In Silico Clinical Trials in the Orthopedic Device Industry: From Fantasy to Reality?

Philippe Favre,[1] Ghislain Maquer,[2] Adam Henderson,[2]
Daniel Hertig,[2] Daniel Ciric,[2] and Jeffrey E. Bischoff[3]

[1]Zimmer Biomet, Zählerweg 4, 6300 Zug, Switzerland [2]Zimmer Biomet, Sulzerallee 8, 8404 Winterthur, Switzerland; and
[3]Zimmer Biomet, 1800 West Center St, Warsaw, IN 46580, USA

Associate Editor Joel Stitzel oversaw the review of this article.

Abstract—The orthopedic device industry relies heavily on clinical evaluation to confirm the safety, performance, and clinical benefits of its implants. Limited sample size often prevents these studies from capturing the full spectrum of patient variability and real-life implant use. The device industry is accustomed to simulating benchtop tests with numerical methods and recent developments now enable virtual "in silico clinical trials" (ISCT). In this article, we describe how the advancement of computer modeling has naturally led to ISCT; outline the potential benefits of ISCT to patients, healthcare systems, manufacturers, and regulators; and identify how hurdles associated with ISCT may be overcome. In particular, we highlight a process for defining the relevant patient risks to address with ISCT, the utility of a versatile software pipeline, the necessity to ensure model credibility, and the goal of limiting regulatory uncertainty. By complementing—not replacing—traditional clinical trials with computational evidence, ISCT provides a viable technical and regulatory strategy for characterizing the full spectrum of patients, clinical conditions, and configurations that are embodied in contemporary orthopedic implant systems.

ABBREVIATIONS

CE	Conformité Européenne—European Conformity
FE	Finite element
ISCT	In silico clinical trials
MDR	Medical Devices Regulation
Clinical investigation (trial, study)	Systematic investigation in one or more human subjects, undertaken to assess the clinical performance, effectiveness or safety of a medical device
Clinical data	Safety, clinical performance and/or effectiveness information that is generated from the clinical use of a medical device
Clinical evaluation	Systematic and planned process to continuously generate, collect, analyze and assess the clinical data pertaining to a device in order to verify the safety and performance, including clinical benefits, of the device when used as intended by the manufacturer
Clinical evidence	Clinical data and clinical evaluation results pertaining to a device of a sufficient amount and quality to allow a qualified assessment of whether the device is safe and achieves the intended clinical benefit(s), when used as intended by the manufacturer

Address correspondence to Philippe Favre, Zimmer Biomet, Zählerweg 4, 6300 Zug, Switzerland. Electronic mail: philippe.favre@zimmerbiomet.com

Legacy device	Device previously CE marked under the European Medical Devices Directive 93/42/EEC (MDD) or Active Implantable Medical Devices Directive 90/385/EEC (AIMDD)
Validation	The process of determining the degree to which a model or a simulation is an accurate representation of the real world[7]
Verification	The process of determining that a computational model accurately represents the underlying mathematical model and its solution from the perspective of the intended uses of modeling and simulation[7]

INTRODUCTION

Medical device manufacturers aim to improve the quality of life for patients by restoring function and alleviating pain. This industry is highly regulated, as implants must be designed to demonstrate quality and present minimal risk to patients. Every manufacturer is expected to provide data demonstrating the long-term safety, performance and clinical benefits of their products. Such data support device introductions and maintain existing devices on the market, and are currently obtained from four sources: (1) clinical studies, (2) the literature, (3) complaints, and (4) registries.[26]

Since introduction of the Medical Devices Regulation (MDR)[43] in Europe and the new version of ISO 14155,[35] the orthopedic industry is facing new challenges. With, for instance, regular analyses of the risk–benefit ratio through periodic safety update reports and increased scrutiny on "device equivalence", more emphasis has been put on the need for clinical studies to fill any safety data gaps. Yet, these studies have practical challenges of their own. The typically limited sample size reduces the chance to obtain data on rare demographics, clinical indications, surgical procedures, or low-volume implant variants. Additionally, a significant number of patients can be lost to follow-up over the duration of the trial.

The device industry is accustomed to simulating benchtop tests with numerical methods in the setting of pre-clinical testing. It has been proposed to use computer-generated data from "in silico clinical trials" (ISCT) to address challenges faced by clinical studies.[27,42,47,48] Published population level studies have demonstrated the technical feasibility to assess a specific patient risk on a large number of virtual patients.[1,2,12,31,38,40,44] Further, regulators are increasingly open to use of computer-generated evidence.[46,47] ISCT has therefore the potential to become a new viable source for credible data on clinical device performance (Fig. 1).

There are a number of challenges associated with successful deployment of ISCT in support of new devices. The relevant patient risks in scope of the ISCT must be determined. However, there currently exists no guidance on how to identify those risks. Ensuring sufficient credibility of the various models within an ISCT is necessary, to provide confidence to all stakeholders in the utility of the predictions. While recent guidance has emerged on credibility expectations for modeling of devices,[7] there is no consensus on what constitutes sufficient validation evidence for ISCT applications.

PURPOSE

The purpose of this article is to highlight the relevant parameters and associated considerations when planning an ISCT to support the development and commercialization of a medical device. The regulatory environment, the current use of simulation in the industry, and the benefits of ISCT will be outlined. The remaining practical barriers to the use of ISCT will be described, and suggestions on how to overcome these barriers will be provided.

This work will focus on examples from the orthopedic medical device industry, but the general principles laid out in this work are applicable to other implantable medical devices (e.g., stents, heart valves, pacemakers).

EVOLVING REGULATORY EXPECTATIONS ON CLINICAL EVIDENCE

A Stricter Regulatory Environment

The Medical Devices Regulation (MDR 2017/745) is replacing the Medical Devices Directive (MDD 93/42/EEC)[43] with the goal to better protect the patient and enforce more transparency. MDR requirements pose several challenges to manufacturers and the notified bodies, some directly related to the clinical evaluation of orthopedic devices.[26] Some of these challenges include:

– *Reclassification*: Most orthopedic implants remain Class III medical devices under MDR, but some of them, such as spinal devices, may be reclassified.
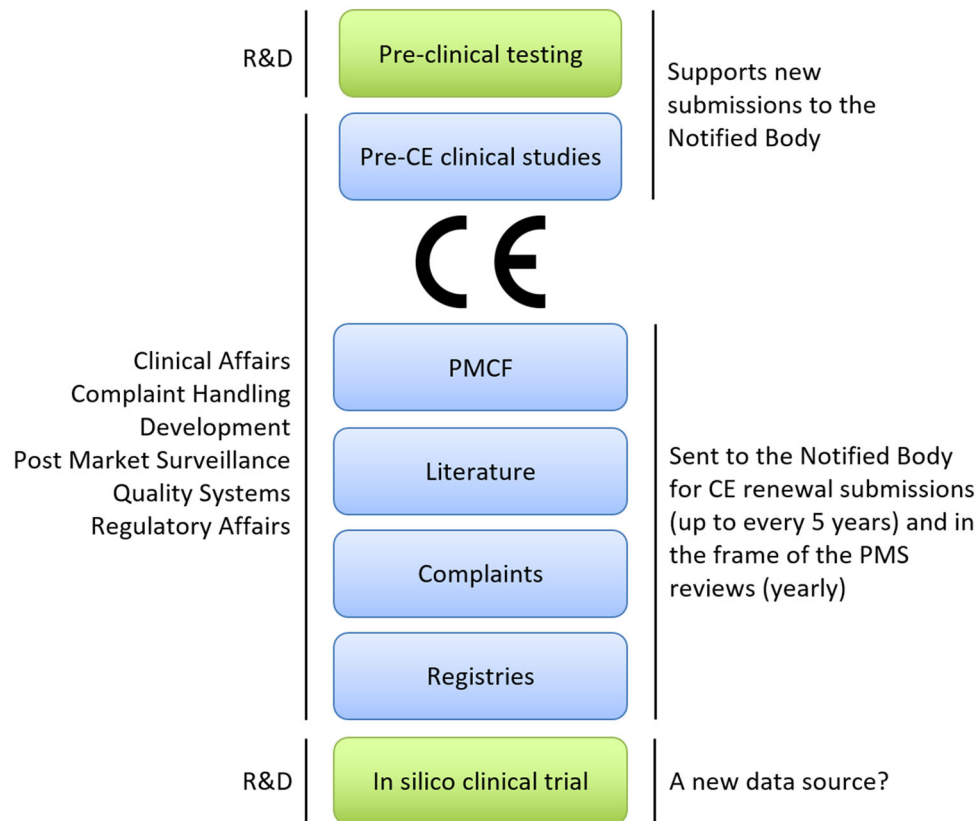
FIGURE 1. Safety and performance are first established by pre-clinical testing through computer modeling and bench testing, provided by Research and Development (R&D) (green). A pre-CE clinical study is often necessary, when a medical device or its use are deemed novel. When conducting clinical evaluations for a device, four traditional sources can be included (blue): 1. Clinical studies (pre-CE clinical studies or post-market clinical follow-up studies—PMCF), 2. Literature, 3. Complaints, and 4. Registries. After receipt of the CE mark by the regulator, the product can be launched to market provided a clear plan for collecting clinical data is available. In parallel with market launch, PMCF studies are established to proactively collect data. The scientific literature, in particular peer-reviewed publications, are screened to ensure that the device offers the intended benefits and to detect potential adverse events. All implant-related complaints directly received by the manufacturer or reported in internal clinical studies or published literature studies are analyzed in post-market surveillance (PMS). Registries are also a great source of long-term data as they follow large cohorts and cover several implant designs.[26] Despite these potential sources, data may be lacking for rare implants, demographics, and indications. ISCT may fill these gaps, and may become the fifth data source.

- *No grandfathering*: All products, including legacy devices certified under MDD, must meet the MDR requirements, i.e., new clinical data must likely be collected by the manufacturer.
- *Device equivalence*: The investigational device must have the same technical, biological, and clinical characteristics as the intended equivalent device. If equivalence cannot be demonstrated on all three levels, device-specific clinical data must be collected by the manufacturer.
- *Device variant*: To deliver operative flexibility to the surgeons, many orthopedic devices are available in several variants (such as cemented and cementless versions of a hip stem). If similar clinical outcomes cannot be demonstrated through other means, variant-specific clinical data must be collected by the manufacturer.

- *Clinical indication*: Instructions for Use (IFU) must clearly state the medical conditions the device is intended to support. If a device is intended for different indications (e.g., primary total hip replacement and revision arthroplasty) and similar clinical outcomes cannot be assumed, indication-specific clinical data must be collected by the manufacturer.
- *Post-market surveillance (PMS)*: The need for proactive data collection (e.g., clinical studies) is emphasized, and reports demonstrating the safety and performance claims of a device with these data are now also reviewed annually by the regulator in PMS reviews.

The main consequences of these challenges for the manufacturer are that clinical data for different device variants and for different indication may be needed. This reinforces the role of a proactive data collection

via PMCF studies, as opposed to passive data collection activities. Based on the MDR requirements, several and/or larger PMCF studies may have to be conducted to capture the required data for new devices. Devices that were already on the market before MDR are also impacted. In many cases, device equivalence is either practically impossible (equivalent device is from a competitor so patient level data is not obtainable) or void (equivalent device meets MDD but not MDR requirements). Literature may be scarce on well-established devices that are no longer scientifically interesting. Older clinical studies may not have been designed to collect data on all variants and indications or might not comply with the new ethical and clinical practices.[32] Finally, national and hospital registries can only provide data on the devices sold in the local market.

Shortcomings of Post-Market Clinical Studies

Under MDR, much emphasis is put on the role of proactive data collection via post-market studies, and rightly so. However, proactive data collection studies have known shortcomings. PMCF trials for orthopedic devices are mostly prospective and observational, and typically follow patients up to 10 years after the surgery. Over such a long period of time, a significant number of participants are "lost to follow-up" due to patient relocation, patient opt-out of follow-up examinations because they are doing well, or patient death. As a consequence, sample size is often exaggerated in order to compensate for the expected loss to follow-up, requiring substantial financial and time investments without the guarantee of finally satisfying the needs of the clinical study. Loss to follow-up has been exacerbated due to the recent Covid-19 pandemic. Hospitals have postponed activities deemed non-vital (including clinical follow-ups), and elderly patients who often belong to risk categories prefer avoiding hospital visits to limit potential exposure to Covid-19.

As mentioned in "A Stricter Regulatory Environment" section, clinical data may now be requested for each implant variation and indication, if similar clinical outcomes across variants and indications cannot be assumed. However, obtaining clinical data can be difficult for certain implant variants tailored for rare demographics (e.g., coxa vara hips), or for rare indications (e.g., hip hemi arthroplasty). Overpowered or multiple clinical trials are possible, but represent suboptimal and costly solutions. Retrospective post-market data collections are an option for legacy devices, but retrospective studies offer less control on the collected data. Further, because such retrospective data were generally collected for other purposes, they may have limited applicability to the device at hand.

Potential Consequences

With these factors in mind, acquiring sufficient clinical evidence to support device certification or recertification under MDR can be prohibitive and in the end, manufacturers may simply decide to retire device variants or abandon indications for use. Similarly, the Notified Bodies, which traditionally verify that a product meets the European conformity (CE) certification criteria, must now respond to a new controller role that requires more workforce and slows down the certification process of new devices. Additionally, the scope of new devices may be limited to the most common sizes and configurations. The treatments for rare demographics, rare indications or rare procedures are most at risk, leaving patients without any appropriate treatment options.

There is thus a need for a new technical and regulatory strategy—a way to demonstrate clinical performance of devices, across the full intended spectrum of patient conditions, without elevating the burden of clinical data acquisition to unsustainable levels. Modeling and simulation can play an important role in this new strategy.

EVOLUTION OF MODELING IN THE MEDICAL DEVICES INDUSTRY

Augmenting clinical trials with data from computer simulations, so called ISCT, has been proposed as a potential solution to address these challenges while satisfying regulatory requirements.[47] Computational modeling is used regularly throughout the product life cycle to evaluate the performance, safety, and effectiveness of medical devices. In order to understand the needs for ISCT in industry, it is important to have an overview of the current use of modeling and its recent evolution, going from the simulation of the product alone to simulating a patient population. These developments have mainly been driven by increases in computational power, advances in modeling techniques, and standardization in establishing model credibility.

Traditional Use of Modeling and Simulation in Industry

Simulation methods have traditionally been used for early proof-of-concept evaluations and identification of worst-case conditions for design verification activities (Fig. 2). Use of modeling and simulation for the purposes of worst-case determination has been codified within several specific modeling and testing standards, including hip stem fatigue strength in total hip arthroplasty[9] and tibial tray fatigue strength in total

knee arthroplasty.[10] These standard analyses are, by definition, performed under very controlled conditions to remove any sources of variability. Variability in bone quality is eliminated by utilization of artificial foam bone with standardized material properties.[8] Variability in loading is eliminated by applying a given load magnitude and orientation. Variability from the surgical procedure is removed by prescribing implant potting level, orientation and position. Such tests provide an excellent basis for understanding the physics of the device, and for comparing the device to a predicate device that has been tested in the same controlled manner, thus simplifying the regulatory review process.

By simulating these tests, significant time and money can be saved. A physical fatigue test can typically take weeks to months, depending on the number of test samples, test cycles and frequency, and number of available load frames. Also each new test involves costs and time for test parts and fixtures. Simulations can often be performed much faster than the physical test, and involve fixed cost only for licenses and hardware. In this context, simulation is most often used for comparison of the different sizes, configurations and loadings within an implant family, to identify the situation most likely to fail. This worst case can then be physically tested to ensure satisfactory performance of the full implant family. In the end, this allows for both potential design optimization and for a drastic reduction in physical testing. In some situations, the simulation allows bypassing the physical test altogether. For example, when the peak stress levels in
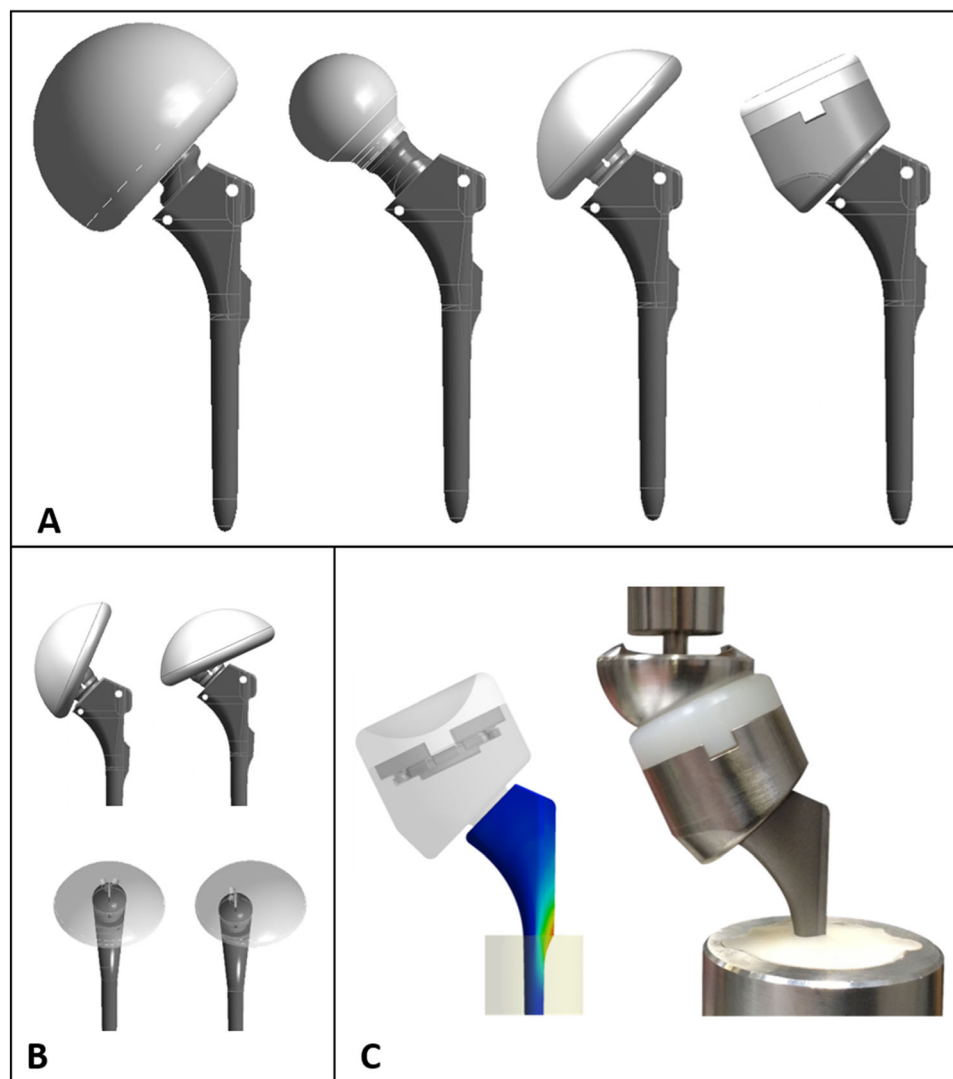

FIGURE 2. Traditional use of modeling to identify the worst case from multiple configurations (a) and component positions (b). Variability in loading is also typically considered. The worst case can then be physically tested (c). All test conditions are very controlled.

the device are predicted to be much lower than fatigue run out stresses, or are shown by direct comparison to be less than those of a predicate device, there is often no benefit or regulatory expectation to conduct physical testing on the device. Such simulation applications frequently require a closer examination of the processes used to manufacture the device (as well as the predicate device, when applicable), to ensure the effects of, for example, application of surface coatings on the fatigue strength of the substrate material, are well understood.

The main limitations to this approach, namely the simulation of precisely specified bench tests, are that the very controlled and standardized conditions neglect patient and surgical variability, and some extreme conditions may be overlooked. Further, the clinical validity of the standardized methods is not consistently demonstrated, and may not sufficiently challenge features of new systems.

Patient-Specific Modeling and Simulation

When the influences of patient and surgical variability need to be understood, benchtop cadaveric and patient specific modeling studies can be deployed (Fig. 3). Such studies can address the influence of anatomy, tissue properties, surgical approach or in vivo loading. For example, in recent years, both cadaveric testing[23] and computer models[21] have been used to understand the effect of bone quality and daily life activities on implant primary stability, in order to provide guidelines to healthcare professionals for optimal patient selection and rehabilitation.

Though cadaveric testing provides the opportunity to understand (across a small cohort) the effect of specimen variability to an extent that cannot be seen with standardized testing, these tests are costlier than the corresponding standardized testing, and are generally more time consuming due to the necessary specimen preparation. Also, cadaveric specimen availability is naturally limited. As such, the potential gain in time and costs of modeling is therefore even larger than for standardized testing. However, such models also become more complex, with the notable need for patient anatomy and adequate tissue material properties characterization. Fortunately, bone segmentation tools have eased the creation of patient specific 3D bone models from CT data[16] and conversion of greyscale data into material properties.[29]

Limitations of such patient-specific studies are that they typically include a limited set of patients, or implement average input parameters under the assumption that they will produce results representative of the average populations, an assumption that has been shown not to systematically hold true.[17] As such,

they do not encompass the full in vivo variability. Further, these models are generally focused on a single aspect of device performance. Finally, though the practice of converting greyscale data into material properties is easily accessible using contemporary software, nevertheless a variety of algorithms in the literature have been proposed for this conversion, with varying levels of complexity and with direct impact on simulation results.[15,29,30] It is thus essential that the sensitivity of the intended conclusion from the computational study on this critical model form assumption is well understood, taking into account whether the study is intended to make patient-specific predictions or to address a patient risk across a larger cohort.

Application at a Population Level

Modern methods can simulate a large number of patient-specific FE models by relying on highly automated modeling pipelines and high-performance computing (Fig. 4). Such techniques fully embrace variability in ways that were not possible before. The anatomical datasets are provided from either a large database of patient scans, or from statistical shape modeling methods.[13,28] These pools of 3D Computer Aided Design (CAD) bone models can be used directly for morphologic and implant fit analysis. For example, population-level morphological analyses have been used to incorporate anatomical variability and surgical factors in the design of orthopedic implants so that they can better fit a broader patient population.[19] Statistical intensity models can generate three-dimensional bone density distributions from 2D DXA scans, and can be coupled with statistical shape modeling to generate bone models of realistic shape and bone density distribution.[4,13,32]

Each of these anatomical models can now be converted to patient specific FE bone models. Moreover, the number of cases can be effectively increased by perturbing implant position or sizing, thereby simulating surgical variability, or by varying bone quality. For example, starting from 16 bone models and varying implant position and bone quality, a total of 2000 unique models were generated to study the primary stability of cementless femoral stems,[2] an order of magnitude larger than the number of patients typically included in a clinical study.

Clinical variability is a confounding factor that in vivo clinical trials cannot fully capture. Patients vary in their anatomy, with bones of different shapes, sizes, and quality, but also vary in their level of activity, which will put very different loading conditions on the device. The surgery can vary as different surgeons may cut the bone in dissimilar manners, and may size or position the implant differently. Population-level sim-
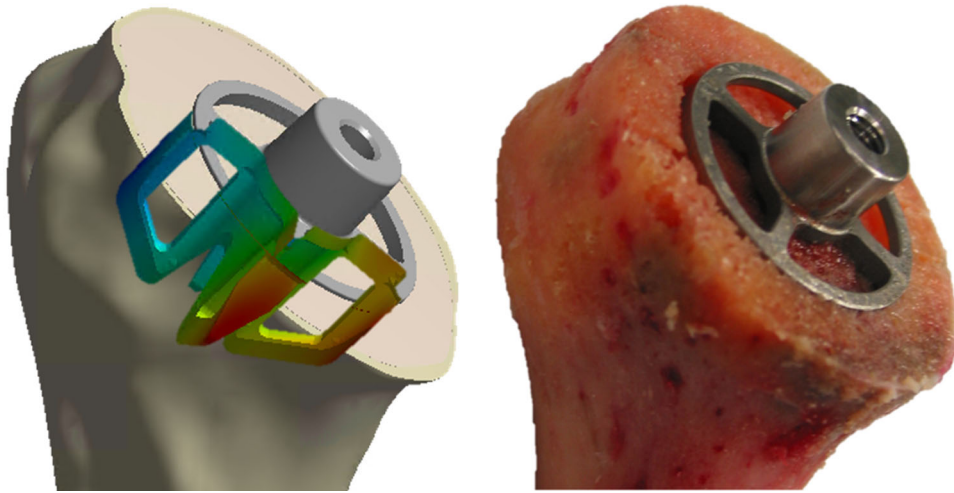
FIGURE 3. Modeling a physical benchtop test performed on a cadaver. Such modeling includes patient characteristics (bone density, bone geometry).
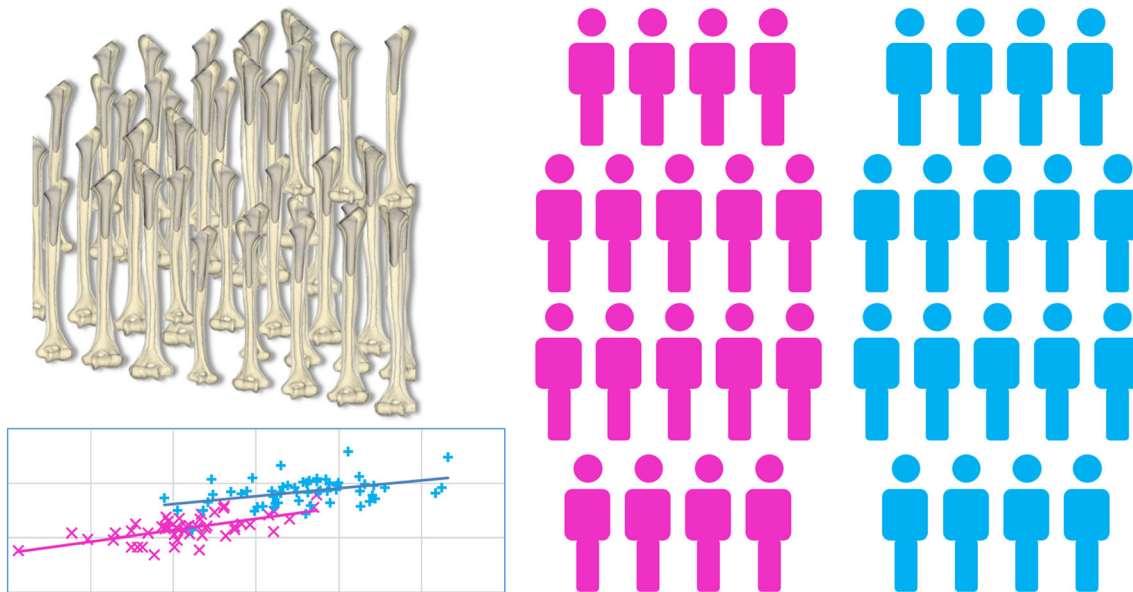

FIGURE 4. Modeling to simulate a whole population allowing statistical comparison of an outcome between two groups (pink and blue).

ulations may provide a means to fully capture the influence of all these sources of variability on the patient risk(s) of interest, and to do so in a parametric fashion.

POTENTIAL BENEFITS OF ISCT

An ISCT pathway for generating credible evidence for the expected clinical performance of a medical device has benefits to all constituents in the device ecosystem: patients, healthcare system, device manufacturers and regulators. It must be noted that key benefits hit multiple stakeholders.

Potential Benefits for the Patient

Simulating clinical trials is a natural application of population-level models. ISCT can complement traditional clinical trials by filling potential data gaps and addressing some of their common shortcomings ("Shortcomings of Post-Market Clinical Studies" section) with simulated clinical data. Beyond just complementing traditional clinical trials, ISCT could positively impact the design and safety of products due

to increased pre-clinical rigor. Indeed, ISCT enables deeper a level of inquiry into variables that can impact patient care, and could help to ensure the patient receives the best possible solution. Virtual scenarios permit exploration of the influence of different parameters such as the influence of surgical variability, variability due to patient activity, and bone quality variability. This allows better evaluation of extreme cases, not just the standard/normal patients. Modelling also allows the analysis of different scenarios within a single patient, permitting a one to one comparison. Confounding factors can be understood by isolating the influence of given parameters, which is impossible *in vivo*. All of these considerations may reasonably result in a device that could contribute to improved outcomes for the patient.

Potential Benefits for the Healthcare System

The elevated costs required to meet current clinical data requirements have direct impact on the healthcare systems, which are already under cost pressure in many countries. Many well-established and cost-effective orthopedic devices may be retired because they do not generate enough revenue to justify the expense of transitioning them to MDR requirements. If recognized by the regulator, computed-generated data could address some of these gaps in a more cost-effective way, thereby minimizing cost escalation in the healthcare system while preserving access to a full set of device options.

One goal of proactive (e.g., clinical trials) and reactive (e.g., complaints) post-market surveillance is to monitor the safety of a medical device by detecting the incidents generated by its use, whether they are device-related, instrument-related, or procedure-related. ISCT is a tool that could detect potential incidents on virtual populations, thus triggering solutions in advance of clinical use. ISCT could thus indirectly reduce healthcare costs due to hospitalization of patients that suffer from complications, if factors contributing to such complications could be identified and corrected pre-clinically.

Potential Benefits for the Device Manufacturer

By augmenting clinical trials, ISCT may reduce clinical sample size and lower associated costs. By detecting potential incidents in advance of clinical use, ISCT may reduce the costs of remediating defective product. By providing compelling scientific data, ISCT may drive sales. These financial benefits are offset by the costs to develop and execute ISCT studies (which may include specialized personnel, software licenses, data storage, and acquisition of clinical data for the validation).

The most significant benefit to a manufacturer is the potential to launch innovative products more quickly, by reducing the time of clinical trials without compromising on the credibility of the clinical evidence. As an example, early migration evaluated *via* Roentgen Stereogrammetric Analysis (RSA) two years after surgery is able to identify implants at risk of revision before actual failure,[36] thus shortening the window of clinical observation from 10 years to 2 years. In several European countries (e.g., Sweden), a 2-year RSA study can now be used as pre-market evaluation of new implant designs, significantly reducing the time to market (a PMCF study is still required as part of post-market surveillance).[39] Here, early implant migration measured by RSA serves as a surrogate for long term implant loosening. A similar impact could be provided by ISCT, namely ISCT data can be used as a surrogate for long term clinical assessment of patient risks. With sufficient validation against long-term clinical data, ISCT could reduce the time to market for new devices.

Potential Benefits for the Regulatory Bodies

The regulator has the responsibility of ensuring products meet the essential regulatory requirements, and approving or rejecting access to the marketplace based on that evaluation. This decision is based on documents provided by the manufacturer, of which the clinical data play a key role. By filling gaps in clinical data, increasing the statistical power of the study, and providing a better understanding of the implant's behavior, ISCT brings additional clinical evidence that can increase confidence by the regulator in the expected clinical performance of the device, and reduce the number of revision rounds during conformity assessment.

BARRIERS TO ISCT APPLICATIONS

Despite these numerous motivations, four main barriers still prevent large-scale use of ISCT in support of a regulatory submission: the large number of patient risks typically included in a clinical study, various challenges of a technical nature, establishing adequate model credibility and regulatory uncertainty.

Barrier 1: Large Number of Potential Patient Risks

Published population level studies have paved the way for ISCT by demonstrating the technical feasibility to study a specific patient risk on a large number of virtual patients.[1,2,12,31,38,40,44] These analyses generally

considered one failure mode in complete isolation; however, clinical studies encompass the large number of patient risks that reflect the total performance of the medical device. Determination of the potential patient risks in scope of an ISCT is necessary to ensure the device performance is sufficiently addressed while maximizing the utility of the ISCT. Unfortunately, there exists currently no guidance on which risks can or should be addressed within an ISCT. For example, a single study has simultaneously investigated femoral neck fractures, prosthesis fracture and aseptic loosening of a prosthesis,[40] but did not justify the decision path to the choice of these specific failure modes.

The spectrum of possible complications included in a clinical study may vary depending on the considered orthopedic implant. While evaluation of infection or pain are generally part of any clinical study, scapular notching is for example a failure mode specific to a reverse shoulder implant. On the other hand, some implant types have a more limited range of complications. For example, trauma implants may need to stay in the body only as long as the bone needs to heal, so many long-term risks that are included in a clinical study for joint replacement would not need to be addressed here. Similarly, the clinical study of a dental implant would require an assessment of primary stability, but does not have a need for range of motion.

One proposed approach for identifying the appropriate potential patient risks for inclusion in an ISCT could balance (1) the severity and probability of the patient risk (2) the impact of the implant design on the patient risk and (3) the technical feasibility of evaluating the patient risk measure *via* ISCT (Fig. 5).

First, ISCT may not be required to address lower risks to the patient, enabling focus on higher patient risk. The risk for the patient is generally measured as a function of severity and probability of the harm. Severity can be defined by following existing guidance,[33] but it is more difficult to identify an appropriate probability of occurrence of a harm. If the device is very new, this may turn out to be even more challenging. National registries, the literature or complaint data can provide valuable data from similar devices. For example, within shoulder arthroplasty, humeral stem fracture has a generally low probability of occurrence based on national registry data.[11] Because of that low occurrence (and therefore low patient risk), augmenting the clinical study with an ISCT evaluation of stem fracture is not likely to impact patient care. On the other hand, implant loosening is a failure mode that occurs more frequently than stem fracture,[11] and can lead to implant revision (relatively high severity); as such, examination of implant loosening within an ISCT would be a meaningful augment to clinical data.
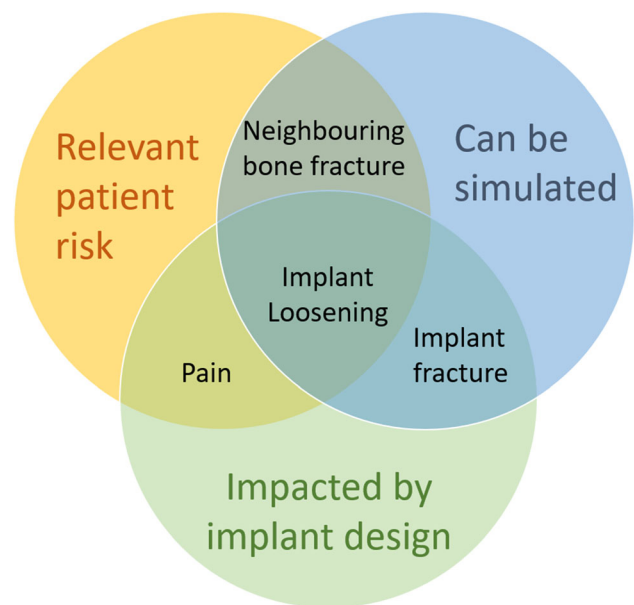

FIGURE 5. Classification of risks depending on relevance, ability to simulate and impact from design.

Second, the impact of the implant design on the potential complication can be considered. For a shoulder replacement for example, risks such as stress shielding[37,45] or poor range of motion can be impacted by implant design,[6,22] and as such could be considered for inclusion in the ISCT. On the other hand, the influence of implant design on risks such as hematoma or delayed wound healing has not been reported to date, and thus inclusion in an ISCT would not add new information on the expected device performance.

Third, critical examination of the modeling basis for the different patient risks is required. For example, stem fracture, cortex perforation or implant range of motion can be appropriately simulated with contemporary modeling techniques (Table 1). On the other hand, the state-of-the-art simulations for other risks such as pain or metallosis are either not sufficiently mature or non-existent, and thus would have insufficient credibility within an ISCT to provide sound predictions on device performance.

A key limitation to this three-tiered approach is that it relies heavily on patient risks that have already been identified, either based on the observed clinical history of predicate devices intended for the same patient indications and/or through a risk analysis on the candidate device. Once identified, the risks can be triaged following the considerations above. Those risks that have not been identified, however, would generally not be addressed *via* ISCT; though it is possible that a computational model developed to study one particular patient risk might reveal information to focus attention on an alternative and previously unidentified

risk, this serendipity should generally not be expected. As such, ISCT data generally will enrich but not replace clinical data, preserving the important role of the clinical study to identify potential new patient risks.

Barrier 2: Technical Challenges

For an ISCT platform to be efficient, models need to be effectively developed and executed. Key features are therefore multi-functionality, adequate software and hardware architecture, and high levels of automation. Also, the right experts need to be part of the project.

For a company offering replacement solutions for many joints in the body, an important technical aspect is that an ISCT platform should be multi-functional, and be able to handle applications as different as a knee, an elbow, or an ankle replacement. Also, the initial investments in building an ISCT platform being high, the platform should be built to be reusable from the beginning. To that end, a possible solution consists of designing the software architecture to comprise a software core that handles the generic aspects of such simulations (Fig. 6). Specifically, the core may include algorithms for positioning the implant in the bone, the definition of the different possible contacts between the bone and the implant (e.g., bonded, frictional), the different force application types (joint reaction force,

individual muscle forces), application of material definitions to the different components, etc. Modules can then identify the model configuration and prescribe the input parameters for the specific implants, such as the angles and distances for positioning the implant within the bone, the coefficient of friction, the force magnitude and orientation, the modulus of elasticity of the components, etc. Such an infrastructure ensures the high costs needed to support development and validation of the initial ISCT application are significantly reduced for subsequent applications.

A key advantage of a virtual patient is that a potential patient risk that may take years to assess *in vivo* can be simulated in hours or less *in silico*. However, in order to process a large number of models (= number of anatomies x number of configurations x number of potential patient risks) in a reasonable amount of time, it is essential that the simulations are automated and optimized as much as possible. Removing manual work not only allows faster simulations, it also provides the advantage that all models are processed in a consistent way, removing the risk of human errors that would invariably be associated with the creation, analysis, and post-processing of thousands of models. Patient-modeling technologies are evolving at a fast pace. Sampling strategies can be implemented to capture the population variability with
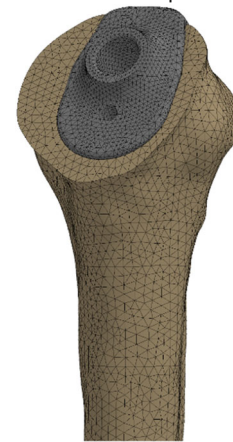

TABLE 1. Examples of typical measurements in clinical studies (left), corresponding modeling results (right), and representative reference.

| In vivo patient risk	In silico output
Aseptic loosening	Micromotion/Interface strain[1]
Implant subsidence	Implant motion[14]
Stress shielding	Change in pre- and post-surgery bone stress/strain[31]
Intra/post-surgery bone fracture	Bone stress/strain[18]
Implant fracture	Implant stress[9]
Impingement	Range of motion[22]
Joint stability	Compressive/shear force ratio to dislocation[24]
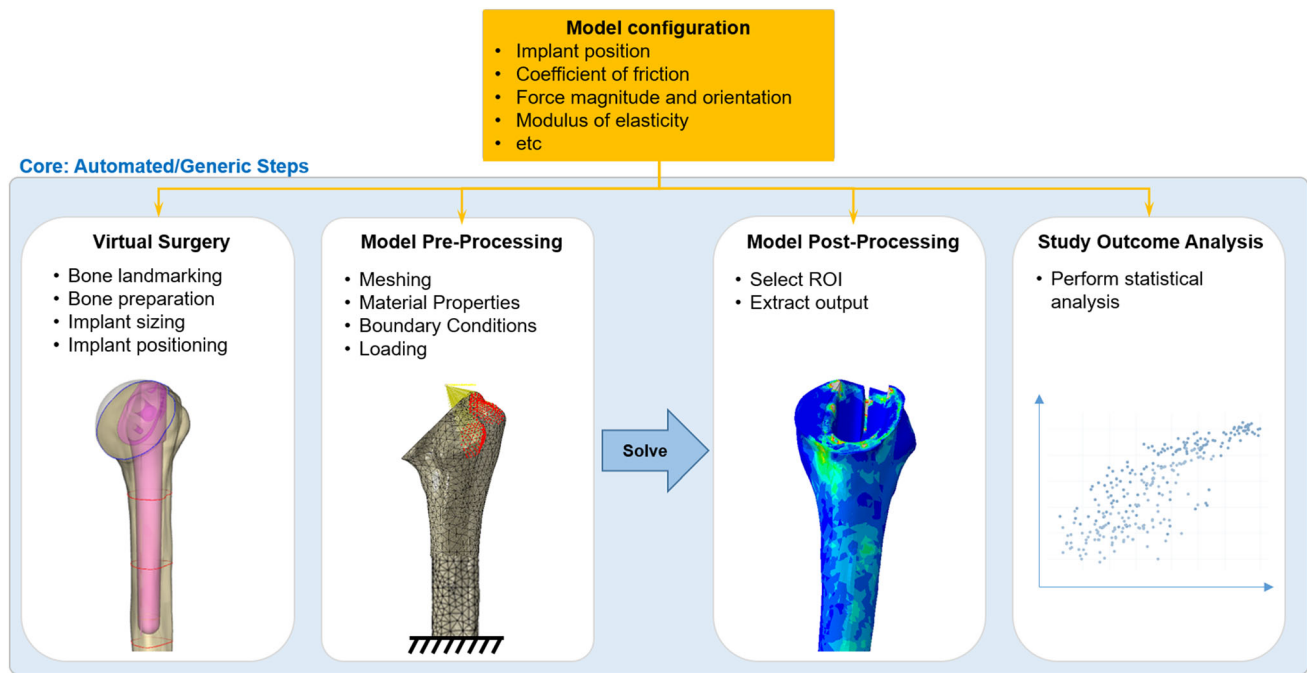
FIGURE 6. Software architecture to permit modularity of the ISCT platform. The model configuration contains the parameters specific to the implant at hand. The core contains the generic steps common to the simulations of any implant type.

a select subset of patient models.[5,41] Mechanistic simulations are already being replaced by surrogate modeling i.e., machine learning algorithms trained on ISCT datasets.[3] The calculation time can also be significantly reduced by carefully designing the software architecture, and optimizing hardware configurations using high-performance computers. For example, if model preparation and solving can be parallelized instead of being run in series, computational time can be reduced drastically. Parallel computing is directly limited by the number of available software licenses, motivating an appropriate balance between license costs and computational speed. The problem of a limited number of available software licenses may be solved by using open source software, where licenses are available at no cost. While this may look like the obvious universal solution, the benefit may be offset by the time needed for the validation of computer software. Computerized system validation involves demonstrating that the computerized system does exactly, consistently and reproducibly what it is designed to do.[34] One needs to demonstrate that the software program is installed properly, and that the results of various benchmark tests against analytical solutions or previous results are appropriate. Proof of software validation must also be provided for any non-commercial code that would be developed internally (e.g., algorithms developed in house), or externally (open source), and requires an additional substantial amount of code testing, code peer review, and proper docu-

mentation. This represents such an extensive work that larger companies often have a dedicated computerized software validation department. The time required for these activities should not be underestimated, keeping in mind that it needs to be updated with any new software version release. For this reason, commercial off-the-shelf software (or well-established open source software) is often preferred, as a major part of software quality assurance is already provided by the software provider. Also, complexity and software validation can be reduced by using as few different software programs as possible.

Barrier 3: Model Credibility

An erroneous decision on performance and/or safety of a medical device due to wrong model assumptions may have direct impact on patient safety. It is therefore of utmost importance to ensure that computational models appropriately represent reality, a goal accomplished by model verification and validation. Establishing and demonstrating sufficient model credibility must be done for each and every patient risk to be included in an ISCT.

A framework to do so is provided by the American Society of Mechanical Engineers guidelines for verification and validation (V&V) of computational models.[7] The purpose of verification is to ensure that the mathematical model is implemented correctly and solved accurately. The verification component of V&V

can be addressed through activities including mesh convergence on the quantities of interest in the context of use, sensitivity study on any non-default key solver parameters, and peer review of key model inputs/outputs. Verification also requires the use of validated software, as described earlier.

The use of ASME V&V 40 guidelines for validation of ISCT is much more limited, as the ISCT application differs from traditional use of computer models in several important ways. Traditional modeling applications are generally focused on assessing worst case configuration within a precisely defined benchtop test needed for design verification, whereas ISCT applications are focused on simulating the clinical use environment which includes significant variability due to patient, surgeon, use conditions, and other clinical variables. Also, by virtue of the connection with benchtop testing, sufficient validation of traditional modeling applications can generally occur *via* well-characterized and well-controlled benchtop tests, whereas sufficient validation of ISCT applications generally entails use of clinical data (such as complaint data, international registries, and clinical data published in the literature) which are far less characterized and controlled.

ASME V&V 40 provides a framework to demonstrate that a model captures the physics of the device by comparison to benchtop testing. Use of a model in the context of ISCT requires additionally to show that the model reproduces clinical findings. This can be accomplished through use of a clinical comparator. However, guidance on the appropriate use of a clinical comparator for purposes of ensuring sufficient clinical model credibility is still lacking. Several efforts are underway to provide such guidance, including a working group of ASME V&V 40 that is specifically focused on application of the V&V 40 credibility strategy to clinical comparators, and a working group on good simulation practices that aims to create the equivalent of good clinical practices for in silico clinical trials within the Avicenna Alliance (https://avicenna-alliance.com). Such documents will allow the developer to ensure sufficient model credibility is reached, using both benchtop and clinical comparators, and will guide both the device manufacturer and the regulator in judging model credibility.

Barrier 4: Regulatory Uncertainty

Any study used in support of a regulatory submission is subject to regulatory review. Signals of ISCT acceptance are not global, and there is regulatory uncertainty for those devices that are targeted for global commercialization. However, several actions can be taken to increase the chances of success.

First, regulatory agencies will look more favorably at ISCT if it follows established international standards. These include, for example, guidelines on clinical studies,[35] numerical modeling[25] and software validation.[34] Adherence to these standards will assist the regulator in ensuring that applicable procedures have been followed.

Second, it is important to consider global utilization of the medical device. While there are signals that the regulatory bodies in the United States[20] and Europe are open to the ISCT paradigm, key regulatory agencies in other target markets have not communicated on the topic. As long as there is no global regulatory acceptance of an ISCT approach, global developers must meet the most demanding bar for clinical data evidence, thereby compromising some of the benefits of ISCT development and deployment. The work of international societies such as the Avicenna Alliance to establish consensus standards will encourage worldwide adoption of ISCT.

CONCLUSIONS

The orthopedic industry has traditionally been using numerical methods to simulate benchtop tests that are performed in a controlled manner. Recent techniques now enable simulation of how the product interacts with a patient, or in the population, encompassing the variability representative of *in vivo* use. In that sense, simulation is now bridging the gap between pre-clinical and clinical testing. By complementing clinical data with computational evidence, known challenges in acquiring clinical data across the spectrum of patients, indications, and configurations that characterize contemporary medical devices may be overcome. Potentially less risky, less expensive, faster, more efficient safety and performance evaluations could provide benefit to patients, healthcare systems, manufacturers and regulators.

However, before ISCT can be used successfully in the medical device industry, four main barriers must be overcome. First, the potential patient risks to include in an ISCT need to be identified; focusing on the failure modes that represent a greater risk for the patient, that are impacted by the implant design, and that can be accurately simulated maximizes the benefit of augmenting a clinical study with an ISCT. Second, given the large number of complex models to solve, an ISCT platform needs to run efficiently, and given the required investment, a platform should be (re)usable for many different products; this can be done by optimizing the software and hardware architectures. Third, model credibility must be provided; existing standards can be used as a basis, but a validation

strategy against clinical data still needs to be developed. Fourth, the regulatory uncertainty is an inevitable risk; however, following relevant existing standards and actively promoting ISCT globally can increase the chances of success.

New areas of medicine such as personalized medicine, where treatment can be better tailored to each individual patient, and the use of a digital twin, where a personalized model for an individual patient is continuously adjusted based on tracked health and lifestyle parameters to predict potential problems, can greatly benefit from advances in ISCT. The barriers that we identified here first need to be addressed before the possibilities offered by these more advanced applications of models can be fully realized.

CONFLICT OF INTEREST

The authors are (PF, GM, AH, DH, DC, JEB) or were (DC) employees of Zimmer Biomet, and therefore they have received/will receive benefits for personal or professional use from Zimmer Biomet related directly or indirectly to the subject of this manuscript.

REFERENCES

[1]Al-Dirini, R. M. A., S. Martelli, D. Huff, J. Zhang, J. G. Clement, T. Besier, and M. Taylor. Evaluating the primary stability of standard vs lateralised cementless femoral stems—a finite element study using a diverse patient cohort. *Clin. Biomech.* 59:101–109, 2018.

[2]Al-Dirini, R. M. A., S. Martelli, D. O'Rourke, D. Huff, J. Zhang, J. G. Clement, T. Besier, and M. Taylor. Virtual trial to evaluate the robustness of cementless femoral stems to patient and surgical variation. *J. Biomech.* 82:346–356, 2019.

[3]Al-Dirini, R. M. A., S. Martelli, and M. Taylor. Computational efficient method for assessing the influence of surgical variability on primary stability of a contemporary femoral stem in a cohort of subjects. *Biomech. Model. Mechanobiol.* 19:1283–1295, 2020.

[4]Aldieri, A., M. Terzini, A. L. Audenino, C. Bignardi, and U. Morbiducci. Combining shape and intensity dxa-based statistical approaches for osteoporotic HIP fracture risk assessment. *Comput. Biol. Med.* 127:104093, 2020.

[5]Ali, A. A., C. W. Clary, L. M. Smoger, D. A. Dennis, C. K. Fitzpatrick, P. J. Rullkoetter, and P. J. Laz. Computational framework for population-based evaluation of TKR-implanted patellofemoral joint mechanics. *Biomech. Model. Mechanobiol.* 19:1309–1317, 2020.

[6]Arenas-Miquelez, A., R. J. Murphy, A. Rosa, D. Caironi, and M. A. Zumstein. Impact of humeral and glenoid component variations on range of motion in reverse geometry total shoulder arthroplasty: a standardized computer model study. *J. Shoulder Elbow Surg.* 30(4):763–771, 2020.

[7]ASME-V&V40. Assessing Credibility of Computational Modeling through Verification and Validation, 2018.

[8]ASTM-F1839-08. Standard Specification for Rigid Polyurethane Foam for Use as a Standard Material for Testing Orthopaedic Devices and Instruments, 2016.

[9]ASTM-F2996-20. Standard Practice for Finite Element Analysis (FEA) of Non-Modular Metallic Orthopaedic Hip Femoral Stems, 2020.

[10]ASTM-F3334-19. Standard Practice for Finite Element Analysis (FEA) of Metallic Orthopaedic Total Knee Tibial Components, 2019.

[11]Australian Orthopaedic Association National Joint Replacement Registry. Hip, Knee & Shoulder Arthroplasty: 2019 Annual Report.

[12]Awadalla, M., R. M. A. Al-Dirini, D. O'Rourke, L. B. Solomon, M. Heldreth, and M. Taylor. Influence of varying stem and metaphyseal sleeve size on the primary stability of cementless revision tibial trays used to reconstruct AORI IIA defects. A simulation study. *J. Orthop. Res.* 36:1876–1886, 2018.

[13]Bischoff, J. E., Y. Dai, C. Goodlett, B. Davis, and M. Bandi. Incorporating population-level variability in orthopedic biomechanical analysis: a review. *J. Biomech. Eng.* 136:021004, 2014.

[14]Bischoff, J. E., O. C. O'Reilly, J. V. Nepola, and B. M. Patterson. The influence of over-reaming on stem stability in reverse shoulder arthroplasty. *JSESArthro* 30:123–131, 2020.

[15]Blanchard, R., C. Morin, A. Malandrino, A. Vella, Z. Sant, and C. Hellmich. Patient-specific fracture risk assessment of vertebrae: a multiscale approach coupling X-ray physics and continuum micromechanics. *Int. J. Numer. Method Biomed. Eng.* 2016. https://doi.org/10.1002/cnm.2760.

[16]Chandran, V., G. Maquer, T. Gerig, P. Zysset, and M. Reyes. Supervised learning for bone shape and cortical thickness estimation from CT images for finite element analysis. *Med. Image Anal.* 52:42–55, 2019.

[17]Cook, D. D., and D. J. Robertson. The generic modeling fallacy: average biomechanical models often produce non-average results!. *J. Biomech.* 49:3609–3615, 2016.

[18]Dahan, G., N. Trabelsi, O. Safran, and Z. Yosibash. Verified and validated finite element analyses of humeri. *J. Biomech.* 49:1094–1102, 2016.

[19]Dai, Y., G. R. Scuderi, C. Penninger, J. E. Bischoff, and A. Rosenberg. Increased shape and size offerings of femoral components improve fit during total knee arthroplasty. *Knee Surg. Sports Traumatol. Arthrosc.* 22:2931–2940, 2014.

[20]Faris, O., and J. Shuren. An FDA Viewpoint on Unique Considerations for Medical-Device Clinical Trials. *N. Engl. J. Med.* 376:1350–1357, 2017.

[21]Favre, P., and A. D. Henderson. Prediction of stemless humeral implant micromotion during upper limb activities. *Clin. Biomech.* 36:46–51, 2016.

[22]Favre, P., B. Moor, J. G. Snedeker, and C. Gerber. Influence of component positioning on impingement in conventional total shoulder arthroplasty. *Clin. Biomech.* 23:175–183, 2008.

[23]Favre, P., J. Seebeck, P. A. Thistlethwaite, M. Obrist, J. G. Steffens, A. R. Hopkins, and P. A. Hulme. In vitro initial stability of a stemless humeral implant. *Clin. Biomech.* 32:113–117, 2016.

[24]Favre, P., M. Senteler, J. Hipp, S. Scherrer, C. Gerber, and J. G. Snedeker. An integrated model of active glenohumeral stability. *J. Biomech.* 45:2248–2255, 2012.

[25]FDA Guidance for Industry and Food and Drug Administration Staff. Reporting of Computational Modeling Studies in Medical Device Submissions.

[26]Fennema, P., and A. Hassan. Getting your devices ready for MDR compliance–a clinical approach and orthopaedic device manufacturers' perspective. *AO* 5:70–75, 2019.

[27]Geris, L. Regenerative orthopaedics: in vitro, in vivo…in silico. *Int. Orthop.* 38:1771–1778, 2014.

[28]Heimann, T., and H. P. Meinzer. Statistical shape models for 3D medical image segmentation: a review. *Med. Image Anal.* 13:543–563, 2009.

[29]Helgason, B., E. Perilli, E. Schileo, F. Taddei, S. Brynjolfsson, and M. Viceconti. Mathematical relationships between bone density and mechanical properties: a literature review. *Clin. Biomech.* 23:135–146, 2008.

[30]Hellmich, C., C. Kober, and B. Erdmann. Micromechanics-based conversion of CT data into anisotropic elasticity tensors, applied to FE simulations of a mandible. *Ann. Biomed. Eng.* 36:108–122, 2008.

[31]Heyland, M., S. Checa, D. Kendoff, and G. N. Duda. Anatomic grooved stem mitigates strain shielding compared to established total hip arthroplasty stem designs in finite-element models. *Sci. Rep.* 9:482, 2019.

[32]Humbert, L., Y. Martelli, R. Fonolla, M. Steghofer, S. Di Gregorio, J. Malouf, J. Romera, and L. M. Barquero. 3D-DXA: assessing the femoral shape, the trabecular macrostructure and the cortex in 3D from DXA images. *IEEE Trans. Med. Imaging* 36:27–39, 2017.

[33]IMDRF. "Software as a Medical Device": Possible Framework for Risk Categorization and Corresponding Considerations IMDRF/SaMD WG/N12FINAL:2014, 2014.

[34]ISO-13485. Medical devices—Quality management systems—Requirements for regulatory purposes. Geneva: ISO, 2016.

[35]ISO-14155. Clinical Investigation of Medical Devices for Human Subjects—Good Clinical Practice, 3rd ed. Geneva: ISO, 2020.

[36]Karrholm, J., R. H. Gill, and E. R. Valstar. The history and future of radiostereometric analysis. *Clin. Orthop. Relat. Res.* 448:10–21, 2006.

[37]Langohr, G. D. G., J. Reeves, C. P. Roche, K. J. Faber, and J. A. Johnson. The effect of short-stem humeral component sizing on humeral bone stress. *J. Shoulder Elbow Surg.* 29:761–767, 2020.

[38]Lerch, M., H. Windhagen, A. E. Kurtz, S. Budde, B. A. Behrens, A. Bouguecha, and A. Almohallami. 'Pre-launch' finite element analysis of a short-stem total hip arthroplasty system consisting of two implant types. *Clin. Biomech.* 61:31–37, 2019.

[39]Malak, T. T., J. A. Broomfield, A. J. Palmer, S. Hopewell, A. Carr, C. Brown, D. Prieto-Alhambra, and S. Glyn-Jones. Surrogate markers of long-term outcome in primary total hip arthroplasty: a systematic review. *Bone Jt. Res* 5:206–214, 2016.

[40]Martelli, S., F. Taddei, E. Schileo, L. Cristofolini, N. Rushton, and M. Viceconti. Biomechanical robustness of a new proximal epiphyseal hip replacement to patient variability and surgical uncertainties: a FE study. *Med. Eng. Phys.* 34:161–171, 2012.

[41]O'Rourke, D., M. Bottema, and M. Taylor. Sampling strategies for approximating patient variability in population-based finite element studies of total hip replacement. *Int. J. Numer. Method. Biomed. Eng.* 35:e3168, 2019.

[42]Pappalardo, F., G. Russo, F. M. Tshinanu, and M. Viceconti. In silico clinical trials: concepts and early adoptions. *Brief. Bioinform.* 20:1699–1708, 2019.

[43]Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC.

[44]Sas, A., P. Pellikaan, S. Kolk, P. Marty, T. Scheerlinck, and G. H. van Lenthe. Effect of anatomical variability on stress-shielding induced by short calcar-guided stems: automated finite element analysis of 90 femora. *J. Orthop. Res.* 37:681–688, 2019.

[45]Schmidutz, F., Y. Agarwal, P. E. Muller, B. Gueorguiev, R. G. Richards, and C. M. Sprecher. Stress-shielding induced bone remodeling in cementless shoulder resurfacing arthroplasty: a finite element analysis and in vivo results. *J. Biomech.* 47:3509–3516, 2014.

[46]Viceconti, M., C. Cobelli, T. Haddad, A. Himes, B. Kovatchev, and M. Palmer. In silico assessment of biomedical products: the conundrum of rare but not so rare events in two case studies. *Proc. Inst. Mech. Eng. H* 231:455–466, 2017.

[47]Viceconti, M., A. Henney, and F. Morley-Fletcher. Avicenna Consortium—In Silico Clinical Trials: How Computer Simulation will Transform the Biomedical Industry. Research and Technological Development Roadmap. Brussels: Avicenna Consortium, 2016.

[48]Viceconti, M., F. Pappalardo, B. Rodriguez, M. Horner, J. Bischoff, and F. Musuamba Tshinanu. In silico trials: Verification, validation and uncertainty quantification of predictive models used in the regulatory evaluation of biomedical products. *Methods* 185:120–127, 2020.