ASME V&V 40-2018

# Assessing Credibility of Computational Modeling Through Verification and Validation: Application to Medical Devices



AN INTERNATIONAL STANDARD

The American Society of Mechanical Engineers

ASME V&V 40-2018

# Assessing Credibility of Computational Modeling Through Verification and Validation: Application to Medical Devices



AN INTERNATIONAL STANDARD

The American Society of Mechanical Engineers

Two Park Avenue • New York, NY • 10016 USA

Date of Issuance: November 19, 2018

This Standard will be revised when the Society approves the issuance of a new edition.

ASME issues written replies to inquiries concerning interpretations of technical aspects of this Standard. Interpretations are published on the Committee web page and under http://go.asme.org/InterpsDatabase. Periodically certain actions of the ASME V&V Committee may be published as Cases. Cases are published on the ASME website under the V&V Committee Page at http://go.asme.org/VnVcommittee as they are issued.

Errata to codes and standards may be posted on the ASME website under the Committee Pages to provide corrections to incorrectly published items, or to correct typographical or grammatical errors in codes and standards. Such errata shall be used on the date posted.

The V&V Committee Page can be found at http://go.asme.org/VnVcommittee. There is an option available to automatically receive an e-mail notification when errata are posted to a particular code or standard. This option can be found on the appropriate Committee Page after selecting "Errata" in the "Publication Information" section.



The American Society of Mechanical Engineers
Two Park Avenue, New York, NY 10016-5990

# CONTENTS

**Tables**

# FOREWORD

Computational models have been used to support the design of medical devices for many years, without any specific guidance on how to assess their credibility. Device manufacturers therefore use internal approaches and best practices for model verification and validation (V&V). This has created challenges for regulatory agencies to develop consistent, structured approaches for evaluating the legitimacy of model results used to support device safety and/or effectiveness.

In recognition of the challenges facing the device industry, the U.S. Food and Drug Administration (FDA) hosted the first in an annual series of workshops on computational modeling for medical devices in 2008. The intent of this series was to bring together researchers, medical device manufacturers, and regulatory agencies to present advanced research, review best practices, and address barriers to the use of computational modeling for the design, development, and evaluation of medical devices. Based on several years of input, it became clear that guidance on V&V for computational models was necessary to support and promote appropriate use of computational modeling in medical device design, development, and evaluation. Due to the growing interest in V&V of computational modeling for medical devices within the ASME V&V subcommittees, the ASME V&V Standards Committee proposed the development of a new subcommittee focused on this area.

The proposal for a new V&V subcommittee focused on medical devices was presented at various device-related conferences over the course of several years, with increasing interest from the medical device community. In 2011, the ASME V&V 40 Subcommittee on Verification and Validation of Computational Modeling for Medical Devices was officially approved. The Subcommittee is composed of members representing a broad cross-section of the medical device community, including device manufacturers, academic groups, consultants, software developers, and government agencies (primarily the FDA). The breadth of knowledge of the Subcommittee members spans solid mechanics, fluid dynamics, electromagnetics, kinematic modeling, and other physics-based modeling.

At the initiation of the ASME V&V 40 Subcommittee, standardization of the V&V process had already been addressed by the first two ASME V&V subcommittees (V&V 10 Verification and Validation in Computational Solid Mechanics, and V&V 20 Verification and Validation in Computational Fluid Dynamics and Heat Transfer). The V&V 40 Subcommittee therefore set out to provide guidance on the application of V&V practices for medical devices. The anticipated guidance would provide a level of standardization for V&V practice that would encourage sound use of modeling to support device development and facilitate objective and consistent evaluation of model credibility by device manufacturers and regulatory agencies.

Medical devices are classified by the FDA based on risk to patients, which requires a greater level of evidence to demonstrate the safety and effectiveness of medical devices that pose a higher risk to patients. Analogously, the V&V 40 Subcommittee focused on developing a risk-based approach to determine the level of V&V needed to support the use of a computational model for evaluating device safety and/or effectiveness. The concept of risk is also foundational to NASA-STD-7009, which predated the V&V 40 Subcommittee and informed the Subcommittee's perspective. However, NASA-STD-7009 explicitly links the required level of V&V activities to each risk level. In contrast, the consensus perspective of the V&V 40 Subcommittee was that the individual organization (e.g., a medical device manufacturer) should have the authority and responsibility to associate a certain level of risk with a certain set of V&V activities, and that the individual organization should justify this association to internal and external stakeholders, including regulatory agencies. Therefore, instead of defining specific credibility criteria, the V&V 40 Subcommittee developed a framework that allows users to determine the appropriate level of credibility required for their computational model.

Several foundational materials for the subcommittee (e.g., NASA-STD-7009, as well as the Predictive Capability Maturity Model introduced in SAND2007-5948) prescribe matrix frameworks. The V&V 40 Subcommittee also started with two matrices: the risk assessment matrix (RAM) and the credibility assessment matrix (CAM). The RAM focused on determining the level of risk for a computational model, while the CAM focused on the level of credibility (achieved through V&V activities) needed to satisfy that level of risk. Case studies conducted in 2013 that used the RAM and CAM exposed a number of practical and functional challenges with these matrices across the spectrum of medical devices, manufacturers, and model applications. Therefore, the V&V 40 Subcommittee revised the RAM/CAM framework, enabling users to define appropriate gradations and levels for risk and credibility. The culmination of these efforts is a risk-informed credibility assessment framework, reflecting the core principle that model credibility is commensurate with the risk associated with decisions influenced by the computational model.

Under the jurisdiction of the ASME Board on Standardization and Testing, ASME V&V 40-2018 was approved by the ASME V&V 40 Subcommittee and the ASME V&V Standards Committee on November 29, 2017. It was approved as an American National Standard by the American National Standards Institute (ANSI) on August 16, 2018.

# ASME V&V COMMITTEE
# Verification and Validation in Computational Modeling and Simulation

(The following is the roster of the Committee at the time of approval of this Standard.)

## STANDARDS COMMITTEE OFFICERS

**T. M. Morrison,** *Chair*
**B. H. Thacker,** *Vice Chair*
**R. L. Crane,** *Secretary*

## STANDARDS COMMITTEE PERSONNEL

**M. B. Benedict,** U.S. Air Force Research Laboratory
**J. Bischoff,** Zimmer Biomet
**R. L. Crane,** The American Society of Mechanical Engineers
**S. W. Doebling,** Los Alamos National Laboratory
**K. Dowding,** Sandia National Laboratories
**C. J. Freitas,** Southwest Research Institute
**M. Horner,** ANSYS, Inc.
**H. Lee,** Bettis Laboratory
**D. M. Moorcroft,** Federal Aviation Administration
**T. M. Morrison,** U.S. Food and Drug Administration
**L. J. Peltier,** Bechtel Nuclear, Security, & Environmental
**S. Rachuri,** U.S. Department of Energy
**R. R. Schultz,** Consultant
**B. H. Thacker,** Southwest Research Institute

## V&V 40 SUBCOMMITTEE — VERIFICATION AND VALIDATION IN COMPUTATIONAL MODELING OF MEDICAL DEVICES

**T. M. Morrison,** *Chair,* U.S. Food and Drug Administration
**J. Bischoff,** *Vice Chair,* Zimmer Biomet
**M. Horner,** *Vice Chair,* ANSYS, Inc.
**R. L. Crane,** *Secretary,* The American Society of Mechanical Engineers
**P. Afshari,** Depuy Synthes Spine
**D. Bardot,** Medical Device Innovation Consortium
**A. Bestelmeyer,** BD
**J. P. Bodner,** Medtronic, PLC
**S. Cheng,** Quidel Corp.
**B. D. Choules,** Embry-Riddle Aeronautical University
**R. Chow,** Boston Scientific Corp.
**J. C. Coburn,** U.S. Food and Drug Administration
**C. Corrales,** Baxter Healthcare Corp.
**K. K. Debus,** Siemens PLM Software
**M. Dharia,** Zimmer Biomet
**S. Eswaran,** Abbott Vascular
**C. M. Funkhouser,** Baxter Healthcare Corp.
**K. Genc,** Synopsys, Inc.
**M. Goodin,** SimuTech Group, Inc.
**I. Guler,** Boston Scientific Corp.
**A. Gupta,** Google, Inc.
**P. Hariharan,** U.S. Food and Drug Administration
**W. Hary,** HeartFlow
**H. Jin,** Medtronic, PLC
**A. Kiapour,** 4WEB Medical, Inc.
**L. Knudsen,** Syncroness, Inc.
**S. Kulkarni,** VEXTEC Corp.
**D. Levine,** Zimmer Biomet
**X. Li,** Abbott Structural Heart
**X. Liu,** Stryker
**B. A. Lurie,** W. L. Gore & Associates
**R. Marinescu,** Smith & Nephew
**L. Mulugeta,** InSilico Labs, LLC
**W. A. Olson,** Ethicon Endo-surgery
**C. Popelar,** Southwest Research Institute
**A. C. Rau,** Exponent, Inc.
**N. Rebelo,** Dassault Systemes Simulia Corp.
**T. L. Rossman,** Mayo Clinic
**P. Saffari,** Engage Medical Device Services, Inc.
**C. Scotti,** W. L. Gore & Associates
**R. Swift,** Cook Research, Inc.
**P. Tomaszewski,** DePuy Synthes Joint Reconstruction
**T. Zhao,** Edwards Lifesciences
**C. Basciano,** *Alternate,* BD
**P. Briant,** *Alternate,* Exponent, Inc.
**S. Sastry,** *Alternate,* W. L. Gore & Associates

# CORRESPONDENCE WITH THE V&V COMMITTEE

**General.** ASME Standards are developed and maintained with the intent to represent the consensus of concerned interests. As such, users of this Standard may interact with the Committee by requesting interpretations, proposing revisions or a case, and attending Committee meetings. Correspondence should be addressed to:

Secretary, V&V Standards Committee
The American Society of Mechanical Engineers
Two Park Avenue
New York, NY 10016-5990
http://go.asme.org/Inquiry

**Proposing Revisions.** Revisions are made periodically to the Standard to incorporate changes that appear necessary or desirable, as demonstrated by the experience gained from the application of the Standard. Approved revisions will be published periodically.

The Committee welcomes proposals for revisions to this Standard. Such proposals should be as specific as possible, citing the paragraph number(s), the proposed wording, and a detailed description of the reasons for the proposal, including any pertinent documentation.

**Proposing a Case.** Cases may be issued to provide alternative rules when justified, to permit early implementation of an approved revision when the need is urgent, or to provide rules not covered by existing provisions. Cases are effective immediately upon ASME approval and shall be posted on the ASME Committee web page.

Requests for Cases shall provide a Statement of Need and Background Information. The request should identify the Standard and the paragraph, figure, or table number(s), and be written as a Question and Reply in the same format as existing Cases. Requests for Cases should also indicate the applicable edition(s) of the Standard to which the proposed Case applies.

**Interpretations.** Upon request, the V&V Standards Committee will render an interpretation of any requirement of the Standard. Interpretations can only be rendered in response to a written request sent to the Secretary of the V&V Standards Committee.

Requests for interpretation should preferably be submitted through the online Interpretation Submittal Form. The form is accessible at http://go.asme.org/InterpretationRequest. Upon submittal of the form, the Inquirer will receive an automatic e-mail confirming receipt.

If the Inquirer is unable to use the online form, he/she may mail the request to the Secretary of the V&V Standards Committee at the above address. The request for an interpretation should be clear and unambiguous. It is further recommended that the Inquirer submit his/her request in the following format:

| | |
|---|---|
| Subject: | Cite the applicable paragraph number(s) and the topic of the inquiry in one or two words. |
| Edition: | Cite the applicable edition of the Standard for which the interpretation is being requested. |
| Question: | Phrase the question as a request for an interpretation of a specific requirement suitable for general understanding and use, not as a request for an approval of a proprietary design or situation. Please provide a condensed and precise question, composed in such a way that a "yes" or "no" reply is acceptable. |
| Proposed Reply(ies): | Provide a proposed reply(ies) in the form of "Yes" or "No," with explanation as needed. If entering replies to more than one question, please number the questions and replies. |
| Background Information: | Provide the Committee with any background information that will assist the Committee in understanding the inquiry. The Inquirer may also include any plans or drawings that are necessary to explain the question; however, they should not contain proprietary names or information. |

Requests that are not in the format described above may be rewritten in the appropriate format by the Committee prior to being answered, which may inadvertently change the intent of the original request.

Moreover, ASME does not act as a consultant for specific engineering problems or for the general application or understanding of the Standard requirements. If, based on the inquiry information submitted, it is the opinion of the Committee that the Inquirer should seek assistance, the inquiry will be returned with the recommendation that such assistance be obtained.

ASME procedures provide for reconsideration of any interpretation when or if additional information that might affect an interpretation is available. Further, persons aggrieved by an interpretation may appeal to the cognizant ASME Committee or Subcommittee. ASME does not "approve," "certify," "rate," or "endorse" any item, construction, proprietary device, or activity.

**Attending Committee Meetings.** The V&V Standards Committee regularly holds meetings and/or telephone conferences that are open to the public. Persons wishing to attend any meeting and/or telephone conference should contact the Secretary of the V&V Standards Committee. Future Committee meeting dates and locations can be found on the Committee Page at http://go.asme.org/VnVcommittee.

INTENTIONALLY LEFT BLANK

# Assessing Credibility of Computational Modeling Through Verification and Validation: Application to Medical Devices

## 1 EXECUTIVE SUMMARY

Computational modeling can be used throughout the product life cycle to provide information about technical performance, safety, and effectiveness of medical devices. Computational models can also be used to assess aspects of in vivo performance without subjecting patients (or animals) to potential harm or unnecessary risk. Establishing the credibility of a computational model to assess performance is important because of the potential risk a device presents to patients and/or healthcare providers.

Model credibility can be established through verification and validation (V&V) activities. Although methods for V&V are becoming well established, guidance is lacking on assessing the relevance and adequacy of the V&V activities for computational models used to support medical device development and evaluation. Given the inherent risk of using a computational model as a basis for predicting medical device performance, the ASME V&V 40 Subcommittee has developed a risk-informed credibility assessment framework. The framework centers on establishing that model credibility is commensurate with the risk associated with the decisions influenced by the computational model. Thus, the intent of this Standard is to provide guidance on how to establish and communicate risk-informed credibility of computational models used in the evaluation of medical devices.

## 2 INTRODUCTION

### 2.1 Motivation

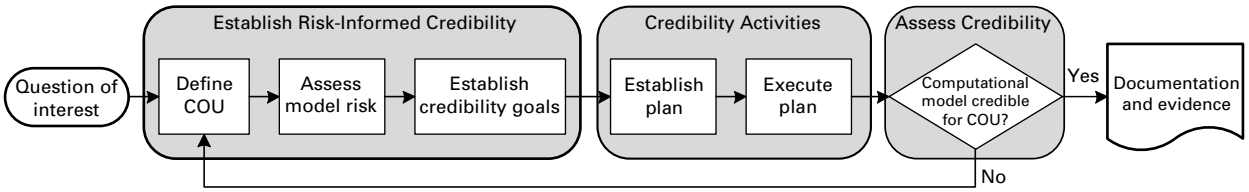
Computational modeling can be used to provide information that supports decisions related to the technical performance, safety, and/or effectiveness of medical devices. Computational models can be used throughout the total product life cycle of medical devices, from validating initial concept, design, and development, to supporting nonclinical and clinical activities, to providing postmarket surveillance. Medical device manufacturers may use computational models to augment in vitro and in vivo evaluations or to simulate such evaluations when they are unjustifiably invasive or prohibitive, and/or are deemed unreasonable. Moreover, computational models may also be used for evaluations that are not possible experimentally or clinically.

Decisions about the performance and/or safety of medical devices have potentially significant consequences, such as patient harm. Because computational modeling plays an increasingly important role in these decisions, there is an increased need to ensure that computational models appropriately represent reality. This can be accomplished through V&V. A considerable body of work on V&V and uncertainty quantification exists and continues to mature. ASME V&V 10 (ref. [1]) presents a general framework for V&V for computational solid mechanics. Additionally, ASME V&V 20 (ref. [2]) outlines a V&V procedure for computational fluid dynamics and heat transfer, both of which are generally applicable to physics-based computational models. As described in the referenced standards, the aim of V&V is to assess the degree to which the computational model is an accurate representation of the reality of interest through the comparison of simulation results with theory, carefully designed and controlled experiments, or other sources of relevant information. However, the relevance and adequacy of the V&V activities, and thus the computational model credibility, are subjective. This can create a lack of common understanding of expectations between stakeholders on what constitutes a *sufficiently* verified and validated computational model. Moreover, while ASME V&V 10 and ASME V&V 20 mention credibility, neither offers guidance on how to establish credibility.

The aim of this Standard is to present a framework for assessing the credibility of a computational model. The framework integrates concepts from two foundational documents: SAND2007-5948 (ref. [3]) and NASA-STD-7009 (ref. [4]). The predictive capability maturity model (PCMM) method of SAND2007-5948 describes different levels of model maturity but does not link maturity with how the computational model could be used to support a decision. NASA-STD-7009 defines the risk associated with using a computational model as a combination of the influence the simulation results have on the decision and the consequence of making a wrong decision. Based on the risk assessment results and programmatic priorities, NASA-STD-7009 specifies a quantitative and/or qualitative level of credibility that needs to be achieved for each modeling and simulation activity.

This Standard provides a risk-informed credibility assessment framework to empower the medical device industry to determine and justify the appropriate level of credibility for using a computational model to inform a decision. The decision could be internal to an organization or part of a regulatory activity, e.g., research or review. Therefore, this

**Figure 2.4-1 Process Diagram of the Risk-Informed Credibility Assessment Framework**




Standard may also be used by regulatory bodies to evaluate the appropriateness and adequacy of credibility activities and the overall model credibility.

## 2.2 Purpose

The purpose of this Standard is to provide a framework for assessing the relevance and adequacy of completed V&V activities that establish credibility of a computational model. The credibility should be commensurate with the degree to which the computational model is relied on as evidence of device performance, functional characteristic, and/or safety to support a decision, and the consequences of that decision being incorrect. This Standard will help users communicate the value of the completed V&V activities and establish the associated credibility of the computational model to support a decision.

## 2.3 Scope

The scope of the Standard encompasses physics-based computational models used for medical device applications. This Standard augments other standards that present V&V methodologies, such as ASME V&V 10 and ASME V&V 20. Therefore, this Standard is intended for the practitioner who is familiar with V&V terminology. It does not present a method for incorporating user expertise or modeler pedigree, nor does it describe the specific V&V activities and rigor that are needed to establish credibility for a particular application and/or device. Instead, this Standard presents a framework for the practitioner to make that assessment using sound engineering judgment. This Standard is not a step-by-step guide, nor is it intended to present a quantitative method for establishing model credibility. While the framework was developed specifically for medical devices, the V&V 40 Subcommittee considers this Standard to be general enough to be applied to other disciplines.

## 2.4 Overview of the Risk-Informed Credibility Assessment Framework

This Standard presents a framework for establishing and assessing model credibility, which is the trust, obtained through the collection of evidence, in the predictive capability of a computational model for a *context of use* (COU). The COU is the specific role and scope of the computational model used to address a question of interest. The framework, referred to as the risk-informed credibility assessment framework, is presented in Figure 2.4-1. The foundational element of the framework is *model risk*, which is the possibility that the computational model leads to an incorrect decision that results in an adverse outcome, such as patient harm or device malfunction. Model risk is a combination of the influence of the computational model relative to other contributing evidence for making a decision, and the consequence for the patient or end users if a decision is incorrect. Model risk is then used to establish the required level of adequacy of the credibility activities for the COU.

The risk-informed credibility assessment framework begins with *identifying a question of interest*, which describes the specific question, decision, or concern that is being addressed. The next step is to *define the COU*, which is a statement that describes the role and scope of the computational model used to inform that decision in relation to other evidence (see section 3). Then, *model risk is assessed* for the COU, which takes into account the role of the computational model to inform the decision and the potential consequence of an incorrect decision (see section 4). Model risk is then used to *establish the goals for each credibility factor*. The credibility factors are elements of the process used to establish the credibility of the computational model for a COU; the factors include verification, validation, and applicability (see section 5). The goals for the credibility factors are used to *plan the activities that establish credibility* (see section 6). Once the activities are defined, the *plan is executed*. After the credibility activities are completed, an assessment is performed to determine if the computational model is *credible for the COU* (see section 7). If sufficient credibility is not achieved, then the risk-informed credibility portion of the framework can be revisited, as indicated by the return arrow in Figure 2.4-1. If sufficient credibility is not achieved, corrective actions may be taken as outlined in section 7. If sufficient credibility is achieved for the COU, then the computational model can be used to inform the decision. Finally, the credibility activities and findings should be

*summarized* (see section 8). To further support the framework, Nonmandatory Appendix A provides an introduction to the Phenomena Identification and Ranking Table, and Nonmandatory Appendix B provides six device-specific examples.

The risk-informed credibility assessment framework may be used throughout the planning, development, and evaluation phases of a project. For instance, a team may use this Standard to assess the risks associated with using a computational model in place of other data sources, or to identify necessary activities and resources before creating a V&V plan. The details of the risk-informed credibility assessment framework are presented in sections 3 through 8.

Illustrations throughout this Standard present key concepts. The illustrations are based on a variety of computational modeling disciplines, which support decision-making for a range of medical devices.

## 3 CONTEXT OF USE

The COU defines the specific role and scope of the computational model used to address the question of interest. It should include a detailed statement of what will be modeled and how the outputs from the computational model will be used to answer or inform the question of interest. It is important to note that the COU is distinct from the "indications for use" or "intended use" of a medical device, which are descriptions of how a device is intended to be used in clinical practice.

A COU for medical device evaluation might involve characterizing or investigating some aspect of technical performance. For example, simulation results from the computational model may facilitate geometry optimization, comparisons to other devices, decisions about bench-testing boundary conditions, or determination of physiologically motivated performance criteria. Alternatively, the simulation results may support patient inclusion criteria for a clinical trial. To establish the scope of the computational model, the COU should include a description of other supporting evidence, such as data from in vitro and/or in vivo studies or other forms of analysis, in its description of the relative contribution of the computational model.

**Illustration 1: Context of Use**

*Medical Device:* A new posterior-stabilized total knee arthroplasty assembly (see Nonmandatory Appendix B, para. B-2.5)

*Question of Interest:* Does the proposed locking mechanism have sufficient strength to prevent liftoff?

*Context of Use:* Finite element analysis (FEA) will be used to determine if the locking mechanism of a new posterior-stabilized total knee arthroplasty assembly has sufficient strength to prevent liftoff, i.e., separation of the tibial component from the metal baseplate, under a variety of loading conditions. Tibial component liftoff is evaluated exclusively using the computational model. All device configurations will be simulated. No predicate device exists to compare with the computed results. No bench testing will be performed for this device.

## 4 MODEL RISK

Model risk is the possibility that the use of the computational model leads to a decision that results in patient harm and/or other undesirable impacts. It reflects the risk the decision maker incurs when using a computational model to support a decision. Model risk is the combination of the influence of the computational model (model influence) and the consequence of an adverse outcome resulting from an incorrect decision (decision consequence).

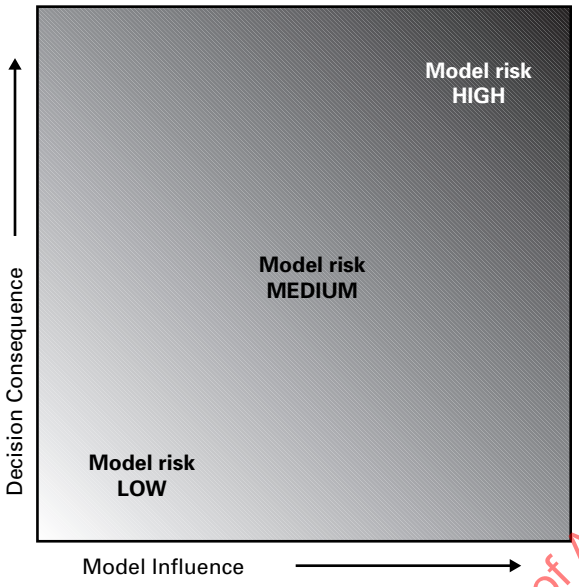
### 4.1 Model Influence

Model influence is the contribution of the computational model relative to other contributing evidence in making a decision.

Model influence can be characterized according to a classification system that may be specific to an organization. The following is an example gradation of model influence from lowest to highest risk:

*(a)* Simulation outputs from the computational model are a minor factor in the decision.
*(b)* Simulation outputs from the computational model are a moderate factor in the decision.
*(c)* Simulation outputs from the computational model are a significant factor in the decision.

**Figure 4.2-1 Schematic of How Model Influence and Decision Consequence Determine Model Risk**




GENERAL NOTE: Darker shades indicate greater model risk.

## 4.2 Decision Consequence

Decision consequence is the significance of an adverse outcome resulting from an incorrect decision.

Consequences are typically considered in the context of potential harm to the patient. However, non-patient-related impacts may also be considered, such as delayed patient access to medical devices, impact on the clinician, financial loss, or increased time to market.

A relevant procedure for assessing and managing the risk of a medical device (e.g., ref. [5]) may be used to identify the severity and probability of occurrence of patient harm from the device, and therefore may inform the decision consequence of the computational model.

Decision consequence can be characterized according to a classification system that may be specific to an organization. The following is an example gradation of decision consequence from lowest to highest risk:

*(a)* An incorrect decision would not adversely affect patient safety or health, but might result in a nuisance to the physician or have other minor impacts.

*(b)* An incorrect decision could result in minor patient injury or the need for physician intervention, or have other moderate impacts.

*(c)* An incorrect decision could result in severe patient injury or death, or have other significant impacts.

Figure 4.2-1 shows the relationship of model risk to model influence (*x*-axis) and decision consequence (*y*-axis). Model influence and decision consequence are independent factors, and an increase in either factor increases model risk. It is incumbent on each organization to develop the relationship of model influence and decision consequence to the overall model risk.

Because the credibility of the computational model should be commensurate with model risk, model risk drives the selection of V&V activities and goals for the credibility factors. The credibility factors are described in section 5.

**Illustration 2: Model Risk**

*Medical Device:* Centrifugal blood pump for circulatory support (see Nonmandatory Appendix B, para. B-2.1)

*Question of Interest:* How is pump-related hemolysis affected by component dimensional tolerances?

*Context of Use:* A computational fluid dynamics (CFD) model will be used to evaluate the sensitivity of pump-induced hemolysis to variations in component dimensions, with the goal of identifying the dimensional tolerances that most likely contribute to increased hemolysis levels. Based on the CFD results, physical pumps with components of varying dimensions will be fabricated and tested.

*Model Influence:* The model influence is MEDIUM because testing will be used to confirm some of the results.

*Decision Consequence:* An incorrect decision to alter the key pump feature's dimensional tolerances could impact hemolysis levels during clinical use. Patient injury could result and require immediate intervention of the clinician to monitor patient hemolysis levels and/or replace the pump. Therefore, the decision consequence is HIGH.

*Model Risk:* The model risk is determined to be MEDIUM-HIGH.

## 5 MODEL CREDIBILITY

Model credibility refers to the trust in the predictive capability of a computational model for the COU. Trust can be established through the collection of evidence from the credibility activities. The process of establishing trust includes performing V&V and then demonstrating the applicability of the V&V evidence to support the use of the computational model for the COU. The collection of V&V evidence includes the following activities: verification studies of the code and calculation, validation studies of the computational model with a comparator, and the associated validation assessment. Each of these activities is evaluated using the credibility factors shown in the right-hand column of Table 5-1. The practitioner can use the credibility factors to determine the rigor needed for each step in the V&V process and to demonstrate applicability.

**Table 5-1 Verification, Validation, and Applicability Activities and Their Associated Credibility Factors**

| Activity (Paragraph) | Credibility Factor (Paragraph) |
|---|---|
| Verification (5.1) | |
| Code (5.1.1) | Software quality assurance (5.1.1.1) |
| | Numerical code verification (5.1.1.2) |
| Calculation (5.1.2) | Discretization error (5.1.2.1) |
| | Numerical solver error (5.1.2.2) |
| | Use error (5.1.2.3) |
| Validation (5.2) | |
| Computational model (5.2.1) | Model form (5.2.1.1) |
| | Model inputs (5.2.1.2) |
| Comparator (5.2.2) | Test samples (5.2.2.1) |
| | Test conditions (5.2.2.2) |
| Assessment (5.2.3) | Equivalency of input parameters (5.2.3.1) |
| | Output comparison (5.2.3.2) |
| Applicability (5.3) | Relevance of the quantities of interest (5.3.1) |
| | Relevance of the validation activities to the COU (5.3.2) |

Associated with each credibility factor is a gradation of activities that describes progressively increasing levels of investigation into each factor. The gradations can be adapted for each COU. The gradations assist with planning and comparison of the activities that can impact model credibility. Example gradations are provided in paras. 5.1 through 5.3 for each credibility factor. Note that some gradations rely on identifying key parameters, which are parameters that meaningfully contribute to the output as appropriate for the COU.

It is incumbent upon the organization performing the V&V activities and applicability assessment to determine goals for each credibility factor such that the overall model credibility is commensurate with the model risk. The rationale for the credibility goals should support the desired confidence in the computational model for the COU. It is recommended that the participants who help establish credibility goals have the appropriate knowledge and experience to assess computational model credibility. A Phenomena Identification and Ranking Table (PIRT) is a tool that can help to identify and provide rationale for setting the goal for each credibility factor (see Nonmandatory Appendix A for more details).

NOTE: It may be valuable for stakeholders to consider how exceeding or missing a specific credibility factor goal would change the overall credibility of the computational model.

Some organizations may want to assign numerical values for each credibility factor gradation. While the numerical values or an overall numerical credibility may support internal decision making, this Standard does not prescribe quantification of the credibility factor gradations. If the credibility of individual factors and/or the entire model are quantified (e.g., through averaging or weighting schemes), then such quantification should not replace the critical thinking needed for a well-informed credibility assessment.

Paragraphs 5.1 through 5.3 describe the credibility factors listed in Table 5-1 in more detail.

## 5.1 Verification

A computational model is the numerical implementation of an underlying mathematical model. The objective of verification is to ensure that the mathematical model is implemented correctly and then accurately solved. Verification is composed of two activities: code verification and calculation verification (ref. [1]).

**5.1.1 Code Verification.** The goals of code verification are to identify and remove errors in the source code and numerical algorithms of the computational software. Documented results from verification studies conducted by the software developer may be referenced to support code verification. However, the verification studies from the software developer may not encompass all aspects of the software used for the COU, and thus additional code verification specific to the COU may be required. Code verification activities include software quality assurance and numerical code verification.

**5.1.1.1 Software Quality Assurance (SQA).** The objective of SQA is to ensure that the software is functioning correctly and produces repeatable results on a specified computer resource in a specified software environment. Types of computational model software include, but are not limited to, off-the-shelf (OTS), modified off-the-shelf (MOTS), and user-developed. SQA is achieved through software validation on OTS and MOTS software and software quality development assurance on MOTS and user-developed software (refs. [3] and [6] through [9]).

For the selected software, it is important to understand unresolved anomalies and their potential effect(s) on the COU, as well as any workarounds, before starting software validation. If user-developed code is used, it is also important to understand the anomaly list for the software development environment, such as compilers and libraries applicable to the computational model.

The following is an example gradation of activities, listed from lowest to highest credibility, that reflects the rigor of SQA:

*(a)* Very little or no SQA procedures were specified or followed.

*(b)* SQA procedures were specified and documented.

*(c)* SQA procedures were specified and documented; 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.

**5.1.1.2 Numerical Code Verification (NCV).** The objective of NCV is to demonstrate correct implementation and functioning of the numerical algorithms (ref. [1]). NCV relies on careful investigation of numerical aspects, such as spatial and temporal convergence rates, spatial convergence in the presence of discontinuities, independence to coordinate transformations, and symmetry tests related to various types of system conditions. NCV is typically conducted by comparing numerical solutions to exact benchmark solutions that are analytical or semi-analytical in nature, as might be generated using the method of manufactured solutions.

The following is an example gradation of activities, listed from lowest to highest credibility, that reflects the rigor of 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.

**5.1.2 Calculation Verification.** The objective of calculation verification is to estimate the numerical error in the quantities of interest (QOIs) due to spatial and temporal discretization of the model (ref. [1]). Calculation verification helps to ensure that the spatial and temporal convergence behavior of the solution of the computational model is analyzed and quantified by refining the discretization parameters and solver convergence tolerances. Additionally, it helps to ensure that practitioner errors are not corrupting the simulation results. Calculation verification involves the estimation of discretization error, numerical solver error, and identification of use error.

**5.1.2.1 Discretization Error.** Discretization error refers to the error associated with solving the computational problem at a finite number of spatial and/or temporal grid points.

The following is an example gradation of activities, listed from lowest to highest credibility, that reflects the rigor of the discretization error analysis:
*(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 and their respective convergence behaviors were observed to be stable, but the discretization error was not estimated.
*(c)* Applicable grid or time-step convergence analyses were performed and discretization error was estimated.

**5.1.2.2 Numerical Solver Error.** Numerical solver error refers to the errors originating from the numerical solution based on the selection of solver parameters [e.g., convergence tolerance(s)].

The following is an example gradation of activities, listed from lowest to highest credibility, that reflects the rigor of the numerical solver error analysis:
*(a)* No solver parameter sensitivity was performed.
*(b)* No solver parameter sensitivity was performed. Solver parameters were established based on values from a previously verified computational model.
*(c)* Problem-specific sensitivity study was performed on solver parameters, confirming that changes in simulation results due to changes in the solver parameters were negligible relative to the model accuracy goal.

**5.1.2.3 Use Error.** Use error refers to errors accrued in the simulation results by the practitioner (e.g., typographical errors).

The following is an example gradation of activities, listed from lowest to highest credibility, that reflects the rigor of the use error investigation:
*(a)* Inputs and outputs were not verified.
*(b)* Key inputs and outputs were verified by the practitioner.
*(c)* Key inputs and outputs were verified by internal peer review.
*(d)* Key inputs and outputs were verified by reproducing simulations as part of an external peer review.

## 5.2 Validation

Validation is the process of assessing the degree to which the computational model is an appropriate representation of the reality of interest. Therefore, validation activities are principally concerned with demonstrating the correctness of the underlying model assumptions and the degree to which sensitivities and uncertainties of the computational model and the associated comparator(s) are understood.

Validation is generally demonstrated by comparing the computational model predictions with the results from the comparator(s), which might be in vitro (e.g., bench testing) and/or in vivo (e.g., clinical trials or animal experiments). Therefore, appropriate validation activities require attention to both the computational model and the comparator, with an appropriately rigorous evaluation of the simulation results. Paragraphs 5.2.1 through 5.2.3 describe aspects of the validation process in more detail.

**5.2.1 Computational Model.** The two credibility factors for the computational model are model form and model inputs, which encompass four components of a computational model: governing equations, system configuration, system properties, and system conditions. The governing equations are the mathematical descriptions of the phenomena being modeled. System configuration could be the geometry of the device, the computational domain, the structure of a physiological control system, or the in vitro test apparatus that is modeled. System properties are the biological, chemical,

and physical properties used in the computational model. System conditions are the constraints that are imposed on the system, such as boundary conditions, loading conditions, and initial conditions.

**5.2.1.1 Model Form.** Model form refers to both the conceptual and mathematical formulation of the computational model (ref. [10]). It includes not only the form of the governing equations but also the form of the system configuration, system properties, and system conditions. Model form is established or selected based on assumptions that will enable the computational model to achieve the desired predictions within the COU. The assumptions that give rise to a model form may be evaluated by preliminary modeling studies to identify the important contributors to model form uncertainty. This might also be accomplished by methods such as scale analysis, sensitivity analysis, and/or by completing a PIRT (see Nonmandatory Appendix A). Any prior knowledge on the success or limitations of the selected model form for the problem types/physics relevant to the COU may be referenced.

The following is an example gradation of activities, listed from lowest to highest credibility, that reflects the extent to which model form assumptions can be evaluated:

*(a)* Influence of model form assumptions was not explored.

*(b)* Influence of expected key model form assumptions was explored.

*(c)* Comprehensive evaluation of model form assumptions was conducted.

**5.2.1.2 Model Inputs.** Model inputs refer to the values for parameters used in the governing equations, system configuration, system properties, and system conditions. The assessment of model input parameters is subdivided into the quantification of sensitivities and quantification of uncertainties.

**5.2.1.2.1 Quantification of Sensitivities.** This component of the credibility factor examines the degree to which the computational model outputs are sensitive to the model inputs.

The following is an example gradation of activities, listed from lowest to highest credibility, that reflects the rigor of the quantification of sensitivities:

*(a)* Sensitivity analysis was not performed.

*(b)* Sensitivity analysis on expected key parameters was performed.

*(c)* Comprehensive sensitivity analysis was performed.

**5.2.1.2.2 Quantification of Uncertainties.** This component of the credibility factor examines the degree to which known or assumed uncertainties in the model inputs are propagated to uncertainties in the simulation results.

The following is an example gradation of activities, listed from lowest to highest credibility, that reflects the rigor of the quantification of uncertainties:

*(a)* Uncertainties were not identified.

*(b)* Uncertainties on expected key inputs were identified and quantified, but were not propagated to quantitatively assess the effect on the simulation results.

*(c)* Uncertainties on all inputs were identified and quantified, and were propagated to quantitatively assess the effect on the simulation results.

**5.2.2 Comparator.** Comparators provide the data against which simulation results are evaluated. Comparators can be in vitro and/or in vivo studies, such as laboratory tests and clinical trials. The comparator might be designed or selected to optimize a balance of resources and relevance to the COU.

The two credibility factors for the comparator are the test samples (e.g., the medical device) and the test conditions (e.g., physiologic loading). These factors are further subdivided into the following four components: quantity, range of characteristics, measurements, and measurement uncertainty. The measurements made to characterize the comparator test samples and test conditions may be used as inputs to the computational model. The measurement data also enable quantification of the uncertainty in the computational model inputs, thereby enabling quantification of the uncertainty in the computational model outputs. The measurement data may also be used to examine the equivalency of the inputs used in the computational model and comparator during the validation assessment. Each component of test samples and test conditions impacts the extent to which the comparator may support model credibility and should be considered separately.

**5.2.2.1 Test Samples**

**5.2.2.1.1 Quantity of Test Samples.** This component of the credibility factor examines the number of samples used in the comparator study. Increased credibility is generally achieved with a larger number of samples.

The following is an example gradation of activities, listed from lowest to highest credibility, that reflects the rigor of the quantity of samples used in the comparator study:

*(a)* A single sample was used.

*(b)* Multiple samples were used, but not enough to be statistically relevant.

*(c)* A statistically relevant number of samples were used.

**5.2.2.1.2 Range of Characteristics of Test Samples.** This component of the credibility factor examines the range of each test sample characteristic of interest included in the comparator study. For example, if the length of the test sample is a characteristic of interest, this factor addresses the range of the lengths studied. Increased credibility is generally achieved with a broader range of test sample characteristics studied.

The following is an example gradation of activities, listed from lowest to highest credibility, that reflects the rigor of the range of test sample characteristics for the comparator study:

*(a)* One or more samples with a single set of characteristics were included.

*(b)* Samples representing a range of characteristics near nominal were included.

*(c)* Samples representing the expected extreme values of the parameters were included.

*(d)* Samples representing the entire range of parameters were included.

**5.2.2.1.3 Measurements of Test Samples.** This component of the credibility factor examines the rigor with which the measurement data characterize each test sample. This component includes characterizations for comparator inputs (e.g., test sample dimensions, material properties) as well as characterization of comparator outputs (e.g., test sample yield strength).

The following is an example gradation of activities, listed from lowest to highest credibility, that reflects the rigor of the test sample measurement characterization:

*(a)* Test samples were not measured and/or characterized.

*(b)* One or more key characteristics of the test samples were measured.

*(c)* All key characteristics of the test samples were measured.

**5.2.2.1.4 Uncertainty of Test Sample Measurements.** This component of the credibility factor examines the analysis of the uncertainty associated with the tools and methods used to obtain the measurements characterizing the samples.

The following is an example gradation of activities, listed from lowest to highest credibility, that reflects the rigor of the analysis of the measurement uncertainty:

*(a)* Samples were not characterized or were characterized with gross observations, and 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.

### 5.2.2.2 Test Conditions

**5.2.2.2.1 Quantity of Test Conditions.** For a given test method, this component of the credibility factor examines the number of test conditions imposed and characterized in the comparator study. For example, the method could specify measuring the test sample strength at multiple strain rates under tensile loading at multiple temperatures. Increased credibility is generally achieved with a larger number of test conditions.

The following is an example gradation of activities, listed from lowest to highest credibility, that reflects the rigor of the number of test conditions used in the comparator study:

*(a)* A single test condition was examined.

*(b)* Multiple (two to four) test conditions were examined.

*(c)* More than four test conditions were examined.

**5.2.2.2.2 Range of Test Conditions.** For a given test method, this component of the credibility factor examines the range of test conditions included in the comparator study. For example, if the test condition is temperature, this factor addresses the range of temperatures studied. Increased credibility is generally achieved by examining a broader range of test conditions.

The following is an example gradation of activities, listed from lowest to highest credibility, that reflects the rigor of the range of the test conditions for the comparator study:

*(a)* A single test condition was examined.

*(b)* Test conditions representing a range of conditions near nominal were examined.

*(c)* Test conditions representing the expected extreme conditions were examined.

*(d)* Test conditions representing the entire range of conditions were examined.

**5.2.2.2.3 Measurements of Test Conditions.** This component of the credibility factor examines the rigor with which the measurement data characterize the test conditions.

The following is an example gradation of activities, listed from lowest to highest credibility, that reflects the rigor of the test condition measurements:

*(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.

**5.2.2.2.4 Uncertainty of Test Condition Measurements.** This component of the credibility factor examines the analysis of the uncertainty associated with the tools and methods used to obtain the measurements characterizing the test conditions.

The following is an example gradation of activities, listed from lowest to highest credibility, that reflects the rigor of the analysis of the measurement uncertainty:

*(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.

**5.2.3 Assessment.** An assessment of the accuracy of the simulation output can be performed after the outputs from the V&V activities are obtained and compared. The credibility factors associated with this assessment, as shown in Table 5-1, are the equivalency of the input parameters and the rigor of the output comparison.

**5.2.3.1 Equivalency of Input Parameters.** Equivalency between the type and range of the input parameters of the computational model and those of the comparator leads to increased credibility.

The following is an example gradation of activities, listed from lowest to highest credibility, that reflects the equivalency of the input parameters:

*(a)* The types of some inputs were dissimilar.

*(b)* The types of all inputs were similar, but the ranges were not equivalent.

*(c)* The types and ranges of all inputs were equivalent.

**5.2.3.2 Output Comparison.** Equivalency between the types of output from the computational model and those from the comparator leads to increased credibility. Increased quantification and incorporation of uncertainties in the output also lead to increased credibility. Credibility relies on both experimental uncertainty and computational uncertainty, and an acceptable comparison error.

Paragraphs 5.2.3.2.1 through 5.2.3.2.4 provide example gradations of activities, listed from lowest to highest credibility, that reflect the rigor of the output comparison.

**5.2.3.2.1 Quantity**

*(a)* A single output was compared.

*(b)* Multiple outputs were compared.

**5.2.3.2.2 Equivalency of Output Parameters.** This component refers to the type of output, not the values of the output.

*(a)* Types of outputs were dissimilar.

*(b)* Types of outputs were similar.

*(c)* Types of outputs were equivalent.

**5.2.3.2.3 Rigor of Output Comparison.** This component refers to the method used to compare the QOIs from the computational model to those from the comparator.

*(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.

**5.2.3.2.4 Agreement of Output Comparison.** This component refers to the qualitative or quantitative agreement between the QOIs from the computational model and those from the comparator.

*(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.

NOTE: A satisfactory level of agreement may be assessed based on criteria established for the COU by the practitioner.

**Illustration 3: Rigor of Output Comparison and Agreement of Output Comparison**

*Medical Device:* Centrifugal blood pump for circulatory support (see Nonmandatory Appendix B, para. B-2.1)

The proposed gradations for the rigor of output comparison and agreement of output comparison are combined into a single gradation, from lowest to highest rigor, as follows:

| Level | Description |
|---|---|
| 1 | Visual comparison concludes good agreement. |
| 2 | Comparison by measuring the difference between computational results and experimental data. Differences are less than 20%. |
| 3 | Comparison by measuring the difference between computational results and experimental data. Differences are less than 10%. |
| 4 | Comparison with uncertainty estimated and incorporated from the comparator or computational model. Differences between computational results and experimental data are less than 5%. Includes consideration of some uncertainty, but statistical distributions for uncertainty quantification are unknown. |
| 5 | Comparison with uncertainties estimated and incorporated from both the comparator and the computational model, including comparison error. Differences between computational results and experimental data are less than 5%. Statistical distributions for uncertainty quantifications are known. |

Based on a MEDIUM-HIGH model risk for the blood pump, as shown in Illustration 2, the validation activities should demonstrate the model accuracy is within 5% and must include a consideration of uncertainty, corresponding to Level 4.



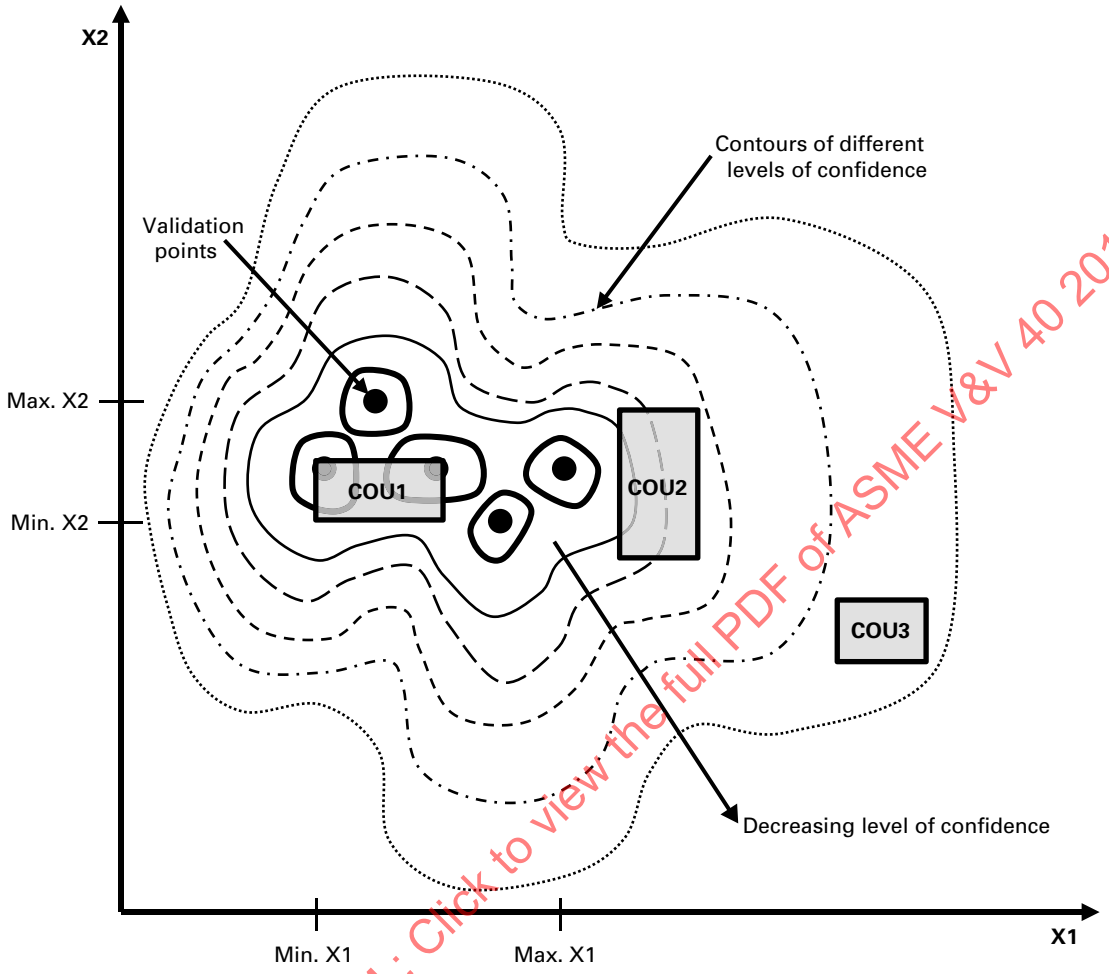
## 5.3 Applicability of the Validation Activities to the COU

Applicability is the relevance of the validation activities to support the use of the computational model for a COU. The applicability of the validation activities is governed by two factors: the relevance of the QOIs used in the validation activities to the QOIs of the COU, and the relevance of the validation conditions relative to those of the COU.

The measured QOIs of the validation activities are not always identical to the QOIs for the COU because the QOIs for the COU are not always directly measurable, might not be measured without unduly perturbing the intended test conditions, and/or might not be obtained within acceptable ranges of uncertainty and error. Therefore, the measured QOIs of the validation activities may be surrogates for the QOIs for the COU, with varying degrees of applicability.

Applicability of the validation activities and the inferred confidence are illustrated in Figure 5.3-1 for a two-parameter (X1, X2) computational model. In this example, validation was performed using the X1–X2 values at the five points labeled "validation points." The greatest level of applicability occurs where the COU overlaps one or more validation points (see COU1 in Figure 5.3-1). However, the opportunities to fully replicate the COU conditions can be limited for medical devices (see COU2 and COU3 in Figure 5.3-1). If the COU is not completely bounded by the conditions used in the validation, the confidence in the predictive capability of the computational model beyond the validation points may only be inferred. As the COU conditions extend a greater distance beyond the conditions used in the validation, there will be less confidence in the predictive capability of the model. And while the practitioner may have more confidence in a prediction when the COU conditions are in close proximity to the validation points, the close proximity does not mean the prediction is credible. The

**Figure 5.3-1 Illustrative Examples of Three COUs Relative to the Validation Points for a Two-Parameter (X1, X2) Computational Model**




GENERAL NOTE: The parameters of the computational model could represent loading conditions, component sizes, etc. Min. and max. X1 and X2 represent the range of the parameter values in the validation activities. The greatest level of model confidence occurs at the validation points, and the inferred confidence decreases away from the validation points. Note that the quantification of confidence contours is extremely involved and is rarely performed in practice.



agreement of the output comparison at the validation points addresses the adequacy of the validation activities related to the COU.

The credibility factors associated with establishing applicability of the computational model to the COU are the relevance of the validation activities to the COU and the relevance of the validation QOIs.

**5.3.1 Relevance of the QOIs.** This factor compares the QOIs from the validation activities to the QOIs for the COU.

The following is an example gradation of activities, listed from lowest to highest credibility, that reflects the relevance of the QOIs for the COU:

*(a)* The QOIs from the validation activities were related, though not identical, to those for the COU.

*(b)* A subset of the QOIs from the validation activities were identical to those for the COU.

*(c)* The QOIs from the validation activities were identical to those for the COU.

**5.3.2 Relevance of the Validation Activities to the COU.** This factor summarizes the relative proximity of the COU to the validation points.

The following is an example gradation of activities, listed from lowest to highest credibility, that reflects the relevance of the validation activities to the COU:

*(a)* There was no overlap between the ranges of the validation points and the COU (COU3 in Figure 5.3-1).

*(b)* There was partial overlap between the ranges of the validation points and the COU (COU2 in Figure 5.3-1).

*(c)* The COU encompassed some of the validation points (COU1 in Figure 5.3-1).

*(d)* The COU encompassed all validation points (not shown in Figure 5.3-1), and the validation points spanned the entire COU space.

**Illustration 4: Relevance of the Validation Activities to the COU**

*Medical Device*: Plate-and-screw system for fracture fixation (see Nonmandatory Appendix B, para. B-2.4)

*Question of Interest:* What is the maximum temperature in the tissue near a plate-and-screw system due to the presence of a spinal fixation device during a magnetic resonance imaging (MRI) scan?

*Context of Use:* The COU of the computational model is to evaluate multiple configurations of the proposed plate-and-screw system in order to identify the worst-case configuration, based on the predicted temperature increase in surrounding tissue. The resulting worst-case configuration will then be physically tested to quantify the temperature increase.

*Relevance of the Validation Activities to the COU:* Validation for the QOI of temperature is not possible in a clinical setting for this device. Therefore, a phantom is used to validate the temperature-rise prediction in the vicinity of an implanted device during an MRI exam per an industry standard. However, the phantom is not directly applicable to the COU for predicting maximum temperature in the surrounding tissue in humans because it uses simplified geometries and materials that are not representative of the patient.

## 6 THE PLAN

The purpose of the plan is to define the appropriate activities and acceptable results for each credibility factor that establishes model credibility commensurate with the model risk. It is incumbent upon the organization performing the V&V and applicability assessment to define appropriate activities to meet each credibility goal, along with the criteria that demonstrate that each goal has been met. This will likely rely on the relationship between model influence and decision consequence to the overall model risk, and the translation of that risk into the credibility goals. The plan does not necessarily define protocols for executing the activities. Development of a plan facilitates communication among the stakeholders. The stakeholders may review the plan such that, upon completion, the overall credibility will be sufficient to use the computational model for the COU and the associated model risk.

A plan for assessing the credibility of the computational model may contain the following information:

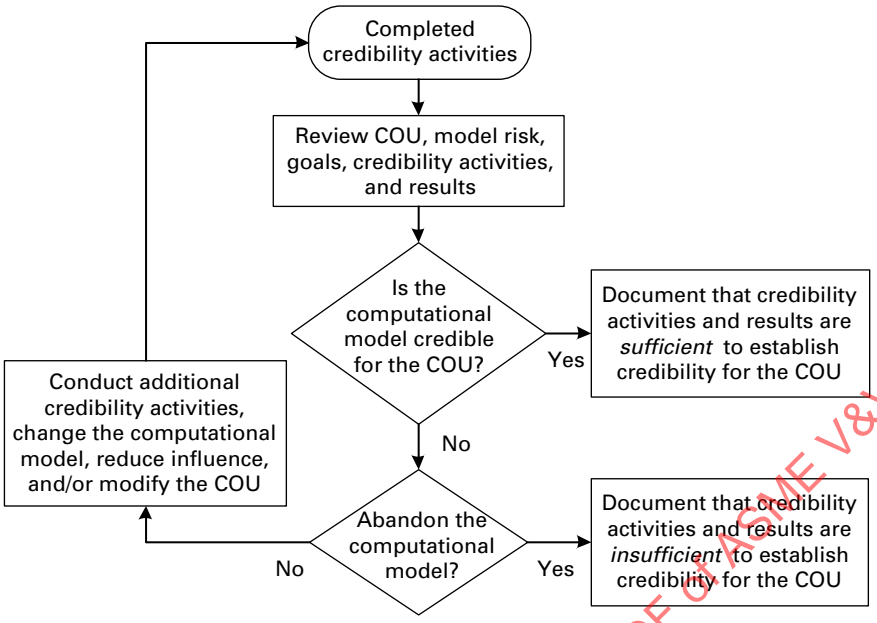
*(a)* purpose of the credibility activities
*(b)* description of the computational model
*(c)* COU of the computational model
*(d)* model risk assessment
*(e)* credibility factor goals
*(f)* activities and rationale for each credibility factor

Once the plan has been executed, the credibility of the computational model can be assessed according to the methodology described in section 7.

## 7 CREDIBILITY ASSESSMENT

The credibility of the computational model for the COU is determined through a review of the V&V results (i.e., completed activities and outcomes), with consideration given to the COU, model risk, established credibility factor goals, and any additional knowledge gained during the V&V process. A computational model credibility assessment flowchart is provided in Figure 7-1. A review of the rationale indicating that the completed activities are sufficient to establish model credibility commensurate with the model risk is integral to the credibility assessment. This should include the individual assessment of each credibility factor. It is recommended that the reviewing participants have the appropriate knowledge and experience to assess computational model credibility. It is possible that upon completion of the credibility activities, the credibility goals might not have been met as initially planned; however, the computational model may still be sufficiently credible for decision-making based on the rationale developed. It is therefore suggested that an organization's internal review process be used to facilitate this assessment.

**Figure 7-1 Example Workflow for Assessing Computational Model Credibility**




Once the review of the COU, model risk, credibility goals, and V&V outcome is completed, and if the activities are deemed adequate, then document that the credibility activities and the V&V outcome are sufficient to establish the credibility for the COU (see section 8). Simulation practices (e.g., mesh convergence, quantification of model input uncertainties) established during the credibility activities should then be followed when the model is applied to the COU. Additionally, the model use should not deviate from the COU that motivated the credibility activities without further consideration.

If the credibility activities are deemed inadequate to justify model credibility for the COU, then the computational model may be abandoned. Given the available information and resources, an adequate computational model might not be possible, and data from physical tests or clinical studies might be better alternatives to support decision-making. Alternatively, the following steps may be pursued to attain a computational model that is credible for the COU:

*(a) Conduct Additional Credibility Activities.* Completion of additional credibility activities can improve the assessed credibility of the computational model. Such activities may include adding another comparator, or improving test conditions control, sample characterization, or data uncertainty.

*(b) Change the Computational Model.* Changing the computational model might involve modifying the code, solution formulation, system configuration, system properties, boundary conditions, and/or governing equations. Building modifications of the computational model into the V&V plan can alleviate excessive revisions of the plan.

*(c) Reduce the Influence of the Computational Model.* The influence of the computational model, and thus model risk, may be reduced by performing other activities (e.g., physical tests, clinical studies) that provide additional evidence. Reducing the influence of the computational model reduces the credibility needed for the COU because it lowers the model risk.

*(d) Modify the COU.* The COU could be modified to lower model risk and thereby reduce the credibility of the model needed for the COU.

**Illustration 5: Computational Model Is *Not* Credible for COU**

*Medical Device:* A new posterior-stabilized total knee arthroplasty assembly (see Nonmandatory Appendix B, para. B-2.5)

*Computational Model Is Not Credible for COU:* Validation activities were completed and the computational model was determined not credible for the initial COU (COU1 in the matrix below). COU1 has the highest model risk because the influence of the model is the greatest for determining liftoff. For the other COUs, the model influence is lower (and thus model risk is also lower) because additional supporting data are available; COU4 has the lowest risk and most available relevant data to support credibility.

**Matrix of Proposed COUs**

| Benchtop Testing | Existence of Predicate Device | |
|---|---|---|
| | No | Yes |
| None | COU1 | COU3 |
| Worst case | COU2 | COU4 |

In this study, because the model was not credible for COU1, the COU was modified to COU3 to reduce the influence of the computational model by incorporating predicate device data.

## 8 DOCUMENTATION AND EVIDENCE

Substantiating the computational model as appropriate for the COU requires documentation of the activities performed for verification, validation, and applicability. The documentation should describe the computational model and the decision being informed by the computational model, and the relevant aspects of the verification, validation, and applicability assessment activities, and should include the evidence that establishes the credibility of the computational model for the COU. The following is guidance for documenting the credibility activities and evidence supporting the credibility of the computational model:

*(a) Background.* Information that describes the device, process, or system feature(s) being modeled. This may include information about the basic operation of the device, process, or system. It may also include a description of the clinical application as it relates to the COU.

*(b) COU of the Computational Model.* A description of the COU for the computational model that includes information regarding the decision that is being informed by the computational model results, as well as a description of any other sources of supporting evidence that are informing the decision.

*(c) Computational Model Details.* Documentation describing the relevant details of the computational model for the COU.[1]

*(d) Model Risk.* Documentation of the overall model risk, including an evaluation of the computational model influence and decision consequence, and an overall statement regarding model risk determination.

*(e) Credibility Activities, Results, and Computational Model Credibility Assessment.* Documentation of the credibility assessment activities, including a description of the goals for the credibility factors, the activities conducted, and the evidence supporting the credibility of the computational model. This includes the relevant details of the computational model and each comparator (in vitro and/or in vivo studies) used for the V&V activities. It is also recommended that documentation for any comparator used during the V&V activities follow established best practices for describing tests or studies within the appropriate technical field.

*(f) Conclusions.* A summary of the overall credibility of the computational model for the COU as evidenced by the credibility activities.

[1] Guidance is available regarding the reporting of computational models used in U.S. FDA regulatory submissions for medical devices (ref. [11]).

# MANDATORY APPENDIX I
# REFERENCES

[1] ASME V&V 10-2006 (R2016), Guide for Verification & Validation in Computational Solid Mechanics, The American Society of Mechanical Engineers (ASME), New York
[2] ASME V&V 20-2009 (R2016), Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer, The American Society of Mechanical Engineers (ASME), New York
[3] SAND2007-5948, Predictive Capability Maturity Model for Computational Modeling and Simulation, Sandia National Laboratories, Albuquerque, NM
[4] NASA-STD-7009 (2008), Standard for Models and Simulation, National Aeronautics and Space Administration (NASA), Washington, DC
[5] ISO 14971:2007, Medical devices — Application of risk management to medical devices, International Organization for Standardization (ISO), Geneva, Switzerland
[6] SAND2003-3769, Verification, Validation, and Predictive Capability in Computational Engineering and Physics, Sandia National Laboratories, Albuquerque, NM
[7] AAMI TIR36:2007, Validation of software for regulated processes, Association for the Advancement of Medical Instrumentation (AAMI), Arlington, VA
[8] ANSI/AAMI/IEC 62304:2006, Medical device software — Software life cycle processes, Association for the Advancement of Medical Instrumentation (AAMI), Arlington, VA
[9] General Principles of Software Validation: Final Guidance for Industry and FDA Staff, January 11, 2002, U.S. Food and Drug Administration (FDA), Silver Spring, MD
[10] Oberkampf, W. L., and Roy, C. J. (2006), Verification and Validation in Scientific Computing, Cambridge University Press, New York
[11] Reporting of Computational Modeling Studies in Medical Device Submissions; Final Guidance for Industry and Food and Drug Administration Staff, September 21, 2016, U.S. Food and Drug Administration (FDA), Silver Spring, MD
[12] World Health Organization (2003), Medical Device Regulations, Global Overview and Guiding Principles, World Health Organization (WHO), Geneva, Switzerland

# MANDATORY APPENDIX II
# GLOSSARY

*anomaly:* anything observed in the documentation or operation of software that deviates from expectations based on previously verified software products or reference documents. Examples include bugs, defects, errors, exceptions, and faults.

*applicability:* the relevance of the validation activities to support the use of the computational model for a context of use.

*calculation verification:* the process of determining the solution accuracy of a calculation. Also called solution verification.

*code verification:* the process of identifying errors in the numerical algorithms of a computer code.

*comparator:* the test data that are used for validation, which may be data from in vitro or in vivo studies. The selection of the comparator should be based on the context of use.

*computational model:* the numerical implementation of the mathematical model performed by means of a computer.

*context of use (COU):* a statement that defines the specific role and scope of the computational model used to address the question of interest.

*credibility:* the trust, established through the collection of evidence, in the predictive capability of a computational model for a context of use.

*decision consequence:* the significance of an adverse outcome resulting from an incorrect decision.

*determination:* the process of establishing something exactly, typically by calculation or research.

*effectiveness:* efficacy in the real-world environment. A device is clinically effective when it produces the effect intended by the manufacturer relative to the medical condition(s) (ref. [12]).

*goal:* an aim or desired outcome.

*governing equation:* the mathematical relationship that describes the phenomenon of interest.

*key parameters:* the parameters that meaningfully contribute to the output as appropriate for the context of use.

*key test conditions:* the test conditions that meaningfully contribute to the output as appropriate for the context of use.

*mathematical model:* the mathematical equations, boundary conditions, initial conditions, and modeling data needed to describe the conceptual model.

*medical device:* an instrument, apparatus, implement, machine, contrivance, implant, in vitro reagent, or other similar or related article, including a component part, or accessory that is

*(a)* recognized in the official National Formulary, or the U.S. Pharmacopoeia, or any supplement to them

*(b)* intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in man or other animals, or

*(c)* intended to affect the structure or any function of the body of man or other animals, and which does not achieve its primary intended purposes through chemical action within or on the body of man or other animals and which is not dependent upon being metabolized for the achievement of any of its primary intended purposes

*model:* a mathematical, physical, or logical description or representation of a system, entity, phenomenon, or process. Any data that go into a model are considered part of the model.

*model influence:* the contribution of the computational model relative to other contributing evidence in making a decision.

*model risk:* the possibility that the computational model and the simulation results may lead to an incorrect decision that would lead to an adverse outcome.

*off-the-shelf (OTS) software:* a ready-made software that is available to the general public through commercial license or open source agreement.

*output:* the quantities of interest generated by a simulation and/or comparator.

*quantity of interest (QOI):* the calculated or measured result from a computational model or comparator, respectively.

*question of interest:* the specific question, decision, or concern that is being addressed.

*rigor:* as related to the context of use, the quality of being extremely thorough, exhaustive, or accurate.

*simulation:* the imitation of the characteristics of a system, entity, phenomenon, or process using a computational model; a specific "run" of the computational model with one set of parameters that results in the quantity of interest or multiple quantities of interest.

*technical performance:* the performance considerations of a medical device that include technical functions in addition to (clinical) effectiveness (ref. [12]).

*uncertainty:* a potential deficiency in any phase or activity of the modeling, computation, or experimentation process that is due to inherent variability or lack of knowledge (ref. [1]).

*validation:* the process of determining the degree to which a model or a simulation is an accurate representation of the real world.

*verification:* the process of determining that a computational model accurately represents the underlying mathematical model and its solution from the perspective of the intended uses of modeling and simulation (ref. [2]). See also *calculation verification* and *code verification*.

# NONMANDATORY APPENDIX A
# PHENOMENA IDENTIFICATION AND RANKING TABLE

## A-1 INTRODUCTION TO PIRT

With respect to computational modeling, the Phenomena Identification and Ranking Table (PIRT) provides a systematic approach to compiling the phenomena associated with the QOIs being modeled, and then ranking them in the order of importance required to satisfy the COU.

The phenomena are aspects of the system that might influence the QOIs; the ranking can be thought of as a qualitative sensitivity analysis. At this stage, sound engineering judgment, rather than a formal sensitivity analysis, is used to rank the important aspects. The PIRT exercise can help identify key processes, and then the ranking and associated rationale can help inform the selection of the goals for the credibility factors in parallel with the development of the V&V plan. Other important aspects that complement the identification and ranking of phenomena are determining the amount of knowledge about those phenomena, and then determining how much confidence one has in capturing those phenomena. The latter is typically a function of the former. These concepts are presented in Table A-1-1.

## A-2 COMPILING THE PIRT

### A-2.1 Classification of Phenomena

Phenomena can be classified according to their importance once they have been identified and presented in a table. Classification helps to determine which aspects need further investigation or research (e.g., phenomena with high importance but with low knowledge and/or confidence).

The following is an example gradation, listed from highest to lowest credibility, that reflects the importance of the phenomena to the COU:

*(a)* High (H) implies that the phenomenon, model, or parameter has a controlling impact on the COU. Simulation of experiments and/or analytic modeling with a high degree of accuracy is critical.

*(b)* Medium (M) implies that the phenomenon has a moderate impact on the COU and only a moderate degree of accuracy is required for analytic modeling or measurements.

*(c)* Low (L) implies that the phenomenon has a minimal or zero impact on the COU.

### A-2.2 Knowledge/Confidence Levels

Knowledge/confidence level summarizes the user's understanding of how appropriately each phenomenon, model, or parameter is calculated or used in determining the COU. The following is an example gradation, listed from highest to lowest credibility, that reflects the user's knowledge/confidence level regarding each phenomenon:

*(a)* Known (K) implies fully or almost fully known (e.g., more than 75% of the knowledge base is established).

*(b)* Partially known (P) implies the knowledge base is moderate (e.g., 25% to 75% of the knowledge base is established).

*(c)* Unknown (U) implies that the knowledge base is low (e.g., less than 25% of the knowledge base is established).

These concepts are summarized in Table A-2.2-1.

This process can then be taken one step further to determine the mitigation of uncertainty or low knowledge/confidence, as shown in Table A-2.2-2.

**Table A-1-1 A Sample PIRT**

| Phenomenon | Description | Importance | Knowledge/Confidence |
|---|---|---|---|
| | | | |
| | | | |

**Table A-2.2-1 An Example Gradation of Knowledge/Confidence Level and Importance**

| Knowledge Level | Importance of Phenomenon to COU [Note (1)] | | |
|---|---|---|---|
| | H | M | L |
| K | | | |
| P | Needs further research | | |
| U | Needs further research | Needs further research | |

NOTE: (1) Empty cells indicate opportunities to describe the importance of the phenomenon.

**Table A-2.2-2 A Sample PIRT Including a Mitigation Column**

| Phenomenon | Description | Importance | Knowledge/Confidence | Mitigation of Uncertainty |
|---|---|---|---|---|
| | | | | |
| | | | | |

In summary, the PIRT can be used to assess and understand key processes in the computational model and their importance in the outcome, and to document rationale and mitigation strategies for uncertainty and low confidence. The PIRT also enables open communication between the stakeholders and can guide resource allocation.

## A-3 REFERENCES

Diamond, D. J. (2006), "Experience Using the Phenomena Identification and Ranking Technique for Nuclear Analysis," BNL-76750-2006-CP, Brookhaven National Laboratory, Upton, NY

# NONMANDATORY APPENDIX B
# EXAMPLES OF RISK-INFORMED CREDIBILITY ASSESSMENT CONCEPTS

## B-1 INTRODUCTION

This Standard presents a framework for establishing the credibility goals for a computational model related to medical devices based on the risk associated with the COU. This Appendix illustrates how establishing risk-informed credibility (see sections 3 through 5) may be put into practice, with the following objectives:

*(a)* Provide examples based on a variety of medical device types and involving a range of governing physics.

*(b)* Present examples that demonstrate model risk and credibility approaches that are consistent with this framework.

*(c)* Illustrate how the gradations for each credibility factor can be adapted for different applications.

The examples in this Appendix are intended to illustrate selected elements of the risk-informed credibility assessment framework, as indicated in Figure B-1-1.

All examples in this Appendix provide a question of interest and a COU as the risk-informed credibility assessment framework is anchored by these concepts. Each example then addresses a subset of credibility factors within the context of that example. Table B-1-1 provides the credibility factors that are addressed in each example. Complete assessment of the credibility of a computational model should address all factors.

Each example is framed around a specific device type and governing physics, as shown in Table B-1-2. However, it is expected that key attributes of each example may be illustrative for other device types or physics beyond those that are in a specific example.

## B-2 EXAMPLES

The following considerations apply to the examples in this Appendix:

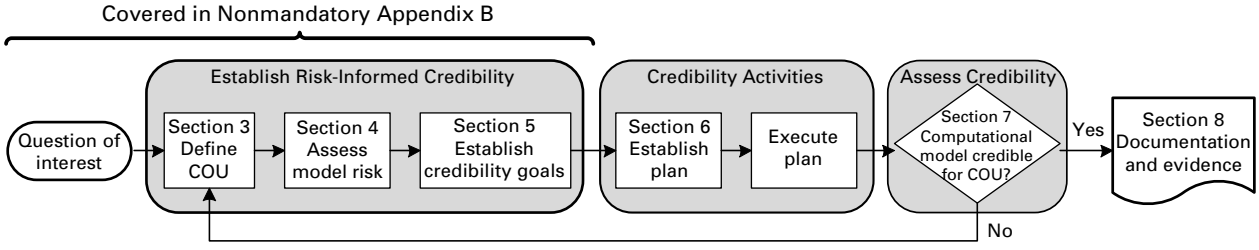
*(a)* Each example highlights specific aspects of the risk-informed credibility assessment framework without providing an end-to-end illustration of the risk-informed credibility process. In practice, a complete assessment of the credibility of a computational model application should address all credibility factors.

*(b)* No example is intended to be so prescriptive that it can be taken in its current form. Rather, each example is intended to illustrate the philosophy and practice of the risk-informed credibility assessment framework.

*(c)* Each example is an illustration that may assist in applying the risk-informed credibility assessment framework, but should not be considered "industry approved" or "regulatory approved."



**Figure B-1-1 Elements of the ASME V&V 40 Risk-Informed Credibility Assessment Framework Illustrated in Nonmandatory Appendix B**



## Table B-1-1 Mapping of Examples to Selected Credibility Factors

| Activity (Paragraph) | Credibility Factor (Paragraph) | Example [Note (1)] 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| **Model Risk (4)** | Model influence (4.1) | X | X | X | X | X | X |
| | Decision consequence (4.2) | X | X | X | ... | ... | X |
| | Model risk assessment | X | X | X | X | X | X |
| **Model Credibility (5)** | | | | | | | |
| Verification (5.1) | | | | | | | |
| Code (5.1.1) | Software quality assurance (5.1.1.1) | ... | ... | X | ... | ... | ... |
| | Numerical code verification (5.1.1.2) | X | ... | X | ... | X | ... |
| Calculation (5.1.2) | Discretization error (5.1.2.1) | X | X | ... | ... | ... | ... |
| | Numerical solver error (5.1.2.2) | X | ... | ... | ... | ... | ... |
| | Use error (5.1.2.3) | ... | ... | X | ... | ... | ... |
| Validation (5.2) | | | | | | | |
| Computational model (5.2.1) | Model form (5.2.1.1) | X | ... | ... | ... | X | ... |
| | Model inputs (5.2.1.2) | ... | X | X | X | X | X |
| Comparator (5.2.2) | Test samples (5.2.2.1) | X | X | ... | ... | X | ... |
| | Test conditions (5.2.2.2) | X | X | X | ... | ... | X |
| Assessment (5.2.3) | Equivalency of input parameters (5.2.3.1) | ... | X | ... | ... | X | X |
| | Output comparison (5.2.3.2) | X | X | X | X | ... | ... |
| **Applicability (5.3)** | Relevance of the quantities of interest (5.3.1) | ... | X | ... | ... | ... | X |
| | Relevance of the validation activities to the COU (5.3.2) | X | X | ... | X | X | ... |

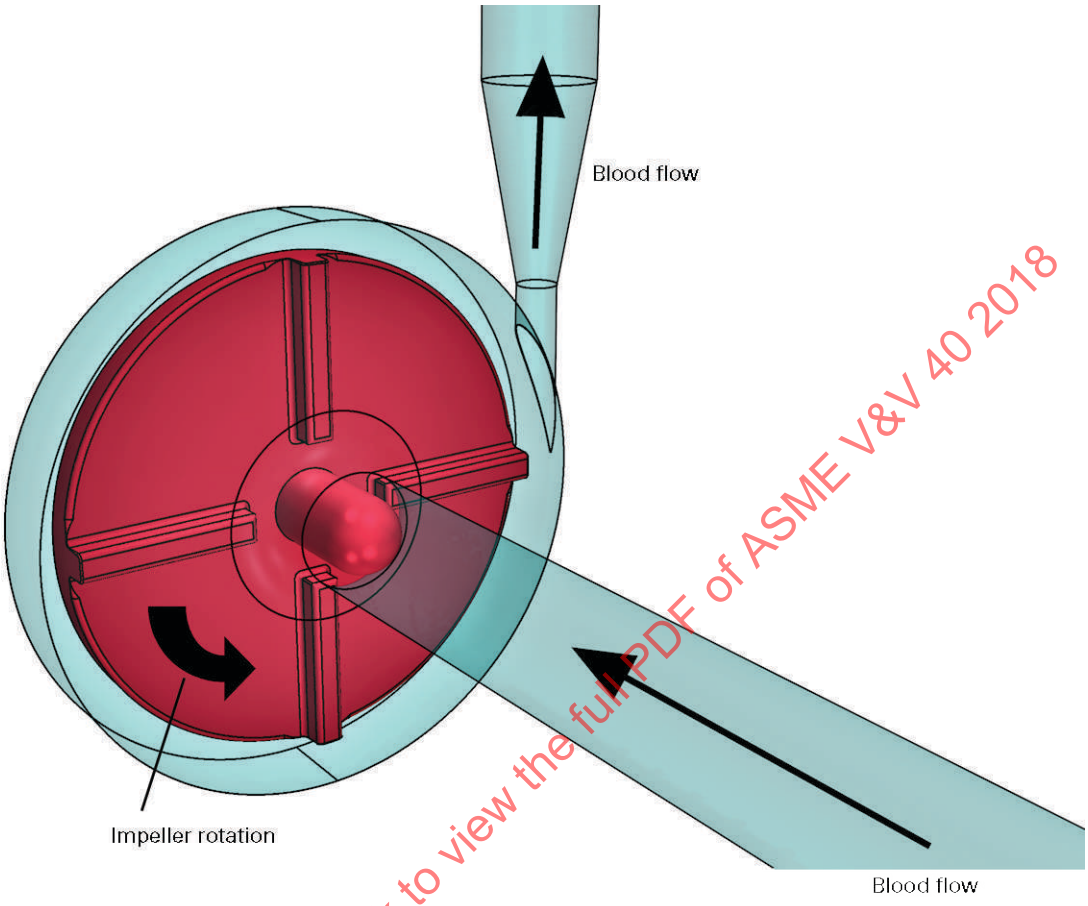
NOTE: (1) See paras. B-2.1 through B-2.6 for the examples.

## Table B-1-2 Mapping of Examples to Device Type and Modeling Approach

| Example [Note (1)] | Device Type | Governing Physics | Of Special Interest |
|---|---|---|---|
| 1 | Centrifugal blood pump | Fluid mechanics | Risk assessment |
| 2 | Aneurysm flow diverter | Fluid mechanics | In vitro test data and preclinical evidence |
| 3 | Hospital bed | Rigid body mechanics | Single computational model supports multiple COUs |
| 4 | Implanted plate/screw system | Electromagnetics | Different comparators |
| 5 | Total knee arthroplasty system | Solid mechanics | Family of designs |
| 6 | Interbody fusion device | Solid mechanics | Comparator testing per industry standard |

NOTE: (1) See paras. B-2.1 through B-2.6 for the examples.

**Figure B-2.1.1-1 Illustration of a Centrifugal Blood Pump Design**




GENERAL NOTE: Adapted from Benchmark 2: Blood Pump in "Computational Fluid Dynamics Round Robin Study" by P. Hariharan, U.S. Food and Drug Administration.



## B-2.1 Example 1: Assessing Hemolysis in Centrifugal Blood Pumps

This example focuses on a single COU, related to assessing the blood damage that can result from the use of a centrifugal blood pump. A risk assessment and specific activities intended to establish the credibility of the computational model are reviewed. Particular focus is given to the validation comparator, providing an example of how carefully conducted in vitro testing can support the use of the computational model for the COU. Additionally, the necessary model accuracy, as demonstrated through the validation activities, is dependent on the model risk.

**B-2.1.1 Background.** Centrifugal blood pumps (see Figure B-2.1.1-1) are often used to maintain a patient's blood flow during cardiopulmonary bypass surgery. In this example, a pump design is in the final stages of product testing. Production-quality pumps were manufactured using production components and assembly procedures. During the final stages of in vitro testing, while the production pumps were being tested, elevated plasma-free hemoglobin levels were observed, indicating increased hemolysis — damage to red blood cells — associated with the pump function. The hemolysis levels were higher than those measured in earlier hemolysis testing using pumps fabricated from prototype components. The production pump hemolysis levels were also higher than those for the predicate device that was included in the study. Potential reasons for the higher-than-expected hemolysis levels were identified, with the most likely reason determined to be component dimensional tolerances.

**B-2.1.2 Question of Interest.** How is pump-related hemolysis impacted by component dimensional tolerances?

**B-2.1.3 Context of Use.** A computational fluid dynamics (CFD) model is used to evaluate the sensitivity of pump-induced hemolysis to variations in component dimensions, with the goal of identifying the current dimensional tolerances that are most likely contributing to the increased hemolysis levels. Based upon the CFD results, physical pumps with

different dimensional configurations will be fabricated using components of varying dimensions. For comparison purposes, the component dimensions for each pump tested will be measured to ensure that the actual physical dimensions match those used in the computational model.

For the computational model and the in vitro testing, hemolysis will be quantified using the modified index of hemolysis (MIH). The MIH is a measure of the quantity of hemoglobin released into plasma as blood is pumped through the test circuit. As part of the in vitro testing, MIH will be determined from plasma-free hemoglobin measurements. In the model, MIH will be calculated using an empirical function of blood shear stress and shear exposure time.

Following successful validation, the computational model will be used to guide future dimensional design and tolerance changes that will again be confirmed through in vitro hemolysis testing.

**B-2.1.4 Model Risk.** To assess the model risk associated with this COU, classifications for both model influence and decision consequence are proposed.

**B-2.1.4.1 Model Influence.** The proposed classification system for model influence is as follows:

| Model Influence | Description |
|---|---|
| LOW | The output of the model has a small influence on a design or safety decision. |
| MEDIUM | The output of the model has an important role in a design or safety decision. |
| HIGH | The output of the model has a dominant role in a design or safety decision. |

The intent of the computational model is to identify the key pump components or features whose dimensional variation could lead to increased hemolysis, which will then be directly assessed through in vitro testing. Additionally, results from in vitro testing of the new centrifugal pump will be compared against results from a predicate device and static controls to take into account the variability in blood sample fragility. Based on these considerations, the model influence is ranked as MEDIUM.

**B-2.1.4.2 Decision Consequence.** The proposed classification system for decision consequence is as follows:

| Decision Consequence | Description |
|---|---|
| LOW | A poor decision may result in increased clinician monitoring, no increased patient risk. |
| MEDIUM | A poor decision may result in short-term patient risk and increased clinician monitoring. |
| HIGH | A poor decision may result in immediate danger to the patient (e.g., injury or death), thus requiring significant clinician intervention up to and including immediate replacement of device. |

An incorrect decision to alter the pump component dimensional tolerances could adversely impact the plasma-free hemoglobin levels during clinical use. This could result in patient injury and require immediate intervention of the clinician to monitor patient hemolysis levels and/or replace the pump. As such, the decision consequence associated with this COU is ranked as HIGH.

The model influence and decision consequence are mapped to a five-level risk schema, as shown in Figure B-2.1.4.2-1. Based on this risk analysis, the COU has a model risk of MEDIUM-HIGH, which corresponds to Level 4 in the Model Risk Matrix.

To guide the next steps of establishing credibility goals and planning how to achieve those goals, it is helpful to relate the specific risk levels to tangible outcomes. The outcomes will then be examined during the credibility assessment that follows validation. Table B-2.1.4.2-1 provides an example of how to relate each risk level to adequate validation outcomes. The accuracy targets are intended to represent the need for increased accuracy with increased model risk. For this example, because the COU has a risk level of 4, the uncertainty must be estimated from the comparator or the model.

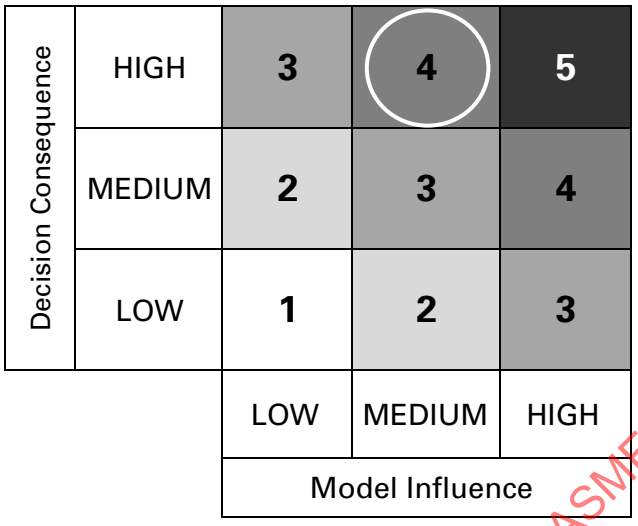
**B-2.1.5 Establish Credibility Goals.** For this COU, the following aspects of establishing credibility goals are highlighted, based on the unique challenges associated with measuring and predicting hemolysis: numerical code verification (NCV), discretization error, governing equations, and consideration of several aspects of the comparator.

**B-2.1.5.1 Verification**

**B-2.1.5.1.1 Code Verification — NCV.** The following scale is used to guide verification activities:

| Credibility | Description |
|---|---|
| A | NCV is not performed. |
| B | The numerical solution is compared to an accurate benchmark solution from another verified code. |
| C | Discretization error is quantified by comparison to an exact solution, and a grid convergence study is carried out to show that the numerical solution asymptotically approaches the exact solution as the discretization is refined. However, the observed order of accuracy is not quantified. |
| D | In addition to the quantification of discretization error and the execution of a grid convergence study, the observed order of accuracy is quantified and compared to the theoretical order of accuracy. |

**Figure B-2.1.4.2-1 Model Risk Matrix for Example 1**




Credibility level C is chosen based on the risk associated with the COU. To achieve this level for this example, assuming the MIH is computed using a modified off-the-shelf CFD code, discretization error is quantified by comparison to an exact solution on a simplified geometry. Additionally, a grid convergence study is carried out to show that the numerical solution asymptotically converges to the exact solution as the discretization is refined.

**B-2.1.5.1.2 Calculation Verification — Discretization Error.** The following scale is used to guide verification activities:

| Credibility | Description |
|---|---|
| A | No grid convergence analyses are performed. |
| B | Applicable grid convergence analyses are performed; conservation equation balances are not checked. |
| C | Applicable grid convergence analyses are performed, but not for problem-specific QOIs. Conservation equation balances are checked; no estimation of discretization error is performed. |
| D | Conservation equation balances are checked; estimation of discretization error is performed for problem-specific QOIs. |

Credibility level D is chosen based on the risk associated with the COU. To achieve this level, mesh sensitivity studies are conducted with a focus on refining the mesh in regions with elevated shear stress and extended shear exposure times. Mesh quality in these regions is also assessed. Further, the mesh verification studies are conducted directly on the primary variable of interest (MIH), as well as impeller torque and peak wall shear stress in critical locations. Finally, conservation of mass and momentum are verified.

**Table B-2.1.4.2-1 Corresponding Risk Levels for the Credibility Factors That Address Rigor of Output Comparison and Agreement of Output Comparison, With the Addition of Validation Metric in Figure B-2.1.4.2-1**

| Risk Level | Validation Metric |
|---|---|
| 1 | Visual comparison concludes good agreement. |
| 2 | Comparison by measuring the difference between computational results and experimental data. Differences are less than 20%. |
| 3 | Comparison by measuring the difference between computational results and experimental data. Differences are less than 10%. |
| 4 | Comparison with uncertainty estimated and incorporated from the comparator or the computational model. Differences between computational results and experimental data are less than 5%. Includes consideration of some uncertainty, but statistical distributions for further uncertainty quantification are unknown. |
| 5 | Comparison with uncertainties estimated and incorporated from both the comparator and the computational model, including comparison error. Differences between computational results and experimental data are less than 5%. Statistical distributions are known for rigorous treatment of uncertainty. |

**B-2.1.5.1.3 Calculation Verification — Numerical Solver Error.** The following scale is used to guide verification activities:

| Credibility | Description |
|---|---|
| A | No solver convergence tolerance sensitivity study is performed. No justification is provided for choosing a specific convergence criterion. |
| B | No solver convergence tolerance sensitivity study is performed. Solver convergence tolerances are established based on values used in previously verified computational models. |
| C | Problem-specific sensitivity study is performed on solver convergence tolerance. Sensitivity study shows that the changes in global conservation quantities due to changes in the convergence criteria are negligible relative to the model accuracy goal. |
| D | Problem-specific sensitivity study is performed on solver convergence tolerance. Sensitivity study shows that the changes in global conservation quantities and problem-specific quantities due to changes in the convergence criteria are negligible relative to the model accuracy goal. |

Credibility level D is chosen based on the risk associated with the COU. To achieve this level, key global quantities such as mass and momentum imbalances, and specific local quantities, including impeller torque, peak shear stress, and MIH, are monitored to ensure their values are independent of convergence criteria.

### B-2.1.5.2 Validation

**B-2.1.5.2.1 Computational Model — Model Form.** The following scale is used to guide validation activities associated with the governing equations:

| Credibility | Description |
|---|---|
| A | Little or no attempt is made to explore the influence of model form. |
| B | Key modeling assumptions are identified. |
| C | Comprehensive evaluation of model form assumptions is completed. |

Credibility level C is chosen based on the risk associated with the COU. In the current example, key equations of the computational model are the turbulence model and empirical model for MIH. Ensuring that the resolution of the mesh elements along the wall is within the recommended range is one way to ensure that the model-form requirements of the turbulence model are being met. Though the hemolysis model used to calculate MIH may be substantiated in the literature, the empirical coefficients used in the hemolysis generation expression are also varied to assess their impact on the hemolysis index prediction (input parameter uncertainty). To achieve the highest level of credibility, the impact of the model form — in this case, the hemolysis generation expression — on the model results is also quantified.

**B-2.1.5.2.2 Comparator.** In this example, in vitro laboratory tests using pumps manufactured at component dimensions predicted by the computational model to result in elevated hemolysis are used to validate the model. This set of tests serves two purposes: to determine the effect of dimensional tolerances on hemolysis and to validate the model's ability to predict the in vitro test results. To further isolate and assess the effects of dimensional changes and variability in blood fragility, the hemolysis tests are conducted on a commercially available predicate device and include static control samples (blood samples that will not be exposed to the pumping circuit) in addition to the production pump samples.

For each trial pump, sample-to-sample variation is quantified by running multiple experiments using different blood samples. Multiple blood samples are taken periodically throughout each test to assess the potential time dependence of hemolysis. The credibility of this comparator is assessed as follows:

*(a) Comparator — Test Samples.* The following scale is used to guide sample measurement activities:

| Credibility | Description |
|---|---|
| A | Key test sample properties are identified but not quantified. |
| B | Key test sample properties are identified and quantified, but the uncertainty of the measurement is not quantified. |
| C | Key test sample properties and the uncertainty of their measurement are quantified, but the statistical distributions of the properties are unknown. |
| D | Key test sample properties and the uncertainty of their measurement are quantified, and statistical distributions of the properties are known. |

Credibility level C is chosen based on the risk associated with the COU. To achieve level C for this credibility factor, components are dimensionally characterized at appropriate locations during trial pump manufacturing. The uncertainty associated with dimensional measurements is also characterized.

*(b) Comparator — Test Conditions*

*(1)* The following scale is used to guide characterization of measurement data on test conditions:

| Credibility | Description |
|---|---|
| A | Test conditions are qualitatively characterized. |
| B | A single key characteristic of the test condition is measured. |
| C | All key characteristics of the test condition are measured. |

Credibility level C is chosen based on the risk associated with the COU. To achieve level C for this credibility factor, the following key characteristics are measured:

*(-a)* rotational speed of trial and predicate pumps to achieve desired blood flow rates
*(-b)* blood temperature (maintained at body temperature using a water bath)
*(-c)* pump outlet pressure (set at a clinically relevant arterial pressure)
*(-d)* blood age and physical properties (density and viscosity)
*(-e)* blood plasma–free hemoglobin, total hemoglobin, and hematocrit levels
*(-f)* blood volume in circuit
*(-g)* duration of test, timing of sampling

*(2)* The following scale is used to assess the uncertainty of measurements for characterizing test conditions:

| Credibility | Description |
|---|---|
| A | Measurements are qualitative observations (e.g., imaging without quantification) with limited spatial/temporal monitoring; measurement uncertainty is not quantified. |
| B | Measurements are obtained from instruments with known accuracy and monitored at critical locations. |
| C | Measurements are obtained from instruments with known uncertainty and monitored against specific tolerances at critical locations. |

Credibility level B is chosen based on the risk associated with the COU. To achieve level B for this credibility factor, instruments with known accuracy are used to measure the quantities needed for model inputs (such as blood viscosity and blood flow rate) and the quantities used to calculate the MIH for the comparator (such as plasma-free hemoglobin).

**B-2.1.5.2.3 Assessment — Output Comparison.** In this example, the model risk is connected to the extent of agreement between the model and comparator outputs as described in Table B-2.1.4.2-1. Thus, based on the risk assessment for this COU, a credibility level of 4 was selected such that the validation activities demonstrate model accuracy to within 5% and uncertainty has been estimated.

**B-2.1.5.3 Applicability: Relevance of the Validation Activities to the COU.** The following scale is used to assess the relevance of the validation activities to the COU:

| Credibility | Description |
|---|---|
| A | There is no overlap between the ranges of the validation points and the COU. |
| B | There is partial overlap between the ranges of the validation points and the COU. |
| C | The COU encompasses some of the validation points. |
| D | The COU encompasses all of the validation points. |

The model and test configurations where the validation model and validation comparator were evaluated were chosen to be at the extreme dimensional tolerances expected to result in increased hemolysis. These extreme dimensions also represent the device configurations most directly connected to the COU. As such, there were minimal differences between the validation activities and the COU. Therefore, the validation activities are highly applicable to the COU, achieving a credibility level D. Reduced applicability would be achieved, for example, if the validation model and validation comparator configurations were at nominal dimensions or at dimensional tolerances that did not result in increased hemolysis.

**B-2.1.5.4 Summary.** Table B-2.1.5.4-1 summarizes the rigor selected for each credibility factor and the credibility level for each V&V activity. Ellipses (...) indicate credibility factors that were not reviewed in this example. The activity credibility levels range from MEDIUM to HIGH with an overall model credibility of MEDIUM-HIGH, which is commensurate with the model risk of Level 4, MEDIUM-HIGH.

**Table B-2.1.5.4-1 Credibility Factors Summary**

| Activity | Credibility Factor (Paragraph) | Level of Rigor | | Credibility Level |
|---|---|---|---|---|
| | | Selected | Maximum | |
| Verification | | | | |
| Code | Software quality assurance | ... | ... | ... |
| | Numerical code verification (B-2.1.5.1.1) | C | D | MEDIUM-HIGH |
| Calculation | Discretization error (B-2.1.5.1.2) | D | D | HIGH |
| | Numerical solver error (B-2.1.5.1.3) | D | D | |
| | Use error | ... | ... | ... |
| Validation | | | | |
| Computational model | Model form (B-2.1.5.2.1) | C | C | HIGH |
| | Model inputs | ... | ... | ... |
| Comparator | Test samples: Measurement uncertainty [B-2.1.5.2.2(a)] | C | D | MEDIUM |
| | Test conditions | | | |
| | Measurements [B-2.1.5.2.2(b)(1)] | C | C | |
| | Measurement uncertainty [B-2.1.5.2.2(b)(2)] | B | C | |
| Assessment | Equivalency of input parameters | ... | .... | .... |
| | Output comparison (B-2.1.5.2.3) | 4 | 5 | MEDIUM-HIGH |
| Applicability | Relevance of the quantities of interest | ... | ... | ... |
| | Relevance of the validation activities to the COU [B-2.1.5.3] | D | D | HIGH |

### B-2.1.6 References

Harihana, P., Computational Fluid Dynamics Round Robin Study, U.S. Food and Drug Administration, Silver Spring, MD (https://fdacfd.nci.nih.gov/interlab_study_2_blood_pump)

Pathmanathan, P., Gray, R. A., Romero, V. J., Morrison, T. M. (2017), "Applicability analysis of validation evidence for biomedical computational models," *J. Verif. Valid. Uncert.*, vol. 2, issue 2, 021005-021005-11 (DOI: 10.1115/1.4037671)

## B-2.2 Example 2: Predicting the Performance of Flow Diverters in the Treatment of Brain Aneurysms

This example describes two COUs and their associated computational model. In the first COU, the computational model results have a high influence on the device design decision because the model results provide the only evidence for decision-making. In the second COU, the design decision is supported by in vitro test data and preclinical evidence, thus lowering the model influence.

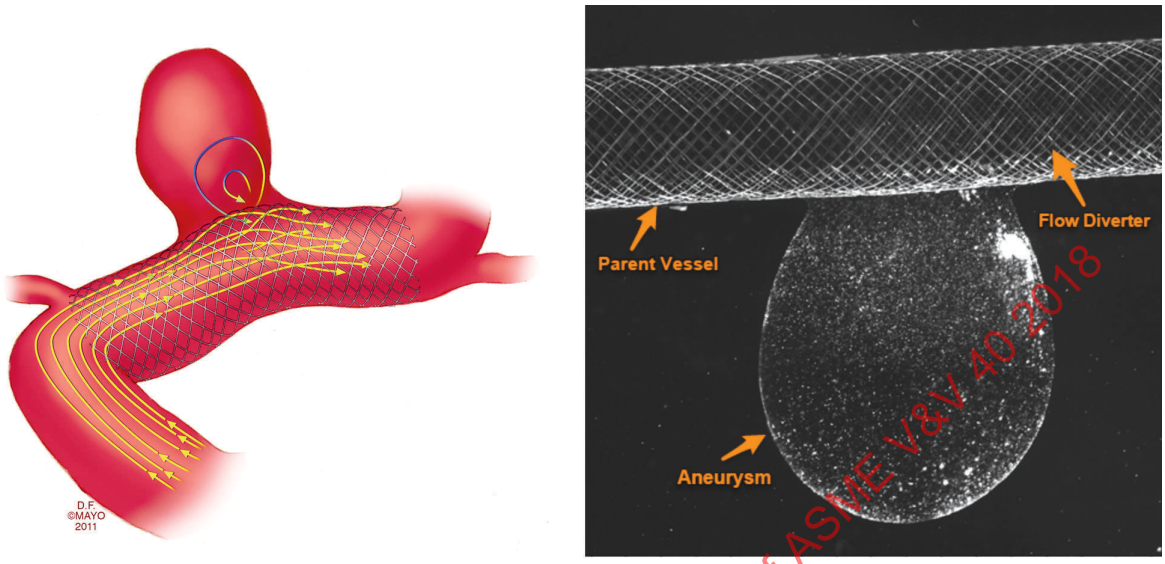
**B-2.2.1 Background.** An intracranial aneurysm is the result of degradation and bulging of the wall of a blood vessel that supplies blood to the brain. Aneurysm rupture, or even leakage of blood out of an aneurysm into brain tissue, has a very high mortality rate. A flow diverter (see Figure B-2.2.1-1) is a wire mesh tube that is placed in the aneurysm's parent vessel. This device redirects blood flow away from the aneurysm, promoting aneurysm occlusion and parent vessel healing. Large, nonspherical, and wide-necked aneurysms are commonly treated with flow diverters.

Diverter performance is commonly assessed by the percent reduction in the rate of blood flow entering the aneurysm after the device is implanted. Poor flow diversion may delay clot formation inside an aneurysm, prolonging the risk of rupture, and thus is a safety concern. Computational models of the blood flow across a flow diverter into a treated aneurysm can be used to predict diverter effectiveness, to evaluate new diverter designs, and to better understand clinical outcomes (see Figure B-2.2.1-2).

**B-2.2.2 Question of Interest.** Is the flow diversion performance of a next-generation flow diverter equivalent to or better than the performance of a predicate device for which the safety and effectiveness have been proven through clinical use?

**B-2.2.3 Contexts of Use.** Two COUs assessing the flow diversion performance of the next-generation device are given below. These COUs have different levels of influence on the decision of whether the flow diversion performance of the new device is equivalent to or better than that of the predicate device. In both cases, the computational model is used to predict a flow diversion performance metric, defined as the percent reduction in the time-averaged aneurysm inflow rate after the flow diverter is deployed across the aneurysm neck. Evaluations are based on a set of patient-specific geometries obtained from clinical cases where the predicate flow diverter device has been used and successful clinical outcomes have

**Figure B-2.2.1-1 An Example of a Flow Diverter Placed in a Parent Vessel With a Side-Wall Aneurysm**

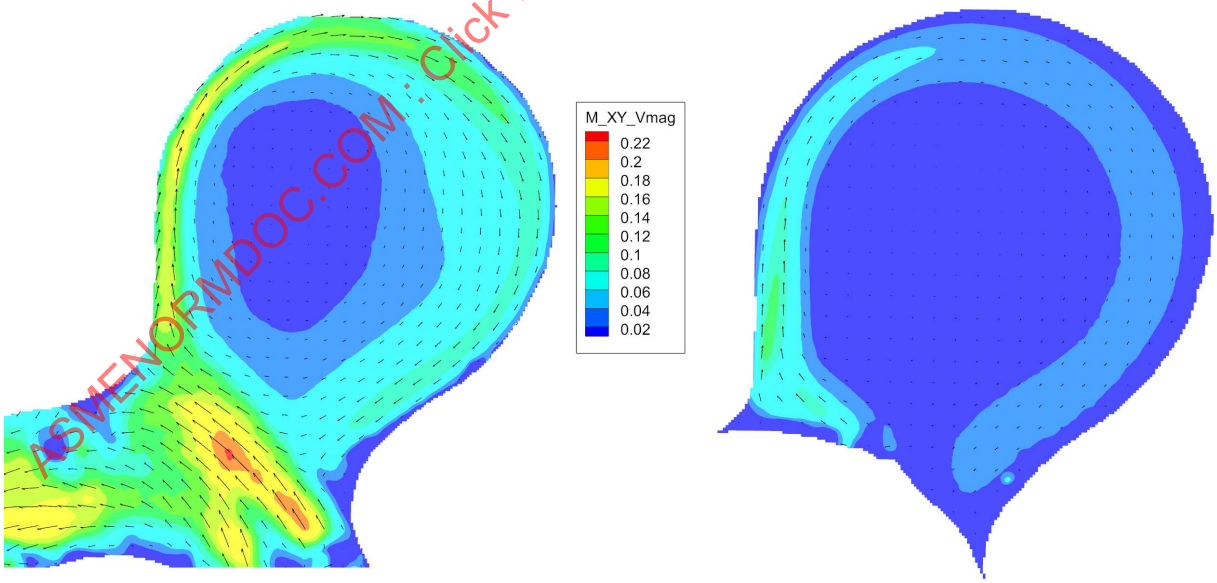



(a) Flow Pattern Through a Vessel With an Implanted Flow Diverter [Note (1)]

(b) Image From an In Vitro Test Platform Designed to Test Flow Diverter Performance

NOTE: (1) Illustration (a) used with permission of Mayo Foundation for Medical Education and Research. All rights reserved.

**Figure B-2.2.1-2 The Flow Patterns Before and After the Placement of a Flow Diverter, Highlighting the Significant Reduction in Blood Flow Within the Aneurysm After Diverter Placement**

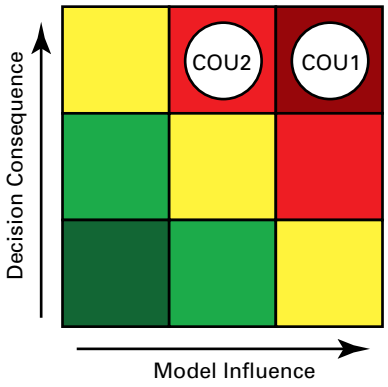



(a) Before Placement of Diverter

(b) After Placement of Diverter

GENERAL NOTE: The color scale is adjusted to illustrate the effectiveness of the flow diverter and the behavior of residual flow still entering the aneurysm.

**Figure B-2.2.4.2-1 Model Risk Matrix for Example 2**




been demonstrated by long-term follow-up. While patient-specific geometries are used as inputs to the computational modeling studies, neither COU intends to predict patient-specific outcomes as a result of clinical use of this device.

*(a) COU1 — Performance Evaluation With Simulation Only.* Computational modeling is used to evaluate the relative flow diversion performance of a next-generation flow diverter with respect to a predicate device. There is no supporting data from in vitro testing available for flow diversion performance of the new device.

*(b) COU2 — Performance Evaluation With Simulation and Additional Supporting Data From In Vitro and In Vivo Testing.* In addition to computational modeling studies, in vitro testing and in vivo preclinical studies are conducted to support the determination of whether the flow diversion performance of the new device is equivalent or better than the predicate device.

### B-2.2.4 Model Risk

**B-2.2.4.1 Model Influence.** The following three-level classification system is used to assess model influence in this example:

| Model Influence | Description |
|---|---|
| LOW | Results from the computational model are a minor factor in the decision. |
| MEDIUM | Results from the computational model and other supporting evidence play an equal role in the decision. |
| HIGH | Results from the computational model are a significant factor in the decision. |

Based on this classification system, COU1 has a HIGH influence because the computational model results are the only data informing the decision. COU2 has a MEDIUM influence because supporting data from in vitro testing and in vivo preclinical studies complement the computational modeling studies.

**B-2.2.4.2 Decision Consequence.** The following three-level classification system is used to assess decision consequence in this example:

| Decision Consequence | Description |
|---|---|
| LOW | An incorrect decision based on the computational model results will not result in patient harm. |
| MEDIUM | An incorrect decision based on the computational model results would result in minor patient injury or potentially require physician intervention or have other moderate impacts. |
| HIGH | An incorrect decision based on the computational model results could result in patient harm in clinical use and would have a negative impact on product development costs. |

Based on this classification system, both COUs would have a HIGH consequence because an incorrect decision could cause patient harm.

The model risk is a combination of the model influence and decision consequence. In this example, because COU1 has a HIGH model influence and a HIGH decision consequence, the model risk is HIGH. In contrast, COU2 has a MEDIUM model influence and a HIGH decision consequence, leading to a MEDIUM-HIGH model risk. The model risk associated with each COU is depicted in Figure B-2.2.4.2-1.

**B-2.2.5 Establish Credibility Goals.** The COUs described above require different levels of model credibility due to the different levels of model risk. Selected V&V activities and their relevance to establishing the credibility of the computational model for each COU are discussed in paras. B-2.2.5.1 through B-2.2.5.3.

**B-2.2.5.1 Verification: Calculation Verification — Discretization Error.** A grid convergence study is performed to estimate discretization error. Since the model risk associated with COU1 is higher relative to COU2, a smaller spatial and temporal discretization error is necessary for COU1 to achieve a more accurate numerical solution.

#### B-2.2.5.2 Validation

**B-2.2.5.2.1 Computational Model — Model Inputs.** Aspects regarding geometry were considered for this credibility factor. Flow diverters are typically oversized with respect to the diameter of the parent vessel to ensure good wall apposition and anchoring. In the case of wide-necked aneurysms, the deployed device may have a larger diameter at the aneurysm neck because of the absence of wall support in these locations. The deployed diameter affects the geometry of the openings between the wires, which in turn affects the local area porosity and flow resistance. For both COUs, performing sensitivity analyses to assess the effect of variation of the deployed device diameter on the computational model results would increase model credibility for this factor.

#### B-2.2.5.2.2 Comparator

*(a) Test Samples.* For both COUs, the performance evaluations are based on a set of patient-specific geometries obtained from clinical cases. For COU2, nominal dimensions of the implanted device are assumed. However, for COU1, which requires increased credibility, the installed predicate device in the in vitro tests is characterized using tools such as microcomputed tomography (micro-CT) and three-dimensional (3D) reconstruction methods, thus enabling the actual device dimensions to be used in the computational model.

*(b) Test Conditions.* A higher credibility can be achieved for COU1 by increased monitoring of test conditions, such as the inlet/outlet flow conditions applied during the in vitro studies, enabling the uncertainty of the flow conditions to be quantified. Increasing control over the flow conditions can also minimize uncertainty.

#### B-2.2.5.2.3 Assessment

*(a) Equivalency of Input Parameters.* Equivalent model and comparator inputs result in higher model credibility as compared to inputs that are similar. For COU2, sufficient credibility may be achieved by ensuring comparable input flow rates (similar inputs) in both the validation comparator and validation model. For the higher-risk COU1, it may be appropriate to ensure consistency in the spatial and temporal variations in the input waveforms (equivalent inputs), which may be obtained using particle-image velocimetry (PIV) or other image-based methods.

*(b) Output Comparison.* The primary output of this study is the aneurysm neck velocity field. In particular, it is important to ensure that the spatial and temporal resolutions of this velocity field are sufficient and consistent between the computational model and the comparator, so that the inflow-rate calculations from the model and the in vitro test will be accurate and comparable. This may be accomplished for COU2 by comparing two-dimensional (2D) PIV measurements at selected points on the plane of measurement to model predictions. To achieve the increased credibility required for COU1, it may be appropriate to quantify differences in the volumetric velocity field based on 3D PIV measurements, and also to include the uncertainty of the inlet/outlet flow conditions.

#### B-2.2.5.3 Applicability

*(a) Relevance of the QOIs.* In this example, the QOIs of the validation study (the percent reduction in blood flow after flow diverter implantation) are identical to what was specified for both COUs.

*(b) Relevance of the Validation Activities to the COU.* Even though the model validation is conducted using multiple patient-specific device and vessel/aneurysm geometries (i.e., validation points), the COU space extends beyond these validation points. That is, the model could be used to make predictions for geometries in a range that is different from the geometry range of the validation points. As a result, the model credibility for this factor depends on how well the patient-specific geometries used for validation align with the geometries that may be experienced in clinical use.

### B-2.3 Example 3: Stability and Adjustability of Hospital Beds

This example is primarily intended to highlight how a single computational model framework can have multiple COUs to support different questions of interest, each of which has a unique level of consequence and therefore a unique risk profile. The impact of the patient consequence (and risk) differential on the rigor of V&V activities is considered. This example also illustrates how the risk assessment can reflect factors other than patient harm.

**B-2.3.1 Background.** Hospital beds (see Figure B-2.3.1-1) are designed based on the needs of the patient and caregiver while they are in the hospital or at home. One such need is to maintain stability, which ensures that the bed will not tip over under any circumstance. Another need is to manipulate the bed sleep surface, which may be achieved with embedded

**Figure B-2.3.1-1 Schematic of a Hospital Bed**

GENERAL NOTE: Courtesy of Hill-Rom, Chicago, IL.

actuators. The bed is capable of moving the sleep surface up and down. The sleep surface can also be extended and retracted in both the lateral and longitudinal directions to increase the sleep surface area. The bed may also need to be lowered to ease patient entry or exit, thus reducing the risk of patient falls. The sleep surface sections may also be raised to achieve different positions for patient comfort or medical procedures. Computational modeling can be used to predict bed stability in different configurations and to predict the ability of the actuators to articulate the bed.

A bed typically has four casters. Instability of the bed is defined as a condition for which both casters opposite the load lift off the floor. It is a condition that may result in harm to the patient or caregiver, as well as damage to the equipment. Thus, ensuring stability requires that the summation of the caster reaction forces on the opposite side of the applied load is greater than zero. In this example, a kinematic model of the mechanical system is used to address this problem. Within this model, weldments and other subassemblies are considered rigid bodies with mass. Patient weight and any other applied loads are considered as lumped forces. Constraints are added at pivot joints, sliding joints, or contact points between bodies to represent the fixed and free degrees of freedom.

In this example, two separate questions of interest, Q1 and Q2, are posed, each having its own COU: one is related to bed stability, and the other is related to the ease of bed articulation.

### B-2.3.2 Questions of Interest and Contexts of Use

#### B-2.3.2.1 Bed Stability

*(a) Q1.* Does the bed meet stability requirements?

*(b) COU1.* The purpose of the computational model is to predict whether the bed is expected to tip. Stability test data are available on early iterations of the current design to support a decision on the stability of the final design. However, no physical testing of the final design is intended. Therefore, this decision will be based on computational model predictions of the caster reaction forces.

#### B-2.3.2.2 Ease of Bed Articulation

*(a) Q2.* What is the design margin of the selected actuator(s) for the bed articulation requirement?

*(b) COU2.* In this COU, the computational model is used to evaluate the actuator loads needed to articulate the bed, which guides the selection of the appropriate actuators to incorporate into the bed. The ability of the selected actuators to sufficiently articulate the bed is evaluated through physical testing of the current design.

### B-2.3.3 Model Risk

**B-2.3.3.1 Model Influence.** The following scale is used to assess model influence in this example:

| Model Influence | Description |
|---|---|
| Negligible | Results from the computational model are a negligible factor in the decision. Results are used in research projects that have no direct bearing on the decision. |
| Minor | Results from the computational model are a minor factor in the decision. Ample test data for the real system in the real environment are available, and computational model results are used as supplementary information. |
| Moderate | Results from the computational model are a moderate factor in the decision. Limited test data for the real system in the real environment are available, or ample test data for similar systems in similar environments are available. |
| Significant | Results from the computational model are a significant factor but not the sole factor in the decision. No test data are available for the real system in the real environment. Limited test data for similar systems in similar environments are available. |
| Controlling | Results from the computational model are the controlling (sole) factor in the decision. No test data are available. |

Based on this scale, the model influence is moderate for both COU1 and COU2 due to the presence of relevant physical test data (either on the current system earlier in the design process or on the current design) to supplement model predictions.

**B-2.3.3.2 Decision Consequence.** The following scale is used to assess decision consequence in this example:

| Decision Consequence | Description |
|---|---|
| None | The decision is not linked to hazards in the device risk assessment or system failure modes and effects analysis. Additionally, there is no consequence of the decision on the patient or caregiver, or on the performance of the equipment. |
| Minor | A poor decision would not adversely affect personal safety or health and/or would not result in damage to the equipment beyond normal use. |
| Moderate | A poor decision may result in minor injury to the patient or caregiver and/or minor damage to the equipment. |
| Critical | A poor decision may result in severe injury or death to the patient or caregiver and/or major damage to the equipment. |

For Q1–COU1, in which sufficient bed stability is assessed, the decision consequence is critical because the patient could suffer a severe injury or death if the bed fails this requirement. Additionally, the equipment could suffer major damage. In contrast, for Q2–COU2, if the wrong actuator is chosen, the only result will be that the motor fails to extend and/or retract the motor shaft. The actuator is not expected to break since the static load capability is typically higher than its nominal load rating. Therefore, the decision consequence is minor for this COU.

The overall risk level for Q1–COU1 is therefore greater than for Q2–COU2 due to the more severe decision consequence. Accordingly, more rigorous V&V activities are required to establish appropriate model credibility for COU1 than for COU2.

**B-2.3.4 Establish Credibility Goals.** Factors associated with model credibility that will be described in the context of this example are code verification, calculation verification, validation model, validation comparator, and output assessment. The approach to choosing the credibility goals for each factor is also discussed.

#### B-2.3.4.1 Verification

**B-2.3.4.1.1 Code Verification — Software Quality Assurance (SQA).** The computational model uses off-the-shelf (OTS) software. The following is an example gradation of activities for OTS software:

*(a)* No SQA procedures are documented.

*(b)* SQA procedures from the vendor are referenced.

*(c)* A supplier audit is conducted with the vendor to confirm that quality procedures are conducted and documented during the software development process.

*(d)* Benchmark verification test cases, provided by the vendor, are run on the user's computer platform. The results are compared to vendor results and documented.

The computational model software for COU1 and COU2 is tested through a documented SQA process from the vendor. The process consists of ensuring that benchmark test cases replicate previously established results and comparing the error with analytical solutions. This level of credibility may be deemed appropriate for both COUs. However, since COU1 has elevated risk relative to COU2, the end user may perform additional benchmarks to support COU1.

**B-2.3.4.1.2 Code Verification — Numerical Code Verification (NCV).** Both COUs use the same kinematic analysis software, in which the underlying physics of a mechanical system are modeled through the solution of nonlinear numerical equations. The following is an example gradation of NCV activities:

*(a)* NCV is not performed.

*(b)* The numerical solution is compared to an accurate benchmark solution from another verified code.

*(c)* Discretization error is quantified by comparison to an exact solution.

COU2 has a lower risk, and so a comparison of the numerical solution to an accurate benchmark solution from another verified code is deemed acceptable. COU1 has a higher risk, which requires quantifying the discretization error from the simulation time-step, conducting a convergence study to show a reduction of error in the numerical solution with smaller time-step size, and comparing the results of that convergence study to an analytical solution.

**B-2.3.4.1.3 Calculation Verification — Use Error.** The following is an example gradation of activities for this credibility factor:

*(a)* Inputs and outputs are not verified.

*(b)* Key inputs and outputs are verified by the practitioner.

*(c)* Key inputs and outputs are verified by internal peer review.

*(d)* Key inputs and outputs are verified by reproducing important simulations as part of an external peer review.

Since the model risk is higher for COU1, use error is addressed by an internal peer review that verifies key inputs and outputs. This ensures that the critical model input parameters (such as bed weight) are confirmed with the design team, that the correct modeling approach is used, and that the results are interpreted properly. For the lower-risk COU2, the practitioner needs only to verify key inputs and outputs against reference inputs and analytical solutions, respectively, to ensure adequate credibility.

### B-2.3.4.2 Validation

**B-2.3.4.2.1 Computational Model — Model Inputs.** Aspects of boundary and loading conditions were considered for this credibility factor. The major sources of uncertainty are location of the applied load and the variation in the product center of gravity. COU1 has a higher model risk, and so the uncertainty is quantified for these factors. The uncertainty in load position is accounted for by conducting a study of load positions and comparing the computational model results with the test results. The uncertainty in the bed center of gravity is quantified by accounting for variation in material density and comparing the computational model results with the test results. Simply using the nominal values for these parameters is acceptable for COU2 since the risk is lower and physical testing is performed on the final product.

**B-2.3.4.2.2 Comparator —Test Conditions.** For this example, all detachable accessories are removed from the bed, and patient loads are applied using weight bags. The caster reaction forces are measured using load cells that sit underneath each caster. Two potential sources of variability in the comparator data are the point of application of the weight bags and the orientation of the load cells underneath the bed. For the higher-risk COU (COU1), it is appropriate to evaluate the uncertainty associated with the load location and load cell orientation through a reliability and repeatability study involving multiple stability loading conditions to better understand the sensitivity of the output to those inputs. For the lower-risk COU (COU2), such an uncertainty assessment is not required.

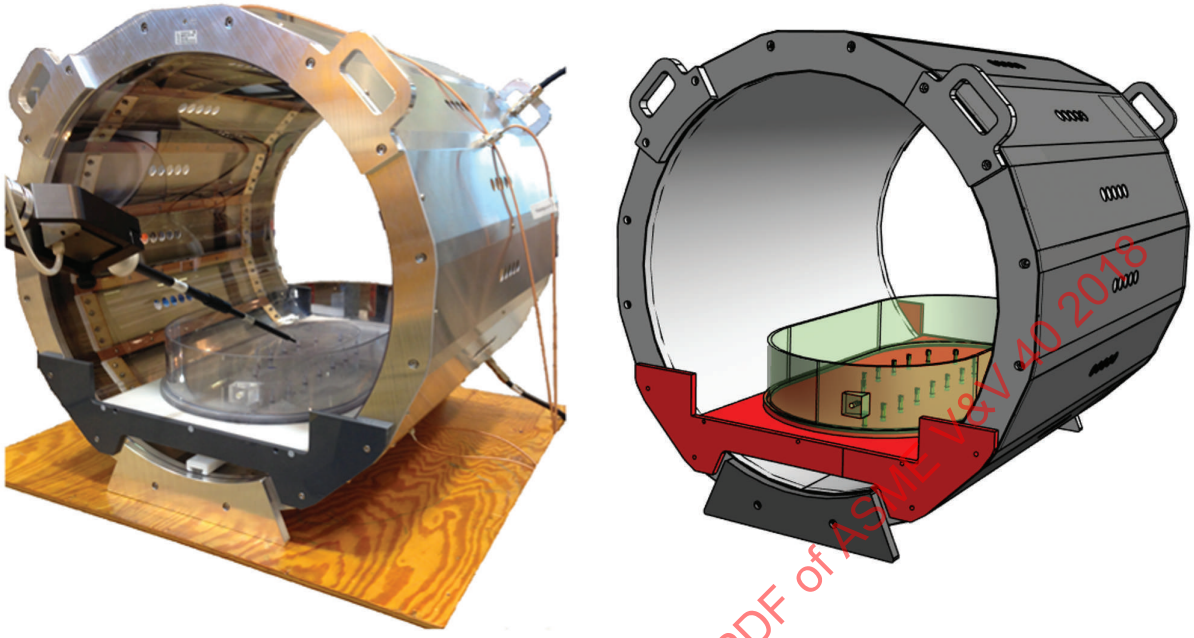
**B-2.3.4.2.3 Assessment — Output Comparison.** Key outputs are equivalent between the validation model and the validation comparator for both COUs. For COU1, the key outputs of the model and comparator are the reaction forces at the caster locations. For COU2, the key outputs of the model and comparator are the actuator reaction forces.

As the risk associated with the COU increases, the rigor of the output comparison is increased by including model-to-comparator load comparisons at multiple validation points. For COU1, the rigor of the output comparison is further increased by incorporating the uncertainty of the surrogate patient load application location and bed center of gravity in both the model and the comparator. However, for COU2, it is sufficient to compare results between the model and comparator assuming nominal values for the load application location and caster orientation.

## B-2.4 Example 4: Radiofrequency-Induced Temperature Rise in Patients During Magnetic Resonance Imaging

This example focuses on a computational model with two potential COUs, each of which has a different amount of computational model influence on the question of interest and therefore presents a different model risk. Two unique comparators are considered, and the discussion focuses on the relationship between model risk and the selection of the appropriate validation pathway.

**Figure B-2.4.1-1 Physical Test Set-Up and Computational Model Representation of a Gel Phantom Inside an MRI**



GENERAL NOTE: Courtesy of Zurich MedTech (ZMT), Zurich, Switzerland.

**B-2.4.1 Background.** Magnetic resonance imaging (MRI) is a widely used radiological imaging technique with over 39 million estimated scans performed in the United States in 2016 (see para. B-2.4.6, ref. [1]). The success of MRI is due to its clinical versatility, the use of non-ionizing radiation, and the high soft-tissue contrast (see para. B-2.4.6, ref. [2]). However, the radiofrequency (RF) field used to produce the images may generate excessive tissue heating, potentially resulting in permanent injury. Bench testing studies have been conducted to study RF heating during MRI (see para. B-2.4.6, ref. [3]), but the specific geometry, loading conditions (e.g., position of patient within the coil), and multiple types of MRI sequences make it difficult to identify specific conditions where excessive heating will occur a priori. Computational modeling can be used to identify potentially dangerous levels of tissue heating, specifically in the presence of implantable medical devices. These studies can be performed using a non-anatomical gel phantom (see para. B-2.4.6, refs. [4] and [5]) and/or anatomically accurate models (see para. B-2.4.6, refs. [6] and [7]).

This example focuses on the evaluation of tissue heating in the presence of a trauma plate and screw system, which is intended for fixation of bone fractures. The computational model and experimental setup are based on the gel phantom approach (see Figure B-2.4.1-1 and para. B-2.4.6, ref. [8]). Consideration is given to some of the primary sources of variability associated with RF heating in the presence of this class of medical devices, specifically multiple sizes of plates and screws, multiple screw locations in each plate, and multiple screw trajectories for each location (see para. B-2.4.6, ref. [5]).

**B-2.4.2 Question of Interest.** What is the maximum temperature increase in the tissue near a plate-and-screw system, due to the presence of the device, during an MRI scan?

**B-2.4.3 Contexts of Use.** For both COUs, the computational model is validated first against experimental results.

*(a) COU1.* The COU of the computational model is to evaluate multiple configurations of the proposed plate-and-screw system to identify the worst-case configuration, which is the configuration with the largest predicted temperature increase in the surrounding tissue. The resulting worst-case configuration will then be physically tested to quantify the temperature increase in the phantom. That is, physical testing will be part of the design decision.

*(b) COU2.* The COU of the computational model is to evaluate multiple configurations of the proposed plate-and-screw system to identify the worst-case temperature increase. No additional physical testing of the worst-case configuration will be performed.

**B-2.4.4 Model Risk.** The risk assessment for the two COUs is differentiated strictly by the model influence, since the decision consequence (permanent tissue damage due to excessive tissue heating) is identical for both COUs. In this example, model influence is categorized as follows:

| Model Influence | Description |
|---|---|
| Supporting | The results from the computational model have a supporting role in the decision; additional experimental or clinical evidence exists. |
| Primary | The results from the computational model have a primary role in the decision; limited additional experimental or clinical evidence exists. |
| Exclusive | The results from the computational model are the sole influence on the decision; no additional experimental or clinical evidence exists. |

Based on this categorization, the model influence for COU1 is primary because of the intention of gathering experimental data beyond the computational model predictions. The model influence for COU2 is exclusive because no additional experimental data will be acquired. Therefore, the overall model risk associated with COU2 is greater than that for COU1.

**B-2.4.5 Establish Credibility Goals.** The role of the computational model varies when mitigating the patient risk associated with the plate-and-screw assembly in an MRI environment. Whether the computational model plays a supporting, primary, or exclusive role changes the rigor with which the credibility of the model must be established. That, in turn, influences the verification and validation activities needed to support the COU.

**B-2.4.5.1 Validation**

**B-2.4.5.1.1 Computational Model — Model Form.** Aspects of governing equations were considered for this credibility factor. High-frequency electromagnetic simulations used to determine energy absorption by the gel phantom are sufficient for COU1. However, for the increased credibility that is appropriate for COU2, electromagnetic simulations coupled to thermal analyses provide the energy absorption and the temperature rise, respectively.

**B-2.4.5.1.2 Computational Model — Model Inputs.** Aspects of system configuration and system conditions were considered for this credibility factor. For COU1, the computational model is used to identify worst-case condition(s), which guides subsequent physical testing. In COU2, however, the predicted temperature increase from the computational model is evaluated in an absolute sense. Therefore, the sensitivity of the simulation output to changes in geometry is assessed in COU2, as this assessment increases the credibility of the validation model. Similarly, assessment of the sensitivity of the model with respect to the system properties, namely electrical properties (electrical conductivity and permittivity of the phantom and device) and thermal properties (thermal conductivity, density, and heat capacity of the phantom and device), increases the credibility of the validation model to levels appropriate for the higher-risk COU.

**B-2.4.5.1.3 Comparator — Test Conditions.** One contemplated comparator, Comparator 1, is a gel phantom built as specified in ASTM F2182 (see para. B-2.4.6, ref. [4]) and subsequently scanned using MRI. Electromagnetic field magnitude data are collected at various locations inside the MRI coil. For validation, the phantom can be modeled using the same configuration as the physical test and results compared to experimental data collected at the same locations. In this case, the validation model predicts energy absorption throughout the phantom as a function of time, and the comparator yields discrete measurements in the physical system as a function of time. Therefore, Comparator 1 may be appropriate for the validation of COU1's computational model.

An enhanced comparator, Comparator 2, is Comparator 1 supplemented to include a representative plate-and-screw construct. The computational model predicts the temperature rise in addition to energy absorption throughout the gel phantom, including in the immediate vicinity of the device construct. This comparator yields similar measurements but only at discrete locations relative to the device construct. Given the higher model risk in COU2, the test conditions proposed for Comparator 2 provide increased model credibility.

**B-2.4.5.1.4 Assessment — Output Comparison.** The risk associated with the COU can drive the rigor with which the outputs from the computational model are compared to those from the comparator. For COU1, the model influence is moderate, and therefore comparing the total energy absorption at selected locations inside the phantom is sufficient. For COU2, the model influence is increased, and the need for additional credibility justifies greater rigor in the output comparison. In particular, a comparison of energy absorption at all critical locations inside the phantom and in the gap between the phantom and the coil is required (see para. B-2.4.6, ref. [8]).

**B-2.4.5.2 Applicability: Relevance of the Validation Activities to the COU.** The two comparators are unique in their applicability to the COU. For Comparator 1, there are significant differences between the validation points and the COU for both system configuration (no representative device construct was included) and system properties (because the device construct was not included, the relevant thermal properties of the device materials were not included). For Comparator 2, both of these differences are addressed by including a representative (though not necessarily worst-case) plate-and-screw construct in both the computational model and comparator. Therefore, the credibility associated with the validation activities to the COU is higher for Comparator 2 than for Comparator 1.

The reduced credibility associated with Comparator 1 may be commensurate with the moderate risk profile for COU1, while the higher credibility of Comparator 2 would be appropriate for COU2. Note that the credibility of Comparator 2, associated specifically with the applicability of the validation points to the COU, can be further increased by addressing any additional differences in system configuration between the validation activities and the COU. Only a single potential configuration of the device and a single location of the device in the gel was used in the validation activities, whereas the COU encompasses a range of configurations and locations. The need for additional levels of credibility is dictated by the absolute model risk (incorporating both model influence and decision consequence) attributed to the COU.

### B-2.4.6 References

[1] IMV MR 2017 Benchmark Report, IMV Medical Information Division, Inc., Des Plaines, IL (http://www.imvinfo.com/index.aspx?sec=mri&sub=dis&itemid=200085)

[2] Blamire, M. (2008), "The technology of MRI — The next 10 years?" *Brit. J. Radiol.*, vol. 81, pp. 601–617

[3] Song, T., Zhiheng Xu, Z., Iacono, M. I., Angelone, L. M., Rajan, S. S. (2018), "Retrospective analysis of RF heating measurements of passive medical implants," *Magn. Reson. Med.* (DOI: 10.1002/mrm.27346)

[4] ASTM F2182-11a, Standard Test Method for Measurement of Radio Frequency Induced Heating on or Near Passive Implants During Magnetic Resonance Imaging, American Society for Testing and Materials (ASTM International), West Conshohocken, PA

[5] Assessment of Radiofrequency-Induced Heating in the Magnetic Resonance (MR) Environment for Multi-Configuration Passive Medical Devices: Final Guidance for Industry and Food and Drug Administration Staff, March 22, 2016, U.S. Food and Drug Administration (FDA), Silver Spring, MD

[6] Christ, A., Kainz, W., Hahn, E. G., et al. (2010), "The virtual family — Development of surface-based anatomical models of two adults and two children for dosimetric simulations," *Phys. Med. Biol.*, vol. 55, pp. N23–N38

[7] Murbach, M., Neufeld, E., Kainz W., Pruessmann, K. P., Kuster, N. (2014), "Whole-body and local RF absorption in human models as a function of anatomy and position within 1.5T MR body coil," *Magn. Reson. Med.*, vol. 71, pp. 839–845

[8] Lucano, E., Liberti, M., Mendoza, G., Lloyd, T., Iacono, M. I., Apollonio, F., Wedan, S., Kainz, W., Angelone, L. M. (2016), "Assessing the electromagnetic field generated by a radiofrequency body coil at 64 MHz: Defeaturing vs. accuracy," *IEEE Trans. Biomed. Eng.*, vol. 8, pp. 1591–1601 (DOI: 10.1109/TBME.2015.2506680)

## B-2.5 Example 5: Evaluation of the Locking Mechanism Strength of a Posterior-Stabilized Total Knee Arthroplasty Design

This example focuses first on delineating several different COUs for a computational model, each of which has different amounts of model influence on the question of interest. Then, selected aspects of establishing credibility goals are reviewed, focusing on ways in which the sensitivity of the model and comparator to the system configuration and boundary conditions may be quantified. Finally, potential applications (applicability) of the V&V activities to the COU are discussed based on the extent to which the design family of interest is different from the design family used during the V&V activities.

**B-2.5.1 Background.** Many total knee arthroplasty (TKA) systems use a polyethylene tibial component that is locked into a metal tibial baseplate (see Figure B-2.5.1-1). During in vivo use, the tibial component could dissociate from the tibial baseplate if the locking mechanism between the two does not have sufficient strength to withstand physiological loading applied through the femoral component. In this example, the locking mechanism strength is evaluated by measuring liftoff distance of the tibial component from the tibial baseplate when subjected to physiological loading. A smaller liftoff distance, whether measured experimentally or predicted computationally, is thus indicative of a stronger locking mechanism.

Figure B-2.5.1-1 presents a schematic of a posterior stabilized TKA assembly, with the set of boundary conditions that are assumed for this example. In particular, the femoral component is assumed to load the tibial component spine from the anterior side (thus exerting a posteriorly directed force on the tibial component), resulting in anterior liftoff of the tibial component from the baseplate.