Transactions on Additive Manufacturing Meets Medicine
Trans. AMMM, 2023, Vol. 5, No 1, Article ID 804, 13 Pages

Infinite Science Publishing

Original Research Article

Towards credible computational models: Application of a risk-based framework for establishing credibility

A.-K. Carl[1*], M. Kirillov[1], D. Hochmann[1], and Eric Quadrat[2]

[1] *Biomechatronics Research Laboratory, FH Münster University of Applied Sciences, Steinfurt, Germany*
[2] *Simq GmbH, Grafing near Munich, Germany*
[*] *Corresponding author, email: a.carl@fh-muenster.de*

Abstract: The use of computational modeling and simulation (CMS) as a tool for gaining insight into the technical performance and safety of medical devices has emerged continuously over the past years. However, to rely on information and decisions derived from model predictions, it is essential to establish model credibility for the specific context of use. Limited regulatory requirements and lack of consensus on the level of verification and validation activities required result in rare use of CMS as a source of evidence in the medical device approval process. The American Society of Mechanical Engineers (ASME) developed a risk-informed framework to establish appropriate credibility requirements of a computational model: the ASME V&V 40-2018 standard. This paper aims to outline the concepts of this standard and to demonstrate its application using an example from the orthotics field. The necessary steps to establish model credibility for a custom-made 3D printed wrist hand orthosis (WHO) are presented. It is shown that the credibility requirements of each verification and validation activity depend on model risk by applying two different contexts of use to the same computational model.

I. Introduction

In recent years, computational modeling and simulation (CMS) has become an essential tool in providing information about the technical performance, safety, and effectiveness of medical devices [1]. Before entering the market, manufacturers must prove the safety and performance of their medical devices [2]. In the past, relevant data was obtained mainly from three sources: (i) controlled benchtop tests, (ii) animal testing, or (iii) clinical trials [3]. By using CMS these conventional data sources can be complemented or in some cases even be replaced. Insights from data can be gained or data can be generated that cannot be obtained through traditional testing methods [3]. CMS allows the safety and performance of a medical device to be evaluated without exposing individuals to potential harm or unnecessary risks [1]. In the future, widespread application of *in silico* as a source of scientific evidence in the approval process of medical devices is expected [4]. Due to the use of computational models, for example, the number of participants as well as the duration of clinical trials can be reduced and thus the costs incurred [2, 5–7]. CMS provides the ability to comparably easy and quick conduct multiple simulations and evaluate the safety and performance of a medical device under different conditions [8]. An advantage over traditional testing methods. This is especially important as emerging technologies such as additive manufacturing (AM), also known as three-dimensional (3D) printing, continue to evolve and thus present new challenges. AM is nowadays not used solely for the production of prototypes, but also for the manufacture of final products [9]. Therefore, characterizing the mechanical properties and behavior of 3D printed devices through appropriate testing is increasingly important. In addition, AM allows unprecedented customization to each patient's anatomy, generating medical devices tailored to a single individual. However, this means that all device configurations must be evaluated for safety and performance, making traditional test methods difficult to apply in practice and therefore uneconomical. A potential solution is CMS. Through *in silico* models, it is possible to assess patient-specific medical devices on a case-by-case basis.

Even though advantages of CMS are widely recognized, there is one critical aspect that needs to be adequately taken into account before utilizing *in silico* models to support medical device development and evaluation: assessment of the model credibility [5, 10]. To rely on information and decisions based on model predictions, it must be demonstrated that: (i) the mathematical model is appropriately implemented and solved precisely, (ii) the model is an adequate representation of the intended reality, and (iii) the sensitivities and uncertainties of the model and its relevant comparator are evaluated. To put it another way, computational models need to be verified and validated [10]. Verification, validation and uncertainty quantification activities help to build confidence in the predictive capability of the *in silico* model, commonly referred to as model credibility [3]. Provided that simulation outputs are an integrated part of the decision-making process, they could cause potential harm to patients and/or healthcare providers. Therefore, it is of great importance to establish the credibility of the computational model [1]. Thus, verification and validation activities aim to ensure that the computational model adequately represents the reality of interest. This can be achieved by comparing simulation results, for example, with controlled experiments or other relevant information sources. However, relevance and adequacy of those activities, and consequently the model credibility, are subjective. This can cause discrepancies between stakeholders on what constitutes a sufficiently verified and validated computational model [1]. A clear regulatory framework procures a common understanding for the application of CMS for medical devices and for assessing model credibility. At present, the lack of regulatory guidance as well as of consensus on the evidential value for model validation causes the infrequent use of CMS as source of evidence in the approval process of medical devices [11, 12].

The Center for Devices and Radiological Health (CDRH), a department of the U.S. Food and Drug Administration (FDA), identified CMS as a regulatory science priority just under a decade ago [13]. Since then, the FDA has made numerous attempts in developing computational modeling technologies to support regulatory decision making and addressing the regulatory uncertainties. As a result, the CDRH issued a guidance document in 2016 to provide recommendations on FDA-compliant reports of CMS studies that are used to support medical device submissions [14]. With this document, the FDA supports the use of CMS in medical device submissions, because these *in silico* models can provide valuable information that cannot be obtained with conventional test methods, or even be used as replacement for bench testing. The guidance is intended "to improve the consistency and predictability of the review of CMS studies" [14], as well as ease comprehensive interpretation and full review of those studies.

Tasked with the same mission, the American Society of Mechanical Engineers (ASME) V&V 40 Subcommittee for Verification and Validation (V&V) in computational modeling of medical devices developed a risk-based credibility assessment framework: the so-called technical standard ASME V&V 40-2018 "Assessing Credibility of Computational Modeling through Verification and Validation: Application to Medical Devices". It guides the process of evaluating model credibility when using computational models in the field of medical technology [1].

In Europe, the Avicenna Alliance, a global non-profit organization founded in 2016, aims to make *in silico* medicine standard practice in the health care sector. It has its origin in the "Virtual Physiological Human (VPH)", a European initiative to support the development of computational models of human physiology. The Association was first assigned by the European Commission to develop a publicly available "Roadmap for in silico medicine". Today, the goal of the Avicenna Alliance is to put this document into practice. Establishment of a well-functioning legal framework for the use of *in silico* medicine is a priority in this context [15]. In 2021, the Association published a position paper [16] to increase acceptance of computational evidence in the regulatory decision-making process for medical devices. The paper lists benefits of CMS, defines common terminology related to CMS, and provides an overview of a risk-based approach to establish model credibility. Recommendations for reporting *in silico* studies and results obtained are also specified. The document of the Avicenna Alliance is based on specifications of the FDA and the ASME standard V&V 40-2018, respectively. Although the Avicenna Alliance exists, the European Union lacks any regulatory framework regarding the use of CMS as evidence for medical device submissions and the assessment of model credibility. In recent years, the European Union has revised the legislation on medical devices and in vitro diagnostics. Since May 2021, the Medical Device Regulation (EU) 2017/745 (MDR) [17] has been in force. Its objective is to ensure high public health standards, improved traceability and greater patient safety, among others. Achieving this requires more in-depth evidence to support safety and quality claims. Extensive pre-clinical evaluations, namely *in vitro* and *in vivo* studies as well as clinical trials will become a necessity in the approval process of many medical devices. Using CMS can help manufacturers comply with MDR requirements and remain competitive.

To the authors' best knowledge, there are few examples in scientific literature of the application of the ASME V&V 40-2018 standard [11, 18–20], but none in the field of orthotics. The aim of this paper is to present a risk-based approach for assessing model credibility for custom-made 3D printed wrist hand orthoses (WHOs) according to ASME V&V 40-2018 used in the regulatory decision-making process. An overview of the necessary

steps to establish model credibility is outlined. In accordance with the defined contexts of use and the extent to what the computational model is considered in the decision-making process, a detailed analysis of the model risk is described. To provide useful information for those verifying and validating a computational model for testing 3D printed WHOs or similar devices, the article also presents relevant bench tests and discusses their use in the validation process.

II. Application of ASME V&V 40-2018

The following article describes the practical application of the ASME V&V 40-2018 standard using the example of a custom-made 3D printed WHO. The explanations given refer to data collected within the interdisciplinary research project "Simulationsgestützte Medizintechnikplattform zur individuellen 3D-Hilfsmittelversorgung (SIGMA3D)"[1] on digitalization of orthotics, funded by the German Federal Ministry of Education and Research (BMBF). It should be noted that the article does not provide general approaches for computational solid mechanics, 3D printed orthoses, experimental comparisons, or acceptance criteria for model accuracy. Specific recommendations are beyond the scope of this publication. Instead, it is intended to show what is needed to ensure sufficient model credibility associated with the V&V 40-2018 framework.

Details provided in this section include: (i) an overview of the ASME V&V 40-2018 standard, (ii) a description of the custom-made 3D printed WHO, (iii) the assessment of the model risk according to the V&V 40-2018 framework, (iv) a description of the computational model for predicting the functional and safety parameters, (v) the physical testing comparators, and (vi) the selected credibility goals depending on model risk.

II.I. Overview of the ASME V&V 40 standard

The ASME V&V 40-2018 standard provides a risk-informed framework for establishing appropriate credibility requirements of a computational model. It gives guidance for assessing the relevance and adequacy of completed verification and validation activities that ensure credibility of a computational model. The technical standard supplements other standards regarding V&V methods, such as ASME V&V 10-2019 [21] and ASME V&V 20-2009 [22]. It describes how much verification, validation, and uncertainty quantification is needed to establish trust in the predictive capability of a computational model for decision-making in a particular context of use [23]. The ASME V&V 40-2018 standard envisages to comply the credibility requirements of an *in silico* model with the risks for a specified context of use. Simplified, the V&V 40-2018 standard comprises the following steps, among others:

1. **Identification of the question of interest:** the question of interest describes the specific question, decision or concern that is being addressed;
2. **Definition of the context of use (COU):** the COU defines the specific role and scope of the computational model used to address the question of interest;
3. **Assessment of the model risk:** model risk is the possibility that the use of the computational model may lead to a decision that results in patient harm or other undesirable outcome(s). It is the combination of the influence of the computational model in decision-making (model influence) and the consequence of an adverse outcome resulting from a false or incorrect decision (decision consequence); and
4. **Establishment of model credibility** through the verification, validation and applicability activities and their associated goals for each credibility factor (see Table 1), in accordance with the model risk [1].

Further steps include collecting V&V evidence and any additional knowledge gained during the V&V process, and then demonstrating the applicability of this evidence to support the use of the computational model for the COU. Credibility assessment ensures that the credibility activities performed and the results obtained are sufficient to establish credibility of the computational model. According to the ASME V&V 40-2018 standard, the verification, validation, and applicability assessment activities carried out must be documented, whereby documentation should include evidence supporting the credibility of the computational model for the COU [1].

Table 1: Verification, validation, and applicability activities and their associated credibility factors as presented in the ASME V&V 40-2018 standard [1].

Activities	Credibility Factors
Verification	
Code	Software quality assurance
	Numerical code verification
Calculation	Discretization error
	Numerical solver error
	Use error
Validation	
Computational model	Model form
	Model inputs
Comparator	Test samples
	Test conditions
Assessment	Equivalency of input parameters
	Output comparison
Applicability	Relevance of the quantities of interest
	Relevance of the validation activities to the COU

[1] »Simulation based medical technology platform for individual 3D aid supply«

II.II. Custom-made 3D printed wrist hand orthosis (WHO)

The medical device described is a custom-made 3D printed WHO, shown in Figure 1. It was methodically developed and manufactured by Care Center Deutschland GmbH within the BMBF-funded project SIGMA3D. The orthosis is used in patients with lesions around the wrist. It is worn by patients in a variety of everyday situations, immobilizing the wrist while providing maximum freedom of movement for the fingers and thumb. The custom-made 3D printed WHO is individually adapted to the anatomy of each patient.

Figure 1: Dorsal (left), ulnar (middle), and radial (right) view of the custom-made 3D printed wrist hand orthosis (WHO), designed by Care Center Deutschland GmbH.

II.III. Model risk

To assess custom-made 3D printed WHOs computational solid mechanics is used. Based on the questions of interest, model credibility requirements are evaluated for two different COUs. The questions of interest addressed in the example given are:

1. "How do decisions regarding the material and design influence the functional parameters of the custom-made 3D printed WHO?" and
2. "How do decisions regarding the material and design influence the occurring strains and stresses on the custom-made 3D printed WHO?".

CMS will be used instead of physical testing to characterize the mechanical properties of the 3D printed WHO. The orthosis is individually adapted to each patient's needs. Therefore, all configuration options across the proposed product portfolio need to be assessed for safety and performance. CMS is the only practical approach to evaluate the functional and safety parameters of the individualized WHO on a case-by-case basis. The following is a presentation of the model risk assessment in accordance with V&V 40-2018. As said in section II.I, model risk is a combination of model influence and decision consequence. It determines the credibility requirements of the V&V activities. Table 2 provides an overview of both COUs and the corresponding model risk assessments.

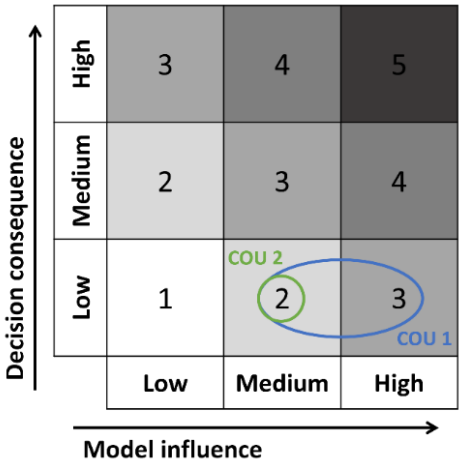

Figure 2: Model Risk Matrix for the given example of a custom-made 3D printed WHO, indicating the risk for COU 1 at a level 2/3 and for COU 2 at level 2.

II.IV. Computational Model

The computational model is created using various commercially available software packages. The starting point is a 3D scan in STL format of the patient's forearm anatomy, including the hand. The design of the orthosis is then adapted to this anatomy, see Figure 3. The geometry of the forearm is cut and intersected to match the setup of the test rig. This was done using a commercial CAD software (Ansys® SpaceClaim, Version 2022R1; ANSYS, Inc., United States). A finite element representation of all the geometry parts was created using Ansys® ICEM CFD, Version 2022R2 (ANSYS, Inc., United States).

Isotropic, linear elastic material models are selected for different parts of the forearm geometries. The orthosis has an isotropic material model that takes plasticity into account. The material parameters were determined from literature and from experimental tests. A frictional contact is defined between the forearm and the orthosis. Experimental data were used to determine the coefficient of friction. The straps with a hook and loop fastener on the orthosis, which fix the orthosis to the anatomy of the forearm, have been modeled with springs.

The boundary conditions were defined as an axis corresponding to the wrist rotation axis. Rotation around the axis was chosen to simulate flexion and extension. A static simulation without dynamic effects was performed using Ansys® Mechanical, Version 2022R2 (ANSYS, Inc., United States).

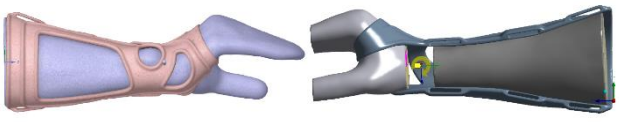

Figure 3: The initial 3D scan with the designed orthosis (left) and the resulting simulation model with the corresponding axis of rotation (right).

Table 2: Model risk assessment for the contexts of use (COUs) to address the questions of interest: (i) How do decisions regarding the material and design influence the functional parameters of the custom-made 3D printed wrist hand orthosis (WHO)? and (ii) How do decisions regarding the material and design influence the occurring strains and stresses on the custom-made 3D printed WHO?

	COU 1	**COU 2**
Context of Use	Performance evaluation of the functional properties of the custom-made 3D printed WHO	Superiority evaluation of the strain distribution of the custom-made 3D printed WHO
	– The computational model will be used to evaluate the functional properties of custom-made 3D printed WHO across all sizes in the proposed product portfolio. – CMS will be a replacement for physical testing. – The computational model will compute the *Range of Motion (ROM)* and the *Stiffness* of the WHO. – No supporting benchtop test data or consideration of a predictive device will be generated. – The final approval for the device design will be based on the simulation results only.	– The computational model will be used to predict the occurring strains over the operating range of the 3D printed WHO. – CMS will be a replacement for physical testing. – The successfully tested worst case design will serve as validation basis for the strain predictions. Simulation outcomes are benchmarked against test rig results of the successfully tested worst case design. – Simulation results will be the primary factor for the final approval of the device design.
Model Risk Assessment		
Model Influence	Model influence is moderate to high since the output from the computational model is the sole source to inform the decision regarding the final product; there is no supporting evaluation with bench testing. → **MEDIUM to HIGH**	Model influence is medium since the output from the computational model is the sole source to inform the decision regarding the final product, but complementary benchtop testing of the worst case design is also conducted to evaluate the safety parameter *Strains* of the orthosis. → **MEDIUM**
Decision Consequence	In case of an incorrect decision, the impact of a 3D printed WHO on the health or safety of patients is negligible. Due to the quality assurance of the medical supplier, each final WHO is checked by the orthopedic technician prior to its dispensing to the patient. Thus, the probability of occurrence of an incorrect decision is minimal. → **LOW**	
Model Risk	Consequently, the model risk is **MEDIUM to HIGH – LOW (level 2/3)**, see Figure 2.	Consequently, the model risk is **MEDIUM – LOW (level 2)**, see Figure 2.

For the assessment of model credibility, the reaction moments of rotation and local strains on the surface of the orthosis are compute. The reaction moments provide information about the stiffness of the orthosis. In addition, conclusions about the range of motion can be derived (question of interest 1). The calculated local strains provide information about the strain distribution in the orthosis during flexion and extension (question of interest 2).

II.V. Validation Comparators

Physical testing of the functional characteristics. Physical testing was used to replicate flexion and extension of the human wrist in order to evaluate the functional properties of the 3D printed WHO. Using a force-controlled test rig (shown in Figure 4), the mechanical characteristics of the orthosis were obtained. The tests were performed using a universal material testing machine (20 kN Allround Table Top Zwick/Roell, Germany) at defined traverse speed of 4 mm/s to obtain the moment-angle relationship of the orthosis. The translational movement of the universal testing machine was converted into a rotational movement around the joint. An individually adapted forearm model was used for physical testing. Application of the WHO was controlled by measuring the tensile force in the straps with two belt force sensors, with the application force set at 20 N. Physical testing of each orthosis was repeated two times with five test cycles each and previously established test parameters [24].

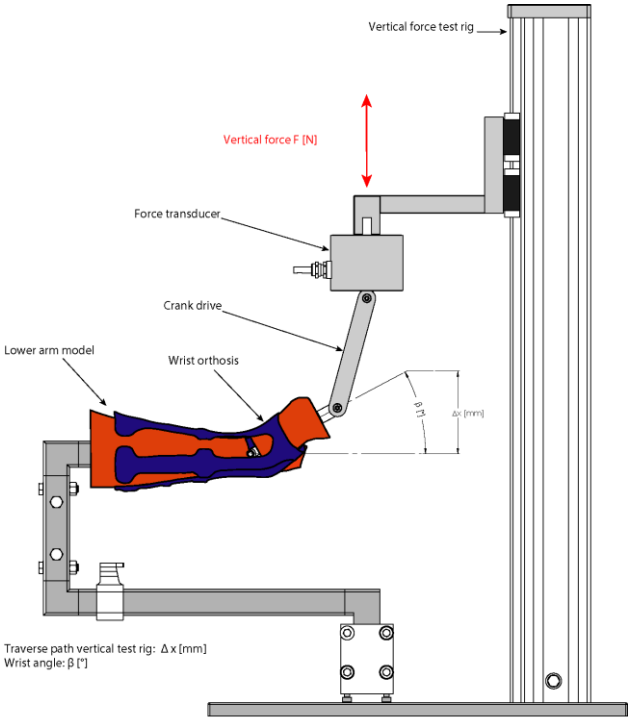

Figure 4: Schematic set-up of the test rig for wrist hand orthoses (WHOs).

Testing of simulated safety parameters. Strain distributions occurring on the surface of the WHO were evaluated using a 3D measurement system (ARAMIS SRX, GOM Metrology, Germany). In addition, strain gages positioned at four predefined locations on the WHO

were used to validate stresses and strains, see Figure 5. Testing with the non-contact optical measuring system was performed twice with five test cycles each, to measure the strain distributions both on the dorsal side (strain gages 1 to 3) and on the palmar side (strain gage 4). Additional repeatability tests were completed for the strain gages. Thus, the measurement of strains recorded by strain gages was repeated six times with five test cycles each. A total of 30 test cycles were recorded.

Acceptance criteria for performance predictions. For the 3D printed WHO, the acceptance criteria for the question of interest regarding the performance evaluation were (ii) the ROM and (ii) the stiffness of the orthosis, which is described by the slope of the moment-angle characteristic curve. The stiffness of the 3D printed WHO was determined by linear regression. The acceptance criteria were established by comparing the simulation results of the 3D printed WHO to the experimental data.

Acceptance criteria for strain predictions. The simulated safety parameters were evaluated with the maximum strains occurring as acceptance criteria. For COU 2, the mean strains in the preferred measurement direction of the strain gages, as well as the minor and major strains on the surface of the orthosis were experimentally obtained and compared with the simulation results.

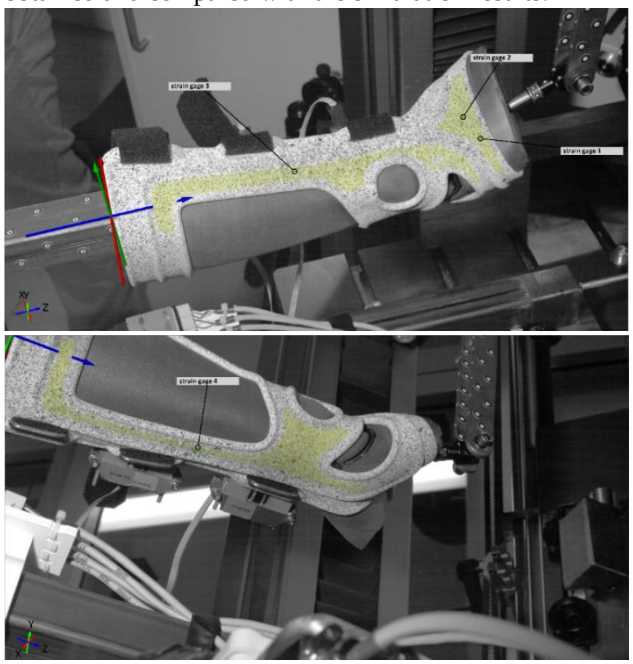

Figure 5: Positioning of the strain gages on the 3D printed WHO to validate the occurring strains.

II.VI. Credibility Goals

The following section describes the selected credibility goals, including its justification for a subset of credibility factors defined in V&V 40-2018. Evaluation of each credibility factor and the overall assessment of model credibility are presented in section *0. Results*. Note that this article does not represent a full review of model credibility according to the ASME V&V 40-2018 standard, an evaluation of all credibility factors would be needed.

To establish trust in the computational model, extensive V&V activities must be performed. According to the ASME Standards Committee on Verification and Validation in Computational Solid Mechanics, verification is “the process of determining that a computational model accurately represents the underlying mathematical model and its solution” [25]. Thus, the aim of verification is to ensure that the model is implemented correctly and performs as designed [1]. Verification must precede validation. Because, if a computational model is validated using a code that contains (serious) errors, it may lead to incorrect conclusions about the validity of the model. Validation is “the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model” [25]. Therefore, the primarily purpose of validation activities is to prove the correctness of the underlying model assumptions and demonstrate the extent to which sensitivities and uncertainties of the computational model and the associated comparator(s) are known [1]. In addition, according to the ASME V&V 40-2018 standard, the applicability assessment is an important evaluation step to demonstrate the relevance of the V&V evidence to the COU.

The credibility factors associated with verifying and validating the computational model for the given example of a custom-made 3D printed WHO as well as the applicability assessment are outlined in Table 3. Supplementary, for each credibility factor, a sample list of activities with increasing levels of rigor is presented, with the credibility goal set depending on the model risk. The model risk for the selected COUs in the given example is shown in Figure 2.

III. Results and discussion

The final step in the risk-based credibility assessment according to ASME V&V 40-2018 is to determine if the computational model is credible for the intended COU. Based on all information and evidence collected during model development, from physical testing, and other V&V activities, model credibility is assessed. The following section outlines the results obtained in the example of a custom-made 3D printed WHO and how the credibility goals were met for each COU.

Table 3: Goals and rationale for credibility factors associated with verification, validation, and applicability assessment for the given example of a custom-made 3D printed WHO according to the ASME V&V 40-2018 standard with an exemplary gradation of activities, listed from lowest to highest credibility [1].

Verification	Credibility goals and rationale
Code Verification	
Software Quality Assurance (SQA) (a) Very little or no SQA procedures were specified or followed. (b) SQA procedures were specified and documented. (c) In addition to (b); the software anomaly list and the software development environment were fully understood, and the impact on the COU was analyzed and documented; quality metrics were tracked.	Defined credibility goal for the given example: COU 1: (b); COU 2: (b) The gradation of activities is taken from the V&V 40-2018 standard. Since the CMS will replace the physical tests, the model risk in this example is assessed as level 2/3 (COU 1) and level 2 (COU 2). According to ASME V&V 40-2018, this is referred to as moderate model risk.
Numerical Code Verification (NCV) (a) NCV was not performed. (b) The numerical solution was compared to an accurate benchmark solution from another verified code. (c) Discretization error was quantified by comparison to an exact solution, and a grid convergence study demonstrated that the numerical solution asymptotically approached the exact solution as the discretization was refined. (d) In addition to the quantification of discretization error and the execution of a grid convergence study as described in (c), the observed order of accuracy was quantified and compared to the theoretical order of accuracy.	Defined credibility goal for the given example: COU 1: (b); COU 2: (b) The gradation of activities is taken from the V&V 40-2018 standard. Since the CMS will replace the physical tests, the model risk in this example is assessed as level 2/3 (COU 1) and level 2 (COU 2). According to ASME V&V 40-2018, this is referred to as moderate model risk.
Calculation Verification	
Discretization Error is the error associated with the chosen mesh, element type, and the level of mesh refinement (a) No grid or time-step convergence analysis was performed to estimate the discretization error. (b) Applicable grid or time-step convergence analyses were performed; convergence behaviors were observed to be stable; no estimation of discretization error. (c) Applicable grid or time-step convergence analyses were performed, and discretization error was estimated.	Defined credibility goal for the given example: COU 1: (b); COU 2: (b) The gradation of activities has been taken from the V&V 40-2018 standard. Since the CMS will replace the physical tests, the model risk in this example is assessed as level 2/3 (COU 1) and level 2 (COU 2). According to ASME V&V 40-2018, this is referred to as moderate model risk.
Validation	
Comparator	
Test Conditions	
Quantity of Test Conditions (a) A single test condition was examined. (b) Multiple (two to four) test conditions were examined. (c) More than four test conditions were examined.	Defined credibility goal for the given example: COU 1: (a); COU 2: (b) Deviating from the given V&V 40-2018 grading for moderate model risk, a single test condition was examined for COU 1. Stiffness is the key functional parameter to evaluate orthoses and commonly used. It is described by the slope of the moment-angle characteristic curve. The test setup provides the moments for given rotations. Thus, the single test condition is sufficient to obtain credibility. Activity gradation for COU 2 was taken from the ASME V&V 40-2018 standard in relation to model risk. A single measurement is not sufficient for a complete overview of the strain distributions.
Measurements of Test Conditions (a) Test conditions were qualitatively measured and/or characterized. (b) One or more key characteristics of the test conditions were measured. (c) All key characteristics of the test conditions were measured.	Defined credibility goal for the given example: COU 1: (b); COU 2: (b) The gradation of activities has been taken from the V&V 40-2018 standard. Since the CMS will replace the physical tests, the model risk in this example is assessed as level 2/3 (COU 1) and level 2 (COU 2). According to ASME V&V 40-2018, this is referred to as moderate model risk.
Uncertainty of Test Condition Measurements (a) Test conditions were not characterized or were characterized with gross observations; measurement uncertainty was not addressed. (b) Uncertainty analysis incorporated instrument accuracy only. (c) Uncertainty analysis incorporated instrument accuracy and repeatability (i.e., statistical treatment of repeated measurements). (d) Uncertainty analysis incorporated a comprehensive uncertainty quantification, which included instrument accuracy, repeatability, and other conditions affecting the measurements.	Defined credibility goal for the given example: COU 1: (b); COU 2: (b) The gradation of activities has been taken from the V&V 40-2018 standard. Since the CMS will replace the physical tests, the model risk in this example is assessed as level 2/3 (COU 1) and level 2 (COU 2). According to ASME V&V 40-2018, this is referred to as moderate model risk.

Assessment	
Output Comparison	
Quantity (a) A single output was compared. (b) Multiple outputs were compared.	Defined credibility goal for the given example: COU 1: (b); COU 2: (b) The gradation of activities has been taken from the V&V 40-2018 standard. Since the CMS will replace the physical tests, the model risk in this example is assessed as level 2/3 (COU 1) and level 2 (COU 2). According to ASME V&V 40-2018, this is referred to as moderate model risk.
Rigor of Output Comparison (a) Visual comparison was performed. (b) Comparison was performed by determining the arithmetic difference between computational results and experimental results. (c) Uncertainty in the output of the computational model or the comparator was used in the output comparison. (d) Uncertainties in the output of the computational model and the comparator were used in the output comparison.	Defined credibility goal for the given example: COU 1: (c); COU 2: (a) For COU 1, the gradation of activities has been taken from the V&V 40-2018 standard. Since the CMS will replace the physical tests, the model risk in this example is assessed as level 2/3 (COU 1) and level 2 (COU 2). According to ASME V&V 40-2018, this is referred to as moderate model risk. For COU 2, visual comparison was considered sufficient due to the high number of comparison data.
Agreement of Output Comparison (a) The level of agreement of the output comparison was not satisfactory for key comparisons. (b) The level of agreement of the output comparison was satisfactory for key comparisons, but not all comparisons. (c) The level of agreement of the output comparison was satisfactory for all comparisons.	Defined credibility goal for the given example: COU 1: (b); COU 2: (b) The gradation of activities has been taken from the V&V 40-2018 standard. Since the CMS will replace the physical tests, the model risk in this example is assessed as level 2/3 (COU 1) and level 2 (COU 2). According to ASME V&V 40-2018, this is referred to as moderate model risk.
Applicability	
Relevance of the Quantities of Interest (a) The quantities of interest from the validation activities were related, though not identical, to those for the COU. (b) A subset of the quantities of interest **from** the validation activities were identical to those for the COU. (c) The quantities of interest from the validation activities were identical to those for the COU.	Defined credibility goal for the given example: COU 1: (c); COU 2: (b) For COU 1, a higher gradation of activities was possible to achieve higher credibility, since the moment and rotation were fully measured in the validation tests and simulations. For COU 2, the gradation of activities has been taken from the V&V 40-2018 standard. Strains could only be measured for certain areas on the surface of the orthosis. In these areas, validation was possible. Successful validation in the selected areas provides sufficient confidence for the remaining areas.
Relevance of the Validation Activities to the COU (a) There was no overlap between the ranges of the validation points and the COU. (b) There was partial overlap between the ranges of the validation points and the COU. (c) The COU encompassed some of the validation points. (d) The COU encompassed all validation points, and the validation points spanned the entire COU space.	Defined credibility goal for the given example: COU 1: (b); COU 2: (b) The gradation of activities has been taken from the V&V 40-2018 standard. Since the CMS will replace the physical tests, the model risk in this example is assessed as level 2/3 (COU 1) and level 2 (COU 2). According to ASME V&V 40-2018, this is referred to as moderate model risk.

III.I. Assessment of model credibility (COU 1)

Verification. For both code and calculation verification, all credibility goals listed in Table 3 were achieved. Simulation was performed using the off-the-shelf software Ansys® Mechanical, Version 2022R2 (ANSYS, Inc., United States). The software provider has a quality management system certified according to ISO 9001 and including the software development process and quality assurance. The certificate is valid. The objective for the software quality assurance has thus been achieved. A number of benchmark tests are provided by the software vendor. These include tests against analytical calculations and tests against other benchmarked solutions. In order to achieve the goal for code verification, appropriate test cases were selected and computed. A convergence study with three different meshes was performed for all quantities of interest to verify the calculation.

Validation. The assessment of each credibility factor associated with validation is presented in Table 4. All specified credibility goals listed in Table 3 were achieved.

Table 4: Credibility assessment of COU 1 based on the selected credibility goals for validation

Comparator	
Test Conditions	Based on preliminary studies of maximum wrist moments, the moments in extension and flexion were determined for testing the custom-made 3D printed WHO. Pre-testing was performed for a number of different combinations of moments in both extension and flexion. The test parameters were set at 5 Nm in extension and 10 Nm in flexion because several orthosis designs failed at higher moments.

	Application of the WHO was monitored by means of two belt force sensors. Before each test, the WHO was applied to the forearm model with a 20 N application force. During the physical test, wrist angle and wrist moment were calculated from the force applied to the orthosis and the resulting vertical travel of the universal testing machine (20 kN Allround Table Top Zwick/Roell, Germany). A universal amplifier was used to record the measurement data over the course of the test. By repeating the test six times with five test cycles each and monitoring the application of the WHO to the forearm model before every test, the goal for the credibility factor *Comparator → Test Conditions* was achieved.
Assessment	
Output Comparison	
Quantity	The intended goal for this credibility factor was achieved through the comparison of the *stiffness* in extension and flexion. Due to different control mechanisms between the computational model (angle-driven) and the physical test (force-controlled), a comparison of the functional parameter *ROM* was not possible. However, as part of the stiffness characteristic comparison, the endpoints of the curves were evaluated.
Rigor of Output Comparison	The intended credibility goal was to quantify the uncertainties of the measured angles and moments, respectively the resulting stiffness, in the experimental data, and to compare these results with simulation outcomes. Thus, physical testing of each orthosis was repeated six times with five test cycles each. In addition, the total uncertainty of the physical test method was estimated according to the *»General requirements for the competence of testing and calibration laboratories« (ISO/IEC 17025:2017)*. Results were then compared to corresponding simulation data (a representative stiffness comparison is shown in Figure 6).
Agreement of Output Comparison	The comparison between experimental and computational moment-angle characteristics is shown in Figure 6. Qualitative assessment (level (a) for *Rigor of Output Comparison*) shows that the curve shapes agree reasonably well: For the most part, the computational model curve lies within the range of the coefficient of variation of the experimental data. For the quantitative comparison of the stiffness by determining the slope of the regression lines, level 2 was set for the *Rigor of Output Comparison and Agreement of Output Comparison*. Meaning differences between computational results and experimental data are less than 20 %. Except for flexion angles above $\alpha = 8{,}81°$, the determined differences in stiffnesses are less than 20 %. Therefore, the computational model is sufficiently credible, up to a flexion angle of $\alpha = 8{,}81°$, to evaluate the functional properties of custom-made 3D printed WHO across all sizes in the proposed product portfolio.

Applicability. Applicability is the relevance of the validation activities in support of the use of the computational model for a COU [1]. Therefore, it is assessed based on the relevance of the quantities of interest and the relevance of the validation activities. Table 5 shows how the selected credibility goals for *Applicability* were achieved for COU 1.

Table 5: Applicability assessment of COU 1 based on the selected credibility goals.

Relevance of the Quantities of Interest	The stiffness of the 3D printed WHO is a quantity of interest directly applicable to the question of interest. Stiffness (slope of the moment-angle characteristic) is a common parameter used to evaluate the functional properties of orthoses and is therefore an acceptable validation metric.
Relevance of the Validation Activities to the COU	Applicability is judged by how much "overlap" there is between the conditions used in the validation and the COU conditions. For the given example of a custom-made 3D printed WHO, the goal for this credibility factor only requires a partial overlap. To achieve this, the moments applied to the WHO are within the COU and hence the possible operating range.

Overall, for COU 1, the computational model was sufficiently credible to evaluate the functional parameter stiffness of the custom-made 3D printed WHO across all sizes in the proposed product portfolio, but only up to a flexion angle of $\alpha = 8{,}81°$. Conclusions about the accuracy of predictions of the ROM were not possible due to different control mechanisms between computational model (angle-driven) and physical test (force-controlled).

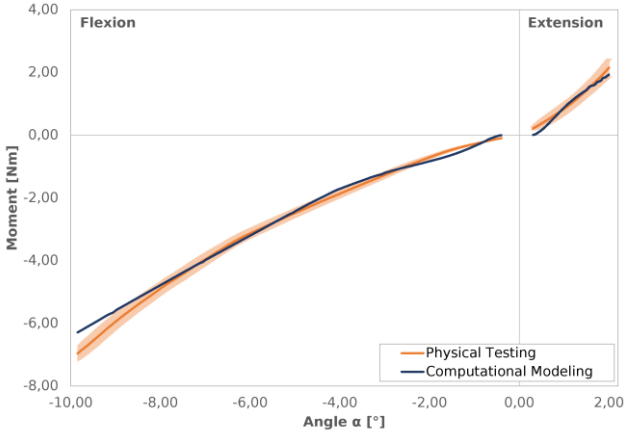

Figure 6: Comparison of the moment-angle characteristic curve between simulation and experimental data. The pale orange area around the curve shows the coefficient of variation for the experimental data.

III.II. Assessment of model credibility (COU 2)

Verification. All of the verification activities listed for COU 1 also apply to COU 2.

Validation. The credibility assessment regarding the validation activities for COU 2 is shown in Table 6. All intended credibility goals listed in Table 3 were met.

Table 6: Credibility assessment of COU 2 based on the selected credibility goals for validation.

Comparator	
Test Conditions	Mirroring COU 1, the test parameter for the wrist moment was set at 5 Nm in extension and 10 Nm in flexion. Application of the WHO was also monitored using two belt force sensors and a 20 N application force. The universal amplifier and the 3D measurement system were synchronized using a trigger signal from the universal material testing machine to compare the data collected with the strain gages and the non-contact optical measuring system. The strain gages detected strains in the material of the WHO prior to measurement caused by bonding the strain gages to a curved, non-planar surface, applying the orthosis to the forearm model and closing the straps. These strains are defined as an offset and are subtracted from the data obtained with the strain gages. By repeating the test six times with five test cycles each and monitoring the application of the WHO to the forearm model before every test, the goal for the credibility factor *Comparator → Test Conditions* was achieved.
Assessment	
Output Comparison	
Quantity	The intended goal for this credibility factor was achieved by comparing the *maximum strains occurring* and the *major strain characteristic* in extension and flexion. For the evaluation of the major strains, the results of the computational model are compared with those of a non-contact optical measurement system, since strain gages only allow determination of local strains. Due to the low measurement signal and the high signal-to-noise ratio, the significance of the *major strain data* is limited. Therefore, the *major strain characteristic* is not considered when evaluating the validity of the computational model.
Rigor of Output Comparison	The intended credibility goal was achieved by visually comparing the strains calculated by the simulation with the experimental data. To quantify the uncertainties of the measured strains, physical testing of each orthosis was repeated six times with five test cycles each. The standard deviation was determined for the data recorded with the strain gages and the non-contact optical measuring system, shown as the area around the curves in Figure 7 and Figure 8, respectively.
Agreement of Output Comparison	The comparison between experimental and computational strain characteristics is shown in Figure 7 and in Figure 8 for the major strain characteristic, respectively. Load phases in extension (orange) and flexion (light blue) are highlighted in the figures. The white spaces represent the release phase, which was not simulated. Qualitative assessment (level (a) for *Rigor of Output Comparison*) shows that the characteristic curve of the recorded maximum strains is almost identical for all test methods. For the major strain characteristic, the qualitative evaluation is similar, see Figure 8.

Applicability. For the applicability assessment, the achievement of the credibility goals is shown in Table 7.

Table 7: Applicability assessment of COU 2 based on the selected credibility goals.

Relevance of the Quantities of Interest	The strains occurring in the 3D printed WHO are a quantity of interest directly applicable to the question of interest for COU 2. They are used to measure deformations on the surface of the device and thus are an assessment of material stressing. Therefore, strains are an acceptable validation metric.
Relevance of the Validation Activities to the COU	Considering the model risk for COU 2, the goal for this credibility factor required a partial overlap between the conditions used in the validation and the COU conditions. As for COU 1, this is accomplished by ensuring that the parameters for testing the 3D printed WHO are within the COU and thus the possible operating range.

III.III. Reporting of Computational Model and Credibility Assessment Results

As a final step, the V&V 40-2018 standard recommends documenting the activities performed for verification, validation, and applicability in a report. It should include background information on the product or process being modeled, but also key details and rationales regarding the question of interest and the COU, the computational model including model risk, selected credibility goals, V&V activities, and the credibility assessment. As mentioned in the introduction, the FDA has published a guidance document on this subject to provide recommendations on FDA-compliant reports of computational and modeling studies that are used to support medical device submissions [14].

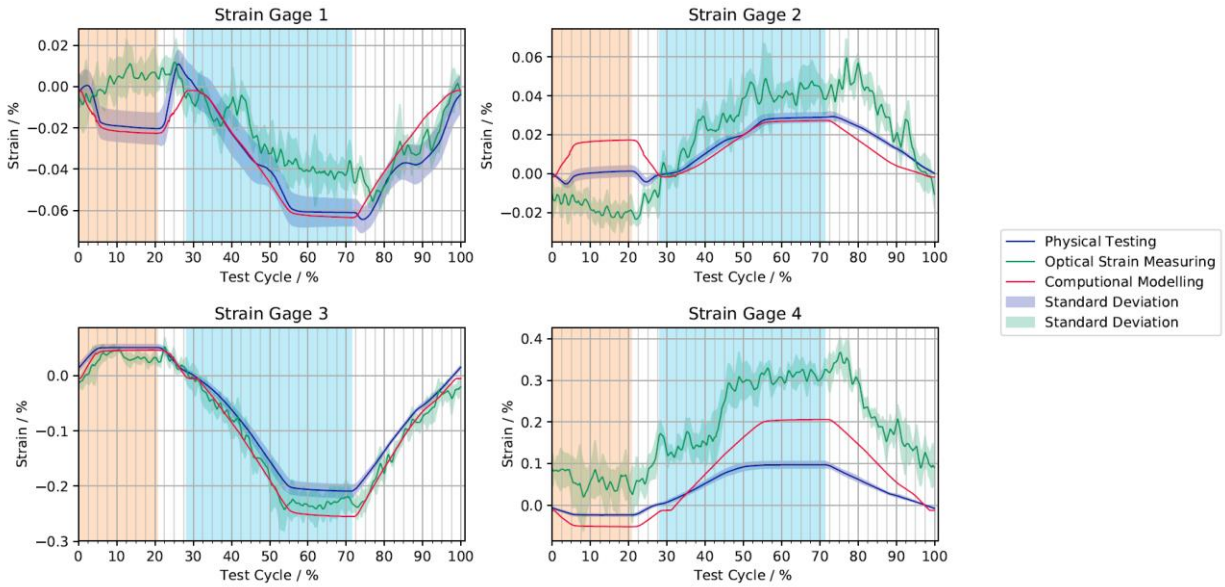

Figure 7: Representation of the characteristic curve of the strains depending on the test cycle for strain gage 1 to 4, respectively. Compared are the results of the strain gages (blue), the non-contact optical measurement system (green), and the computational model (red). The colored area around the curve represents the standard deviation. The load phase in extension (orange) and in flexion (light blue) is highlighted in color, whereas the release phase is shown in white.

III.IV. Discussion

Over the past few years, CMS has emerged as an essential tool in providing information about the technical performance, safety, and effectiveness of medical devices. It has the potential to significantly impact development processes and clinical evaluation of devices in the medical technology industry. However, limited regulatory requirements and lack of consensus on the level of V&V activities required to establish sufficient model credibility for decision-making, result in rare use of CMS as a source of evidence in the approval process of medical devices. The ASME V&V 40-2018 gives guidance for establishing appropriate credibility requirements of a computational model based on risk.

The main objectives of this paper were to outline the concepts of the risk-based framework for establishing appropriate credibility requirements of a computational model – the ASME V&V 40-2018 standard – and to demonstrate its application using an example from the orthotics field. Although the intent was to present a detailed use case of the standard, not all credibility factors were listed and evaluated in this article for lack of space. For more information, readers are referred to ASME V&V 40-2018, whose annex contains six application examples.

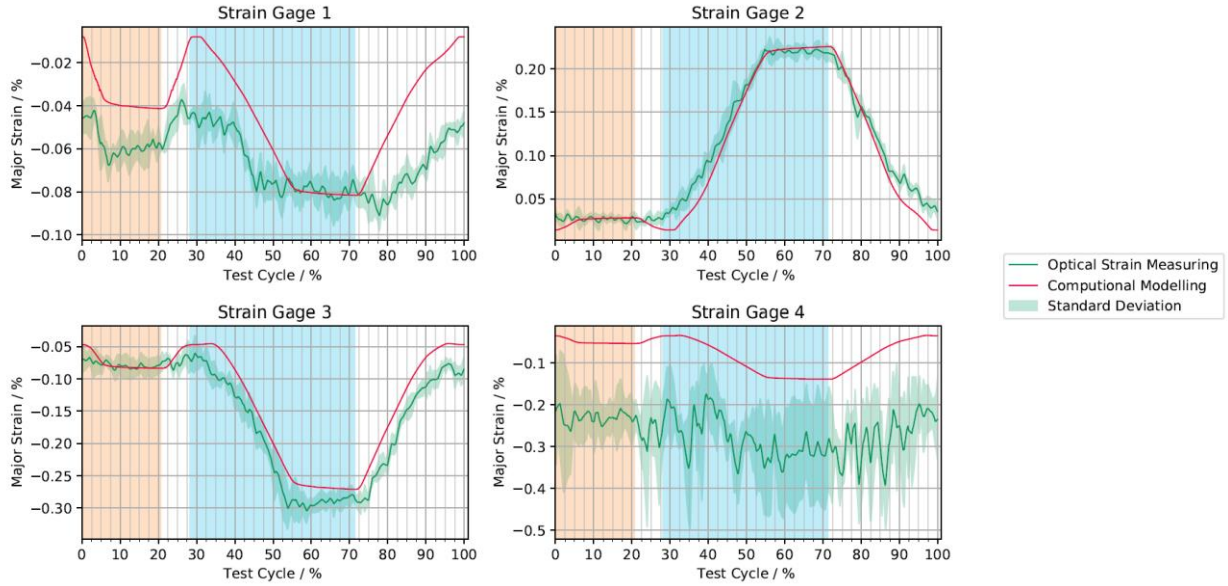

Figure 8: Representation of the major strain characteristic depending on the test cycle for strain gage 1 to 4, respectively. Compared are the results of the non-contact optical measurement system (green) with the computational model (red). The colored area around the curve represents the standard deviation. The load phase in extension (orange) and in flexion (light blue) is highlighted in color, whereas the release phase is shown in white.

The findings presented were collected as part of the BMBF-funded project SIGMA3D. However, the given example is based on a medical device that is placed on the market as custom-made. Those devices qualify for an exemption to pre-market approval standards in many jurisdictions and go through a simplified approval process. According to the MDR, custom-made devices are not required to carry the CE marking. Therefore, the data shown have not yet been used as scientific evidence in regulatory device submissions. Nonetheless, the steps outlined to establish model credibility as well as the results and conclusions, represent a real-life use case and are not hypothetical. The process described has shown that different COUs can be applied to the same computational model and that model risk determines the credibility requirements of each V&V activity. Because the computational model has been determined to be credible for the intended COUs, the subsequent application must use the best practices established during the validation process, when making predictions with the model. Furthermore, the use of the computational model should be consistent with the chosen COUs to maintain appropriate credibility.

The COU of computational models used in the medical device industry is often directly related to patient outcomes or device performance in clinical practice. Therefore, it might seem inadequate to validate and evaluate the relevant aspects of the reality of interest with rig tests as comparator. However, especially in the field of orthotics, there is lack of normative guidelines for testing devices, and testing is context-sensitive. In the example given, the physical test methods were developed according to VDI guideline 5703 "Systematical development for a model-based testing of medical devices" [26]. Test parameters were defined based on literature data or, when no data were available, based on measurements performed. Given the limitations, the selected metrics reflect the actual operating range in which patients use the orthosis. The developed test bench procedure was characterized in terms of repeatability, reproducibility and robustness. Accordingly, the bench test presented is a suitable comparator for assessing the potential impact on patients.

The approach described in ASME V&V 40-2018 may be understood as an iterative process in practice. If the credibility activities and V&V results are deemed insufficient to establish model credibility for the specified COU, the standard identifies appropriate steps for revising aspects of the process. These include, but are not limited to, (i) performing additional credibility activities, such as reviewing and adjusting the gradation levels for each credibility factor, (ii) modifying the computational model, (iii) reducing the influence of the computational model on decision-making and thus the model risk and credibility goals, respectively, or (iv) adjusting the COU. Users may use any or a combination of the above to achieve sufficient model credibility for the COU.

IV. Conclusions

To conclude, the ASME V&V 40-2018 standard provides a detailed framework for establishing appropriate credibility requirements of a computational model based on its risk. It offers an indication for users to justify *»how much«* rigor is required, rather than providing a step-by-step instructions on *»how to«* perform corresponding V&V activities. Together with the FDA guideline »Reporting of Computational Modeling Studies in Medical Device Submissions«, the ASME V&V 40-2018 standard provides guidance on the use of CMS as scientific evidence in regulatory submissions.

There are no comparable regulatory requirements in Europe. The MDR, in effect since May 2021, explicitly calls for the use of computational models, but does not provide specific requirements or guidance. Similarly, the position paper published by the Avicenna Alliance in 2021 does not state its own specifications, but refers to the requirements of the ASME V&V 40-2018 standard [1] and the FDA guidance document [14], respectively. Therefore, medical device manufacturers in Europe using CMS as a source of evidence in the approval process should refer to the guidelines given by the United States in order to demonstrate sufficient model credibility.

Acknowledgments

We gratefully acknowledge the support provided by the Mecuris GmbH (Munich, Germany) throughout this research, especially by Mr. Ylli Binakaj. Their input, commitment and collaboration significantly enhanced the quality of our research.
Research funding: Funded by the German Federal Ministry of Education and Research (BMBF).

Author's statement

Conflict of interest: Ann-Kathrin Carl, Maxim Kirillov, and David Hochmann state no conflict of interest. Eric Quadrat is an employee of Simq GmbH, Grafing near Munich, Germany.

References

[1] American Society of Mechanical Engineers, V&V 40-2018: Assessing Credibility of Computational Modeling through Verification and Validation: Application to Medical Devices.
[2] P. Favre, et al., In Silico Clinical Trials in the Orthopedic Device Industry: From Fantasy to Reality?, Annals of Biomedical Engineering, 49(12), pp. 3213–3226, 2021.
[3] B. Stott, et al., A Critical Comparison of Comparators Used to Demonstrate Credibility of Physics-Based Numerical Spine Models, Annals of Biomedical Engineering, 51(1), pp. 150–162, 2023.
[4] Thomas O'Connor: Perspective on the Validation of Computational Models for Establishing Control Strategies, April 2019.
[5] C. Curreli, et al., A Credibility Assessment Plan for an In Silico Model that Predicts the Dose-Response Relationship of New Tuberculosis Treatments, Annals of Biomedical Engineering, 51(1), pp. 200–210, 2023.
[6] F. Pappalardo, et al., In silico clinical trials: concepts and early adoptions, Briefings in Bioinformatics, 20(5), pp. 1699–1708, 2019.
[7] M. Viceconti, et al., Possible Contexts of Use for In Silico Trials Methodologies: A Consensus-Based Review, IEEE Journal of Biomedical and Health Informatics, 25(10), pp. 3977–3982, 2021.
[8] M. Taylor and P. J. Prendergast, Four decades of finite element analysis of orthopaedic devices: where are we now and what are the opportunities?, Journal of Biomechanics, 48(5), pp. 767–778, 2015.

[9] M. R. Khosravani, et al., Characterization of 3D-printed PLA parts with different raster orientations and printing speeds, Scientific Reports, 12(1), pp. 1016, 2022.
[10] A. E. Anderson, B. J. Ellis, and J. A. Weiss, Verification, validation and sensitivity studies in computational biomechanics, Computer Methods in Biomechanics and Biomedical Engineering, 10(3), pp. 171–184, 2007.
[11] T. M. Morrison, et al., Assessing Computational Model Credibility Using a Risk-Based Framework: Application to Hemolysis in Centrifugal Blood Pumps, ASAIO (American Society for Artificial Internal Organs) Journal, 65(4), pp. 349–360, 2019.
[12] Medical Device Innovation Consortium: Landscape Report & Industry Survey on the Use of Computational Modeling & Simulation in Medical Device Development, 2023. https://mdic.org/wp-content/uploads/2023/01/CMS_Landscape_Report.pdf. Accessed 25 May 2023.
[13] U.S. Food and Drug Administration: Regulatory Science Priorities (FY2016), October 2015. https://www.fda.gov/media/94312/download. Accessed 13 June 2022.
[14] U.S. Food and Drug Administration, Guidance for Industry and Food and Drug Administration Staff: Reporting of Computational Modeling Studies in Medical Device Submissions, September 2016. https://www.fda.gov/media/87586/download. Accessed 19 May 2022.
[15] Avicenna Alliance - Association for Predictive Medicine: In silico medicine and the Avicenna Alliance. https://avicenna-alliance.com/files/user_upload/PDF/P80303001246-0421_AVI_In_Silico_Flyer_4P_Web.pdf. Accessed 25 May 2023.
[16] M. Horner, M. Reiterer, and C. F. Rousseau: Avicenna Alliance Position Paper, Ensuring the Quality of in silico Evidence: Application to Medical Devices, 2021. https://avicenna-alliance.com/files/user_upload/Avicenna_Alliance_Position_paper_in_silico_evidence_application_to_Medical_Devices.pdf. Accessed 25 May 2023.
[17] 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, MDR.
[18] F. Berti, et al., How to Validate in silico Deployment of Coronary Stents: Strategies and Limitations in the Choice of Comparator, Frontiers in Medical Technology, 3, pp. 702656, 2021.
[19] B. Parvinian, et al., Credibility Evidence for Computational Patient Models Used in the Development of Physiological Closed-Loop Controlled Devices for Critical Care Medicine, Frontiers in Physiology, 10, pp. 220, 2019.
[20] P. Pathmanathan and R. A. Gray, Validation and Trustworthiness of Multiscale Models of Cardiac Electrophysiology, Frontiers in Physiology, 9, pp. 106, 2018.
[21] American Society of Mechanical Engineers, V&V 10-2019: Standard for Verification and Validation in Computational Solid Mechanics.
[22] American Society of Mechanical Engineers, ASME V&V 20-2009: Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer.
[23] M. Horner: ASME V&V 40 Overview, 2019. https://cstools.asme.org/csconnect/FileUpload.cfm?View=yes&ID=57744. Accessed 5 June 2023.
[24] T. Grigartzik, 2019: Entwicklung eines Prüfverfahrens zur Prüfung und Bewertung von Handgelenkorthesen. Bachelorarbeit, FH Münster.
[25] L. E. Schwer, An overview of the PTC 60/V&V 10: guide for verification and validation in computational solid mechanics, Engineering with Computers, 23(4), pp. 245–252, 2007.
[26] VDI-Richtlinie 5703 Blatt 1:2021-12: Systematische Entwicklung modellbasierter Prüfungen für Medizinprodukte - Grundlagen.