On validation of multibody musculoskeletal models

Lund, Morten Enemark; de Zee, Mark; Andersen, Michael Skipper; Rasmussen, John

Published in:
Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine

DOI (link to publication from Publisher):
10.1177/0954411911431516

Publication date:
2012

Document Version
Publisher's PDF, also known as Version of record

Special Issue Article

Institution of MECHANICAL ENGINEERS
Journal of ENGINEERING IN MEDICINE

Proc IMechE Part H:
J Engineering in Medicine
226(2) 82–94

DOI: 10.1177/0954411911431516
pih.sagepub.com
SAGE

On validation of multibody musculoskeletal models

Morten Enemark Lund[1], Mark de Zee[2], Michael Skipper Andersen[1] and John Rasmussen[1]

Abstract

We review the opportunities to validate multibody musculoskeletal models in view of the current transition of musculoskeletal modelling from a research topic to a practical simulation tool in product design, healthcare and other important applications. This transition creates a new need for justification that the models are adequate representations of the systems they simulate. The need for a consistent terminology and established standards is identified and knowledge from fields with a more progressed state-of-the-art in verification and validation is introduced. A number of practical steps for improvement of the validation of multibody musculoskeletal models are pointed out and directions for future research in the field are proposed. It is hoped that a more structured approach to model validation can help to improve the credibility of musculoskeletal models.

Keywords

Musculoskeletal modelling, validation

Date received: 25 April 2011; accepted: 9 November 2011

Introduction

Multibody musculoskeletal simulation and model based estimation of muscle forces is gaining a central role in decision-making processes within product design[1–3] and orthopaedics.[4,5] This development necessitates a justified confidence that the models are adequate representations of the systems they simulate, and it raises the question of how multibody musculoskeletal models can be critically evaluated. Which tests are necessary to clarify the assumptions, limitations and uncertainties of the model simulation and how can the results of such tests be assessed? This is the central question of the present work, and we shall review literature from the biomechanics field and from entirely different fields that also rely on the accuracy of models and have progressed further in their insights on the topic.

Computer simulations and models of the human musculoskeletal system have evolved over the past decades. Development of multibody musculoskeletal modelling has been driven not only by advances in scientific research,[6] but also by the advent of commercial musculoskeletal modelling systems. System developments such as AnyBody,[7] Lifemodeler,[8] SIMM[9] and OpenSIM[10] have recently made musculoskeletal models available to science and industry.

The use of computational engineering for practical applications adapted to the individual patient holds significant promise and enormous challenges.[11] Individually targeted medical care was recently declared by the National Academy of Engineering to be one of the major challenges for engineering in the 21st century.[12] The expectations of what can be achieved in healthcare with the help of computational models continue to grow; fuelled by the increasing speed and capabilities of computer systems and by the impact computer simulations have had in traditional areas of engineering.

Recent studies have shown the potential of musculoskeletal simulation tools to help clinicians answer questions or simulate quantities, which were otherwise impractical or impossible to obtain experimentally.[13,14] Output from patient-specific musculoskeletal simulations may, for instance, determine the strategy

[1]Department of Mechanical and Manufacturing Engineering, Aalborg University, Denmark
[2]Department of Health Science and Technology, Aalborg University, Denmark

Corresponding author:
John Rasmussen, Department of Mechanical and Manufacturing Engineering, Aalborg University, Fibigerstraede 16, Aalborg East 9220, Denmark.
Email: jr@m-tech.aau.dk

for a surgical procedure or directly influence diagnostics and decisions on treatment.[15–18] Simulation results can also determine ergonomic product design or be used for design optimization of orthopaedic implants and orthoses.[1,4,5]

A consequence of the transition from science to clinical applications is that the results of musculoskeletal simulations attain a direct influence on applications and processes where mistakes and misjudgements are potentially critical. Diagnostics and surgical planning are obvious examples where consequences for the patient are immediate, but use of musculoskeletal simulations in product design may also have serious consequences, albeit typically on a longer time scale.

Computer-aided engineering

Computer-aided engineering (CAE) has conquered nearly every aspect of traditional science and engineering. Today, it is inconceivable to design a car, build a bridge or create an advanced electrical circuit without the use of computational models. Simulation of musculoskeletal systems can be interpreted as yet another field for CAE, and it may be valuable to consider the development of musculoskeletal simulation in light of the evolutions of related CAE fields.

The argument for the practical use of a computational model is the opportunity to acquire data or knowledge that is difficult to obtain experimentally. This means that the computational model will reach its full potential only when it is considered reliable enough not to require experimental evidence of the validity of each new result. In other words, the model and underlying methods must have been sufficiently proven or evaluated in a general sense to provide confidence in individual results.

The history of CAE contains many examples of application fields that have evolved from scientific developments, often through setbacks, so that it has now reached a level of maturity where they are widely used as a basis for critical design decisions: structural and solid mechanics are examples of application areas where CAE has revolutionized engineering. Three decades ago, predictions by finite element models (FEMs) were usually accompanied by experimental results to support computational predictions. Today, many prototypes in the automotive and civil engineering industries are exclusively built and tested with computer-aided tools. Physical experiments are only used in cases where FEM codes are known to be less reliable.[19]

Verification and validation

Traditional fields of CAE would not be the important tools they are today in the absence of the tremendous efforts spent in scrutinizing the models. The totality of these efforts, procedures, methodologies and processes has been formalized into the field of verification & validation (V&V). V&V is the process of assessing the accuracy and validity of computer models and computational predictions. The subject spans from philosophical issues to practical procedures on how model reliability is quantified.

Basically, model verification deals with how the model is implemented and the numerical accuracy of the solution. Model validation, on the other hand, is the process of determining how well the model represents the real world and experimental data. Roache[20] in his book on V&V tries to simplify the concepts by stating that verification deals with mathematics whereas validation deals with physics, but, as we shall see later, making such a clear distinction in musculoskeletal modelling can be problematic.

The fundamental goal of a validation process is to assess the computer simulation's ability to predict certain variables of interest. Thus, the main topic of validation is the comparison of computational results with experimental data to quantify uncertainty and errors in the computational model. The validation process has often been divided into three distinct parts:[21]

(a) comparison between computational and experimental results;
(b) extrapolating the model prediction to conditions where the model is intended to be used;
(c) determining whether the accuracy is sufficient for the intended use.

In musculoskeletal modelling, the first issue covers subjects such as validation experiments and model calibration, which usually is the topic of publications reporting validation studies in biomechanics. However, the American Society of Mechanical Engineers (ASME) guideline on validation in solid mechanics considers all three issues to be a part of the validation process.[22] This broader definition of the validation process emphasises that validation is also the responsibility of the final user of the model. Given users' different applications of the same models, there is no definite way of judging what level of accuracy is sufficient for a model to be called 'validated'. The accuracy requirement depends on the application and not the model itself. Thus, when existing musculoskeletal models in generally available modelling systems are applied to specific problems, it is up to the user to complete the validation process and assess whether the model is accurate enough to be used for the intended purpose.

Figure 1 illustrates the process of validation in the context of musculoskeletal modelling. After a model has been proposed to address a specific question, the process is split into modelling and experimental branches. Figure 1 shows the close interaction required between design of the computational model and design of the validation experiment. At this point, it is important to stress that the interaction does not entail changing either the model or the experiment in response to output. In fact, the output of the model and experiment

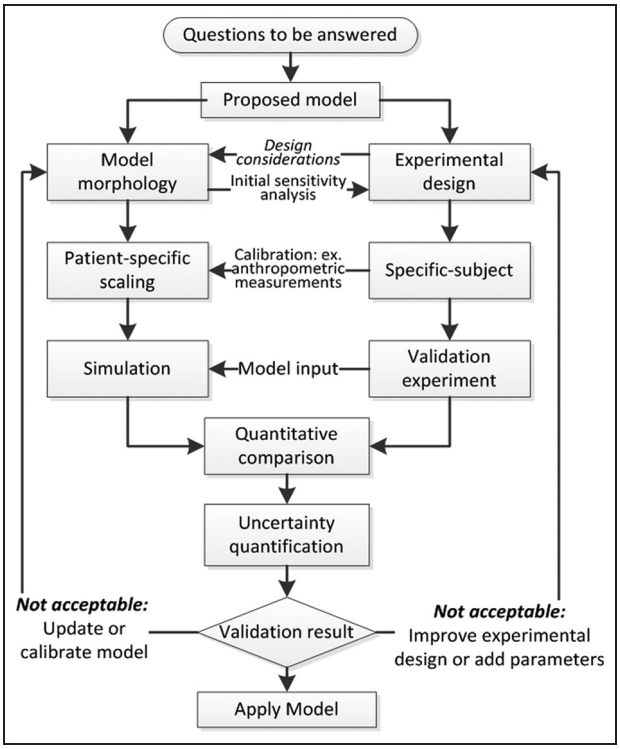

Figure 1. Illustration of the validation process in the context of musculoskeletal modelling. The validation process begins with the question to be explored by the model. The activities of validation are split into an experimental and a modelling branch. Despite dependencies between the branches, the output of each branch should not be revealed to avoid either branch being adjusted to match the results. When outputs are ready, the model and experiment are quantitatively compared and, if possible, the uncertainty of the result is quantified. This validation result can then be compared to the requirements of a specific application before the model is applied. If validation results and requirements do not match, then the model may be revised, or the experiment improved to reduce experimental uncertainty. The activities are inspired by the ASME 'Guide for verification and validation in computational solid mechanics'.[22]

should not be revealed until data are ready for comparison. We return to the concept of blind validation in the section 'Model calibration versus blind validation'. However, the model and experiment may influence each other in other ways. For example, an *initial sensitivity analysis* (Figure 1) can suggest the most important variables which must be controlled and measured in the validation experiment. Likewise, *design considerations* (Figure 1) may influence model design in cases where experimental limitations prevent precise knowledge of certain variables, for example, allowing the model to handle intervals for which certain parameters vary in the experiment or across the population. *Uncertainty quantification* is the last important step in Figure 1 before validation results are compared with the requirements of a given model. Uncertainty information can come from repeated measurements in the experimental case and from sensitivity and reliability analysis in case of model output.

It is important to realize that it is not the computational model that is validated, but rather the mathematical or conceptual model behind it, which relates to a scientific theory or hypothesis. Thus, any errors and uncertainties caused by implementation issues, discretization errors or software bugs are usually not defined to be part of the validation process, but belong instead to the verification process. It is a point of discussion as to whether this distinction between validation and verification can be made with the same precision in musculoskeletal modelling. For instance, the muscle configuration in a musculoskeletal model can be viewed as a discretization of continuous muscles, and discretization is usually not considered to be a part of the validation process.[23] However, excluding the muscle configuration would be contrary to how validation is usually perceived in musculoskeletal modelling. Another case, where it is difficult to distinguish validation from verification, is checking that the code follows the basic laws of physics. In general, we adopt an approach that considers both these cases as validation because they deal with the correspondence between the model and a physical/physiological reality.

Finally, while areas such as solid mechanics or fluid dynamics offer good opportunities to establish gold standards for correct results through analytical models of simple problems and accurate experimental methods, the same is not the case for multibody musculoskeletal modelling simply because it is very challenging to accurately measure internal mechanical parameters in a living organism.

Although a clear distinction between verification and validation may be difficult, it is important to realize that verification is a prerequisite for validation. Verification provides the evidence that the computer code correctly solves the underlying mathematical model. Absence of verification creates a risk of mixing modelling errors and errors caused by implementation issues.

Model verification consists of two parts: the first, sometimes called solution verification, quantifies how the mathematical model is different from the discretized model. The second part, called code verification, investigates the actual software implementation of the algorithm. Software bugs are found by subjecting the algorithm to a series of benchmarking tests against known analytical solutions.

Defining a consistent terminology

For obvious reasons, some areas of engineering have a much more established tradition for V&V than is the case for biomechanics. V&V has been an issue of much debate and studies within areas such as hydrology, nuclear reactor safety, astronautics and aeronautics.[21,24,25] Important general works on V&V in the engineering literature are the review articles by

Oberkampf and Trucano[21,26] Oberkampf et al.[23] and Oberkampf and Barone,[27] the book by Roache[20] and the standards and guidelines published by the American Institute of Aeronautics and Astronautics[28] and ASME.[22,29,30]

A persistent problem in establishing a common view on V&V is the confusion in terminology.[24,25,31,32] Verification and validation have become the names of two distinct processes in V&V, but in disciplines without an established tradition for V&V, concepts and terminology are frequently used interchangeably, and biomechanics is no exception. Confusion also arises since the words verification and validation clash with the layman's view that a validated or verified model is also valid or true in the common sense of the word. Nigg and Herzog[33] advocate that the term *model validation* is dropped completely for musculoskeletal modelling and *model evaluation* is used instead. Some authors[24,25,31] also use the term *corroborated* in parallel with *validation*, which according to philosophers of science is what really happens in science. Karl Popper, the father of modern philosophy of science, argues that postulates and theories are only scientific if they can be disproven or falsified.[34] Thus, a scientific theory/model is never verified/validated, only corroborated. Abandoning the word validation may be desirable, but the term is already widely used in all parts of science and V&V in traditional areas of engineering have already matured so much that it is impractical to define an independent terminology for musculoskeletal modelling.

Biomechanics and musculoskeletal simulation

Established CAE methods are used extensively in biomechanics, for instance in finite element analysis of deformations in bones and soft tissues, and finite volume analysis for simulation of blood flow or respiration. Similarly, musculoskeletal simulation relies heavily on multibody simulation techniques originally devised for vehicle dynamics and other technical applications.

However, life introduces significant additional complexity into the models, and the fact that a method has been accepted as reliable for a technical field does not mean that it enjoys the same fidelity for simulation in biomechanics. An additional challenge in *bio*mechanical modelling is that the complexity of the physical systems can only be reduced to a certain extent, leading to more complicated models. Consequently, if validation activities do not turn out positive, it becomes difficult to specify why this happens and how the model can be improved. Furthermore, there is typically a lack of confidence in the basic inputs of the model and it is typically much more difficult to obtain reliable experimental data for living organisms to which the analysis results can be compared.

The inferior level of maturity of biomechanical simulation compared to simulation in the technical field is likely due to these differences. In the following, we shall therefore review the state-of-the-art of validation of multibody musculoskeletal models and compare to the progress of other fields of CAE that share the need for validation but for a variety of reasons have progressed further. Subsequently, we shall attempt to establish a terminology that is precise enough to support a scientific discourse on the subject of validation of multibody musculoskeletal models and attempt a prediction of initiatives that will mature the field and help realize its' potential.

Validation concepts

There is only limited literature available which explicitly deals with the subject of validation within biomechanics. Griffin[35] considered the problem of validation a decade ago and concluded that 'It is meaningless to refer to a model as being a "validated model"' and that 'It is more helpful to quote the limits to the applicability and the errors of a quantitative model…'. Griffin's review focused on biodynamic vibration models and did not consider V&V methodologies from other areas of engineering. Nevertheless, many of the observations and recommendations regarding mechanistic vibration models are also useful in a more general sense and can be adopted for musculoskeletal modelling. Anderson et al.[36] and Henninger et al.[37] were the first to introduce V&V methodologies in the context of biomechanics, and these works present a few examples of the validation of FEM-based methods. Henninger et al.[37] concluded that a standard for validation in biomechanics is an elusive goal given the diversity of applications and methods. The diversity is also evident even within the narrow scope of multibody musculoskeletal modelling, and the majority of papers contain some aspect of validation. In the book 'Biomechanics of the musculoskeletal system' Nigg and Herzog[33] identified three ways to perform validation of musculoskeletal models: first, direct measurements, second, indirect measurements and third, trend measurements. Even though these three concepts do play a role in many validation studies, there are many more aspects, which are important. This point can be illustrated by considering the review on model-based estimation of muscle forces by Erdemir et al.[6] Of 78 models in the review, 64 included validation, spanning a wide range of methodologies. In the following, we give an overview of the most common concepts and approaches in multibody musculoskeletal modelling.

Quantitative versus qualitative validation

Almost all works in musculoskeletal modelling contain some aspect of validation in the sense that experimental and computational curves or values are compared, either to justify a proposed model or provide confidence in the conclusion of a specific research question. The types of these comparisons vary greatly, and both qualitative and quantitative methods are used. Qualitative

methods are characterized by subjective terms describing agreement between the results as for example 'good', 'fair' or 'poor'. The qualitative approach to validation is usually taken when there is little data to support the results, or when it is necessary to compare different physical quantities. Qualitative validation is, for example, usually reported when time histories of computed muscle activation are compared to electromyography (EMG) recordings. In such cases, qualitative validation may be the only way to assess the quality or performance of a model, and we shall return to the topic of EMG comparisons later in this paper.

Quantitative validation is warranted in all cases where it is possible to define an objective criterion for the difference between the computed and experimental quantity of interest. Such mathematical measures are referred to as validation metrics in the context of V&V.[27] It is only by the use of validation metrics that it is possible to take the next step and quantify the uncertainty of the model prediction.

Stochastic versus deterministic models

The literature on V&V distinguishes between deterministic and stochastic validation metrics. Deterministic measures are single values that do not include the uncertainty of experimental measurements or model calculations. Stochastic validation metrics can incorporate both experimental and model uncertainties. In the ideal case, where the probability distributions of all parameters are known or can be estimated, it would be possible to find the model output as a probability distribution and reject or accept the model based on some confidence level. Such reliability methods are the subject of many V&V papers in traditional areas of engineering.[27,38–40] Rebba and Mahadevan[39] evaluated different statistical approaches for model validation and concluded that the classic null point hypothesis testing using *p*-values is a bad choice for model validation since it does not provide any estimate on the confidence with which to reject or accept model predictions. *P*-values are often misinterpreted as the probability the null hypothesis is true.[41] It can easily be seen that this is not the case, because the *p*-value changes with sample size, whereas the probability of the null hypothesis is a constant although unknown value. Instead Rebba and Mahadevan argued for simple statistical validation metrics of the form $r = P(-\varepsilon < m - c < \varepsilon)$, where m-c is the difference between measured data and computational predictions, and ε is the accuracy requirement. The interpretation of the reliability metric r then becomes straightforward. However, constructing such reliability metrics is far from trivial since m and c both have probability distributions that can be difficult to estimate. Studies on reliability metrics in traditional engineering are limited to the modelling of very simple physical systems, where all parameters can be tightly controlled or precisely measured. The full-blown reliability approach to validation that combines both uncertainties in experiment and modelling may be desirable for musculoskeletal models, but it is probably infeasible due to the large number of parameters and the intra and inter-subject variability, which is unavoidable in biological systems and difficult to describe in terms of probability distributions. In practice, most studies on musculoskeletal models take an approach that use some sort of statistical quantification of the experimental uncertainty, while still considering the model output as being deterministic.

Sensitivity analysis is the first step in recognizing that the model output is not deterministic, due to uncertainties in input parameters. Most sensitivity studies within musculoskeletal modelling so far have looked at the sensitivity of muscle force predictions to perturbations in the muscle–tendon parameters.[42–45] In all these studies, it turned out that muscle force estimates are most sensitive to changes in the tendon slack length, with optimal fibre length and the maximal muscle force also being important. Other studies have indicated that the muscle's line of action (i.e. moment arm) is also a critical parameter[46,47] both in terms of muscle force predictions and the prediction of joint reaction forces.

Indirect versus direct validation

Direct validation is the case where it is possible to directly compare a model output of interest with an experimental measurement of the same quantity. Joint reaction force is an example of a quantity, which is normally impossible to measure in-vivo, but can be measured in special circumstances, where patients have been fitted with instrumented joint replacements.[48–51] Such select cases therefore offer a direct validation of musculoskeletal models.

Although examples of direct validation of musculoskeletal models exist, the opportunities remain rare. Models are built with the intention of predicting a specific scenario or explaining a phenomenon, which is difficult or costly to investigate empirically, or the model is used as a scientific tool upon which it is possible to perform experiments which for ethical or safety reasons cannot involve humans. Thus, validation studies of such models are also inherently difficult, costly and ethically problematic to perform. Since such scenarios are common in musculoskeletal modelling, many studies focus their validation efforts on other variables in the model that are easier to measure experimentally. The motivation for this approach is that any confidence in a model's ability to predict measureable variables may reflect on the confidence of the variable of interest that cannot be measured. A common type of such indirect validation is comparison of EMG measurements with predicted muscle forces to infer confidence in estimation of joint compressive forces.[52,53]

Even though indirect validation may provide some evidence that something in the model is behaving as expected, it does not guarantee that the model can correctly predict the actual quantity of interest. This means

that, even though indirect validation is both quantitative and objective, it will eventually result in a subjective and qualitative assessment, when the confidence in the model's performance is transferred from the indirect measurements to the actual quantity of interest.

Trend versus absolute validation

Should a model be trusted just because it happens to pass the requirements in the validation process? This question is important because a model may be able to make correct predictions while still being wrong at the conceptual level. In other words, it may be correct but for the wrong reasons. This is particularly a danger when the model has been calibrated to match experimental data. What is important then is not the absolute difference between model and experiment, but instead how the underlying variables interact. Only if the underlying conceptual model is physically sound will it be possible to use models without performing validation experiments to confirm results. The investigation of the correctness of variable interaction is called trend validation, and its importance was emphasized by Nigg and Herzog[33] more than a decade ago. Successful trend validation is an important prerequisite if models and simulation are to be used for the so-called 'what–if' scenarios. If we make an intervention by changing a certain parameter, what will happen to the output of interest? Such a question could be very relevant if we want to use a musculoskeletal model for predicting the outcome after a surgical intervention. To answer these kinds of questions, the model parameters must interact correctly with each other.

Model calibration versus blind validation

The differentiation between blind and non-blind validation comes from the idea that validation should be performed against independent data that was not used to construct the model. Knowing the results in advance poses the risk that validation studies degrade into simple calibration of the model. As stressed in a number of publications on V&V in other areas of engineering,[21,24,54] calibration and validation are two different tasks. Calibration is about finding the best parameter values for the model in order to fit measured data, whereas validation is assessing how well the model can predict experimental data. Calibration is a natural part of constructing a new model, but the data used for calibration cannot afterwards be used to validate the model.

Real blind validation is rare in musculoskeletal modelling, since both modeller and experimentalist are often the same person or group of people. It is therefore commendable to be critical about validation results when the analyst has seen the data in advance. Oberkampf and Trucano[21] state that 'Knowing the "correct answer" before hand is extremely seductive, even to a saint'.

Even though blind validation is rare, recent attempts have been made to introduce this methodology in musculoskeletal modelling. The 'Grand Challenge Competition to Predict in vivo Knee Loads' is a project run at the ASME Summer Bioengineering conferences. Starting in 2010 and planned to last until 2016, it provides a series of comprehensive experimental data sets to allow the musculoskeletal modelling community to participate in an annual competition to predict knee compressive forces during different kinds of gait. The in-vivo measured compressive forces are only disclosed to the contestants at the conference when the simulated results had been published.[55] In addition to being an example of true blind validation, the competition is also evolving into a database of freely available validation benchmarks for simulating knee models.

Validation frameworks and benchmarks

When systems increase in scale and complexity, it becomes increasingly difficult to do true validation on the top-level systems. An example from a field in which this problem has been evident for a long time is simulation of catastrophic failures of complex high-consequence systems such as airplanes, nuclear power plants and nuclear storage facilities. These are simulations, which cannot be physically tested. However, it is still necessary to gain confidence in the predictions of the top-level model. The engineering approach to this dilemma was to divide the problem into a hierarchy of parts and sub-systems, which could be handled and validated, individually.

Validation hierarchy

Babuška et al.[56] illustrates the validation hierarchy with respect to FEM methods with a commercial aircraft. On the top level is the full aircraft model, which cannot be validated in full scale. On the tiers below are models of the major components, such as wings and fuselage together with their corresponding accreditation tests. On next lowest tiers are models of the minor components and structural elements and on the lowest level are material models and the corresponding material testing.

The concept of a validation hierarchy is also relevant for musculoskeletal modelling, when validation experiments on the top-level model are not possible. There are two areas in which a hierarchical approach to validation may be especially useful in musculoskeletal modelling. The first is those cases where validation experiments cannot be performed for ethical reasons. This could, for example, involve computational models to investigate injury mechanics or trauma. The second case is patient-specific models, where all models are different to a lesser or greater extent and each of the models is restricted to the prediction of some specific quantity or outcome without having to measure it directly.

Oberkampf and Trucano[21] in their work on validation benchmarks, pointed out that construction of a

validation hierarchy is useful for several key reasons. First, it captures 'a large range of complexities in systems, physics, material and geometries', and opens new possibilities in validation experiments, which could not be done on the top-level model. For musculoskeletal modelling, this entails a need to rethink what constitutes validation experiments. Subsystems and submodels may be validated in-vivo in animal experiments, in-vitro with cadaver specimens or totally different kinds of experiments. Second, Oberkampf and Trucano[21] pointed out that the validation hierarchy requires many different experts to work together to construct the hierarchy of validation studies, which may uncover interactions between subsystems that were not previously taken into account. Finally, it allows the results of validation on one level of the hierarchy to be related in terms accuracy to levels above and below.

The latter point is the more challenging. Musculoskeletal models, especially those included in generally available software systems, are often comprised of a number of subsystems and submodels and rely on particular methods, many of which have previously been published in the scientific literature. These are, for example, body part models, muscle configurations, joint models and methodologies such as different scaling procedures and muscle recruitment criteria. The point that there might also be a considerable publication bias when reporting the performance of new models and methods should not be ignored. The examples selected for publication are naturally those that perform the best. When the same methods/models are later included in musculoskeletal software packages, the context and assumptions under which they were originally validated may be different from the context for which the users decide to apply the models.

Gathering the relevant information for a validation hierarchy and relating the data in a sensible fashion is not trivial. Scientific papers presenting validation results are rarely published with hierarchical validation in mind, and the subsystem of the musculoskeletal system and their interactions are not as simple to characterize as traditional engineering constructs. They are frequently subject to the deficiencies outlined in the previous paragraphs and it may be difficult to extract the necessary quantitative information on model accuracy that can be related to other levels of the validation hierarchy. Ultimately, this means that submodels cannot always contribute directly to an assessment of the accuracy and shortcomings of the top-level model. Despite the clear limitation associated with applying the validation hierarchy outside the scope of traditional areas of engineering, it does provide some benefits and opportunities for musculoskeletal modelling which are emphasized in the section 'Validation hierarchy for musculoskeletal models'.

Validation benchmarks

The previous paragraph listed a number of reasons for scepticism towards model performance reported in the scientific literature. However, models may also improve after they are first published. We must keep in mind that models/methods are rarely one-shot developments, but evolve in an iterative process. Thus, the validation results or performance reported in the first publications on a model may be very different from how the model performs later. Biomechanists may therefore be reluctant to publish validation studies that objectively state the accuracy of a model, since such work remains locked in the scientific literature, while the model may later improve. This problem is especially relevant when working with musculoskeletal modelling packages, which tend to produce major updates every one or two years. By the time a validation study has been performed, analysed and published, a new version with supposedly superior qualities will have been released. The consequence is that the majority of studies that can be found in the scientific literature are based on older modelling packages than what is currently 'state-of-the-art'. Although modelling software usually improves with each new release, the opposite can also be the case. Models are so complex that there is no way of knowing this for sure by just studying the release log.

Commercial code companies in traditional areas of engineering have faced similar challenges with the release of new versions. The solution has been to provide users with benchmark examples that can be run with the previous and new versions and results can document the accuracy and reliability of the physical models and codes. Oberkampf and Trucano[21] surveyed leading CAE software for solid mechanics and found several CAE software products like ANSYS®[58] and ABAQUS®[57] that provide well-documented verification benchmarks to illustrate the code performance and engineering accuracy. Establishment of similar verification benchmarks for musculoskeletal modelling software is an obvious opportunity, albeit with the difference that there are no simple examples in musculoskeletal modelling for which analytical solutions exist for comparison.

There have been a few efforts to provide something equivalent to V&V benchmarks for musculoskeletal modelling. One of the most noteworthy is the OrthoLoad project,[59] which provides a database of measurements from instrumented joint implants performing various daily activities. The database does not contain true validation benchmarks but instead snapshots of valuable empirical data in the sense that they are obtained from particular individuals with particular impairments performing particular tasks. To the extent the test subjects and the selected activities can be modelled in software, they provide an excellent opportunity for a top-level validation of simulated joint reaction forces. Currently the database is lacking the necessary high-quality measurements documenting anthropometry, kinematics and external forces, however, the OrthoLoad homepage[59] states the ambition to add them: 'In the future additional information like ground reaction forces and gait data will be added'. If done carefully, the database could become a very important

source for validation benchmarks for musculoskeletal modelling. Another project with a similar ambition for musculoskeletal modelling is the 'Grand Challenge Competition to Predict in vivo Knee Loads'.[55]

The data sets provided by both of these projects belong to the top level of the validation hierarchy discussed earlier. Top-level data carry the risk that experimental and computational results agree, even though a submodel is still incorrect. This can be caused by the top-level model being insensitive to the incorrect submodel, or by several submodel errors accidentally cancelling one another. The former case is benevolent on the top level, whereas the risk of the second case diminishes with the provision of more data sets for top-level validation and with the execution of trend validation as discussed in the section 'Trend versus absolute validation'. Therefore, large databases of top-level validation benchmarks will be valuable.

It is worth noting that all models are essentially based on assumptions and approximations, for instance, the assumption of rigidity of the bones, and if decomposed sufficiently into subsystems, they will arrive at a point where these assumptions are obviously wrong.

Opportunities for musculoskeletal model validation

It is clear from the previous section that a standard procedure for validation similar to that advocated in solid mechanics may be an elusive goal in musculoskeletal modelling, given the large variability in biological systems, the difficulties in measuring output variables and the diversity of applications. It has been claimed that validation is inherently impossible, since scientific models or hypotheses can only be falsified but never proven valid.[60,61] Although this is a widely accepted point-of-view in philosophy of science, the applicability of engineering models is undisputed. We can therefore conclude that many physical models rely on the confidence within a limited domain that comes from repeated unsuccessful attempts at falsification. It seems obvious that the same principle may apply to the dissemination of musculoskeletal models into practical use.

Even though it might not yet be possible to establish standards or even guidelines for validation of musculoskeletal models, there are a number of practical approaches that may advance the field of musculoskeletal modelling.

Stronger focus on trend validation

The investigations of the so-called 'what–if' scenarios are often mentioned as one of the most promising uses of musculoskeletal models.[6] However, as already indicated, this requires that the model parameters interact with each other in the correct manner, which can be tested using a trend validation. A musculoskeletal model that results in the correct trends but at inaccurate absolute values can still be useful in many cases. This situation is likely to occur because musculoskeletal models have many uncertainties in the input parameters. Despite the importance of predicting the correct trends, systematic trend validations of musculoskeletal models are scarce in the literature. We recommend that, in the hierarchy of validation results, emphasis should be placed upon trend results. This requires carefully planned validation experiments in which one or more parameters can be changed systematically, while monitoring an output measure as a function of those parameters. Changing the same parameters in the simulation enables comparison of trends as well as the absolute values. An example of a trend validation experiment is given by Olesen et al.[62] in which seating posture is systematically changed while measuring the shear forces between the body and seat and subsequently comparing with a computer simulation model.

Better comparison of EMG with data

In the earlier mentioned review by Erdemir et al.,[6] 43 out of 68 attempts to validate multibody musculoskeletal models relied on EMG comparisons. The inability to measure muscle forces directly leaves the researcher with the option of using EMG measurement to obtain information about the active state of a muscle and comparing it with a simulated active state. Unfortunately, the active state computed in musculoskeletal modelling software is not a direct simulation of EMG for dynamic cases. Therefore, many authors compare only the timing of onset and offset of muscles between the model and the experiment, thus failing to take advantage of the richness of the EMG signal.[63–66]

Praagman et al.[67] used the near infrared spectroscopy technique to show proportionality between metabolism and EMG in single muscles for isometric cases. It is not known whether this finding also applies to dynamic muscle contraction, but it sparks the idea of developing a direct EMG simulation, possibly based on metabolism models. Many of the necessary parameters, such as contraction velocity and fibre length, are readily available in the musculoskeletal model, but the problem of predicting a multi-dimensional EMG signal from a single-dimensional computed muscle active state must be solved.

Even if development of an EMG simulation were successful, experimentalists should bear in mind that EMG is not simple to record reliably. The recorded signal from surface electrodes is subject to significant cross-talk and depends on the position of the EMG electrode both over the muscle and in relation to the neuromuscular junction or nerve innervation zone. EMG electrodes placed over the innervation zones record remarkably lower activity, and if innervation zones move under the electrodes, as it happens in dynamic conditions, it may be misinterpreted as changes of muscle activation level.[68] These vary

between muscles and for the same muscle between different subjects. Recording EMG with multiple electrodes over the same muscle may, therefore, be the only way to obtain reliable information. Given the limitation of EMG measurements in validation, the actual role of traditional EMG measurements may be limited to rejecting models which produce a clearly wrong result.

Development of quantitative validation metrics

We have emphasized that for validation to be useful, it must be based as much as possible on quantitative measures. Quantitative validation is warranted in all the cases where it is possible to define validation metrics (i.e. a mathematical measure of the difference between computational and experimental results).

Single values are typically compared in terms of their relative error, and standard statistical methods are well suited for these tasks. However, the same methods are less useful when the computational output is not a single value but is instead a complex transient response. As an example, consider two similar curves shifted slightly along the abscissa. If the curves change rapidly, the relative difference can become very large. The comparison becomes sensitive to inaccuracies on the abscissa, and the magnitude of the quantity itself. Thus, the relative error (or some summation of it) is a poor indicator of model accuracy, and it is likely that the measure will correlate poorly with the subjective impression of experts simply looking at the graphs.

As previously mentioned, other areas of engineering have faced similar challenges. Schwer[69] studied quantitative means to compare complex waveforms with the goal of minimizing subjectivity, while still maintaining correlation with expert opinions. Curve comparison metrics divide the error into two parts: a magnitude error M, which is insensitive to inaccuracies on the abscissa; and a phase error P that does not depend on the amplitude of the curve. The magnitude error and phase error are then combined to produce a single value that represents the difference between curves. Consult Schwer[69] and Lund et al.[70] for a detailed description of the metrics. Schwer[69] showed that this kind of validation metric was able to mimic how 11 experts evaluated a model of seismic wave propagation. Please note that the practical validation against EMG onset and offset in this framework can be interpreted as putting emphasis only on the phase error.

We believe that curve comparison metrics hold a potential to further quantitative methods in musculoskeletal model validation and should be explored in future research.

Validation hierarchy for musculoskeletal models

Development of a carefully planned validation hierarchy for musculoskeletal models may significantly enhance the state-of-the-art in the field by allowing modellers to initiate coordinated V&V on lower levels to facilitate higher level validation.

Figure 2 shows a possible validation hierarchy scheme for musculoskeletal models. The Hill model, often used in musculoskeletal modelling, is an example of the low-level submodels. Validation experiments for submodels, methods and model structures can be performed separately and the quantitative performance may serve as input for validation on systems at the level above. Even though this is not always possible, the validation hierarchy will still serve as a valuable overview and documentation of validation activities on all levels of the model. Thus, it allows others to recreate the validation studies, and makes the model performance and assumptions transparent to the end user.

Top-level validation is likely infeasible without quantitative validation of lower levels. In the following, we therefore mention some obvious opportunities for low-level validation activities.

Mechanical models. Mechanical models of musculoskeletal systems are subject to the laws of mechanics. This offers obvious opportunities for both verification and validation of models based on established equilibrium principles in forces, energy or momentum. The advantage of these approaches is that a single error in a very complicated mechanical system will often create an overall imbalance that can be detected like a checksum.

An example of a typical error detected by energy imbalance is work unintentionally performed by reaction forces between mechanical elements. This can be detected as a difference between the work performed by the external forces and the work performed by the muscles. This may happen in musculoskeletal models when measured joint kinematics are imposed on the model, for instance measured knee kinematics in two or three dimensions,[71] and reaction forces are introduced in the same directions.

Another case is muscles wrapping over bones, where the work of the contact forces against the moving bone may be difficult to compute correctly. Errors in this type of model can also be detected as a difference between the work of internal and external forces.

Behaviour of physiological models. Detailed musculoskeletal models literally contain thousands of parameters, all of which may have complex influences on the results. Checks for non-physiological behaviours of the model may help identify problems in the model and localize the root of the observed errors. Examples are checks for symmetric results when symmetric models are subjected to symmetric loads, or checks of the model-predicted strength of joints against experimental data. Experimentally investigated joint strengths are subject to significant inter-person variation, but the joint strengths as a function of joint angles of, for instance, knee extension or elbow flexion, exhibit some similarities over subjects even though the absolute values may vary, cf. the aforementioned discussion on the value of trend validation.

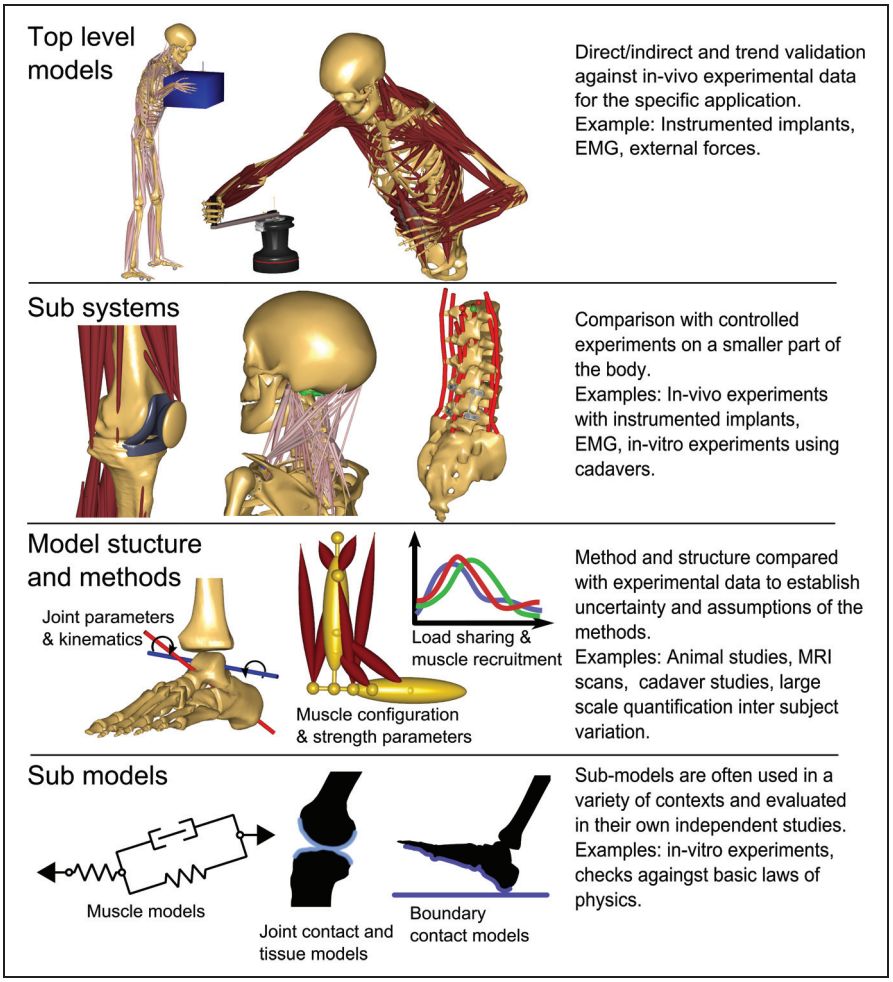

Figure 2. Validation hierarchy for musculoskeletal models. The figure shows some examples of methods and submodels that can be validated separately.

A validation framework for musculoskeletal models

In engineering simulation, leading companies, such as ANSYS[58] and Simulia[57] and independent organizations, such as National Agency for Finite Element Methods and Standards[72] have established verification benchmarks for finite element products. This is done in the realization that published V&V results are snapshots of the situation at a specific time, after which models and software continue to evolve.

The development of musculoskeletal models is likely an open-ended activity in which model and algorithm improvements will continue for the foreseeable future, and this creates a similar need for V&V benchmarks or frameworks that evolve with the methods and allow for comparison not only between simulated and experimental results but also between different versions of software and models.

A validation framework will allow a user to obtain data for the model validity on a number of cases. It then becomes the obligation of the user to interpret the information and evaluate the credibility of the model for a prospective simulation (Figure 1).

Such a validation framework would provide users with the documentation needed to evaluate the performance of the models without having to rely on results in outdated publications. Implementation of a growing number of validation benchmarks and experimental studies into the framework would not only ensure that these studies and data sets continue to be available for users to judge model credibility, but it would also serve as building blocks for a growing validation hierarchy.

There are two main requirements for this vision to become reality. First, the validation framework would need to be maintained by the communities or companies behind the musculoskeletal modelling packages. Second, researchers, who perform validation studies, have to make the data publicly available for others to re-use. These two requirements call for careful design of software to collect and manage the data and allow for interaction with users, submitting data to enrich the framework.

Conclusions

We have focused on how improving the practice and methodology of validation can help increase the

credibility of musculoskeletal models. The following points sum up the key features and concepts behind model V&V experiments.

Essential general findings on validation

1. Validation is an open-ended process tightly linked with the evolution of models and algorithms, and validation experiments are specific to a certain use of the model. It is, therefore, not relevant to refer to a model as being 'validated' in general.
2. Validation can only be carried out with respect to certain applications based on which a user judges whether the model is adequate for a specific task.
3. Model development and validation experiments are closely linked. The model sensitivities to changes in the input parameters decide the parameters to control or vary in the validation experiment and the precision with which they need to be measured. Therefore, methods for quantifying uncertainties in musculoskeletal models need to be developed.
4. Validation experiments are characterized by particular requirements for measurement and documentation of all important parameters and their uncertainty. Thus, not all experiments qualify as validation experiments.
5. Quantitative validation metrics are encouraged in all cases where experimental and simulation results are compared. Further research on these metrics for use with musculoskeletal model validation is needed.
6. Blind validation is essential to avoid bias in the results. Validation against experimental results that are not revealed until the computational results are ready is, therefore, the ideal situation.
7. Models and algorithms should be tested against experimental measurements over a wide range of applications.

Topics for further study in order to improve the validation of musculoskeletal models

1. Establishment of a consistent terminology on musculoskeletal modelling V&V corresponding with what has been adopted in traditional areas of CAE.
2. Establishment of validation hierarchies through which models can be validated at increasing levels of complexity.
3. Development of validation benchmarks open for third-party scrutiny, and documented sufficiently to allow recreation of the validation experiments with other assumptions, parameters or algorithms.
4. Specification of validation experiment requirements. On the lower tiers of the validation hierarchy, there is a possibility to use other approaches than in-vivo experiments, for example cadaver studies, checks against the laws of physics or animal experiments.

Musculoskeletal modelling is undergoing an exciting change from the realms of research into real applications. The potential benefit to medical treatments, product design, health and safety in the workplace and many other applications is tremendous, but it must also be realized that errors, uncertainties and unqualified use may have serious consequences. This situation is similar to the evolution of other CAE domains and similarly serious efforts in model validation will be necessary for musculoskeletal models.

Funding

This work was supported by The Danish National Advanced Technology Foundation.

References

1. Rasmussen J, Tørholm S and De Zee M. Computational analysis of the influence of seat pan inclination and friction on muscle activity and spinal joint forces. *Int J Ind Ergonom* 2009; 39: 52–57.
2. Grujicic M, Pandurangan B, Xie X, et al. Musculoskeletal computational analysis of the influence of car-seat design/adjustments on long-distance driving fatigue. *Int J Ind Ergonom* 2010; 40: 345–355.
3. Wagner D, Rasmussen J and Reed M. Assessing the importance of motion dynamics for ergonomic analysis of manual materials handling tasks using the AnyBody modeling system. *J Passenger Cars: Mech Syst* 2008; 116: 12–14.
4. Williams JL and Gomaa ST. Kinematics of rotating platform total knee replacements - does the position of the rotating platform axis matter? In *The third frontiers in biomedical devices conference and exhibition*, Irvine, CA, 18–20 June 2008 pp. 15–16.
5. Grujicic M, Arakere G, Xie X, et al. Design-optimization and material selection for a femoral-fracture fixation-plate implant. *Mater Des* 2010; 31: 3463–3473.
6. Erdemir A, McLean S, Herzog W and Van den Bogert AJ. Model-based estimation of muscle forces exerted during movements. *Clin Biomech* 2007; 22: 131–154.
7. Damsgaard M, Rasmussen J, Christensen ST, et al. Analysis of musculoskeletal systems in the AnyBody modeling system. *Simul Modell Pract Theory* 2006; 14: 1100–1111.
8. LifeModeler Inc. Lifemodeler. http://www.lifemodeler.-com/ (accessed 22 April 2010).
9. Delp SL and Loan JP. A graphics-based software system to develop and analyze models of musculoskeletal structures. *Comput Biol Med* 1995; 25: 21–34.
10. Delp SL, Anderson FC, Arnold AS, et al. OpenSim: open-source software to create and analyze dynamic simulations of movement. *IEEE Trans Biomed Engng* 2007; 54: 1940–1950.
11. Neal ML and Kerckhoffs R. Current progress in patient-specific modeling. *Brief Bioinform* 2010; 11: 111–126.
12. National Academy of Engineering. Grand challenges for engineering, www.nae.edu (accessed 19 March 2010).

13. Stansfield BW, Nicol AC, Paul JP, et al. Direct comparison of calculated hip joint contact forces with those measured using instrumented implants. An evaluation of a three-dimensional mathematical model of the lower limb. *J Biomech* 2003; 36: 929–936.
14. Kim HJ, Fernandez JW, Akbarshahi M, et al. Evaluation of predicted knee-joint muscle forces during gait using an instrumented knee implant. *J Orthop Res* 2009; 27: 1326–1331.
15. Piazza SJ and Delp SL. Three-dimensional dynamic simulation of total knee replacement motion during a step-up task. *J Biomech Engng* 2001; 123: 599–606.
16. Magermans DJ, Chadwick EKJ, Veeger HEJ, et al. Effectiveness of tendon transfers for massive rotator cuff tears: a simulation study. *Clin Biomech* 2004; 19: 116–122.
17. Saul KR, Murray WM, Hentz VR and Delp SL. Biomechanics of the Steindler flexorplasty surgery: a computer simulation study. *J Hand Surg* 2003; 28: 979–986.
18. Holzbaur KRS, Murray WM and Delp SL. A model of the upper extremity for simulating musculoskeletal surgery and analyzing neuromuscular control. *Ann Biomed Engng* 2005; 33: 829–840.
19. Johnson FT, Tinoco EN and Yu NJ. Thirty years of development and application of CFD at Boeing Commercial Airplanes, Seattle. *Comput Fluids* 2005; 34: 1115–1151.
20. Roache PJ. Verification and validation and computational science and engineering. Albuquerque, NM: Hermosa Publishers, 1998.
21. Oberkampf WL and Trucano TG. Verification and validation benchmarks. *Nucl Engng Des* 2008; 238: 716–743.
22. American Society of Mechanical Engineers. G*uide for verification and validation in computational solid mechanics*. Report V & V 10-2006, 2006. New York: American Society of Mechanical Engineers.
23. Oberkampf WL, Trucano TG and Hirsch C. *Verification, validation, and predictive capability in computational engineering and physics*. Report SAND2003-3769, 2003. Albuquerque, NM: Sandia National Laboratories.
24. Refsgaard JC and Henriksen HJ. Modelling guidelines—terminology and guiding principles. *Adv Water Resour* 2004; 27: 71–82.
25. Babuska I and Oden JT. Verification and validation in computational engineering and science: basic concepts. *Comput Methods Appl Mech Engng* 2004; 193: 4057–4066.
26. Oberkampf WL and Trucano TG. Design of and comparison with verification and validation benchmarks. In: *Benchmarking of CFD codes for application to nuclear reactor safety,* Garching Germany: Committee on the Safety of Nuclear Installations, Nuclear Energy Agency, 2006.
27. Oberkampf WL and Barone MF. Measures of agreement between computation and experiment: validation metrics. *J Comput Phys* 2006; 217: 5–36.
28. American Institute of Aeronautics and Astronautics. G*uide for the verification and validation of computational fluid dynamics simulations*. Report, 1998, AIAA G-088-1998, 1998. AIAA.
29. American Scociety of Mechanical Engineers (ASME). Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer. , 2009, V & V 20-2009 (American Society of Mechanical Engineers, New York, USA).
30. Schwer LE. An overview of the PTC 60/V&V 10: guide for verification and validation in computational solid mechanics. *Engng Comput* 2007; 23: 245–252.
31. Rykiel EJ, Jr. Testing ecological models: the meaning of validation. *Ecol Model* 1996; 90: 229–244.
32. Hills RG, Pilch M, Dowding KJ, et al. Validation challenge workshop. *Comput Methods Appl Mech Engng* 2008; 197: 2375–2380.
33. Nigg BM and Herzog W. *Biomechanics of the musculoskeletal system*. Chichester, UK: Wiley, 1999.
34. Popper KR and Weiss G. The logic of scientific discovery. *Phys Today* 1959; 12: 53–54.
35. Griffin MJ. The validation of biodynamic models. *Clin Biomech* 2001; 16: S81–S92.
36. Anderson AE, Ellis BJ and Weiss JA. Verification, validation and sensitivity studies in computational biomechanics. *Comput Methods Biomech Biomed Engng* 2007; 10: 171–184.
37. Henninger HB, Reese SP, Anderson AE and Weiss JA. Validation of computational models in biomechanics. *Proc IMechE, Part H: J Engng Med* 2010; 224: 801–812.
38. Jiang X, Mahadevan S and Urbina A. Bayesian nonlinear structural equation modeling for hierarchical validation of dynamical systems. *Mech Syst Signal Process* 2010; 24: 957–975.
39. Rebba R and Mahadevan S. Computational methods for model reliability assessment. *Reliab Engng Syst Saf* 2008; 93: 1197–1207.
40. Chen W, Xiong Y, Tsui K and Wang S. Some metrics and a Bayesian procedure for validating predictive models in engineering design. In *ASME international design engineering technical conferences and computers and information in engineering conference* (*DETC2006*), Philadelphia, PA, 10 September 2006.
41. Cohen J. The earth is round ($p < .05$). *Am Psychol* 1994; 49: 997–1003.
42. Xiao M and Higginson J. Sensitivity of estimated muscle force in forward simulation of normal walking. *J Appl Biomech* 2010; 26: 142–149.
43. Scovil CY and Ronsky, JL. Sensitivity of a Hill-based muscle model to perturbations in model parameters. *J Biomech* 2006; 39: 2055–2063.
44. Redl C, Gfoehler M and Pandy MG. Sensitivity of muscle force estimates to variations in muscle-tendon properties. *Hum Mov Sci* 2007; 26: 306–319.
45. De Groote F, Van Campen A, Jonkers I and De Schutter J. Sensitivity of dynamic simulations of gait and dynamometer experiments to Hill muscle model parameters of knee flexors and extensors. *J Biomech* 2010; 43: 1876–1883.
46. Nussbaum MA, Chaffin DB and Rechtien CJ. Muscle lines-of-action affect predicted forces in optimization-based spine muscle modeling. *J Biomech* 1995; 28: 401–409.
47. Raikova RT and Prilutsky BI. Sensitivity of predicted muscle forces to parameters of the optimization-based human leg model revealed by analytical and numerical analyses. *J Biomech* 2001; 34: 1243–1255.
48. Bergmann G, Graichen F and Rohlmann A. Hip joint loading during walking and running, measured in two patients. *J Biomech* 1993; 26: 969–990.
49. D'Lima DD, Townsend CP, Arms SW, et al. An implantable telemetry device to measure intra-articular tibial forces. *J Biomech* 2005; 38: 299–304.
50. Heinlein B, Graichen F, Bender A, et al. Design, calibration and pre-clinical testing of an instrumented tibial tray. *J Biomech* 2007; 40: S4–S10.

51. Bergmann G, Graichen F, Bender A, et al. In vivo glenohumeral contact forces—measurements in the first patient 7 months postoperatively. *J Biomech* 2007; 40: 2139–2149.
52. Hug F. Can muscle coordination be precisely studied by surface electromyography? *J Electromyogr Kinesiology* 2011; 21: 1–12.
53. De Zee M, Dalstra M, Cattaneo PM, et al. Validation of a musculo-skeletal model of the mandible and its application to mandibular distraction osteogenesis. *J Biomech* 2007; 40: 1192–1201.
54. American Society of Mechanical Engineers. *Guide for verification and validation in computational solid mechanics*. Report PTC 60/V&V 10, 2007. New York: American Society of Mechanical Engineers.
55. Fregly BJ, D'Lima DD and Besier T. Grand challenge competition to predict in vivo knee loads. https://simtk.org/home/kneeloads/ (accessed 26 March 2011).
56. Babuška I, Nobile F and Tempone R. Reliability of computational science. *Numer Methods Partial Differ Equ* 2007; 23: 753–784.
57. Simulia. Abaqus verification manual, http://www.simulia.com/support/documentation.html (accessed 4 April 2011).
58. ANSYS. ANSYS verification manual, http://www.ansys.-com/About + ANSYS/Quality + Assurance/Quality + Services (accessed 4 April 2011).
59. Bergmann, G. OrthoLoad, http://www.OrthoLoad.com (accessed 14 April 2011).
60. Konikow LF and Bredehoeft JD. Ground-water models cannot be validated. *Adv Water Resour* 1992; 15: 75–83.
61. Oreskes N, Shrader-Frechette K and Belitz K. Verification, validation, and confirmation of numerical models in the earth sciences. *Science* 1994; 263: 641.
62. Olesen C, De Zee M and Rasmussen J. Experimental validation of a computational seated human model for pressure ulcer research. *J Appl Biomech* 2011 (in review).
63. Anderson FC and Pandy MG. Static and dynamic optimization solutions for gait are practically equivalent. *J Biomech* 2001; 34: 153–161.
64. Klein Horsman MD. The Twente lower extremity model: consistent dynamic simulation of the human locomotor apparatus. Report, University of Twente, Enschede, The Netherlands, 2007.
65. Prilutsky BI and Gregor RJ. Analysis of muscle coordination strategies in cycling. *IEEE Trans Rehab Engng* 2000; 8: 362–370.
66. Prilutsky BI and Zatsiorsky VM. Optimization-based models of muscle coordination. *Exerc Sport Sci Rev* 2002; 30: 32–38.
67. Praagman M, Veeger HEJ, Chadwick EKJ, et al. Muscle oxygen consumption, determined by NIRS, in relation to external force and EMG. *J Biomech* 2003; 36: 905–912.
68. Mesin L, Merletti R and Rainoldi A. Surface EMG: the issue of electrode location. *J. Electromyogr Kinesiology* 2009; 19: 719–726.
69. Schwer LE. Validation metrics for response histories: perspectives and case studies. *Engng Comput* 2007; 23: 295–309.
70. Lund ME, De Zee M and Rasmussen J. Comparing calculated and measured curves in validation of musculuskeletal models. In: *The XIII international symposium on computer simulation in biomechanics*, Leuven Belgium, 30 August 2011.
71. Sandholm A, Schwartz C, Pronost N, et al. Evaluation of a geometry-based knee joint compared to a planar knee joint. *Visual Comput* 2011; 27: 161–171.
72. NAFEMS. NAFEMS, http://www.nafems.org (accessed 6 April 6 2011).