Credibility assessment of implicit finite element models for hip resurfacing: load transfer and stress shielding under ASME V&V 40

Elena P. Marquez, PhD[a], Noah J. Patel, MS[b], Rina S. Kawamura, PhD[c], George M. Leone, MD[1], Hannah L. Briggs, PhD[1]

[a]*Department of Mechanical Engineering, Center for Orthopaedic Simulation, Mid-Atlantic University, Baltimore, MD, USA,*
[b]*Department of Biomedical Engineering, Pacific Tech, San Diego, CA, USA,*
[c]*Department of Orthopaedic Surgery, Harborview Medical Center, Seattle, WA, USA,*

Abstract

Objective: To quantify credibility for two implicit finite element (FE) models of a cemented femoral head resurfacing implant with porous backing under ASME V&V 40, focused on (i) femoral neck load transfer during level walking and (ii) proximal femoral stress shielding during stair ascent. Methods: Model A comprises a patient-average femur with CT-mapped heterogeneous elastic properties, explicit cement mantle, and a detailed implant–cement construct, solving quasi-static contact with cohesive interfaces. Model B employs a geometry-calibrated composite femur with homogenized cortical/trabecular shells and simplified interfaces. A benchtop apparatus applied physiologic hip contact and muscle surrogates to implanted composite femurs. Surface strain gauges and stereo digital image correlation (DIC) provided measurements under level walking (peak 2.6 body weight, BW) and stair ascent (peak 2.9 BW). Solver verification used h-refinement, contact penalty tuning, and nonlinear residual controls to bound numerical effects. Model outputs were compared to measured calcar and superior-neck strains, and to DIC-derived regional strain distributions; change in proximal strain energy density (SED) served as a surrogate for stress shielding. Credibility was graded per model–mechanism pair across six factors on a 0–12 Assurance Index (0–3 low; 4–8 moderate; 9–12 high). Results: For level walking, Model A predicted calcar compressive strain of 1350 microstrain versus 1460 ± 160 microstrain measured (7.5 percent difference) and superior-neck tensile strain of +980 microstrain versus +1090 ± 140 microstrain (10.1 percent). For stair ascent, Model A predicted a 22 percent reduction in proximal SED (regions analogous to Gruen 1–2) versus 19 ± 7 percent measured by DIC. Model B showed larger discrepancies: calcar 1210 microstrain (17 percent difference), superior neck +840 microstrain (23 percent), and 30 percent proximal SED reduction versus 19 ± 7 percent. Mesh- and solver-induced changes in peak von Mises stress dropped below 2.0 percent for both models after refinement; contact traction changes were less than 3.0 percent. Validation test geometry, implant seating, cement mantle thickness (2.0 ± 0.3 mm), and load vectors matched the simulations within pre-specified tolerances, and telemetric hip data guided load magnitudes and directions. Credibility scores (Assurance Index) for Model A were predominantly high for walking (10–11 for numerical effects, 9–10 for structural idealizations, 9 for comparisons, and 9 for setup fidelity), and high-to-moderate for stair ascent (8–10). Model B achieved moderate grades for structural idealizations (6–7) and comparisons (6–7), and high for numerical checks (9–10). Conclusions: For assessing neck load transfer during walking and proximal stress changes during stair ascent, the detailed CT-mapped model demonstrates high credibility for the intended decision context, while the surrogate model supports trend-level inferences with reduced assurance. The validation activities mirrored the simulated device class and loading environment closely enough to inform the decision, and the selected metrics map directly to the failure risks of interest.

Keywords: hip resurfacing, finite element verification and validation, stress shielding

1. Introduction

Resurfacing arthroplasty of the femoral head remains a compelling option for selected patients, promising bone conservation, restoration of proximal femoral mechanics, and facilitated revision pathways compared to stemmed total hip replacements. However, outcomes depend critically on how the resurfacing construct redistributes loads across the femoral neck and proximal metaphysis. Elevated stresses or abnormal stress trajectories can predispose to early neck fractures, while reduced proximal mechanical stimulus can prompt adaptive bone loss that undermines fixation. Computational mechanics models are increasingly used to inform design, preclinical evaluation, and regulatory submissions, yet their role in decision-making must be supported by a transparent and proportional credibility argument.

The ASME V&V 40 framework offers a structured approach to assess whether model credibility is adequate for a given context of use, with attention to the decision risk and to targeted evaluation of model capabilities. Applying this standard in orthopaedics requires explicit mapping from clinical failure mechanisms to quantities that simulations can compute with sufficient reliability, and a disciplined linkage between bench testing, numerical checks, and the anticipated use of model results. The study reported here evaluates the credibility of two implicit finite element (FE) models of a cemented, porous-backed femoral head resurfacing implant against two mechanisms: (i) neck load transfer during level walking and (ii) proximal femoral stress shielding during stair ascent.

Model A is a high-fidelity construct coupling a CT-mapped, heterogeneous femur to an explicit cement mantle and a cobalt-chromium cap with porous backing. It computes quasi-static stress and strain fields, interface tractions, and compartmental load transfer between implant–cement and native bone during physiologic activities. Model B is a reduced-order, geometry-calibrated surrogate with homogenized cortical and trabecular shells and tied interfaces sufficient to predict bulk load sharing and strain distributions at lower computational cost. Both models are assessed against a benchtop experiment that reproduces physiologic hip contact and dominant muscle surrogates while instrumenting the proximal femur with strain gauges and stereo digital image correlation (DIC).

In pursuit of proportionality under ASME V&V 40, the analysis targets factors most determinative of the intended decision: whether predicted neck strain patterns under level walking remain within a clinically acceptable band relative to intact femurae, and whether the change in proximal mechanical stimulus under stair ascent falls below thresholds associated with bone resorption. The evaluation employs a 0–12 Assurance Index with category boundaries at 0–3, 4–8, and 9–12 for low, moderate, and high assurance, respectively, applied to each factor for each model–mechanism pair.

The remainder of the paper details the device and question of interest, articulates the context of use and decision risks, defines the mechanisms and their mapping to measurable outputs, describes the computational models and numerical verification, presents the benchtop methods and measurement protocols, explains the comparison strategy, and consolidates scores with concise justifications. We then interpret the implications for design and preclinical evaluation of resurfacing systems, discuss limitations, and identify priorities for further credibility strengthening.

To anchor the decision-making linkage, we note at the outset that both selected activities and load vectors were drawn from in vivo telemetry in resurfacing patients, with peak hip contact forces of 2.6 body weight (BW) during level walking and 2.9 BW during stair ascent applied in orientations consistent with instrumented-implant datasets. This alignment to patient-measured loading reduces uncertainty in the translation from bench and model to anticipated clinical use.

Relevance of quantities monitored to clinical decision — gradation.

a.
Qualitative narrative maps outputs to clinical concerns with no intermediate surrogate or threshold.

b.
Outputs correspond to recognized surrogates or regions of interest; thresholds derived from literature, not tailored to device class.

c.
Outputs and thresholds are device-class specific; mapping to failure modes is documented and reviewed.

d.
Outputs are directly tied to risk-based acceptance criteria quantified for the device; mapping is endorsed by multiple independent sources.

2. Device description and question of interest

The device considered is a cemented femoral head resurfacing implant consisting of a cobalt-chromium-molybdenum (Co-Cr-Mo) hemispherical cap with a porous titanium backing and a short, central peg. The cap diameter spans 44–54 mm, with a 1.5 mm average porous layer and an internal geometry contoured to accommodate a uniform polymethylmethacrylate (PMMA) cement mantle. The cement mantle target thickness is 2.0 mm circumferentially, with allowances between 1.5–2.5 mm due to surface roughness, porosity, and seating variability. The central peg engages a prepared femoral head pilot hole and provides rotational stability; cement occupies the annulus between peg and host bone.

The porous backing of the cap is intended to distribute interfacial stresses and to modulate the stiffness of the construct, making the combined implant–cement assembly less stiff than a solid Co-Cr cap and therefore attenuating stress perturbations to the femoral neck. However, the overall stiffness of the resurfaced unit remains higher than that of native cancellous bone and cartilage, and the mechanical pathway for load transmission shifts from articular cartilage and subchondral bone to the stiffer cap–cement composite. This shift can reduce load carried by the proximal femoral cortex and metaphyseal trabeculae, especially during activities demanding higher compressive and bending loads, such as stair ascent.

The central question is whether the resurfacing implant, when cemented with a porous backing into a femoral head with standard alignment, preserves physiologic neck load paths under level walking and avoids undue reduction of proximal femoral mechanical stimulus under stair ascent. The model-based evidence will be used to support design confirmation and to inform preclinical performance claims concerning mechanical compatibility with the proximal femur. Specifically, decision-makers require confidence that (i) calcar compressive and superior neck tensile strains under level walking remain within a clinically acceptable band relative to intact conditions, and (ii) reductions in proximal strain energy density under stair ascent do not exceed ranges associated with bone resorption in analogous regions. For clarity in decision linkage, we identified two scalar quantities to evaluate against acceptance targets established before the study:

1. Neck strain indices during level walking: surface strain at the calcar (inferomedial neck) and superior neck (superolateral cortex) measured in microstrain, with acceptance defined as the resurfaced-to-intact ratio lying between 0.75 and 1.25 for each location or absolute differences within 200 microstrain.

2. Proximal stimulus index during stair ascent: relative change in strain energy density (SED) averaged over regions analogous to Gruen zones 1–2 on the femoral side, with acceptance defined as a mean reduction below 25 percent and no region exceeding 35 percent reduction.

These thresholds were informed by published associations between local mechanical stimulus and bone maintenance in the proximal femur and by a device-focused risk assessment. For the failure mode and effects analysis (FMEA) underpinning design control, two high-priority risks were linked to these metrics: femoral neck fracture during early rehabilitation (probability weight 0.12, severity weight 0.3) and proximal bone resorption leading to aseptic loosening (probability weight 0.09, severity weight 0.25). This mapping ensures that the computed outputs directly address the device's highest-risk mechanical concerns.

In light of this linkage, we emphasize a finding central to this study: the metrics we selected for evaluation map directly to the device's dominant mechanical risks, and the numeric thresholds used to judge acceptability derive from the same risk analysis and literature basis that governed the design inputs. This direct tie to decision criteria avoids reliance on tangential outputs and concentrates the assessment on measures that the device team and reviewers have already agreed are determinative.

Quantities of interest selection — gradation.

a.
Outputs are selected post hoc from what the model already computes.

b.
Outputs are pre-specified and include at least one surrogate with literature support.

c.
Outputs are pre-specified, device-class specific, and traceable to risk analysis with acceptance bands.

d.
Outputs are pre-specified, device-specific, risk-weighted, with acceptance thresholds vetted by independent experts.

3. Context of use and model risk rationale

The intended use of the computational evidence is to inform design confirmation of a cemented, porous-backed femoral head resurfacing implant by quantifying load transfer through the femoral neck during level walking and assessing proximal femoral stress changes during stair ascent. The users are engineers responsible for device design and preclinical performance evaluation, and clinicians advising on fixation strategy. In this context, model outputs will be weighed alongside benchtop measurements, material and geometric tolerances, and clinical literature to support claims of mechanical compatibility with the proximal femur. The decision is consequential but not determinative in isolation; the model does not directly drive clinical clearance or replace safety-critical testing, but it does influence the confidence in design adequacy and in the focus of subsequent bench or animal testing.

Risk-informed proportionality under ASME V&V 40 indicates moderate-to-high consequence of error for the walking mechanism: underestimation of neck tensile/compressive strain gradients could obscure elevated fracture risk. For stair ascent, overestimation of proximal stress changes could lead to conservative interpretations, while underestimation could mask resorption risk. The consequence of model error is mitigated by the reliance on independent experimental measurements, by conservative acceptance thresholds, and by the complementary nature of the two models—a high-fidelity heterogeneous model supporting detailed assessment and a surrogate model providing cross-checks and sensitivity insight.

We align the model scope to the decision: only early postoperative activities (level walking and stair ascent) are considered, reflecting the fixation state of a cemented resurfacing construct prior to substantial biological adaptation. Loading vectors and magnitudes derive from instrumented-implant telemetry in resurfacing cohorts, ensuring physiological plausibility. Boundary conditions represent a potted distal femur with constraints consistent across bench and model, avoiding artefactual compliance. The femur geometry reflects a patient-average morphology, and the implant corresponds to available sizes and seating strategies, embedding realistic tolerances for the cement mantle and peg engagement.

To situate the agreement between models and measurements with respect to decision needs, we provide a consolidated statement: the differences observed between model predictions and measured neck strains remained within 8–12 percent at both calcar and superior neck during level walking, and the predicted reduction in proximal mechanical stimulus during stair ascent fell within the measured range (19 ± 7 percent measured versus 22 percent predicted for the detailed model). These outcomes, taken together with correct directional trends and region-wise agreement, indicate that the level of concordance achieved is adequate for the planned decision use, which centers on conformance to predefined acceptance bands rather than on exact reproduction of every local strain feature.

Finally, we note the role of numerical checks and parametric sensitivity—especially for contact parameters and bone material mapping—in bounding the contribution of numerical choices to output uncertainty. Where sensitivity exceeded predefined tolerances, we either refined meshes or justified parameter selections by independent sources. This approach reduced the chance that favorable agreement would be illusory or non-robust to plausible perturbations.

Comparison adequacy for decision — gradation.

a.
Visual or qualitative agreement only; no quantitative targets or tolerances.

b.
Quantitative comparisons at a few locations; tolerances defined post hoc.

c.
Predefined quantitative tolerances with multiple measurement locations or fields; independent uncertainty estimates included.

d.
Predefined quantitative tolerances with coverage of fields, uncertainty propagation, and decisions based on acceptance bands.

4. Mechanisms and quantities of interest

Two mechanisms of interest were identified through design risk analysis and literature mapping of resurfacing failure modes to proximal femoral mechanics. First, femoral neck load transfer during level walking governs the distribution of compressive and tensile strains across the calcar and superolateral neck. Abnormal transfer can increase calcar compression while elevating superolateral tension or vice versa, potentially driving neck fracture when combined with notching, osteopenia, or malalignment. Second, proximal femoral stress shielding during stair ascent reflects reduced mechanical stimulus in bone segments adjacent to the implant–cement construct, potentially triggering remodeling and resorption in those regions. Stair ascent was selected as a high-demand daily activity that accentuates joint reaction and muscle forces while producing characteristic bending moments on the neck and proximal shaft.

For neck load transfer under level walking, the quantities evaluated were uniaxial surface strains at two locations instrumented on the bench: the calcar (inferomedial neck) and the superior neck (superolateral cortex). The selection of these locations follows prior fracture case analyses and instrumentation studies that have associated these regions with peak compressive and tensile stresses, respectively, during the gait cycle. We compute model-based equivalents by sampling the FE surface strains at the centerpoints of the rosettes and by averaging principal strain in the gauge's dominant direction over the gauge footprint to mirror the measurement process.

For proximal stress shielding under stair ascent, the quantity evaluated was the percent change in strain energy density (SED) averaged over three proximal regions adjacent to the resurfacing cap—approximating Gruen zones 1–2 on the femoral side—relative to an intact (pre-implantation) baseline under the same loading path. Although SED is not measured directly on the bench, stereo DIC provides a high-resolution map of surface strains from which relative changes can be inferred; with assumed plane-stress and constitutive parameters, these inform the SED change at the surface, serving as a surrogate for the near-surface subcortex. In addition, modeled SED fields allow calculation of region-averaged reductions throughout the cortical thickness where DIC cannot probe.

Loads for these activities were derived from instrumented-implant telemetry in resurfacing patients and applied as hip contact and dominant muscle surrogates. For level walking, the peak joint reaction of 2.6 BW was applied near mid-stance, with a direction approximating 16 degrees lateral-medial inclination to the femoral shaft axis and 15 degrees anterior tilt in the sagittal plane. Muscle surrogates included abductors summing to 1.8 BW, with hamstrings approximately 0.2 BW and vasti 0.4 BW, aligned to reproduce known moment arms and resultant bending moments at the neck. For stair ascent, peak joint reaction of 2.9 BW was applied at early stance with an orientation of 7 degrees lateral-medial inclination and 25 degrees anterior tilt, with abductors summing to 2.2 BW and hip flexors adding 0.5 BW.

We emphasize that the loading magnitudes and directions were explicitly selected to reflect telemetry-based ranges reported for resurfacing patients, with joint reaction magnitudes of 2.5–2.8 BW for level walking and 2.8–3.2 BW for stair ascent, and with vector orientations within 5–10 degrees of those used here. Adopting these values ensured that both bench and model probed the same physiologic envelope as the clinical activities of interest.

Test condition selection — gradation.

a. Loads or boundary conditions are generic or chosen for convenience.

b. Loads and boundary conditions are drawn from literature but not device class specific.

c. Loads and boundary conditions are device class specific and traceable to measurement data.

d. Loads and boundary conditions are device specific, directly measured in target population, and cover expected variability.

5. Computational models and model form justification

Model A: A patient-average femur geometry was constructed by registering a high-resolution femoral atlas to a de-identified CT dataset representative of the resurfacing population (male, 55 years, T-score 0.5). The CT had 0.6 mm isotropic voxels. The periosteal and endosteal contours were segmented using semi-automatic thresholding and active contour refinement. A volumetric finite element mesh of 10-node tetrahedra (TET10) was generated with curvature-adaptive refinement: minimum edge length 0.7 mm in the head–neck region and 1.8 mm elsewhere, yielding approximately 1.9 million elements and 2.5 million nodes. A separate, more refined mesh (minimum 0.5 mm in the head–neck) was created for solver verification.

Heterogeneous bone material properties were mapped from CT Hounsfield units (HU) to apparent density and elastic modulus. We used a bilinear relationship validated for femoral bone: apparent density rho (g/cc) = 0.001 x HU for HU > 0 and zero otherwise; Young's modulus E (MPa) = 10,200 x rho^2.57 for trabecular regions and 17,000 x rho^2.0 for cortical regions identified by a cortex mask, with Poisson's ratio 0.30 (cortex) and 0.20 (trabeculae). A cap material (Co-Cr-Mo) of E = 210 GPa and nu = 0.30 was modeled, with a porous titanium backing assigned an effective modulus of 15 GPa based on homogenization of a 60 percent porosity lattice. The cement mantle (PMMA) was modeled explicitly with E = 2.3 GPa and nu = 0.35, with thickness prescribed at 2.0 mm nominally and allowed to vary between 1.5–2.5 mm depending on the femoral head surface and cap internal contour. The central peg was included as a 10 mm diameter, 12 mm length feature with a 0.5 mm cement annulus.

Interfaces were treated to represent both adhesion and potential micro-slip. The cap–cement interface was assumed perfectly bonded, consistent with inner porous layer intentionally roughened to interdigitate with cement; bonded contact was enforced via shared nodes in that interface. The bone–cement interface was represented by a cohesive zone with penalty stiffnesses K_n = 2.0 GPa/mm normal and K_t = 0.8 GPa/mm tangential, damage initiation at 2.5 MPa nominal shear and 6.0

MPa nominal normal traction, and exponential softening with fracture energies 0.25 N/mm (shear) and 0.40 N/mm (normal). These values lie within reported ranges for cement–bone interlocks. Contact with a small friction coefficient (mu = 0.2) was also allowed in the tangential direction once decohesion initiated. The distal femur was potted in a 60 mm-high PMMA block, modeled with E = 2.4 GPa and bonded to the cortical shell over the potted length.

Boundary and loading conditions mirrored those used on the bench. Hip contact was applied as a distributed pressure over a 30 mm spherical cap region centered at the femoral head's anatomical center, with direction and magnitude corresponding to level walking and stair ascent peaks. Muscle force surrogates were applied via cylindrical bushings bonded to anatomical landmarks (greater trochanter and lateral femoral shaft) and oriented to reproduce abductors, vasti, and hamstring/flexor resultants, with forces calibrated to achieve target joint reaction and bending moments. The distal pot was fully fixed at its bottom surface, with no constraint applied elsewhere. Quasi-static analyses were conducted by incrementally ramping the hip contact and muscle forces to peak magnitudes.

Model B: A geometry-calibrated composite femur was created by morphing a fourth-generation composite femur geometry to match the patient-average femur's neck-shaft angle (131 degrees), femoral head diameter (50 mm), and cortical thickness distribution. The volumetric mesh comprised approximately 300,000 TET10 elements with minimum edge length 1.5 mm in the head–neck region. The cortex was modeled as an orthotropic homogeneous shell of 2.5 mm mean thickness with longitudinal E = 17 GPa, circumferential E = 12 GPa, radial E = 8 GPa, and Poisson's ratios of 0.33, calibrated to composite femur coupons. The trabecular region was represented as a homogeneous isotropic core with E = 0.6 GPa, nu = 0.20. The porous-backed cap and PMMA cement were represented as a composite layer with an effective modulus of 8.0 GPa in compression and shear, bonded to the prepared femoral head via a tied interface. The peg was represented as a rigid connector transmitting axial force and bending but no torsion, to reproduce the simplified behavior of a central stabilizer.

Boundary and loading conditions matched Model A to isolate the effects of material homogenization and interface simplification. Both models were solved using an implicit Newton–Raphson algorithm with line search and augmented Lagrangian contact, implemented in Abaqus/Standard 2022 (Dassault Systemes) with double-precision arithmetic.

As part of establishing the adequacy of structural idealizations for the intended decision, we observe that the explicit representation of the cement mantle and cohesive bone–cement behavior in Model A, together with CT-based heterogeneity and porous backing stiffness, reproduces the key mechanical discontinuities and stiffness transitions expected to govern neck load flow and proximal stimulus; in contrast, the composite-layer representation in Model B accelerates solution time but blurs stress gradients at the cement–bone interface, reducing local fidelity while preserving bulk bending response.

Model form fidelity — gradation.

Table 1: Principal model parameters for Model A and Model B.

Credibility factor	
Femur geometry and mesh (Model A)	CT-mapped patient-average femur; 1.9 millio
Femur geometry and mesh (Model B)	Composite-morphed geometry;
Bone material mapping (Model A)	rho = 0.001 x HU; E_trab = 10,200 x rho^2.57
Bone material (Model B)	Cortex orthotropic: E_long = 17 GPa, E_circ = 12
Cap and porous backing	Co-Cr-Mo cap E = 210 GPa,
Cement mantle	PMMA E = 2.3 GPa, nu = 0.
Bone–cement interface (Model A)	Cohesive: K_n = 2.0 GPa/mm; K_t = 0.8 GPa/mm; str
Cap–cement interface	
Loads – level walking	Hip contact 2.6 BW at 16 deg lateral-medial, 15
Loads – stair ascent	Hip contact 2.9 BW at 7 deg lateral-me
Boundary	Distal 60 mm
Solver	Implicit Newton–Raphson with lin

a.
Geometry and materials are generic; interfaces are idealized as perfectly rigid or perfectly bonded without justification.

b.
Geometry reflects device and anatomy; materials are homogeneous; interfaces are tied with limited sensitivity analysis.

c.
Geometry reflects device and anatomy; materials include heterogeneity or orthotropy; interfaces include calibrated stiffness or cohesive behavior.

d.
Full device–anatomy representation with validated heterogeneity and interface laws, with uncertainty bounds tied to decision.

6. Numerical solver error characterization for implicit analyses

Solution verification focused on mesh adequacy, contact parameter tuning, and nonlinear convergence criteria. For each model and mechanism, we performed an h-refinement study, a contact penalty sensitivity analysis, and a residual tolerance sweep. The goal was to ensure that numerical artifacts contributed less uncertainty than the experimental variability, and materially less than the acceptance margins for the decision metrics.

Mesh refinement: For Model A, a base mesh with 0.7 mm minimum edge length in the head–neck region and a refined mesh with 0.5 mm minimum edge were compared. For level walking, the change in peak von Mises stress in the superior neck between base and refined meshes was 1.4 percent, the change in calcar compressive strain at the gauge footprint was 1.8 percent, and the change in interface traction peaks at the bone–cement boundary was 2.7 percent. For stair ascent, the corresponding changes were 1.9 percent (peak von Mises), 2.2 percent (calcar strain), and 2.9 percent (interface traction). For Model B, comparing minimum edge lengths of 1.5 mm and 1.0 mm in the head–neck region yielded changes of 1.2 percent (peak von Mises) and 1.6 percent (calcar/superior strains) for walking, and 1.6 percent and 2.0 percent for stair ascent.

Contact penalty sensitivity: For Model A, the augmented Lagrangian normal penalty factor varied between 5 x 10^3 and 2 x 10^4 N/mm and tangential penalty between 1 x 10^3 and 5 x 10^3 N/mm. Within this band, the change in region-averaged SED under stair ascent was less than 1.5 percent and the change in peak contact traction less than 3.0 percent. The cohesive stiffnesses for the bone–cement interface were varied ±25 percent; the change in calcar strain at the gauge footprint was 2.1 percent (walking) and the change in proximal SED reduction was 1.8 percent (stair ascent). For Model B (tied interfaces), a notional penalty stiffness was used to avoid ill-conditioning; varying it by ±50 percent altered peak strains by less than 1.0 percent.

Nonlinear convergence: Convergence tolerances for force residual were tightened from 10^6 to 10^8 times reference force, and displacement increments were allowed to reduce to 10^8 times characteristic length. Under these settings, the change in calcar strain for Model A was 0.6 percent (walking) and 0.9 percent (stair ascent); for Model B, the changes were 0.4 percent and 0.6 percent. Contact stabilization energy was monitored and remained below 0.4 percent of total strain energy across all runs.

Together, these checks support that discretization and algorithmic choices contribute a small fraction of overall output variability and fall below the experimental noise levels observed on the bench. We also note that the patterns of strain and SED changes were stable across these variations, with iso-strain contours shifting by less than 1.0 mm in the head–neck region. This stability under procedural refinements indicates that the numerical backbone of both models is sufficiently robust for the intended comparisons.

Numerical solution verification — gradation.

a.
No mesh or solver sensitivity studies are performed.

b.
Single-parameter sensitivity for mesh density or tolerance with qualitative assessment.

c.
Systematic h-refinement and solver tolerance studies with quantitative convergence metrics.

d.
Systematic studies with quantified numerical error bounds propagated to decision metrics.

7. Experimental methods and test conditions

Specimens: Ten fourth-generation composite femurs (Pacific Research Labs, medium left) were prepared by an experienced surgeon to receive a porous-backed cemented resurfacing cap sized to 50 mm diameter, central peg 10 mm diameter by 12 mm length. Femoral neck-shaft angle averaged 131 ± 1.5 degrees as measured post-preparation. A PMMA cement mantle was targeted at 2.0 mm thickness, with intraoperative technique used to ensure mantle continuity and venting. The range of mantle thickness measured by CT post-implantation was 2.0 ± 0.3 mm circumferentially, with a minimum local thickness of 1.5 mm and a maximum of 2.6 mm. The central peg cement annulus averaged 0.54 ± 0.06 mm.

Instrumentation: Two 350-ohm uniaxial strain gauges (Vishay Micro-Measurements) were bonded to the femoral neck surface with M-Bond 200 adhesive: one at the calcar (inferomedial neck) oriented to capture principal compressive strain under walking, and one at the superior neck (superolateral cortex) oriented to capture principal tensile strain. Gauge length was 3 mm; locations were recorded relative to anatomical landmarks using 3D digitization. Stereo DIC (GOM Aramis) was used to obtain full-field surface strains on the lateral and anterior aspects of the proximal femur. A matte speckle pattern was applied; subset size was 19 pixels, step size 12 pixels, with calibration yielding 14 micrometer in-plane accuracy over a 120 mm field of view.

Fixture and loads: The distal femur was potted in a 60 mm-height PMMA block. The pot's bottom face was bolted to a six-axis load cell that recorded ground-truth support reactions. A femoral head load applicator delivered the hip contact via a spherical indenter with a 30 mm radius contacting the cap surface at the anatomical center. Muscle surrogates were implemented using cable-and-pulley systems aligned to reproduce abductors, vasti, and hamstrings/flexors, with individual load cells monitoring force magnitudes. For level walking, the target peak hip contact was 2.6 BW, with abductors summing to 1.8 BW, hamstrings to 0.2 BW, and vasti to 0.4 BW. The contact vector orientation relative to the femoral shaft axis was set at 16 degrees lateral-medial inclination and 15 degrees anterior tilt, verified by a 3D coordinate system. For stair ascent, the target peak hip contact was 2.9 BW, with abductors at 2.2 BW and hip flexors at 0.5 BW, oriented at 7 degrees lateral-medial inclination and 25 degrees anterior tilt. Forces were ramped quasi-statically over 10 seconds to peak to minimize rate effects.

Measurement protocol: Each specimen underwent three loading cycles for level walking and three for stair ascent. Strain gauge readings were sampled at 100 Hz and averaged over a 0.5 s plateau at peak load. DIC frames were synchronized to load cell readings; strains were extracted at the gauge footprint to cross-validate rosette readings and region-averaged over proximal windows approximating Gruen zones 1–2 for the SED surrogate computations. Calibration of DIC to SED surrogate was achieved by assuming plane stress and using local elastic properties (E_cortex = 17 GPa, nu = 0.33 for the composite cortex) to compute energy density from measured strains. The intact baseline was obtained prior to implantation for each specimen under the same loading cases, allowing computation of percent change post-implantation.

Data reduction and uncertainty: For strain gauges, calibration factors were verified against known bending of a standard bar, yielding ±1.5 percent accuracy. Placement repeatability contributed an estimated ±80 microstrain variability. For DIC, the field measurement uncertainty was ±60 microstrain RMSE over the regions of interest. The repeatability of load application was ±2.5 percent across cycles. Combined standard uncertainty (k=1) for gauge readings was ±160 microstrain, and for DIC-derived regional strain averages was ±95 microstrain. For SED change, uncertainty propagated from DIC strain and modulus assumptions yielded a ±7 percent uncertainty on the regional percent change.

Outcome measures: For level walking, the primary outcomes were calcar compressive strain and superior-neck tensile strain at peak load. For stair ascent, the primary outcomes were percent changes in region-averaged SED for proximal regions adjacent to the cap (three regions per femur), computed relative to intact. Secondary outcomes included DIC-derived field agreement metrics (RMSE and correlation coefficients) between model and measurement and qualitative assessment of iso-strain band locations. The benchtop tests were designed and executed before any model-to-test comparisons to avoid bias.

Test condition execution — gradation.

a.
Loads and boundary conditions are applied approximately with no verification.

b.
Loads and boundary conditions are applied with partial verification; specimen preparation not standardized.

c.
Loads and boundary conditions are applied with full verification; specimen preparation standardized; measurement uncertainty quantified.

d.
As previous plus traceability to primary clinical measurements and cross-laboratory reproducibility evidence.

8. Output comparison procedures

Comparison metrics and acceptance: For level walking, we compared model-predicted calcar compressive and superior-neck tensile strains to gauge measurements. Acceptance was defined a priori as resurfaced-to-intact strain ratios between 0.75 and 1.25 or absolute differences within 200 microstrain. We also computed the RMSE between DIC-derived surface strain fields and model surface strains over the proximal lateral and anterior regions to ensure that localized agreement was not achieved at the expense of broader field discrepancies. For stair ascent, we compared the percent decrease in SED over three proximal regions between model and DIC-derived surrogates, with acceptance if the mean reduction remained below 25 percent and no region exceeded 35 percent reduction.

Model-to-measurement registration: For both models, gauge footprints and DIC regions were projected onto the model surface using anatomical landmarks digitized both on specimens and the model. Local coordinate frames were aligned to the neck axis and shaft axis, with gauge orientations mapped accordingly. Model strains were averaged over the projected gauge footprint to mirror gauge spatial filtering. For DIC comparison, model surface nodes were used to sample the nearest-neighbor measured strains after rigid alignment.

Results – level walking: For Model A, the calcar compressive strain was 1350 microstrain compared to 1460 ± 160 microstrain measured across specimens (7.5 percent difference from the mean), and the superior-neck tensile strain was +980 microstrain compared to +1090 ± 140 microstrain (10.1 percent difference). DIC field RMSE between model and measurement over the lateral and anterior regions was 95 microstrain, with a correlation coefficient of 0.91 for principal strain directions. The resurfaced-to-intact strain ratio at the calcar was 0.89 (model) versus 0.86 ± 0.09 (measured), and at the superior neck 1.12 (model) versus 1.15 ± 0.11 (measured). For Model B, the calcar compressive strain was 1210 microstrain (17 percent difference), and the superior-neck tensile strain was +840 microstrain (23 percent difference). DIC field RMSE for Model B was 145 microstrain with a correlation coefficient of 0.84. The resurfaced-to-intact ratios were 0.80 (calcar) and 1.01 (superior neck), both within the absolute acceptance band but closer to the decision boundary.

Results – stair ascent: For Model A, the region-averaged SED reduction in proximal regions was 22 percent (zones analogous to Gruen 1–2), compared to 19 ± 7 percent measured by DIC-derived surrogates. No region exceeded a 28 percent reduction in the model, and no region exceeded 31 percent by measurement. For Model B, the region-averaged SED reduction was 30 percent, with one region at 36 percent, exceeding the acceptance ceiling; measured DIC-based reductions remained at 19 ± 7 percent. The model–measurement difference for Model A therefore falls within the propagated measurement uncertainty, while Model B overestimated the proximal stimulus decrease.

Compartmental load transfer: Model A computed the fraction of hip contact transmitted through the cap–cement composite versus through the native neck and proximal cortex during level walking. The computed split at peak load was 62 percent through the cap–cement construct and 38 percent through the native neck, with the construct share decreasing to 58 percent at lower load plateaus. Direct measurement of this split at the bench was not feasible; however, indirect indicators such as the bending moment at the neck inferred from distal support reactions and muscle force magnitudes were consistent with the model's predicted moment balance within 5 percent. Model B predicted a 70/30 split, reflecting stiffer cap–cement representation and the tied interface.

Uncertainty handling: Comparisons included both mean differences and the experimental standard deviations. For Model A, the 7.5–10.1 percent differences in gauge readings under walking fell below the combined uncertainty envelopes (±160 microstrain gauges, ±95 microstrain DIC RMSE). For stair ascent, the 22 percent predicted SED reduction matched 19 ± 7 percent measured within uncertainty. For Model B, differences approached or exceeded the uncertainty envelopes, aligning with expectations for a homogenized representation.

Comparison methodology rigor — gradation.

a.

[redacted]

Figure 1: Benchtop test setup. The distal femur is potted in PMMA and mounted to a six-axis load cell. Hip contact is delivered via a spherical indenter centered on the femoral head, with cable-and-pulley muscle surrogates aligned to replicate abductors, vasti, and hamstrings/flexors. Strain gauges are bonded at the calcar and superior neck; stereo DIC cameras view the lateral and anterior surfaces.

Pointwise comparisons with no registration or filtering consistency.

b.
Registration performed; filtering partially matched; uncertainty noted but not used.

c.
Registration and filtering aligned to measurement; uncertainty quantified and reported.

d.
As previous plus uncertainty propagated to acceptance decisions with sensitivity checks.

9. Relevance of the validation activities to the COU

The validation activities were planned to replicate the device configuration, surgical preparation, and loading environments of interest. Specimen preparation used the same resurfacing cap size range and cementation technique intended for clinical use. The cement mantle thickness distribution (2.0 ± 0.3 mm circumferentially) measured by CT on the benchtop specimens matched the variation encoded in the Model A mesh and fell within the expected surgical tolerance. The central peg dimensions and cement annulus mirrored the implant geometry, and the porous backing on the cap was represented in both bench and model as a stiffness-modulating layer.

Loading paths for level walking and stair ascent were selected to match in vivo telemetry in resurfacing patients (2.6 BW and 2.9 BW peaks, respectively), and muscle surrogates were arranged to produce bending moments at the neck consistent with published moment arms and measured joint contact directions. Boundary conditions were harmonized across bench and model: the distal pot length, potting material, and constraint at the pot base were matched within 1.0 mm and were documented to ensure reproducibility. Strain gauges were located at the same anatomical targets used to extract model strains; DIC fields were sampled over the same windows on model and specimen.

Device class applicability was prioritized: the porous-backed cemented cap and central peg geometries were the same as those simulated, and the femoral neck angle and head diameter of specimens were morphed or selected to reflect the patient-average femur in Model A. The stair ascent activity encompasses a load spectrum and vector orientation especially relevant to proximal stress changes, and the level walking activity captures the most frequent daily load with known influence on neck strain. The validation was therefore apposite to the intended decision, especially in how localized surface strains and region-averaged stimulus changes were probed where the model outputs are assessed.

As an integrative observation, the benchtop methods matched the simulated device, implant seating, cement mantle range, and loading vectors closely enough that the same mechanical pathways—cap–cement composite to bone, neck bending under abductors, and proximal cortical engagement—were exercised in both settings. This parallelism strengthens the argument that the comparisons performed are informative for the decisions at hand rather than for a tangential construct or load case.

Validation applicability to decision context — gradation.

a.
Validation is performed on a different device class or loading not relevant to the decision.

b.
Validation covers the device class but with substantial differences in configuration or loading.

c.
Validation replicates the device configuration and key loading features; measurements map to outputs.

d.
Validation mirrors the full decision environment including variability; direct one-to-one mapping to outputs.

10. Use error considerations

Practical use of these models introduces opportunities for misuse or misinterpretation, especially when extrapolating outside the examined range of implant sizes, mantle thicknesses, or alignment angles. Inputting malalignment beyond the 10-degree envelope studied here, or assuming bone quality outside the CT calibration domain, could alter predictions in ways not bounded by this assessment. Similarly, interpreting local stress values at mesh-scale hot spots as failure predictors can be misleading without context. We therefore emphasize disciplined use within the defined scope, and we caution that the models were not designed to evaluate extreme notching, severe varus positioning, or late postoperative remodeling phenomena.

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

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

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

User guidance and misuse mitigation — gradation.

a.
No guidance beyond code defaults is provided to users.

b.
Basic instructions are provided; limitations are qualitatively noted.

c.
Detailed instructions specify input domains, limitations, and checks for common mistakes.

d.
As previous plus automated safeguards and training modules to prevent misuse.

11. Numerical robustness and discretization adequacy (integrated finding)

Across both modeled activities, grid refinement, contact penalty adjustments, and residual tolerance tightening altered the principal outputs by less than 2 percent for peak von Mises stress, less than 2.2 percent for calcar gauge strain, and less than 3 percent for interface traction peaks; stabilization energies remained below 0.4 percent of total strain energy. These small shifts relative to experimental variability indicate that solution procedures and mesh resolution do not materially influence the conclusions for either mechanism.
Solver settings transparency — gradation.

a.
Solver settings are undocumented.

b.
Solver settings are documented but not justified.

c.
Solver settings are documented with sensitivity evidence.

d.
Solver settings are documented, justified by sensitivity, and tied to acceptance tolerances.

12. Alignment of measured and simulated decision metrics (integrated finding)

Agreement to within 8–12 percent for calcar and superior neck strains during level walking was achieved for the detailed heterogeneous model, and the predicted 22 percent reduction in proximal strain energy density during stair ascent fell within the measured 19 ± 7 percent band; the surrogate model reproduced trends but with larger deviations, particularly a 30 percent reduction in proximal stimulus. This level of concordance and correct directional response supports reliance on the detailed model for the intended design decision.
Strength of model–test agreement — gradation.

a.
Sparse metrics with notable discrepancies and no uncertainty.

b.
Multiple metrics agree within loose bounds; uncertainty partially quantified.

c.
Multiple metrics agree within predefined bounds; uncertainty quantified.

d.
Field-level agreement within bounds, uncertainty propagated to conclusions, and trends verified across conditions.

13. Decision linkage of outputs to clinical risks (integrated finding)

The two primary outputs—neck strains under level walking and proximal stimulus reduction under stair ascent—map directly to the highest-weighted risks in the device risk analysis (femoral neck fracture and proximal bone resorption), and thresholds for acceptability derive from those same risk models and published associations between mechanical stimulus and bone maintenance. This direct mapping ensures that the evaluation targets the quantities that matter most to the decision.
Traceability from outputs to risk — gradation.

a.
No traceability is established.

b.
Partial traceability with qualitative mapping.

c.
Full traceability with quantitative thresholds.

d.
Full traceability with thresholds and sensitivity to risk weights.

14. Experimental coverage of the modeled load environment (integrated finding)

The applied hip contact magnitudes (2.6 BW walking, 2.9 BW stair) and orientations (16 and 7 degrees lateral-medial; 15 and 25 degrees anterior tilt) fall squarely within telemetry-reported bands for resurfacing patients, and the muscle surrogate forces reproduced neck bending consistent with in vivo moment arms, establishing that the load cases exercised in both bench and model are representative of the clinical activities used to judge acceptability.
Representativeness of test loads — gradation.

a.
Loads are outside clinically relevant ranges.

b.
Loads are at edges of clinical ranges with limited justification.

c.
Loads are centered within clinical ranges and justified by data.

d.
Loads cover variability within clinical ranges and are stratified by condition.

15. Assurance ladder for model structure and interfaces

The structural representation of the cap–cement–bone assembly and the modeling of interfaces are pivotal to predicting localized strain gradients and proximal load sharing. A graded view of increasing rigor is helpful to situate the two models evaluated here relative to the decision demands.
Structural representation rigor — gradation.

a.
Implant and bone modeled as monolithic elastic blocks; no interfaces.

b.
Implant and bone geometries resolved; interfaces tied; materials homogeneous.

c.
Implant–cement–bone interfaces represented with stiffness or cohesive models; bone heterogeneity or orthotropy included.

d.
Interfaces validated against independent tests; bone heterogeneity validated; porosity modeled explicitly where it influences decision metrics.

16. Credibility factor scoring by model–mechanism pair (0–12 Assurance Index): Model A

The following table summarizes the graded assessment for Model A across the two mechanisms of interest. Each factor is scored on the 0–12 Assurance Index using evidence detailed in the preceding sections. The level column presents the score only; the basis restates a quantitative justification also presented in the prose.
Assurance index interpretation — gradation.

a.
0–3: Evidence does not support reliance for decision; major gaps or misalignment.

b.
4–8: Evidence supports trend-level or bounded use; moderate gaps remain.

c.
9–12: Evidence supports use for the intended decision; residual gaps do not alter conclusions.

d.
Use of the index is accompanied by sensitivity to residual uncertainties and explicit scope limits.

17. Credibility factor scoring by model–mechanism pair (0–12 Assurance Index): Model B

The following table presents the graded assessment for Model B across the same two mechanisms, using the 0–12 Assurance Index. The simplified structural representation yields different assurances compared to Model A, particularly for localized outputs.
Assurance index interpretation (surrogate model) — gradation.

a.
0–3: Not adequate for decision; use only for exploratory trends.

b.
4–8: Supports trend-level guidance with caution; check against high-fidelity model or data.

c.
9–12: Supports decision use within scoped outputs and conditions.

d.
Documented conditions for which surrogate maintains acceptability.

18. Discussion

This study applied ASME V&V 40 principles to assess the credibility of two finite element models of a cemented, porous-backed femoral head resurfacing implant for two clinically motivated mechanisms: femoral neck load transfer during level walking and proximal femoral stress shielding during stair ascent. By explicitly selecting outputs tied to prioritized risks—surface strains at the calcar and superior neck and proximal strain energy density changes—we anchored the modeling activity to the decision needs. A benchtop experiment with physiologically motivated contact and muscle surrogate loads provided measurements for quantitative comparison, while solver verification bounded the influence of discretization and algorithmic choices.

For neck load transfer under level walking, Model A achieved quantitative agreement at the calcar and superolateral neck within 7.5–10.1 percent and reproduced field-level strain patterns with a DIC RMSE of 95 microstrain. These results, coupled with small numerical sensitivities (<2.2 percent for gauge-level strains and <3.0 percent for interface tractions), support reliance on the model's estimates of neck strain trajectories and their ratio to intact values for the considered device and activity. The 62/38 percent split in load transmission through the cap–cement construct versus the native neck is physically plausible and consistent with the observed bending reaction and muscle surrogates; while direct measurement of this split was not feasible, surrogate measures corroborated the modeled moment balance within 5 percent.

For proximal stress changes under stair ascent, Model A predicted a 22 percent reduction in region-averaged SED in Gruen-like proximal regions, within the measured 19 ± 7 percent. No region exceeded a 28 percent reduction in the model or 31 percent in measurement, satisfying the acceptance criteria (mean <25 percent; max <35 percent). This suggests that the porous-backed, cemented cap modulates stiffness sufficiently to avoid severe proximal stress shielding under the tested condition. While DIC-based SED surrogates rely on assumptions (plane stress, elastic modulus), their uncertainty (±7 percent) and alignment with model predictions bolster confidence.

Model B, by construction, aims to provide a faster, trend-level assessment. It demonstrated reasonable alignment for neck load transfer (17–23 percent differences at the calcar and superior neck) but approached acceptance limits and exhibited higher DIC RMSE (145 microstrain). For proximal SED reductions, the simplified cap–cement representation and tied interface produced a 30 percent region-averaged reduction, with one region at 36 percent exceeding the ceiling. This overestimation was anticipated, given the stiffening effect of the composite cap–cement layer and the absence of compliance at the bone–cement interface. Model B therefore functions as a conservative predictor for proximal stress changes and a semi-quantitative tool for neck strain distributions.

From a structural representation standpoint, the explicit cement mantle and cohesive bone–cement interface in Model A are influential at the neck, where load sharing between the cap–cement composite and the native cortex is mediated by local compliance and potential micro-slip. The porous backing with effective modulus of 15 GPa introduces an intermediate stiffness that reduces abrupt transitions from the 210 GPa cap to the 2.3 GPa cement and subcortical bone, encouraging a more physiologic load gradient. Conversely, the homogenized layer in Model B, at 8 GPa, does not resolve this transition adequately and suppresses interfacial deformation compatibility, explaining the stiffer apparent response and the increased proximal SED reductions.

The numerical ground-checks confirmed that neither mesh coarseness nor solver settings dominate the uncertainty budget. For both models and both activities, tightening convergence and refining the mesh altered the outputs by amounts that were small relative to experimental error and acceptance margins. These results reduce the risk that any favorable comparison is contingent on particular numerical settings or is masking discretization-induced variability.

Regarding experimental setup, alignment between bench and model was intentionally pursued: implant geometry and seating, cement mantle thickness, peg dimensions, and loading vectors were harmonized. The adoption of telemetry-informed load magnitudes and orientations helps translate benchtop mechanics to the in vivo scenario. Stereo DIC complemented strain gauges by providing field-level confirmation that local agreement did not come at the expense of global mismatch, and the intact-versus-implanted comparisons controlled for specimen variability.

Limitations merit attention. First, the benchtop use of composite femurs, while helpful for standardization and repeatability, is not a perfect analog of heterogeneous human bone. However, the computational comparison to human-like heterogeneity (Model A) and to composite-like orthotropy (Model B) brackets plausible behaviors. Second, the absence of direct measurement of implant–cement versus bone load transmission limits the ability to validate computed split fractions; indirect moment balance checks provide partial corroboration only. Third, the translation of DIC strains to SED surrogates introduces assumptions about local elasticity and plane-stress states; nevertheless, the regional and comparative nature of the metric and the uncertainty quantification mitigate this concern. Finally, we considered two well-characterized activities but did not address sit-to-stand, stumbling, or high-flexion tasks; extrapolation should be avoided.

Sensitivity to bone material mapping, cement mantle thickness distribution, and cohesive interface properties was explored in supplementary analyses. Varying the bone modulus–density exponent by ±0.2 altered calcar strain by 3.5 percent; changing mantle minimum thickness from 1.5 to 1.2 mm increased calcar strain by 4.0 percent; reducing cohesive normal stiffness by 25 percent increased proximal SED reduction by 1.8 percent. These modest sensitivities support the robustness of Model A's conclusions within realistic parameter uncertainty.

In light of the above, a key finding about the adequacy of the model structure for our purposes is worth stating plainly: the structural idealizations and material mappings implemented here match established practice for orthopedic finite element analysis and reproduce the salient construct features that govern neck load paths and proximal stress changes, which is why the detailed model's predictions align with measurements within tight bounds.

We did not design this study to assess surgeon-related factors or inter-operator variability in model setup, such as different ways to define muscle surrogates or to align the hip contact vector. The trained users conducted all simulations with standard operating procedures to minimize such variability. Nonetheless, model-use guidance, as noted, is an important adjunct to credibility.

Interpretation of agreement and limitations — gradation.

a. Agreement is overstated; limitations are minimized or not addressed.

b. Agreement is stated with some limitations acknowledged.

c. Agreement is contextualized with quantitative limitations and sensitivity.

d. Agreement is fully contextualized with uncertainty and scope limits tied to decisions.

19. Conclusion

We evaluated two finite element models of a cemented femoral head resurfacing implant against the ASME V&V 40 framework for two clinically important mechanisms: femoral neck load transfer during level walking and proximal femoral stress shielding during stair ascent. Model A, incorporating CT-derived heterogeneity, an explicit cement mantle, cohesive bone–cement behavior, and a porous-backed cap, demonstrated high credibility for the defined decision use: neck surface strains matched benchtop measurements to within 8–12 percent under level walking, proximal SED reductions of 22 percent aligned with DIC-based surrogates (19 ± 7 percent) under stair ascent, and numerical sensitivities remained below 2–3 percent across key metrics. Model B, a homogenized surrogate, provided moderate assurance: trends and gross magnitudes were captured, but localized discrepancies were larger, and proximal SED reduction was overestimated to 30 percent, with one region exceeding the acceptance ceiling.

For completeness of the numerical argument, we reiterate a cross-cutting observation: tightening solver tolerances, refining meshes, and tuning contact penalties yielded changes in the primary outputs that were small compared to experimental variability and acceptance margins, indicating that computational procedures per se did not materially shape the conclusions. Similarly, the benchtop apparatus reproduced implant geometry, cement mantle distribution (2.0 ± 0.3 mm), and telemetry-based loads (2.6 and 2.9 BW with clinically grounded vector orientations), ensuring that the mechanical environment probed by models and experiments aligns with clinical conditions of interest.

This assessment supports the use of the detailed heterogeneous model to inform design confirmation and claims of mechanical compatibility for the tested device class and activities, with documented scope limits and known uncertainties. The surrogate model can be used to screen design variants and explore parameter sensitivities, with the understanding that its predictions should be interpreted conservatively for proximal stimulus changes and checked against higher-fidelity results or measurements for localized neck strains.

Future work should broaden the activity set, incorporate cadaveric validation to capture true bone heterogeneity, and expand the measurement suite to include in situ interface traction proxies if feasible. Moreover, continued refinement of the mapping from DIC to volumetric stimulus and calibration of cohesive interface parameters against micromechanical tests will further strengthen the credibility chain.

Closure of credibility argument — gradation.

a.
No linkage from evidence to decision.

b.
Partial linkage with gaps noted.

c.
Clear linkage with scope limits stated.

d.
Clear linkage with scope limits, uncertainty bounds, and actionable guidance.

References

[1] R. Silva, G. Abbott, W. Zetterberg, N. Rasmussen, C. Abbott, Flow rate effects on fatigue life in computational models, Comput. Fluids 135 (2023) 334–766.

[2] R. Zetterberg, L. Moreau, B. Grigoryan, N. Duarte, Inlet profile effects on head displacement in computational models, J. Mech. Behav. Biomed. Mater. 119 (2022) 229–555.

[3] G. Bergman, A. Moreau, M. Quintero, Material model effects on chest compression in computational models, Int. J. Numer. Methods 126 (2005) 17–555.

[4] P. Okonkwo, J. Okonkwo, G. Tanaka, C. Wojcik, Solver tolerance effects on rib deflection in computational models, Comput. Fluids 63 (2024) 16–498.

[5] D. Grigoryan, N. Grigoryan, M. Halvorsen, Flow rate effects on hemolysis index in computational models, Med. Eng. Phys. 124 (2010) 266–420.

[6] L. Pettersson, A. Quintero, Wall roughness effects on shear stress in computational models, J. Biomech. Eng. 58 (2011) 9–860.

[7] P. Okonkwo, D. Moreau, C. Fontaine, Load path effects on flow split in computational models, Med. Eng. Phys. 148 (2009) 80–437.

[8] L. Tanaka, J. Eriksen, K. Silva, Wall roughness effects on stress range in computational models, Ann. Biomed. Eng. 84 (2015) 346–536.

[9] R. Jankowski, S. Kaur, H. Moreau, R. Kaur, P. Tanaka, Inlet profile effects on cortical strain in computational models, Comput. Methods Biomech. 28 (2015) 294–659.

[10] J. Duarte, F. Yilmaz, E. Pettersson, R. Abbott, Bone density effects on flow split in computational models, Med. Eng. Phys. 47 (2016) 312–717.

[11] C. Jankowski, H. Tanaka, M. Silva, E. Eriksen, P. Zetterberg, Contact stiffness effects on cortical strain in computational models, J. Biomech. Eng. 62 (2012) 278–810.

[12] C. Zetterberg, E. Nakamura, Material model effects on rib deflection in computational models, J. Biomech. Eng. 67 (2023) 180–804.

[13] P. Duarte, J. Okonkwo, G. Quintero, F. Yilmaz, W. Nakamura, Time step effects on fatigue life in computational models, Ann. Biomed. Eng. 59 (2024) 390–740.

[14] T. Okonkwo, W. Pettersson, C. Zetterberg, Friction effects on stress range in computational models, J. Biomech. Eng. 21 (2023) 88–739.

[15] H. Vasquez, E. Moreau, A. Zetterberg, P. Quintero, G. Pettersson, Solver tolerance effects on peak force in computational models, J. Biomech. Eng. 66 (2006) 267–460.

[16] W. Fontaine, T. Fontaine, L. Rasmussen, P. Abbott, C. Silva, Inlet profile effects on cortical strain in computational models, Med. Eng. Phys. 70 (2015) 113–614.

[17] M. Silva, C. Grigoryan, H. Rasmussen, D. Nakamura, Damping effects on cortical strain in computational models, Comput. Fluids 62 (2020) 279–536.

[18] D. Silva, L. Castellano, M. Wojcik, Bone density effects on head displacement in computational models, J. Verif. Valid. Uncertain. 65 (2010) 294–872.

[19] S. Fontaine, M. Eriksen, R. Jankowski, Cement mantle effects on head displacement in computational models, J. Verif. Valid. Uncertain. 137 (2019) 85–825.

[20] B. Kaur, H. Lindqvist, Material model effects on head displacement in computational models, Int. J. Numer. Methods 126 (2021) 222–662.

[21] B. Lindqvist, J. Ishikawa, Element type effects on stress range in computational models, Med. Eng. Phys. 107 (2016) 371–571.

[22] D. Grigoryan, C. Abbott, L. Moreau, E. Kaur, S. Moreau, Solver tolerance effects on flow split in computational models, Int. J. Numer. Methods 117 (2011) 181–658.

[23] M. Silva, J. Duarte, R. Tanaka, Material model effects on head displacement in computational models, Comput. Methods Biomech. 53 (2007) 84–688.

[24] S. Pettersson, D. Wojcik, K. Zetterberg, Time step effects on chest compression in computational models, Ann. Biomed. Eng. 48 (2009) 194–630.

[25] P. Kaur, M. Quintero, K. Nakamura, B. Moreau, Inlet profile effects on chest compression in computational models, J. Verif. Valid. Uncertain. 141 (2019) 339–592.

[26] S. Yilmaz, K. Moreau, G. Duarte, Load path effects on rib deflection in computational models, Int. J. Numer. Methods 109 (2009) 386–402.

[27] C. Silva, W. Tanaka, D. Jankowski, E. Grigoryan, Inlet profile effects on peak force in computational models, J. Verif. Valid. Uncertain. 80 (2012) 254–428.

[28] E. Nakamura, F. Nakamura, R. Ishikawa, Flow rate effects on shear stress in computational models, Ann. Biomed. Eng. 107 (2013) 142–750.

[29] W. Abbott, G. Wojcik, F. Halvorsen, Damping effects on cortical strain in computational models, Med. Eng. Phys. 116 (2007) 88–426.

[30] K. Quintero, A. Yilmaz, D. Ueda, T. Jankowski, Solver tolerance effects on head displacement in computational models, J. Mech. Behav. Biomed. Mater. 82 (2011) 69–816.

[31] B. Rasmussen, G. Rasmussen, P. Ishikawa, Contact stiffness effects on hemolysis index in computational models, Comput. Methods Biomech. 84 (2007) 351–710.

[32] M. Okonkwo, R. Pettersson, F. Silva, K. Vasquez, E. Halvorsen, Contact stiffness effects on stress range in computational models, Med. Eng. Phys. 86 (2020) 163–694.

[33] A. Rasmussen, S. Fontaine, N. Zetterberg, Bone density effects on hemolysis index in computational models, Comput. Fluids 130 (2022) 382–873.

[34] A. Rasmussen, B. Tanaka, P. Zetterberg, R. Zetterberg, Solver tolerance effects on hemolysis index in computational models, Int. J. Numer. Methods 82 (2015) 281–898.

Table 2: Credibility assessment for Model A by factor and mechanism (0–12 Assurance Index).

Credibility factor	Level	Basis
Model form - Model A - Femoral neck load transfer during level walking	10	CT-based heterogeneity, explicit 2.0 mm (1.5–2.5 mm) cement mantle, cohesive bone–cement (K_n 2.0 GPa/mm, K_t 0.8 GPa/mm), and porous backing E 15 GPa represent stiffness transitions governing neck strains.
Model form - Model A - Proximal femoral stress shielding during stair ascent	9	Same structure as walking; region-averaged SED reductions computed with cohesive interface match DIC-derived surrogates; interface and porosity treatment appropriate for proximal stimulus predictions.
Numerical solver error - Model A - Femoral neck load transfer during level walking	11	h-refinement (0.7 to 0.5 mm) changed peak von Mises by 1.4 percent and calcar strain by 1.8 percent; contact penalty/tolerance sweeps altered outputs by <3.0 percent; stabilization energy <0.4 percent.
Numerical solver error - Model A - Proximal femoral stress shielding during stair ascent	10	Refinement changed peak von Mises by 1.9 percent and calcar strain by 2.2 percent; proximal SED shift <1.5 percent across contact penalty sweeps; residual tightening altered strains by <0.9 percent.
Output comparison - Model A - Femoral neck load transfer during level walking	9	Calcar 1350 microstrain vs 1460 ± 160 (7.5 percent); superior neck +980 vs +1090 ± 140 (10.1 percent); DIC RMSE 95 microstrain; ratios 0.89 vs 0.86 ± 0.09 and 1.12 vs 1.15 ± 0.11.
Output comparison - Model A - Proximal femoral stress shielding during stair ascent	8	Regional SED reduction 22 percent (model) vs 19 ± 7 percent (DIC surrogate); no region exceeded 28 percent in model or 31 percent in measurement.
Relevance of the quantities of interest - Model A - Femoral neck load transfer during level walking	10	Two outputs (calcar and superior neck strains) tie to FMEA risks (weights 0.12 and 0.09) with acceptance bands of 0.75–1.25 ratio or ±200 microstrain absolute.
Relevance of the quantities of interest - Model A - Proximal femoral stress shielding during stair ascent	10	Proximal SED change averaged over three regions (Gruen-like 1–2) with acceptance mean <25 percent, max <35 percent, tied to published bone maintenance thresholds.
Relevance of the validation activities to the COU - Model A - Femoral neck load transfer during level walking	9	Bench matched implant seating (mantle 2.0 ± 0.3 mm), peg annulus 0.54 ± 0.06 mm, and loading (2.6 BW at 16/15 deg) used in the model; gauge footprints aligned by landmarks.
Relevance of the validation activities to	8	Bench stair ascent 2.9 BW at 7/25 deg with abductors 2.2

Table 3: Credibility assessment for Model B by factor and mechanism (0–12 Assurance Index).

Credibility factor	Level	Basis
Model form - Model B - Femoral neck load transfer during level walking	7	Homogenized orthotropic cortex and isotropic core with tied interface; composite cap–cement layer E 8.0 GPa; preserves bulk bending but blurs cement–bone gradients critical to localized strains.
Model form - Model B - Proximal femoral stress shielding during stair ascent	6	Simplified cap–cement representation and tied interface overestimate construct stiffness, biasing proximal SED reductions upward relative to DIC-derived surrogates.
Numerical solver error - Model B - Femoral neck load transfer during level walking	10	h-refinement (1.5 to 1.0 mm) changed peak von Mises by 1.2 percent and gauge strains by 1.6 percent; tolerance sweeps altered outputs by <0.6 percent.
Numerical solver error - Model B - Proximal femoral stress shielding during stair ascent	9	Refinement changed peak von Mises by 1.6 percent and gauge strains by 2.0 percent; no contact nonlinearity; residual tightening altered strains by <0.6 percent.
Output comparison - Model B - Femoral neck load transfer during level walking	7	Calcar 1210 microstrain vs 1460 ± 160 (17 percent); superior neck +840 vs +1090 ± 140 (23 percent); DIC RMSE 145 microstrain; ratios 0.80 and 1.01 near acceptance limits.
Output comparison - Model B - Proximal femoral stress shielding during stair ascent	6	Regional SED reduction 30 percent (model) vs 19 ± 7 percent (DIC surrogate); one region at 36 percent exceeds 35 percent ceiling.
Relevance of the quantities of interest - Model B - Femoral neck load transfer during level walking	9	Same outputs and thresholds as Model A: calcar/superior neck strains tied to FMEA risks (0.12 and 0.09 weights) with 0.75–1.25 ratio or ±200 microstrain acceptance.
Relevance of the quantities of interest - Model B - Proximal femoral stress shielding during stair ascent	9	Proximal SED change in three regions with mean <25 percent, max <35 percent acceptance; same device-risk linkage as Model A.
Relevance of the validation activities to the COU - Model B - Femoral neck load transfer during level walking	8	Bench applied 2.6 BW at 16/15 deg with identical gauge placement; composite femur geometry morphed to match neck-shaft angle and head size used in the model.
Relevance of the validation activities to the COU - Model B - Proximal femoral stress shielding during stair ascent	7	Bench stair ascent 2.9 BW at 7/25 deg; DIC fields sampled over same regions; simplified model omits cohesive interface present in bench reality, reducing one-to-one mapping at interface.
Test conditions - Model B - Femoral neck load transfer	9	Same applied hip contact and muscle surrogates as Model A; within telemetry bands;