

Deciphering the "Art" in Modeling and Simulation of the Knee Joint: Model Benchmarking

Maryam Nazem¹, Thor E. Andreassen¹, Nancy Kim², Kate Moyle², Jason P. Halloran³, Casey A. Myers¹, Snehal Chokhandre⁴, Thor F. Besier³, Kevin B. Shelburne¹, Ahmet Erdemir⁴, Carl W. Imhauser⁵, Peter J. Laz¹

¹University of Denver, Denver, CO ²Auckland Bioengineering Institute, Auckland, New Zealand

³Washington State University, Spokane, WA ⁴Cleveland Clinic, Cleveland, OH ⁵Hospital for Special Surgery, New York, NY
imhauserc@hss.edu

DISCLOSURES: Nazem (N), Andreassen (N), Kim (N), Moyle (N), Halloran (N), Meyers (N), Chokhandre (N), Besier (Formus Labs), Shelburne (N), Erdemir (N), Imhauser (N), Laz (N)

INTRODUCTION: Reproducibility and credibility of estimates derived from physics-based modeling and simulation tools are crucial for their adoption in the research setting and eventual translation to clinical care. Because of their strong ties to data sharing and the responsible use of resources, reproducible modeling practices carry particular importance. The multicenter KneeHub initiative aims to evaluate how modeling decisions influence estimates of knee mechanics across five independent, expert teams [1]. In this study, the five teams developed and calibrated knee models using the same experimental datasets, [1-3] then tested model performance under benchmarking conditions not used in calibration. The objective was to evaluate the accuracy of model predictions of knee biomechanics for two clinically relevant scenarios. Namely, (1) anterior cruciate ligament (ACL) deficiency and (2) multiplanar loading simulating a common clinical exam of ACL integrity, the pivot shift.

METHODS: Open datasets for two knees were used: the Natural Knee Dataset (DU02) and Open Knee (OKS003). Associated imaging data was used to develop each knee model [2]. Models were then calibrated to reproduce experimental laxity testing in various degrees of freedom and at multiple flexion angles by tuning ligament properties and reference strain [3]. Experimental data were earmarked for use in model benchmarking and thereby were not used in calibration. For the DU02 knee, the model benchmarking cases involved removal of the ACL and evaluation of the knee for three laxity tests: anterior-posterior (AP), internal-external (IE), and varus-valgus (VV) at two flexion angles (15° and 55°), and passive flexion (0° to 120°). For the OKS003 knee, a loading condition that combined IE and VV applied moments at four flexion angles (0°, 30°, 60°, 90°) was utilized. Kinematics for the tibiofemoral joint simulation were compared with the experiment, and root means square (RMS) errors were calculated to evaluate each model and condition. To facilitate comparisons, all model predictions were transformed into the coordinate system of the experiment. RMS errors between the model predictions and experiment were computed for each test with standard deviations capturing variability of predictive accuracy among teams.

RESULTS: Regarding Natural Knee, all models estimated less AP translation (Fig. 1) and less external rotation than the corresponding experimental measurements. Across all tests, IE rotation showed the greatest variability among teams (Table 1). Regarding Open Knee(s), differences between model estimates and experimental measurements under the multiplanar torques were larger in IE than in the VV and AP directions (Table 1). In the OKS003 knee under combined IE and VV loading, the model for one team experienced numerical instability, while the model for another team failed to converge at the 90° flexion angle. These data points were excluded in the calculation of the errors.

DISCUSSION: Models captured general kinematic and kinetic trends but demonstrated systematic differences from experimental data and variability among teams. (Table 1). All data and results for the KneeHub project are publicly available on the SimTK website [4]. Model benchmarking was designed to challenge models beyond the calibration conditions and in clinically relevant ways—such as by removing a key structure (e.g., ACL) or applying combined loading to simulate a pivot shift exam. In the DU02 knee (no ACL), RMS errors for the laxity tests were larger than for passive flexion (Table 1), likely due to the larger applied forces/moments. The calibrated models were observed to be stiffer (Fig. 1), underpredicting the laxity of the experiment. The RMS errors in IE rotation were larger than in VV, even though the applied VV torque (10 Nm) was twice as high as the IE torque (5 N·m); while not directly loaded, results for AP translation showed the smallest errors. Across both knees and all tests, IE rotation showed the largest errors, which may have been influenced by modeler-specific decisions related to the representation and properties of the soft tissue structures. Teams employed distinct calibration strategies—some focused on isolating structures in a particular activity, while others tuned multiple structures across multiple degrees of freedom and flexion angles. Benchmarking demonstrated that extending models beyond calibration conditions can reveal limitations and sources of variability across modeling approaches.

SIGNIFICANCE/CLINICAL RELEVANCE: Benchmarking provides a rigorous framework to test model accuracy in clinically relevant scenarios, such as ACL deficiency and pivot-shift loading. Findings emphasize that models calibrated to one set of conditions should be used cautiously in others. By promoting transparency, reproducibility, and shared data resources, the KneeHub project advances the credibility and clinical translation of computational knee biomechanics.

REFERENCES: [1] Erdemir et al., J Biomech Eng 141, 2019. [2] Rooks et al., J Biomech Eng 143, 2021, [3] Andreassen et al., J Biomech Eng 145, 2024. [4] <https://simtk.org/projects/kneehub>.

ACKNOWLEDGEMENTS: NIH-NIBIB R01EB024573

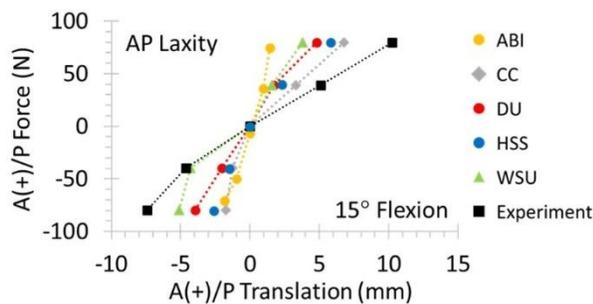


Figure 1: Benchmark laxity responses (no ACL) comparing simulations for the 5 teams with the Natural Knee (DU02) experimental data.

Table 1: RMS errors (mean and SD) comparing kinematic predictions to experiment for all teams. Shaded cells correspond to loaded directions

Model	Test	AP (mm)	IE (deg)	VV (deg)
Natural Knee No ACL	AP Laxity	6.6 ± 2.4	13.5 ± 12.9	3.7 ± 2.0
	IE Laxity	4.7 ± 3.0	8.8 ± 4.2	3.3 ± 1.7
	VV Laxity	5.0 ± 3.4	8.8 ± 7.4	5.3 ± 3.4
	Passive flexion	6.1 ± 5.3	7.4 ± 5.8	4.0 ± 1.9
Open Knee	Combined Loading	1.6 ± 0.8	7.3 ± 3.8	1.7 ± 0.9