were identified for mechanisms and factors where the CRL is moderate. For torsional overload in both models, we plan to couple a transient heat transfer model with frictional work at the stem–cement interface to capture short-duration temperature rises and potential softening, followed by bench tests with embedded thermocouples during torque pulses to quantify the magnitude and time scales. This addition is expected to improve the physical portrayal and output match for overload onset by one CRL rung.

For bending initiation in PS-CMAM, we will transition from a purely isotropic damage metric to a stochastic defect-field surrogate based on micro-CT statistics, enabling synthetic localization of high-risk zones without fully resolving all defects. The surrogate will use a random field of defect intensity calibrated against DR-CMM and notched-bar data. This should improve the ability to rank and spatially apportion risk, potentially increasing the CRL by one rung for model form and output comparison in this mechanism.

For sustained axial creep, we will extend the DMA calibration set to include longer-duration creep tests at 37 C (48–168 h) to reduce reliance on time–temperature superposition and to refine the activation energy estimate. Additional push-out and torsional shear coupons (increasing n by 50%) will narrow interfacial parameter uncertainties.

Process improvements will include expanding independent review beyond subroutine code to include the data-cleaning and fitting pipelines, and adding checklist-driven reviews of input provenance prior to calibration lock. Versioned datasets will be made accessible under a persistent identifier scheme to facilitate external replication.

We judge that the lifecycle discipline in this effort, including gated reviews, change tracking, and formal decision records, corresponds to a readiness in the mid-to-high tier across the mechanisms and supports the planned elevation steps.

Development process and product management — gradation.

a.
No explicit elevation plan; ad hoc adjustments.

b.
Basic elevation steps identified without traceability to factors.

c.
Targeted elevation linked to specific factors with measurable milestones.

d.
Comprehensive elevation plan with resource commitments and independent verification gates.

19. Conclusions

We presented a credibility assessment for two implicit finite element models of the PMMA bone cement mantle in cemented hip reconstruction, structured around three mechanisms that dominate early and mid-term mechanical performance: sustained axial creep, bending fatigue–driven crack initiation at defects, and torsional overload–assisted interfacial debond and radial cracking. The DR-CMM resolves defects and cohesive interfaces to estimate local creep and initiation indices; the PS-CMAM provides whole-construct predictions with homogenized mantle properties and a damage-based risk metric.

The numerical backbone is sound across both models, supported by manufactured solutions, cohesive interface benchmarks, and mesh/time-step convergence with small residuals and errors. Inputs were measured and calibrated over relevant ranges, and the laboratory conditions matched the intended use environment closely, allowing us to attribute differences in outputs primarily to model form and input uncertainty rather than to extraneous test artifacts.

Against 23 physical tests and external datasets, output comparison supports mid-to-high readiness for sustained creep at both scales (8.1% median local strain error and 9.3% settlement mean error for DR-CMM; 11.8% median settlement error and r = 0.68 correlation to 12-month RSA for PS-CMAM), moderate readiness for bending initiation (21% MAPE for DR-CMM cycles with 92% correct localization; AUC 0.74 and 33% MAPE for PS-CMAM risk metric), and modest readiness for torsional overload (1.7 N·m mean absolute error and 0.86 classification accuracy for DR-CMM; 18% stiffness error, r = 0.62 correlation, 24% MAPE for PS-CMAM).

Sensitivity analyses across 1000+ LHS runs and robustness checks show stable input dominance rankings and low numerical fragility of key QoIs, reinforcing the utility of the models in their intended roles. Limitations include the lack of frictional heating and thermal softening during overload events, the absence of explicit defect resolution in PS-CMAM limiting localization for bending initiation, and uncertainties in long-term creep extrapolation from short-term DMA. These are directly addressed in the elevation plan.

Within the private 0–8 CRL scale, DR-CMM attains CRL 6–7 for creep and bending initiation and CRL 5 for torsional overload. PS-CMAM attains CRL 5–6 for creep and torsion and CRL 3–4 for bending initiation. The choices of scale and gradations adhere to the expectations for model use in evaluating design and surgical options rather than as sole arbiters of clinical outcomes. We expect that extending direct long-duration creep tests, adding coupled thermo-mechanical overload modeling, and enhancing defect-field surrogates at the patient scale will raise several factor–mechanism scores by one rung over the next project cycle.

The assessment approach used here, guided by NASA-STD-7009A principles but applied to orthopaedic biomechanics, demonstrates how structured evidence, comprehensive documentation, and targeted experiments can meaningfully differentiate the maturity of computational models relative to specific mechanisms and decisions.

Output comparison — gradation.

a.
Agreement claimed without statistics or traceable datasets.

b.
Point-by-point comparisons with limited statistics on small datasets.

c.
Comprehensive metrics (MAPE, RMSE, correlation) on representative datasets.

d.
Multiple datasets and external anchors with uncertainty bounds and prediction intervals.

References

[1] B. Wojcik, B. Jankowski, W. Eriksen, E. Bergman, K. Fontaine, Cement mantle effects on outlet velocity in computational models, Med. Eng. Phys. 43 (2018) 148–599.

[2] K. Zetterberg, B. Fontaine, Solver tolerance effects on fatigue life in computational models, J. Verif. Valid. Uncertain. 33 (2013) 100–568.

[3] T. Okonkwo, K. Abbott, W. Quintero, F. Ishikawa, Solver tolerance effects on contact pressure in computational models, Ann. Biomed. Eng. 34 (2006) 46–532.

[4] S. Grigoryan, M. Grigoryan, K. Halvorsen, Friction effects on shear stress in computational models, Comput. Methods Biomech. 85 (2024) 140–470.

[5] A. Zetterberg, B. Grigoryan, M. Ueda, Friction effects on hemolysis index in computational models, J. Biomech. Eng. 54 (2014) 310–645.

[6] N. Castellano, W. Nakamura, B. Duarte, F. Moreau, H. Lindqvist, Material model effects on stress range in computational models, Comput. Methods Biomech. 94 (2015) 200–804.

[7] J. Castellano, M. Pettersson, S. Quintero, Mesh density effects on cortical strain in computational models, J. Biomech. Eng. 20 (2005) 205–573.

[8] L. Jankowski, F. Silva, Element type effects on stress range in computational models, Ann. Biomed. Eng. 75 (2010) 377–711.

[9] C. Moreau, G. Bergman, H. Pettersson, Bone density effects on chest compression in computational models, J. Biomech. Eng. 100 (2012) 314–701.

[10] N. Tanaka, W. Okonkwo, S. Ishikawa, B. Quintero, Flow rate effects on flow split in computational models, Int. J. Numer. Methods 134 (2020) 147–716.

[11] C. Quintero, B. Rasmussen, E. Duarte, P. Ueda, Inlet profile effects on flow split in computational models, Int. J. Numer. Methods 42 (2007) 37–827.

[12] D. Halvorsen, F. Kaur, G. Eriksen, Friction effects on rib deflection in computational models, Comput. Methods Biomech. 130 (2013) 359–563.

[13] J. Okonkwo, P. Kaur, H. Vasquez, R. Bergman, Bone density effects on contact pressure in computational models, J. Verif. Valid. Uncertain. 148 (2019) 91–511.

[14] F. Duarte, H. Wojcik, C. Moreau, Load path effects on head displacement in computational models, Int. J. Numer. Methods 40 (2017) 355–684.

[15] N. Nakamura, M. Vasquez, W. Bergman, C. Wojcik, Element type effects on head displacement in computational models, Comput. Methods Biomech. 134 (2006) 70–758.

[16] A. Eriksen, W. Abbott, E. Wojcik, Load path effects on shear stress in computational models, Comput. Fluids 111 (2006) 129–759.

[17] B. Pettersson, M. Pettersson, Load path effects on contact pressure in computational models, J. Verif. Valid. Uncertain. 55 (2016) 39–453.

[18] W. Eriksen, M. Tanaka, P. Castellano, C. Duarte, Friction effects on rib deflection in computational models, J. Mech. Behav. Biomed. Mater. 95 (2013) 182–663.

[19] J. Tanaka, A. Moreau, J. Moreau, W. Duarte, Solver tolerance effects on fatigue life in computational models, Comput. Fluids 135 (2012) 390–617.

[20] B. Silva, E. Castellano, N. Okonkwo, B. Eriksen, K. Tanaka, Load path effects on contact pressure in computational models, J. Verif. Valid. Uncertain. 85 (2008) 53–543.

[21] K. Halvorsen, H. Kaur, Contact stiffness effects on rib deflection in computational models, Comput. Methods Biomech. 55 (2007) 149–786.

[22] M. Nakamura, H. Wojcik, P. Okonkwo, G. Moreau, Flow rate effects on outlet velocity in computational models, J. Biomech. Eng. 40 (2019) 177–900.

[23] K. Eriksen, L. Jankowski, G. Yilmaz, Material model effects on shear stress in computational models, Comput. Methods Biomech. 59 (2012) 312–580.

[24] G. Ishikawa, L. Halvorsen, A. Silva, P. Lindqvist, Material model effects on head displacement in computational models, Med. Eng. Phys. 70 (2018) 36–877.

[25] F. Tanaka, H. Ishikawa, N. Abbott, P. Vasquez, H. Grigoryan, Element type effects on peak force in computational models, J. Verif. Valid. Uncertain. 125 (2014) 396–782.

[26] J. Quintero, H. Rasmussen, H. Castellano, K. Abbott, N. Kaur, Boundary condition effects on hemolysis index in computational models, Comput. Methods Biomech. 139 (2010) 252–733.

[27] L. Eriksen, P. Fontaine, H. Rasmussen, A. Duarte, Damping effects on shear stress in computational models, Comput. Methods Biomech. 146 (2005) 333–536.

[28] D. Nakamura, G. Okonkwo, T. Castellano, L. Jankowski, Time step effects on rib deflection in computational models, Comput. Methods Biomech. 111 (2012) 5–818.

[29] S. Lindqvist, M. Tanaka, W. Jankowski, K. Abbott, Cement mantle effects on cortical strain in computational models, J. Mech. Behav. Biomed. Mater. 74 (2006) 135–801.

[30] S. Tanaka, F. Jankowski, H. Kaur, G. Kaur, J. Lindqvist, Yield stress effects on outlet velocity in computational models, Int. J. Numer. Methods 119 (2008) 349–630.

[31] B. Okonkwo, W. Silva, L. Ueda, J. Abbott, Bone density effects on chest compression in computational models, J. Verif. Valid. Uncertain. 93 (2024) 284–750.

[32] C. Yilmaz, L. Fontaine, E. Quintero, D. Eriksen, Load path effects on stress range in computational models, J. Biomech. Eng. 97 (2018) 354–746.

[33] P. Silva, J. Nakamura, A. Halvorsen, C. Zetterberg, Yield stress effects on shear stress in computational models, Comput. Fluids 85 (2020) 219–787.

[34] E. Zetterberg, R. Pettersson, C. Ueda, R. Jankowski, P. Lindqvist, Boundary condition effects on hemolysis index in computational models, Ann. Biomed. Eng. 38 (2019) 109–794.

Table 1: Credibility factor scoring for DR-CMM across mechanisms using the 0–8 CRL scale.

Credibility factor	Level	Basis
Development process and product management - DR-CMM - Sustained axial loading–induced creep	6	Three formal reviews, 164 commits, 32 tickets, and a requirements trace to QoIs support mid–high lifecycle discipline.
Development process and product management - DR-CMM - Bending fatigue–driven crack initiation at defects	6	Gated plan and independent peer review of subroutines applied uniformly to the fatigue workflow.
Development process and product management - DR-CMM - Torsional overload–assisted interfacial debond and radial cracking	6	Change control and hazard logs cover overload cases; decision records justify omission of thermal coupling.
Model form - DR-CMM - Sustained axial loading–induced creep	6	Burgers viscoelasticity with Arrhenius shift reproduces 24 h and 7-day trends (9.3% mean settlement error).
Model form - DR-CMM - Bending fatigue–driven crack initiation at defects	7	Defect-resolved strain energy density threshold reproduces S–N slope within 0.08 and 21% MAPE on cycles.
Model form - DR-CMM - Torsional overload–assisted interfacial debond and radial cracking	5	Cohesive mixed-mode interfaces capture onset (1.7 N·m mean absolute error) but neglect frictional heating.
Model inputs - DR-CMM - Sustained axial loading–induced creep	6	DMA-based viscoelastic parameters (n=54) with 95% CIs and activation energy 49 ± 7 kJ/mol inform creep.
Model inputs - DR-CMM - Bending fatigue–driven crack initiation at defects	6	Micro-CT defect stats (3.2 ± 1.1% porosity; median 120 μm) and notch geometry calibrate initiation thresholds.
Model inputs - DR-CMM - Torsional overload–assisted interfacial debond and radial cracking	6	Interface strengths 26 ± 5 MPa (stem–cement) and 18 ± 4 MPa (cement–bone), Gc = 0.45 ± 0.08 N/mm used.
Numerical code verification - DR-CMM - Sustained axial loading–induced creep	7	Manufactured solution L2 error <1e3; observed second-order convergence; 24 h creep strain converged within 2%.
Numerical code verification - DR-CMM - Bending fatigue–driven crack initiation at defects	7	Cohesive patch tests within 0.7% of analytic energies; mesh h-refinement changed cycles by <3% at last two levels.
Numerical code verification - DR-CMM - Torsional overload–assisted interfacial debond and radial cracking	7	Torque–angle response stable to solver tolerance tightening (<1% QoI change) and time-step halving.
Output comparison	6	Local DIC strain 8.1%

Table 2: Credibility factor scoring for PS-CMAM across mechanisms using the 0–8 CRL scale.

Credibility factor	Level	Basis
Development process and product management - PS-CMAM - Sustained axial loading–induced creep	6	Same gated lifecycle and traceability used for DR-CMM; whole-construct workflows under the same controls.
Development process and product management - PS-CMAM - Bending fatigue–driven crack initiation at defects	6	Planning, change tracking, and independent review applied; damage parameter calibration documented.
Development process and product management - PS-CMAM - Torsional overload–assisted interfacial debond and radial cracking	6	Hazard analysis and decision records cover cohesive–friction transition modeling for torsion.
Model form - PS-CMAM - Sustained axial loading–induced creep	5	Homogenized viscoelastic law reproduces settlement trends (11.8% median error) without explicit pores.
Model form - PS-CMAM - Bending fatigue–driven crack initiation at defects	4	Isotropic damage risk metric ranks specimens (AUC 0.74) but cannot localize crack nucleation sites.
Model form - PS-CMAM - Torsional overload–assisted interfacial debond and radial cracking	5	Cohesive-to-friction contact captures onset metric (r=0.62 with torque threshold) but lacks thermal effects.
Model inputs - PS-CMAM - Sustained axial loading–induced creep	5	Viscoelastic spectrum from DMA (n=54) and activation energy 49 ± 7 kJ/mol; construct loads from literature.
Model inputs - PS-CMAM - Bending fatigue–driven crack initiation at defects	5	Damage threshold fitted to DR-CMM and bending tests; lacks specimen-specific pore inputs by design.
Model inputs - PS-CMAM - Torsional overload–assisted interfacial debond and radial cracking	5	Cohesive and friction parameters from coupon tests; onset metric calibrated on 12 constructs.
Numerical code verification - PS-CMAM - Sustained axial loading–induced creep	7	Manufactured solution L2 error <1e3; second-order convergence; settlement stable to time-step halving (<0.5%).
Numerical code verification - PS-CMAM - Bending fatigue–driven crack initiation at defects	7	Mesh convergence in proximal mantle changed risk metric distributions by <4% across last two meshes.
Numerical code verification - PS-CMAM - Torsional overload–assisted interfacial debond	7	Torque–angle curves insensitive to solver tolerance tightening (<1% QoI change) and contact regularization.