Received: 17 October 2022 | Revised: 23 April 2023 | Accepted: 30 May 2023

DOI: 10.1002/jor.25652

Journal of Orthopaedic Research®

REVIEW

Reproducibility in modeling and simulation of the knee: Academic, industry, and regulatory perspectives

Carl W. Imhauser[1] | Andrew P. Baumann[2] | Xiangyi (Cheryl) Liu[3] | Jeffrey E. Bischoff[4] | Nico Verdonschot[5,6] | Benjamin J. Fregly[7] | Shady S. Elmasry[1,8] | Neda N. Abdollahi[9,10,11] | Donald R. Hume[12,13] | Nynke B. Rooks[14] | Marco T.-Y. Schneider[14] | William Zaylor[9,10] | Thor F. Besier[14,15] | Jason P. Halloran[16] | Kevin B. Shelburne[12,13] | Ahmet Erdemir[11,17]

[1]Department of Biomechanics, Hospital for Special Surgery, New York, New York, USA

[2]US Food and Drug Administration, Office of Science and Engineering Laboratories, Center for Devices and Radiological Health, Division of Applied Mechanics, Silver Spring, Maryland, USA

[3]Stryker Orthopaedics, Mahwah, New Jersey, USA

[4]Zimmer Biomet, Warsaw, Indiana, USA

[5]Department of Biomechanical Engineering, Technical Medical Institute at University of Twente, Enschede, The Netherlands

[6]Orthopaedic Research Lab, Radboud University Medical Centre, Nijmegen, The Netherlands

[7]Department of Mechanical Engineering, Rice University, Houston, Texas, USA

[8]Department of Mechanical Design and Production, Faculty of Engineering, Cairo University, Cairo, Egypt

[9]Center for Human Machine Systems, Cleveland State University, Cleveland, Ohio, USA

[10]Department of Mechanical Engineering, Cleveland State University, Cleveland, Ohio, USA

[11]Department of Biomedical Engineering, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio, USA

[12]Department of Mechanical and Materials Engineering, University of Denver, Denver, Colorado, USA

[13]Center for Orthopaedic Biomechanics, University of Denver, Denver, Colorado, USA

[14]Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand

[15]Department of Engineering Science, Faculty of Engineering, University of Auckland, Auckland, New Zealand

[16]Applied Sciences Laboratory, Institute for Shock Physics, Washington State University, Spokane, Washington, USA

[17]Computational Biomodeling (CoBi) Core, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio, USA

Correspondence
Carl W. Imhauser, Department of Biomechanics, Hospital for Special Surgery, 535 East 70th St., New York, NY 10021, USA.
Email: imhauserc@hss.edu

Funding information
National Institute of Biomedical Imaging and Bioengineering, Grant/Award Number: R01EB024573

Abstract

Stakeholders in the modeling and simulation (M&S) community organized a workshop at the 2019 Annual Meeting of the Orthopaedic Research Society (ORS) entitled "Reproducibility in Modeling and Simulation of the Knee: Academic, Industry, and Regulatory Perspectives." The goal was to discuss efforts among these stakeholders to address irreproducibility in M&S focusing on the knee joint. An academic representative from a leading orthopedic hospital in the United States described a multi-institutional, open effort funded by the National Institutes of Health to assess model reproducibility in computational knee biomechanics. A

Original article for submission to the *Journal of Orthopaedic Research* summarizing the Workshop held at the 2019 Annual Meeting of the Orthopaedic Research Society.

regulatory representative from the United States Food and Drug Administration indicated the necessity of standards for reproducibility to increase utility of M&S in the regulatory setting. An industry representative from a major orthopedic implant company emphasized improving reproducibility by addressing indeterminacy in personalized modeling through sensitivity analyses, thereby enhancing preclinical evaluation of joint replacement technology. Thought leaders in the M&S community stressed the importance of data sharing to minimize duplication of efforts. A survey comprised 103 attendees revealed strong support for the workshop and for increasing emphasis on computational modeling at future ORS meetings. Nearly all survey respondents (97%) considered reproducibility to be an important issue. Almost half of respondents (45%) tried and failed to reproduce the work of others. Two-thirds of respondents (67%) declared that individual laboratories are most responsible for ensuring reproducible research whereas 44% thought that journals are most responsible. Thought leaders and survey respondents emphasized that computational models must be reproducible and credible to advance knee M&S.

KEYWORDS
credibility, knee, modeling, regulatory, reproducibility, simulation

1 | INTRODUCTION

The scientific method has enabled a vast expansion in human knowledge and far-reaching technological advancements. Reproducibility is a linchpin of this paradigm; if the results and conclusions of a scientific investigation cannot be reproduced, a ground truth has not been established, which precludes the ability to derive a secure, actionable belief about a phenomenon.[1] The inability to reproduce preclinical findings across diverse fields is a serious concern in the scientific community.[2–4] Examples of irreproducible science range from psychology experiments[5] to animal models,[6] to microarray and cell-based studies.[7] These cases merit attention because scientific advancement relies on robust, new knowledge, which provides the underpinnings for novel technologies and products. Moreover, reproducible science garners public confidence and support, which stimulates a virtuous cycle of further investment in scientific endeavors.[8] Thus, scientific advancement is challenged when fundamental research findings cannot be reproduced. Accordingly, these concerning findings have captured the attention of the National Institutes of Health (NIH), which has proposed steps to foster rigor in grant applications submitted to this agency.[3]

Practitioners of modeling and simulation (M&S) in the field of orthopedics have recognized the importance of incorporating verification and validation studies to corroborate specific outcomes of interest (e.g., force, kinematics, strain, contact stress, and so on) with well-controlled experimental data for various musculoskeletal joints and tissues based on a specific context of use.[9–16] Accordingly, guidance documents and regulatory standards for credible practices in M&S have emerged over the last decade serving as a general framework to verify and validate computational models.[17–20] As M&S achieves widespread adoption in orthopedics,[21] reproducibility is also a pressing area of interest across the academic, industry, and regulatory spaces. Numerous subjective decisions of the developer throughout the modeling workflow, the so-called "art" of M&S, may be a critical barrier to achieve reproducible model outcomes, which would inhibit application of M&S in these spaces and compromise the ability to translate these powerful tools to the clinic to improve patient care.[22] To raise awareness on reproducibility challenges in computational knee biomechanics, a community of knee modeling researchers organized a workshop at the 2019 Meeting of the Orthopaedic Research Society (ORS 2019)[21] (see Supporting Information: Appendix 1, Proposal), which was open to all ORS 2019 participants.

The purpose of the workshop was to describe ongoing efforts by stakeholders from the academia, industry, and regulatory spaces and to engage the ORS community in discussion regarding opportunities and challenges in reproducibility in M&S focusing on physics-based modeling of the knee joint. Specific goals were, first, to describe a multi-institutional project funded by NIH aimed at identifying critical aspects of M&S.[21] The second goal was to describe reproducibility challenges and requirements in regulatory submissions when using M&S for device evaluation. The third goal was to discuss considerations for reproducibility of M&S in the preclinical assessment of medical devices. These goals were accomplished through 15-minute presentations by each speaker. The workshop also emphasized involvement from the broader knee biomechanics community via 5 min "flash opinions" provided by thought leaders in the field. The goal of the flash opinions was to stimulate audience participation during an ensuing 30 min question and answer session among all workshop attendees. Attendees provided feedback regarding their

perceptions of reproducibility in M&S of the knee through a survey administered during the workshop. In the present manuscript, we include perspectives of the academic, industry and regulatory representatives, summarize flash opinions of the thought leaders and comments from the audience, and present the survey results.

2 | AN ACADEMIC PERSPECTIVE

Carl Imhauser, PhD, Hospital for Special Surgery, New York, NY.

The reproducibility problem has captured the attention of the NIH and motivated steps to foster rigor in grant applications submitted to this agency.[3] The reproducibility problem is a point of concern in the field of M&S, as these powerful tools become widely adopted for scientific discovery and for clinical care. M&S of the knee has advanced over the last four decades from a small community using code developed in-house to broadly used, commercial and open source software suites that are widely employed in academia and industry.[23–25] As we strive to use M&S of the knee clinically in treatment decision-making, establishing reproducibility of these tools takes on added importance.

The workflow for M&S of the knee involves model development, evaluation, and simulation processes. This workflow relies on objective scientific principles but commonly requires the intuition of the modeler during implementation. This so-called "art of modeling" may consist of subjective decisions acquired through experiential learning and be limited by constraints, such as time, software and hardware resources, and complexity, all with the end application (research question or design criteria) in mind. These decisions introduce uncertainties in the modeling workflow and can impact reproducibility of the mechanical predictions from joint level kinetics to tissue level stresses. Consider, for example, the range of decisions and the number of steps that the modeler faces to describe a single constituent of the knee such as the medial collateral ligament (MCL). First, the modeler must extract geometric information including its shape and insertions via dissections, anatomical atlases, or segmentation of medical scans. Next, the modeler must describe the material formulation of the MCL and its mechanical implementation, for example, three-dimensional elements, two-dimensional membranes, or nonlinear springs, with associated selection of constitutive model form and pretension/prestrain.[26,27] Material properties and joint laxity, which vary widely from person-to-person and with age,[28–31] must also be selected and can range from mean properties extracted from the literature to subject-specific properties via calibration.[10,12,32] Boundary conditions including shape, insertions, and applied loads are also important considerations. Clinical controversy regarding the existence, structure, and biomechanical function of anatomical structures adds to the uncertainty in representations of knee anatomy in M&S.[33] Moreover, the highly complicated and variable nature of knee anatomy further confounds the ability to achieve reproducible practices in M&S of the knee.[34,35] These examples are a subset of the complexity a modeler faces and the myriad decisions a modeler must make to derive a physics-based model of the knee.

To begin addressing reproducibility in M&S of the knee joint, five independent research teams are documenting their knee M&S workflows to address a fundamental question concerning the 'art of modeling' and its potential impact on reproducibility: "Do the predictions of natural knee biomechanics depend on the modeling decisions of separate development teams when the target simulation scenarios and the source data to build models remain the same?" To this end, each team in this NIH-funded research is independently modeling two knees relying on the same data. This cross-institutional effort, which is hosted on the SimTK infrastructure (https://simtk.org/), will reveal the state of reproducibility in M&S of the knee focusing on whole joint and tissue mechanics. Reproducibility of kinematics and kinetics predictions across several clinical scenarios will be used to assess potential for model reuse. Model predictions from each group will also be compared to subject-specific experimental data, to assess how closely model predictions corroborate physical measurements. These simulation scenarios include passive knee flexion, the pivot shift clinical exam, and a weight bearing X-ray. This project involves curation of all aspects of the M&S workflow through its lifecycle including specifications, development and calibration, protocol deviations, source data, intermediate and final outcomes (model components, models, simulation results), and model reuse scenarios. At the time of compilation of this manuscript, this study resulted in several publications and resources.[21,22,36–38]

A key component of this effort is open, prospective documentation of model processes through submission of specifications before initiation of each modeling phase.[37] Subsequent documentation of deviations during each modeling phase from the intended model workflow will reveal differences between what each modeling group intended to do and what was done as the modelers proceed through each phase. Agreement of model predictions among teams would indicate that the simulation approaches are interchangeable and can be used based on need. Heterogeneity of model predictions among teams provides the opportunity to identify contributing factors. This focused effort to understand reproducibility in M&S of the knee may lead to standardization and best practices to facilitate not only reproducibility but also model exchange and repurposing. These efforts are needed to build faith in model predictions and to reinforce robust scientific methodology in this burgeoning field. This work will help to enable clinical translation of M&S expanding its utility from preclinical applications, such as the design and evaluation of orthopedic devices, to direct implementation in clinical care. Active participation from the biomechanical modeling community is encouraged and opportunities for interaction are facilitated through the KneeHub website where all project stages are documented, and all work products are freely available for download.[39]

3 | A REGULATORY PERSPECTIVE

Andrew P. Baumann, PhD, US Food and Drug Administration, Center for Devices and Radiological Health, Office of Science and Engineering Laboratories, Division of Applied Mechanics, Silver Spring, MD.

Medical devices undergo regulatory review before marketing to ensure safety and effectiveness. The level of review and regulatory pathway reflects the device classification and risk level. The Center for Devices and Radiological Health in the United States Food and Drug Administration (FDA) classifies devices as Class I having the lowest risk, Class II having more risk, and Class III having the greatest risk. Subsequent review can range from general controls, to special controls, as well as Premarket Approval (PMA), respectively. Most orthopedic knee devices are reviewed through the 510(k) (special controls) or PMA pathways. These review processes often involve testing to demonstrate preclinical mechanical performance of the device. These mechanical performance tests are traditionally performed "on the bench," using load frames, fixtures, and physical test specimens. Benchtop tests have existed for many years with practices being iteratively refined and improved. The community has published standards and guidance documents specific to knee devices, which detail the accepted way to perform bench tests.[20,40–48] Device manufacturers are familiar with conducting these tests and regulating agencies are familiar with evaluating the results. Thus, there can be a high level of consistency and reproducibility in the tests, which facilitates efficient development, regulatory review, and marketing of knee devices.

Despite its effectiveness, bench testing can be time and resource expensive. Physical test specimens must be manufactured and are often lost to destructive testing. Tests involving fatigue loading also require a significant investment of time. This has led to the use of computational modeling and simulation in regulatory submissions to support and supplement mechanical performance testing. Computational M&S techniques, such as finite element analysis (FEA), allow medical device manufacturers to test devices in a virtual framework. This has the potential to reduce costs and speed testing throughput. M&S also enable rapid changes to test parameters and provides results that would otherwise be inaccessible. For example, in situations where engineers might be limited to measuring strain response with gages and/or digital image correlation on device surfaces, FEA can predict high resolution strain maps through entire device volumes. In addition to being a powerful research, development, and design tool, FEA can be used to test many similar designs within a device family to assist the determination of the worst case for subsequent loading on the bench. Setting up and executing such a parametric analysis is often an advantage of simulations. Moving ahead, FEA will likely see increased use in regulatory evaluation of knee devices due to the benefits it may offer over traditional techniques.

FEA is not without shortcomings. Bench testing has had years of refinement to achieve reproducibility, but FEA is comparatively new to knee device testing. At the time of the 2019 workshop, only one standard existed for FEA of knee components[48] with another since being published.[49] Given this lack of instruction, modelers must often make their own decisions when building a computational framework instead of following an established set of guidelines. The decisions that modelers make vary widely from one practitioner to another, leading to a variety of different simulations and outcomes. This decision making, or "art of modeling," limits the consistency of finite element models. Reproducibility tends to decrease as modeler decision making increases and models diverge. As such, regulators can be burdened by inconsistent modeling techniques that must be interpreted on an individual basis. For example, simulating the performance of a knee device can be complicated by incorporation of knee anatomy. Material models and contact interactions associated with the knee (bone, cartilage, ligaments, tendons, muscles, and so on) all present challenges when evaluating model credibility and associated device performance. With another orthopedic implant (intervertebral body fusion devices), the FDA observed that basic information needed to assess a computational model (code verification, mesh convergence, validation, and so on) was often not included in regulatory submissions.[50] Knee device submissions likely follow a similar trend. Therefore, the computational modeling community should focus on reproducibility and more trustworthy simulations. Regulators cannot rely on simulations before these expectations are met. Current efforts to publish computational modeling standards, guidance documents, and best practices have the potential to increase reproducibility.[20,47] Stakeholders must invest in these tools to make modeling and simulation more reproducible now, such that future simulations may become more consistent and have greater credibility and utility in a regulatory framework.

4 | AN INDUSTRY PERSPECTIVE

Xiangyi (Cheryl) Liu, PhD, Stryker Orthopaedics, Mahwah, NJ.

Reproducibility of clinical outcomes of a given implant or procedure is one of the key credibility assessments of knee models. Knee models are used to evaluate new implants and procedural solutions by simulating the effects of changes in model inputs, such as implant geometries, positioning of the implants or preservation of soft tissues, on model outputs. The model outputs reflect the clinical outcome of current solutions, such as stresses in the implant or tissue, relative motions between implant components, or relative motion between implant and surrounding tissue. From the perspective of model reproducibility, this is mostly demonstrated in the end applications of the knee models. However, there is information in the knee model creation, calibration, and validation process that may be used to assess reproducibility in the proposed end application of the models.

One reason why reproducibility of knee modeling is difficult is because we are solving an indeterminate system. There are numerous soft tissue structures around the knee and each structure has several parameters to describe its behavior. The number of experiments that can be conducted to calibrate these unknowns is smaller than the number of unknowns for cadaveric knee models and even smaller for in vivo knee models. Different labs participating in the project may generate different knee models using the same set of experimental data, because there are many feasible solutions for the indeterminate system.[51] Knowing the model is not the unique solution underscores

the importance of quantifying uncertainties of key input parameters and their effect on key output parameters for the proposed application.[52,53]

Output parameters that industry has historically been interested in are mostly related to durability of the implants such as peak stress, micromotion, contact pressure, and contact area in the implants. In those scenarios, identifying the worst-case condition is the key step before any testing or analysis. With industry putting more and more focus on improving patient satisfaction and quality of life, additional kinematic quantities of interest such as range of motion, stability, and soft tissue balancing are being evaluated in addition to durability factors. These output parameters as well as input parameters for knee models are patient specific. A library of heterogeneous patient-specific models is required to evaluate the clinical outcomes in these scenarios. When building a library of models, it is important to demonstrate reproducibility in the entire process from model creation and calibration to model validation.

5 | "FLASH OPINIONS" ON REPRODUCIBILITY

Jeffrey Bischoff, PhD (Director, Research, Zimmer Biomet).

Nico Verdonschot, PhD (Professor and Scientific Director of the Technical Medical Institute at University of Twente at University of Twente).

Benjamin J. Fregly, PhD (Professor and CPRIT Scholar in Cancer Research, Rice University).

Dr. Bischoff noted the difficulty in clearly documenting work to maximize reproducibility and indicated that even FDA guidance documents and published consensus standards may not be explicit enough to result in clear model and workflow descriptors to ensure reproducibility. In addition, storing and sharing models and derivative data may be ideal to enhance reproducibility, but this may be prevented by conflicts of interests, intellectual property rights, and fear of losing competitive advantage. Finally, Dr. Bischoff stated that driving towards reproducibility too early in the maturity of a model may have a key negative unintended consequence: it may prevent the evaluation of key aspects of model form (e.g., boundary conditions), which originate independently from separate modeling teams. Critical model uncertainties are often identified by considering differences in key model outputs resulting from differences in understanding model intent or in decisions on model form across independent modeling groups. Once identified, the modeling community can collectively drive to a more prescriptive consensus approach towards a model form, which considers both accuracy and reproducibility. For example, this process is central to round robin studies in support of the proposed ASTM modeling standards.[49] Round robin studies highlight the variability in modeling that occurs even when a draft model prescription is in place. Therefore, this approach challenges the standards team to add (and document) more specificity on essential model form attributes based on valid technical justifications.

Professor Verdonschot commented on the inefficiency inherent in many groups using slight variations of the same model and in each lab attempting to build their own versions of the same model. Healthy competition among research groups is an important factor in moving the field forward but the need for the modeling community to share resources, including standardized image MRI scans, remains a challenge to be addressed. Without specific criteria for comparing models, there are likely to be great differences in model predictions. Therefore, assessing reproducibility in simpler scenarios than the native knee may be more appropriate. Even within his research program, models developed from different laboratory members working with the same data set can vary drastically and result in "student-specific" models. Professor Verdonschot observed that the community lacks criteria for contextualizing differences in model predictions across groups (e.g., failure criteria). That is, context must be established to determine what level of differences in an outcome measure is meaningful. Without specific context, ranking may be more relevant than providing exact values. Establishing the purpose of the study is critical, because the modeler needs to make choices and estimations (e.g., in boundary conditions, material formulations for various tissues, choice of activity, and so on) that depend on the context of use. These modeling choices are based on what the modeler wants to assess, such as the behavior of the collateral ligaments, or the cartilage contact stress, or the overall stability in the knee with or without muscle forces. Finally, Professor Verdonschot acknowledged trade-offs between model complexity and feasibility considering time, money, and personnel burden for both developers and end-users.

Professor Fregly emphasized that one must first ask two fundamental questions when considering reproducibility: (1) What specific phenomenon are we trying to reproduce? and (2) What is the intended application of the model? Deciding what model to use, whether the model is "good enough," and what aspects need to be correct depend on the intended purpose and end goal of the model. He stated that model sharing is the most important area affecting reproducibility. Having siloed groups each developing their own models impedes the dissemination and sharing of knowledge. This process of individual laboratories "reinventing the wheel" inhibits reproducibility and prevents the community from building on the work of others. Efforts to provide freely accessible modeling tools such as OpenSim have done the field a great service and greater efforts at model sharing are necessary to accelerate scientific advancement.[54] Professor Fregly acknowledged the fear of losing a competitive advantage through model sharing, but emphasized that open efforts that he has headed, such as the "Grand Challenge Competition to Predict In Vivo Knee Loads,"[51] have been extremely beneficial to his career advancement and, therefore, support the notion that there is ample opportunity for publishing your own work and for model sharing.

Professor Fregly listed three factors that will impact reproducibility. First was the skill of the modeler, which was anecdotally demonstrated by high variability in outcomes in his undergraduate modeling courses. Given the broad spectrum of modeling abilities,

enhancing reproducibility likely requires sharing of knowledge and of best practices to provide modelers the opportunity to obtain similar skill levels. Second was the wide array of decisions made during model creation including choice of geometries (e.g., scaled geometry or image-based), choice of joint models, properties of bones and soft tissues (e.g., cartilage, ligaments, muscle-tendon units), and contact properties. This challenge is compounded by the fact that many parameters are not observable requiring improved methods for assignment and calibration of model properties. Third was variations in numerical implementation of the model such as in choice of integrator and solver settings and in initial conditions. Finally, Professor Fregly emphasized the need to increase model sharing among the musculoskeletal modeling community to enhance knowledge dissemination.

6 | SURVEY OF WORKSHOP ATTENDEES

A survey was disseminated during the workshop to understand the opinions of attendees on reproducibility in M&S (see Supporting Information: Appendix 2, Survey). Among those who attended the workshop, 103 completed the survey, which was 4.2% of the total number of ORS 2019 participants (2436). The survey was modeled after one administered to the broader scientific community.[55]

Nearly all survey respondents (97%) thought reproducibility in M&S was an important issue, whereas 98% liked the workshop and 95% stated that they would attend if it were held again (Table 1, Q1, Q8, and Q9). These findings indicate an interest in both M&S and reproducibility among ORS attendees and a desire to continue discussions on these topics at future ORS meetings. A symposium on practical considerations for model credibility in M&S hosted by the ORS Orthopaedic Implants Section in 2022 indicate increased focus on these topics in the ORS community since the 2019 workshop.

About half of respondents (56%) acknowledged that their lab did not have established procedures to ensure reproducibility of knee M&S (Table 1, Q2). Similarly, 45% have tried and failed to reproduce the work of others (Table 1, Q3). In contrast, in another survey conducted across major research disciplines, 34% of respondents did not have established procedures for reproducibility in their lab and at least 60% of respondents had unsuccessfully tried to reproduce someone else's experiment.[55] Our findings may indicate that the M&S community members who attend the ORS Meeting are lagging the broader research community in addressing the topic of reproducibility.

In our survey, only 8% of respondents reported trying to publish a reproduction attempt (Table 1, Q6). These numbers are even less than what was reported in a larger poll of the scientific community where 24% had published a successful reproduction attempt and 13% had published an unsuccessful reproduction attempt.[55] Our workshop attendees likely also included students and members of industry, which may have reduced this percentage in our cohort. Interestingly, only 16% of those completing our survey had ever been contacted by another lab that had tried to reproduce their work (Table 1, Q7). These findings corroborate the larger poll of the scientific community where <20% said they had ever been contacted by another researcher who was unable to reproduce their work.[55] Anecdotally, these findings may underscore that the burden of reproducing others' work is significant in terms of time and cost; thus, there may be a lack of incentive to see these assessments of reproducibility through to publication. Incentivizing such work through promotion and funding may help address this issue.[36,56]

Interestingly, 43% of respondents believed that rigorous rules for reproducibility would hinder progress/innovation in knee M&S (Table 1, Q4). However, 65% thought that knee M&S publications should provide access to raw data and code. Thus, there appeared to be resistance to establishing formal rules for reproducibility, but a receptiveness to open data sharing.

TABLE 1 Summary results of the survey among 103 participants of the workshop.

Question	Yes	No	Responses
1. Do you consider reproducibility of Knee M&S to be an important issue?	97%	3%	101
2. Does your lab have established procedures for reproducibility?	44%	56%	98
3. Have you tried to reproduce the work of others and failed?	45%	55%	98
4. Do you think rigorous rules for reproducibility in Knee M&S will hinder progress/innovation?	43%	57%	90
5. Should all Knee M&S publications provide access to primary (raw) data and modeling code?	65%	35%	96
6. Have you ever tried to publish a reproduction attempt?	8%	92%	98
7. Have you ever been contacted by another lab that has tried to reproduce your work?	16%	84%	99
8. Did you like the workshop?	98%	2%	100
9. If this workshop were held again, would you attend?	95%	5%	98

Abbreviation: M&S, modeling and simulation.

Regarding the greatest contributors to irreproducibility in knee M&S, respondents to our workshop survey targeted insufficient detail in the methods sections of archival journal publications (66%) and lack of access to source data and code (54%) as the two main culprits (Figure 1). Three quarters (77/102) of respondents indicated that insufficient detail in the methods and/or lack of access to source data were important sources of irreproducibility. These two issues may be effectively resolved by journals and authors, respectively. These factors also were also thought to play a role among the broader scientific community where >80% deemed unavailability of methods/code to always, often, or sometimes contribute to irreproducibility.[55] This finding may reinforce the need to both increase methodological detail in journal publication and to support open data and code sharing in the M&S community.[56] The next greatest perceived cause of irreproducibility was lack of training in proper use of M&S tools (44%). Thus, focused training opportunities in the appropriate use of M&S techniques could improve reproducibility in the orthopedics community. Similarly, in the *Nature* paper, 90% of respondents stated "better mentorship" as an approach that would improve reproducibility.[55] About one-third of workshop attendees deemed selective reporting of the best results (41%), bias against publishing negative findings (36%), low statistical power (35%), and low-quality source data (31%) as important factors. Interestingly, individuals from the broader scientific community deemed selective reporting to always or often contribute to irreproducibility at about 68%.[55] About a quarter of workshop attendees thought that pressure to publish (28%) contributed to irreproducibility. Pressure to publish was a greater concern in the poll of the broader scientific community where about 65% of scientists deemed this factor to always or often contribute to irreproducibility citing "intense competition and time pressure."[55] This discrepancy may stem from the fact that participants in the 2019 ORS workshop included trainees and representatives from industry, who may feel this pressure less acutely than academics. Lack of verified software, poor statistical training, and poor-quality peer review were considered lesser contributors to irreproducibility by audience members (≤17%). Poor-quality peer review was also a lesser-ranked contributor to irreproducibility in the poll of scientists across disciplines, yet 38% of respondents in that larger survey still thought it often or sometimes contributed to irreproducibility.[55]

Regarding the most difficult aspects of knee M&S to reproduce, respondents ranked defining boundary conditions (45%) and material formulations (33%) as the most challenging, while reproducing the joint coordinate system (24%) as one of the least challenging (Figure 2). In contrast, our recent work revealed wide variations in coordinate system definitions including axes orientations and origin locations across modeling groups among five teams but consistent boundary condition definitions.[37] This disparity between perception and reality may highlight an overlooked area driving irreproducibility in Knee M&S and provide an opportunity for improving reproducibility.[21]

About two-thirds of respondents (67%) agreed that the main responsibility for ensuring reproducible research in knee M&S lies with the individual laboratories and the principal investigators (Figure 3). Interestingly, 44% thought that journals bare responsibility in ensuring reproducibility, whereas institutions and funding agencies ranked lower at approximately one-third each. Given restrictions in article length, current journal formats may not allow for enough details to be presented in the methods section of papers, thereby compromising reproducibility potential. Further expansion of online and Supporting Information (e.g., via GitHub, Figshare, and so on) may help remedy this concern. Moreover, journals could play an active role in fostering reproducible practices by encouraging

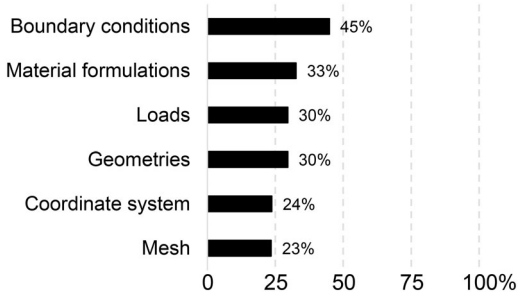

FIGURE 2 Survey results: What aspect of developing a knee model do you find hardest to reproduce?

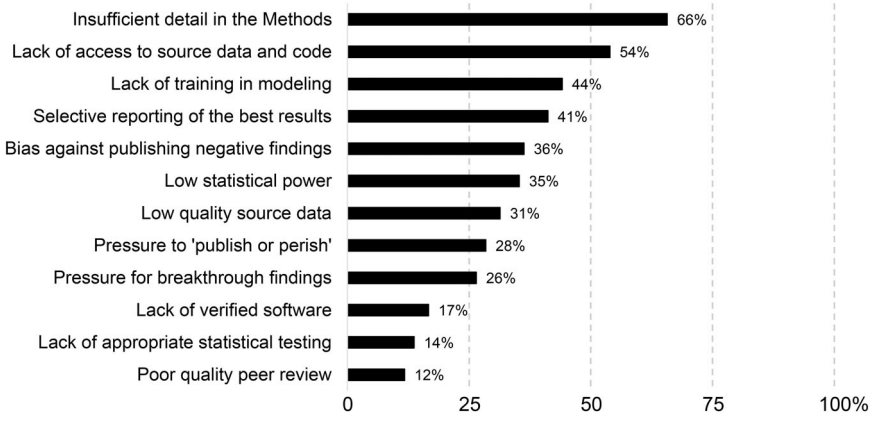

FIGURE 1 Survey results: What are the biggest causes of irreproducibility in Knee modeling and simulation (M&S), in your opinion?

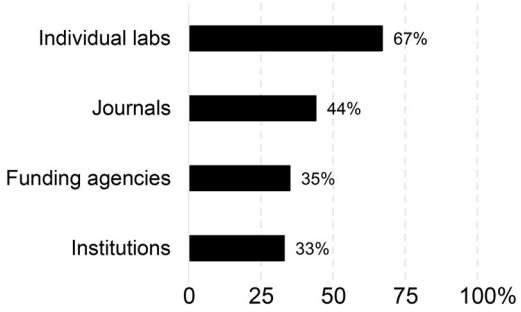

FIGURE 3 Survey results: What entities are most responsible for ensuring reproducible research in Knee modeling and simulation (M&S)?

publication of reproducibility attempts to further incentivize individual laboratories to implement recommendations from the community (Table 1, Q5, Q6).[36] Journals may also have an opportunity to improve reproducibility by asking or requiring authors to share data, methods, and models for each publication. Funding agencies, which support large scale repositories such as the SimTK (https://simtk.org/) for model source data, benchmarking, and sharing, likely aid in shouldering the burden placed on individual laboratories. Training researchers that data sharing is an integral step in the research lifecycle could also help reduce irreproducibility.

Characteristics of the cohort who completed the survey should be considered. Workshop participants are likely to be more receptive to the use of computational modeling in research and, therefore, have concerns regarding reproducibility in M&S. Moreover, the respondents may not be representative of all members of the knee M&S community, who may be less likely to attend the ORS Meeting, because it has historically focused less on M&S.

In the written replies and audience comments, respondents emphasized that reproducibility and credibility practices should be tightly aligned. Specifically, for scientific advancement and clinical translation, computational models must be both reproducible (i.e., have mutually consistent model outputs) and accurate. Model predictions that do not match what happens in the physical world (i.e., experimental measurements) will not advance the field. In comparing model predictions and experimental measurements, errors in the physical experiments must also be considered. Audience members commented that the results of some physical experiments may be more difficult to reproduce than those of computational models due to high sensitivity of outputs to initial conditions (e.g., polyethylene wear studies using gait simulators).[57]

A limitation of this survey is that respondents were not asked to provide their affiliation (academic, industry, regulatory, etc.) and role (student, faculty, staff, and so on), which would help to interpret the responses. Second, we acknowledge that we were not exhaustive in inclusion of perspectives, yet a range of authors are included whose expertise spans M&S in tissue, joint, and neuromuscular biomechanics. Future workshops could focus discussion to specific components of the knee, such as ligament, cartilage, meniscus, and neuromuscular M&S.

7 | CONCLUSION

Concerns regarding irreproducibility of preclinical research findings among the scientific community and funding agencies are highly relevant to physics-based M&S of the knee spanning the academic, regulatory, and industry sectors as these tools achieve broad adoption and widespread use.[21,22] To raise awareness on reproducibility challenges in computational knee biomechanics, a community of knee modeling researchers organized a workshop at ORS 2019. In the academic sector, a rigorous, open, multi-institutional study is underway to determine the state of reproducibility.[21,22,37] In the regulatory space, more work to specify reproducibility requirements in M&S of the knee is needed to foster confidence in the widespread adoption of these tools for device evaluation in regulatory submissions. Concerning the industrial sector, preclinical design and assessment of joint replacement technology will be enhanced by utilizing sensitivity analyses to address irreproducibility stemming from indeterminacy in personalized modeling. Thought leaders in the M&S community emphasized the need to increase collaboration and data sharing to disseminate knowledge and minimize duplication of efforts. Workshop attendees stressed that credibility and reproducibility assessment should be tightly aligned. Nearly all survey respondents (97%) considered reproducibility to be an important issue. They indicated a strong desire to continue discussions at future ORS meetings not only on reproducibility but also on credible practices in M&S. This feedback led to a series of M&S-themed Research Interest Group meetings at subsequent ORS meetings from 2020 to 2022. Survey respondents thought that individual laboratories were most responsible for ensuring reproducible practices in M&S and secondarily were journals. Accordingly, insufficient detail in the methods section of publications and lack of access to source data and code were perceived to be drivers of irreproducibility, yet less than half of respondents belonged to laboratories with well-established procedures for reproducibility. Therefore, irreproducibility may be addressed by individual laboratories via developer's/user's guides, for example, but also through targeted efforts by scientific journals via increased publication opportunities for reproducibility efforts and new or streamlined resources and support for open data sharing of source code and model assets. Forums such as this workshop are an important part of community-based open efforts to enhance model credibility, which is a prerequisite for widespread adoption of these powerful tools to improve patient care.

AUTHOR CONTRIBUTIONS

Carl W. Imhauser, Andrew P. Baumann, Xiangyi (Cheryl) Liu, Jeffrey E. Bischoff, Nico Verdonschot, Benjamin J. Fregly, Jason P. Halloran, Shady S. Elmasry, and Ahmet Erdemir gave substantial contributions to research design, or the acquisition, analysis, or interpretation of data; drafting the paper or revising it critically; approval of the submitted and final versions. Kevin B. Shelburne, Neda N. Abdollahi, Donald R. Hume, Nynke B. Rooks, Marco T.-Y. Schneider, William Zaylor, and Thor F. Besier contributed to drafting the paper or revising it critically; approval of the submitted and final versions.

ACKNOWLEDGMENTS

Research reported in this publication was supported by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under award number R01EB024573.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

ORCID

Carl W. Imhauser http://orcid.org/0000-0003-2445-7112
Andrew P. Baumann http://orcid.org/0000-0001-9872-8822
Shady S. Elmasry http://orcid.org/0000-0003-0193-7711
Marco T.-Y. Schneider http://orcid.org/0000-0002-4918-1389
Thor F. Besier http://orcid.org/0000-0003-0818-7554
Jason P. Halloran http://orcid.org/0000-0001-5866-6583
Kevin B. Shelburne http://orcid.org/0000-0003-3377-2712

REFERENCES

1. Peirce CS. The fixation of belief. *Popular Sci Month*. 1877;12:1-15.
2. Macleod M. Improving the reproducibility and integrity of research: what can different stakeholders contribute? *BMC Res Notes*. 2022;15:146.
3. Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. *Nature*. 2014;505:612-613.
4. Ioannidis JPA. Why most published research findings are false. *PLoS Med*. 2005;2:e124.
5. Carlin SP, Standing LG. Is intelligence enhanced by letter priming? A failure to replicate the results of Ciani and Sheldon (2010). *Psychol Rep*. 2013;112:533-544.
6. Sena ES, van der Worp HB, Bath PMW, Howells DW, Macleod MR. Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. *PLoS Biol*. 2010;8:e1000344.
7. Haibe-Kains B, El-Hachem N, Birkbak NJ, et al. Inconsistency in large pharmacogenomic studies. *Nature*. 2013;504:389-393.
8. Begley CG, Ioannidis JPA. Reproducibility in science: improving the standard for basic and preclinical research. *Circ Res*. 2015;116:116-126.
9. Ellis BJ, Lujan TJ, Dalton MS, Weiss JA. Medial collateral ligament insertion site and contact forces in the ACL-deficient knee. *J Orthop Res*. 2006;24:800-810.
10. Kia M, Schafer K, Lipman J, et al. A multibody knee model corroborates subject-specific experimental measurements of low ligament forces and kinematic coupling during passive flexion. *J Biomech Eng*. 2016;138:051010.
11. Guo H, Santner TJ, Chen T, et al. A statistically-augmented computational platform for evaluating meniscal function. *J Biomech*. 2015;48:1444-1453.
12. Ali AA, Harris MD, Shalhoub S, Maletsky LP, Rullkoetter PJ, Shelburne KB. Combined measurement and modeling of specimen-specific knee mechanics for healthy and ACL-deficient conditions. *J Biomech*. 2017;57:117-124.
13. Garavelli C, Curreli C, Palanca M, Aldieri A, Cristofolini L, Viceconti M. Experimental validation of a subject-specific finite element model of lumbar spine segment using digital image correlation. *PLoS One*. 2022;17:e0272529.
14. Palazzi E, Siegler S, Balakrishnan V, Leardini A, Caravaggi P, Belvedere C. Estimating the stabilizing function of ankle and subtalar ligaments via a morphology-specific three-dimensional dynamic model. *J Biomech*. 2020;98:109421.
15. Kinney AL, Besier TF, D'Lima DD, Fregly BJ. Update on grand challenge competition to predict in vivo knee loads. *J Biomech Eng*. 2013;135:021012.
16. Grassi L, Schileo E, Taddei F, et al. Accuracy of finite element predictions in sideways load configurations for the proximal human femur. *J Biomech*. 2012;45:394-399.
17. Henninger HB, Reese SP, Anderson AE, Weiss JA. Validation of computational models in biomechanics. *Proc Inst Mech Eng Part H*. 2010;224:801-812.
18. Erdemir A, Mulugeta L, Ku JP, et al. Credible practice of modeling and simulation in healthcare: ten rules from a multidisciplinary perspective. *J Transl Med*. 2020;18:369.
19. Viceconti M, Pappalardo F, Rodriguez B, Horner M, Bischoff J, Tshinanu FM. In silico trials: verification, validation and uncertainty quantification of predictive models used in the regulatory evaluation of biomedical products. *Methods*. 2021;185:120-127.
20. ASMEV&V 40, *Assessing Credibility of Computational Modeling through Verification and Validation: Application to Medical Devices*. The American Society of Mechanical Engineers; 2018.
21. Erdemir A, Besier TF, Halloran JP, et al. Deciphering the "art" in modeling and simulation of the knee joint: overall strategy. *J Biomech Eng*. 2019;141:0710021-0710210.
22. Halloran JP, Abdollahi N, Hafez MA, et al. Assessment of reporting practices and reproducibility potential of a cohort of published studies in computational knee biomechanics. *J Orthop Res*. 2022;41:325-334.
23. Maas SA, Ateshian GA, Weiss JA. FEBio: history and advances. *Annu Rev Biomed Eng*. 2017;19:279-299.
24. Seth A, Hicks JL, Uchida TK, et al. OpenSim: simulating musculoskeletal dynamics and neuromuscular control to study human and animal movement. *PLoS Comput Biol*. 2018;14:e1006223.
25. Taylor M, Prendergast PJ. Four decades of finite element analysis of orthopaedic devices: where are we now and what are the opportunities? *J Biomech*. 2015;48:767-778.
26. Galbusera F, Freutel M, Dürselen L, et al. Material models and properties in the finite element analysis of knee ligaments: a literature review. *Front Bioeng Biotechnol*. 2014;2:54.
27. Naghibi Beidokhti H, Janssen D, van de Groes S, Hazrati J, Van den Boogaard T, Verdonschot N. The influence of ligament modelling strategies on the predictive capability of finite element models of the human knee joint. *J Biomech*. 2017;65:1-11.
28. Harner CD, Xerogeanes JW, Livesay GA, et al. The human posterior cruciate ligament complex: an interdisciplinary study. Ligament morphology and biomechanical evaluation. *Am J Sports Med*. 1995;23:736-745.
29. Woo SLY, Hollis JM, Adams DJ, Lyon RM, Takai S. Tensile properties of the human femur-anterior cruciate ligament-tibia complex. The effects of specimen age and orientation. *Am J Sports Med*. 1991;19:217-225.
30. Butler DL, Guan Y, Kay MD, Cummings JF, Feder SM, Levy MS. Location-dependent variations in the material properties of the anterior cruciate ligament. *J Biomech*. 1992;25:511-518.
31. Imhauser CW, Kent, 3rd RN, Boorman-Padgett J, Thein R, Wickiewicz TL, Pearle AD. New parameters describing how knee ligaments carry force in situ predict interspecimen variations in laxity during simulated clinical exams. *J Biomech*. 2017;64:212-218.
32. Zaylor W, Stulberg BN, Halloran JP. Use of distraction loading to estimate subject-specific knee ligament slack lengths. *J Biomech*. 2019;92:1-5.
33. Kumar VD, Sontakke YA, Murugharaj S. Truly existing or hyped up? Unravelling the current knowledge regarding the anatomy, radiology, histology and biomechanics of the enigmatic anterolateral ligament of the knee joint. *Arch Bone Jt Surg*. 2020;8:332-342.
34. LaPrade RF, Ly TV, Wentorf FA, Engebretsen AH, Johansen S, Engebretsen L. The anatomy of the medial part of the knee. *J Bone Joint Surg Am Vol*. 2007;89:2000-2010.
35. Sturnick DR, Vacek PM, DeSarno MJ, et al. Combined anatomic factors predicting risk of anterior cruciate ligament injury for males and females. *Am J Sports Med*. 2015;43:839-847.

36. Erdemir A, Hunter PJ, Holzapfel GA, et al. Perspectives on sharing models and related resources in computational biomechanics research. *J Biomech Eng*. 2018;140:0247011-02470111.
37. Rooks NB, Schneider MTY, Erdemir A, et al. Deciphering the "art" in modeling and simulation of the knee joint: variations in model development. *J Biomech Eng*. 2021;143:061002.
38. Rooks NB, Schneider MTY, Erdemir A, et al. A method to compare heterogeneous types of bone and cartilage meshes. *J Biomech Eng*. 2021;143:111002.
39. Erdemir A. *Reproducibility in Simulation-Based Prediction of Natural Knee Mechanics*. Interagency Modeling and Analysis Group; 2023.
40. ISO. *ISO Standard 14879-1, Implants for surgery – Total knee-joint prostheses, Part 1: Determination of endurance properties of knee tibial trays*. Annual Meeting 2003, Geneva Switzerland; 2000.
41. ISO. *ISO Standard 14243-1, Implants for Surgery – Wear of Total Knee-Joint Prostheses – Part 1: Loading and Displacement Parameters for Wear-Testing Machines with Load Control and Corresponding Environmental Conditions for Test*. International Organization for Standardization; 2009.
42. ASTM. *ASTM Standard F1800. Cyclic Fatigue Testing of Metal Tibial Tray Components of Total Knee Joint Replacements*. ASTM International; 2012.
43. ASTM. *ASTM Standard F1223. Standard Test Method for Determination of Total Knee Replacement Constraint*. ASTM International; 2014.
44. ASTM. *ASTM Standard F1814. Standard Guide for Evaluating Modular Hip and Knee Joint Components*. ASTM International; 2015.
45. ASTM. *ASTM Standard F3140. Standard Test Method for Cyclic Fatigue Testing of Metal Tibial Tray Components of Unicondylar Knee Joint Replacements*. ASTM International; 2017.
46. Guidance for Industry and FDA. *Class II Special Controls Guidance Document: Knee Joint Patellofemorotibial and Femorotibial Metal/Polymer Porous-Coated Uncemented Prostheses*. US Food and Drug Administration; 2003.
47. Guidance for Industry and FDA. *Reporting of Computational Modeling Studies in Medical Device Submissions*. Food and Drug Administration; 2016.
48. ASTM. *ASTM Standard F3161. Standard Test Method for Finite Element Analysis (FEA) of Metallic Orthopaedic Total Knee Femoral Components under Closing Conditions*. ASTM International; 2017.
49. ASTM. *ASTM Standard F3334-19. Standard Practice For Finite Element Analysis (FEA) Of Metallic Orthopaedic Total Knee Tibial Components*. ASTM International; 2019.
50. Baumann AP, Graf T, Peck JH, Dmitriev AE, Coughlan D, Lotz JC. Assessing the use of finite element analysis for mechanical performance evaluation of intervertebral body fusion devices. *JOR Spine*. 2021;4:e1137.
51. Fregly BJ, Besier TF, Lloyd DG, et al. Grand challenge competition to predict in vivo knee loads. *J Orthop Res*. 2012;30:503-513.
52. Navacchia A, Myers CA, Rullkoetter PJ, Shelburne KB. Prediction of in vivo knee joint loads using a global probabilistic analysis. *J Biomech Eng*. 2016;138:4032379.
53. Guo H, Santner TJ, Lerner AL, Maher SA. Reducing uncertainty when using knee-specific finite element models by assessing the effect of input parameters. *J Orthop Res*. 2017;35:2233-2242.
54. Delp SL, Anderson FC, Arnold AS, et al. OpenSim: open-source software to create and analyze dynamic simulations of movement. *IEEE Trans Biomed Eng*. 2007;54:1940-1950.
55. Baker M. 1,500 scientists lift the lid on reproducibility. *Nature*. 2016;533:452-454.
56. Erdemir A, Guess TM, Halloran JP, et al. Commentary on the integration of model sharing and reproducibility analysis to scholarly publishing workflow in computational biomechanics. *IEEE Trans Biomed Eng*. 2016;63:2080-2085.
57. Johnston H, Abdelgaied A, Pandit H, Fisher J, Jennings LM. The effect of surgical alignment and soft tissue conditions on the kinematics and wear of a fixed bearing total knee replacement. *J Mech Behav Biomed Mater*. 2019;100:103386.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Imhauser CW, Baumann AP, Liu X(C), et al. Reproducibility in modeling and simulation of the knee: academic, industry, and regulatory perspectives. *J Orthop Res*. 2023;41:2569-2578. doi:10.1002/jor.25652