Contains Nonbinding Recommendations

Assessing the Credibility of Computational Modeling and Simulation in Medical Device Submissions

---

Guidance for Industry and Food and Drug Administration Staff

Document issued on November 17, 2023.

The draft of this document was issued on December 23, 2021.

For questions about this document, contact Office of Science and Engineering Laboratories (OSEL) by email at OSEL_CDRH@fda.hhs.gov or at (301)-796-2530, or Pras Pathmanathan at (301) 796-3490 or by email pras.pathmanathan@fda.hhs.gov.



U.S. Department of Health and Human Services
Food and Drug Administration
Center for Devices and Radiological Health

Preface

Public Comment

You may submit electronic comments and suggestions at any time for Agency consideration to https://www.regulations.gov. Submit written comments to the Dockets Management Staff, Food and Drug Administration, 5630 Fishers Lane, Room 1061, (HFA-305), Rockville, MD 20852-1740. Identify all comments with the docket number FDA-2021-D-0980. Comments may not be acted upon by the Agency until the document is next revised or updated.

Additional Copies

Additional copies are available from the Internet. You may also send an email request to CDRH-Guidance@fda.hhs.gov to receive a copy of the guidance. Please include the document number GUI01500056 and complete title of the guidance in the request.

Table of Contents

I. Introduction 4
II. Background 5
III. Scope 7
IV. Definitions 8
V. Generalized Framework for Assessing Credibility of Computational Modeling in a Regulatory Submission 10
VI. Key Concepts for Assessing Credibility of Computational Modeling in a Regulatory Submission 15
A. Preliminary steps 15
(1) Question of Interest 15
(2) Context of use (COU) 16
(3) Model risk 17
B. Credibility Evidence 19
(1) Code verification results 22
(2) Model calibration evidence 23
(3) Bench test validation results 23
(4) *In vivo* validation results 27
(5) Population-based validation results 28
(6) Emergent model behavior 29
(7) Model plausibility 29
(8) Calculation verification/UQ results using COU simulations 30
C. Credibility Factors and Credibility Goals 31
D. Adequacy Assessment 33
Appendix 1. Considerations for Each Credibility Evidence Category 36
Appendix 2. Reporting Recommendations for CM&S Credibility Assessment in Medical Device Submissions 40

Assessing the Credibility of Computational Modeling and Simulation in Medical Device Submissions

---

Guidance for Industry and Food and Drug Administration Staff

This guidance represents the current thinking of the Food and Drug Administration (FDA or Agency) on this topic. It does not establish any rights for any person and is not binding on FDA or the public. You can use an alternative approach if it satisfies the requirements of the applicable statutes and regulations. To discuss an alternative approach, contact the FDA staff or Office responsible for this guidance as listed on the title page.

I. Introduction

FDA has developed this guidance document to assist industry and FDA staff in assessing the *credibility* of computational modeling used to support medical device premarket submissions (i.e., Premarket Approval (PMA) Applications, Humanitarian Device Exemption (HDE) Applications, Investigational Device Exemption (IDE) Applications, Premarket Notifications (510(k)s), and De Novo classification requests) or qualification of Medical Device Development Tools (MDDTs). In the context of this guidance, credibility is defined as the trust in the predictive capability of a computational model. Computational models can be used in a variety of ways in medical device regulatory submissions, including to perform '*in silico*' device testing or to influence algorithms within software embedded in a device. Regulatory submissions often lack a clear rationale for why models can be considered credible for the context of use (COU). This guidance provides a general risk-informed framework that can be used in the credibility assessment of computational modeling and simulation (CM&S) used in medical device regulatory submissions. For the purposes of this guidance, CM&S refers to first principles-based (e.g., physics-based or mechanistic) computational models, and not statistical or data-driven (e.g., machine learning or artificial intelligence-based) models. This guidance is intended to help improve the consistency and transparency of the review of CM&S, to increase confidence in the use of CM&S in regulatory submissions, and to facilitate improved interpretation of CM&S credibility evidence submitted in regulatory submissions reviewed by FDA staff. Throughout this

guidance, the terms "FDA," "the Agency," "we," and "us" refer to the Food and Drug Administration and the terms "you" and "yours" refer to medical device manufacturers.

For the current edition of the FDA-recognized consensus standard(s) referenced in this document, see the FDA Recognized Consensus Standards Database.[1] For more information regarding use of consensus standards in regulatory submissions, please refer to the FDA guidance titled "Appropriate Use of Voluntary Consensus Standards in Premarket Submissions for Medical Devices."[2]

In general, FDA's guidance documents do not establish legally enforceable responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word *should* in Agency guidances means that something is suggested or recommended, but not required.

II. Background

The use of CM&S (also referred to as *in silico* methods) in regulatory submissions is well-established and increasing.[3] CM&S of medical devices can streamline development and reduce burdens associated with premarket device evaluation. It can also reveal important information not available from traditional *in vivo* or *in vitro* assessments, such as serious and unexpected adverse events that are undetectable within a study sample but occur frequently enough within the intended population to be of concern. As interest in medical device-related CM&S grows, it will be important to both monitor current usage and identify areas where CM&S might be more broadly leveraged to enhance public health. Use of CM&S to support regulatory submissions necessitates the development of processes and approaches that promote consistency and transparency in the way CM&S is conducted and reviewed.

There are several ways that CM&S can potentially be used to support a regulatory submission, including but not limited to:

1. ***In silico* device testing.** Computational models that simulate medical devices can be used to generate information supporting device safety and/or effectiveness (e.g., *in silico* durability assessment of an implantable stent). Computational models of the device can also be coupled to computational patient models to simulate device performance under representative *in vivo* conditions (e.g., computational electromagnetic models to predict energy absorption of metallic implants). Another possibility is that the physical device itself is tested on an *in silico* patient model, for example hardware-in-the-loop testing of a physiological closed loop control device, where the therapy actuated by the controller is

[1] https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfStandards/search.cfm
[2] https://www.fda.gov/regulatory-information/search-fda-guidance-documents/appropriate-use-voluntary-consensus-standards-premarket-submissions-medical-devices.
[3] Morrison T, Pathmanathan P, Adwan M and Margerrison E. Advancing Regulatory Science With Computational Modeling for Medical Devices at the FDA's Office of Science and Engineering Laboratories. *Frontiers in Medicine,* vol. 5, p. 241, 2018.

converted into an input to the patient model, and the patient model response is converted into a signal passed back to the controller.[4]

2. **CM&S used within medical device software.** Computational modeling may be implemented as device software functions,[5] for example, device software functions that use patient data as inputs to a physics-based computational model to estimate clinical biomarkers such as fractional flow reserve, or device software functions that simulate patient response during surgery for preoperative planning.
3. ***In silico* clinical trials.** *In silico* clinical trials are an emerging application of CM&S where device safety and/or effectiveness is evaluated using a 'virtual cohort' of simulated patients with anatomical and physiological variability representing the indicated patient population. *In silico* clinical trials have a range of possible applications, including but not limited to: augmenting or reducing the size of a real world clinical trial,[6] providing improved inclusion-exclusion criteria, or investigating a device safety concern for which a real world clinical trial would be unethical.
4. **CM&S-based qualified tools**. CM&S-based tools for developing or evaluating a medical device can be submitted to CDRH as a proposal to be considered for the Medical Device Development Tools (MDDT) Program[7] by the FDA as a non-clinical assessment model (NAM) for predicting device safety, effectiveness, or performance (refer to FDA's guidance titled "Qualification of Medical Device Development Tools"[8]).

In all cases, there is a need to demonstrate that the CM&S is credible. For *in silico* device testing and *in silico* clinical trials, final simulation results should be submitted to FDA with supporting credibility evidence so that FDA can assess the credibility of those simulation results. For CM&S in medical device software and MDDTs, example simulation results should be submitted to FDA with supporting credibility evidence so that FDA can assess if future simulations (to be performed post-market or post-tool qualification) are expected to be credible.

Methodologies for model credibility assessment have been established in the scientific literature[9,10] and continue to evolve. Demonstrating model credibility involves various activities that include verification, validation, uncertainty quantification, applicability analysis, as well as adequacy assessment (see Section IV for definitions). The FDA-recognized standard American Society of Mechanical Engineers (ASME) V&V 40 *Assessing Credibility of Computational*

---

[4] Parvinian B, Scully C, Wiyor H, Kumar A, and Weininger S. Regulatory Considerations for Physiological Closed-Loop Controlled Medical Devices Used for Automated Critical Care: Food and Drug Administration Workshop Discussion Topics. *Anesth Analg.,* vol. 126(6), p. 1, 2018.

[5] A device software function is a software function that meets the definition of device in 201(h) of the Federal Food, Drug, and Cosmetic Act. See also https://www.fda.gov/medical-devices/digital-health-center-excellence/device-software-functions-including-mobile-medical-applications

[6] Haddad T, Himes A, Thompson L, Irony T, Nair R, and MDIC Working Group Participants. Incorporation of stochastic engineering models as prior information in Bayesian medical device trials, *J. Biopharm Stat,* vol. 27(6),s pp. 1089-1103, 2017.

[7] https://www.fda.gov/medical-devices/science-and-research-medical-devices/medical-device-development-tools-mddt

[8] https://www.fda.gov/regulatory-information/search-fda-guidance-documents/qualification-medical-device-development-tools

[9] Oberkampf WL and Roy CJ. *Verification and Validation in Scientific Computing*. Cambridge University Press, 2010.

[10] Roache PJ, *Fundamentals of Verification and Validation.* Hermosa Publishers, 2009.

Modeling through Verification and Validation: Application to Medical Devices provides a risk-informed framework for assessing verification, validation, and uncertainty quantification (VVUQ) activities for computational modeling of medical devices. However, most of the validation activities defined in ASME V&V 40 assume the ability to perform well-controlled bench testing to provide data against which simulations' results are evaluated, henceforth referred to as 'traditional validation evidence.' The possibility of using other, non-traditional sources of evidence (e.g., clinical studies, robust model calibration results, or population-based validation results), which may be less controlled but closer to the model context of use, is not explicitly covered in ASME V&V 40-2018, although recent work has considered how to apply ASME V&V 40 to patient-specific computational models.[11] This guidance uses key concepts of ASME V&V 40-2018 but provides a more general framework for demonstrating CM&S credibility in medical device regulatory submissions that incorporate different categories of credibility evidence.

III. Scope

The purpose of this guidance document is to provide a general risk-informed framework for assessing CM&S credibility in medical device regulatory submissions that incorporate both traditional validation evidence and/or other types of supporting data. This guidance document is applicable to first principles-based models (e.g., physics-based or mechanistic models), such as models commonly used in electromagnetics, optics, fluid dynamics, heat and mass transfer, solid mechanics, acoustics, and ultrasonics, as well as mechanistic models of physiological processes. This guidance is not intended to apply to standalone statistical or data-driven models such as standalone regression, machine learning or artificial intelligence-based models. We recognize that there is no clear delineation between first principles and statistical/data-driven models, and that hybrid models using both methods are possible. For hybrid models, this guidance is intended to apply to the first-principles model aspects of the hybrid model only; additional considerations for evaluating statistical/data-driven model aspects are not addressed in this guidance. For information on appropriate evidence to submit for a statistical/data-driven model, including machine learning or artificial-intelligence-based models, we recommend manufacturers seek feedback on their specific device through the Q-submission process (refer to FDA's guidance titled "Requests for Feedback and Meetings for Medical Device Submissions: The Q-Submission Program."[12]) Models that do not involve any simulation, such as purely anatomical models, are not in scope of this guidance.

This guidance document provides recommendations for both planning and reporting model credibility assessment activities. This guidance document does not address methodologies for how to perform modeling studies or technical details for how to gather evidence to support credibility assessment, nor does it provide recommendations concerning the specific level of credibility needed to support regulatory submissions. This guidance is not intended to provide a

[11] Galappaththige S, Gray R, Costa C, Niederer S, Pathmanathan P. Credibility Assessment of Patient-Specific Computational Modeling using Patient-Specific Cardiac Models as an Exemplar, *PLOS Computational Biology*, 2022.

[12] https://www.fda.gov/regulatory-information/search-fda-guidance-documents/requests-feedback-and-meetings-medical-device-submissions-q-submission-program

comprehensive checklist for all CM&S information to support a regulatory submission. Instead, this guidance provides a general framework for how to assess CM&S credibility to support regulatory submissions, and identifies factors a manufacturer should consider when submitting CM&S credibility evidence We recommend that manufacturers seek feedback on their specific use of CM&S through the Q-submission process. Where applicable, other device-specific guidance documents and FDA-recognized standards that include CM&S recommendations may be used in combination with this guidance document. Also, while the general framework is expected to be applicable to *in silico* clinical trials, this is an emerging methodology for which best practices are still being developed, and this guidance does not provide specific recommendations for generating virtual cohorts or executing an *in silico* clinical trial.

IV. Definitions

The definitions listed here are for the purposes of this guidance document and are intended for use in the context of assessing CM&S credibility.

Adequacy assessment: for a given context of use (COU), the process of evaluating the credibility evidence in support of a computational model, together with any other relevant information, possibly including results from the COU simulations, and making a determination on whether the evidence is sufficient considering the model risk. See also prospective adequacy assessment and post-study adequacy assessment.

Applicability: "the relevance of the validation activities to support the use of the computational model for a context of use"[13]

Calculation verification (also called solution verification)**:** "the process of determining the solution accuracy of a calculation"[14]

Code verification: "the process of identifying errors in the numerical algorithms of a computer code"[15]

Comparator: the test data that are used for validation, which may include data from bench testing,[16] *in vivo* studies, or other empirical data[17]

[13] Reprinted with permission of The American Society of Mechanical Engineers from ASME V&V 40-2018 *Assessing Credibility of Computational Modeling through Verification and Validation: Application to Medical Devices*, copyright ASME, Two Park Avenue New York, NY 10016-5990. All rights reserved. No further copies can be made without written permission from ASME. Permission is for this edition only. A copy of the complete standard may be obtained from ASME, www.asme.org.
[14] Reprinted with permission from ASME V&V 40-2018.
[15] Reprinted with permission from ASME V&V 40-2018.
[16] https://www.fda.gov/regulatory-information/search-fda-guidance-documents/recommended-content-and-format-non-clinical-bench-performance-testing-information-premarket
[17] Adapted with permission from ASME V&V 40-2018.

Computational model: "the numerical implementation of the mathematical model performed by means of a computer"[18]

Context of use (COU): "a statement that defines the specific role and scope of the computational model used to address the question of interest"[19]

COU simulations: simulations performed 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"[20]

Credibility evidence: any evidence that could support the credibility of a computational model

Credibility factors: "elements of the process used to establish the credibility of the computational model for a COU"[21]

Decision consequence: the significance of an adverse outcome resulting from an incorrect decision concerning the question of interest[22]

Mathematical model: "the mathematical equations, boundary conditions, initial conditions, and modeling data needed to describe a conceptual model"[23]

Model influence: the contribution of the computational model relative to other contributing evidence in addressing the question of interest (e.g., data from bench testing)[24]

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"[25]

Post-study adequacy assessment: adequacy assessment performed after executing planned credibility assessment activities, and potentially also after conducting the COU simulations, using results from these activities and any other relevant information

Prospective adequacy assessment: adequacy assessment performed before executing planned credibility assessment activities, using selected credibility goals and any other relevant information

[18] Reprinted with permission from ASME V&V 40-2018.
[19] Reprinted with permission from ASME V&V 40-2018.
[20] Reprinted with permission from ASME V&V 40-2018.
[21] Reprinted with permission from ASME V&V 40-2018.
[22] Adapted with permission from ASME V&V 40-2018.
[23] Reprinted with permission from ASME V&V 40-2018.
[24] Adapted with permission from ASME V&V 40-2018.
[25] Reprinted with permission from ASME V&V 40-2018.

Quantity of interest: "the calculated or measured result from a computational model or comparator, respectively"[26]

Question of interest: "the specific question, decision, or concern that is being addressed"[27]

Uncertainty quantification: the process of generating and applying mathematical models to provide a measure of uncertainty in the empirical data or simulation results[28]

Solution verification: see *calculation verification*

Validation: "the process of determining the degree to which a model or a simulation is an accurate representation of the real world"[29]

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."[30] Code verification and calculation verification are two elements of verification.

Note that the terms 'verification' and 'validation' have a variety of meanings in the context of medical device regulation. The above definitions specifically refer to verification and validation of a computational model.

V. Generalized Framework for Assessing Credibility of Computational Modeling in a Regulatory Submission

FDA recommends the following process when developing and assessing the credibility of computational modeling used in a medical device regulatory submission. Detailed information on the key concepts in the framework below are provided in subsequent sections. See Figure 1 for an illustration of the framework using a hypothetical example.

1. Describe the **question(s) of interest** to be addressed in the regulatory submission that will be informed by the computational model. See Section VI.A.(1) for details.
2. Define the **context of use** (COU) of the computational model. See Section VI.A.(2) for details.
3. Determine the **model risk**. See Section VI.A.(3) for details.

[26] Reprinted with permission from ASME V&V 40-2018.
[27] Reprinted with permission from ASME V&V 40-2018.
[28] Reprinted with permission of The American Society of Mechanical Engineers from ASME VVUQ1-2022 *Verification, Validation, and Uncertainty Quantification Terminology in Computational Modeling and Simulation*, copyright ASME, Two Park Avenue New York, NY 10016-5990. All rights reserved. No further copies can be made without written permission from ASME. Permission is for this edition only. A copy of the complete standard may be obtained from ASME, www.asme.org.
[29] Reprinted with permission from ASME V&V 40-2018.
[30] Reprinted with permission from ASME V&V 40-2018.

4. Identify and categorize the **credibility evidence**, either previously generated or planned, which will support credibility of the computational model for the COU. See Section VI.B for a categorization of different types of credibility evidence.
5. Define **credibility factors** for the proposed credibility evidence. For prospectively planned activities, set prospective credibility goals for each credibility factor, with a plan to achieve these goals. For previously generated data, assess the credibility levels achieved. See Section VI.C for a discussion of credibility factors, levels and goals.
6. Perform **prospective adequacy assessment**: if the credibility goals are achieved, will the credibility evidence be sufficient to support using the model for the COU given the risk assessment? See Section VI.D for a discussion of adequacy assessment.
    a. If YES: continue to Step 7. Before proceeding, however, you may wish to utilize the Q-submission process (refer to FDA's guidance titled "Requests for Feedback and Meetings for Medical Device Submissions: The Q-Submission Program" [31]) to receive FDA feedback on the computational model, proposed credibility evidence, plan for generating this evidence, and prospective adequacy assessment. See Appendix 2.
    b. If NO: you may need to modify the model, reduce the model influence, modify the COU or revise the plan to generate credibility evidence. See ASME V&V 40 for a discussion on options. If any changes are made at this stage, go back to Step 2.
7. Generate the credibility evidence by executing the proposed study(ies) and/or analyzing previously generated data.
8. Determine if credibility goals were met and perform **post-study adequacy assessment**: does the credibility evidence support using the model for the COU given the risk assessment? See Section VI.D for a discussion of adequacy assessment.
    a. If YES: continue to Step 9.
    b. If NO: you may wish to modify the model, reduce the model influence, modify the COU or collect additional evidence. See ASME V&V 40 for a more detailed discussion of the various options. If any changes are made at this stage, go back to Step 2.
9. Prepare a CM&S credibility assessment report for inclusion in the regulatory submission. See Appendix 2 for **reporting recommendations.**

FDA recommends this generalized framework, but you can choose to use an alternative approach to demonstrate the credibility of your computational model. If an alternative approach is used, we recommend that you clearly identify the model's COU within the regulatory submission and provide a detailed rationale for why the model can be considered credible for its specific COU. If an alternative approach is planned, we recommend using the Q-submission process to receive FDA feedback on the planned approach and activities, as outlined in Step 6a above.

Relationship between this framework and ASME V&V 40: The framework above is intended to be consistent with ASME V&V 40. Table 1 describes the relationship between each step of the framework and ASME V&V 40. If you plan to perform model validation using well-

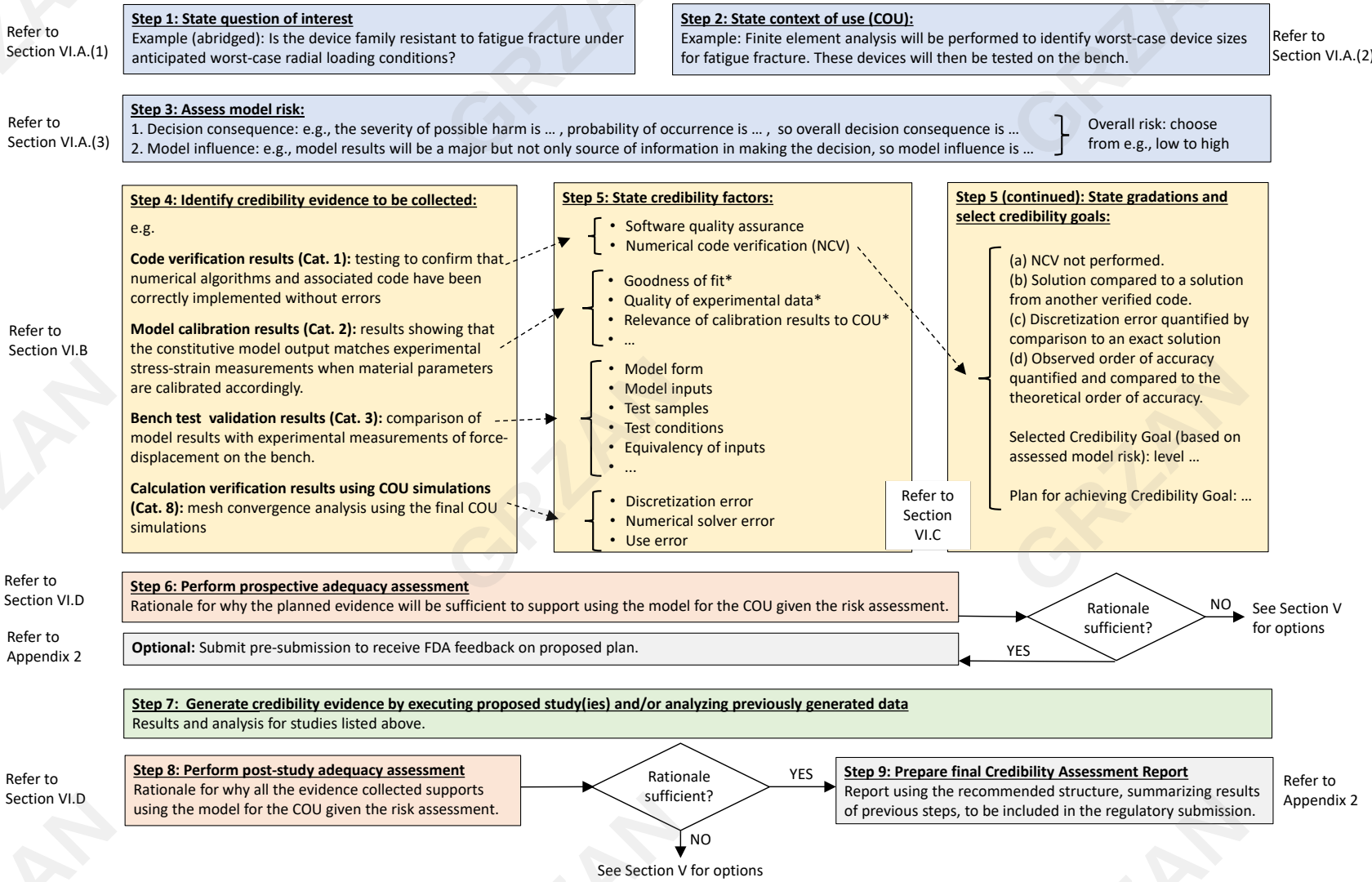
[31] https://www.fda.gov/regulatory-information/search-fda-guidance-documents/requests-feedback-and-meetings-medical-device-submissions-q-submission-program

controlled bench test or *in vivo* experimental data, this guidance framework is fully consistent with ASME V&V 40 but includes additional recommendations in Sections VI.A and VI.D. For cases not within the scope of ASME V&V 40 (i.e., use of other categories of credibility evidence or multiple sources of validation evidence), the framework enables systematic credibility assessment based on the approach of ASME V&V 40.

Table 1: Relationship between guidance framework and ASME V&V 40.

Step	Topic	Relevant section of ASME V&V 40-2018	Comments
1	Question of interest	Section 2.4	Section VI.A.(1) provides additional recommendations for medical device regulatory submissions
2	Context of use	Section 3	Section VI.A.(2) provides additional recommendations for medical device regulatory submissions
3	Model risk	Section 4	Section VI.A.(3) provides additional recommendations for medical device regulatory submissions
4	Credibility evidence	N/A	The term "credibility evidence" does not appear in ASME V&V 40. The categorization of credibility evidence in Section VI.B is unique to this guidance. ASME V&V 40 effectively assumes the following will be generated: code verification results (Cat. 1) with either bench test validation (Cat. 3) or *in vivo* validation (Cat. 4).
5	Credibility factors, gradations and goals	Section 5	ASME V&V 40 defines credibility factors and provides example gradations. These may be used for the cases described in the row above. For other cases (other categories of credibility evidence or multiple sources of validation evidence), users should define appropriate credibility factors. See Section VI.C and Appendix 1 for recommendations.
6	Prospective adequacy assessment	Section 6	The term "adequacy assessment" is not explicitly used in ASME V&V 40. Prospective adequacy assessment, as defined in this guidance, overlaps with Section 6 of ASME V&V 40. Recommendations for prospective adequacy assessment are provided in Section VI.D
7	Generate credibility evidence	N/A	ASME V&V 40 does not address how to perform credibility activities but similarly incorporates evidence generation as part of the overall credibility assessment framework (e.g., "Execute plan" in Figure 2.4-1 of ASME V&V 40-2018).
8	Post-study adequacy assessment	Section 7	The term "adequacy assessment" is not explicitly used in ASME V&V 40. Post-study adequacy assessment, as defined in this guidance, overlaps with Section 7 of ASME V&V 40. Detailed recommendations for post-study adequacy assessment are provided in Section VI.D
9	Credibility assessment report	Section 8	See Appendix 2 for specific recommendations for information to include in a medical device regulatory submission

Figure 1: Overview of generalized framework for assessing model credibility, with an example for each step. Asterisks (*) indicate credibility factors that are defined by the user in this hypothetical example, as they are not defined in ASME V&V 40 'Cat.' (in Step 4) denotes credibility evidence category, as discussed in Section VI.B. Blue boxes are initial steps, yellow boxes are credibility assessment planning steps, red boxes are adequacy assessment steps, grey boxes are steps related to FDA interaction, and the green box is study execution.



VI. Key Concepts for Assessing Credibility of Computational Modeling in a Regulatory Submission

This section describes and discusses the key concepts used in the framework provided above in Section V.

A. Preliminary steps

(1) Question of Interest

Step 1 in the framework is to describe the **question(s) of interest** to be addressed in the regulatory submission that will be informed by the computational model. We recommend describing the question of interest following the recommendations of ASME V&V 40 together with the clarification points below and specific recommendations for medical device regulatory submissions.

The question of interest concerns the decision to be made with input from the computational model and potentially other sources of information. The question of interest should not be confined to the computational model, nor should it be about the computational model. We recommend that the scope of the question of interest describe the question, decision, or concern that is being addressed using the computational model and potentially other sources of information, but nothing more. Therefore, you should avoid overly broad questions of interest such as, "Is the device safe and effective?" For example, a possible question of interest regarding device durability could be, "Is the device resistant to fatigue fracture under anticipated worst-case radial loading conditions?", which might be addressed using a combination of computational modeling and bench testing. To assist in evaluating the decision consequence when assessing the model risk in Section VI.A.(3), it can be helpful to formulate the question of interest in terms of the decision that is to be made and the stakeholder(s) making the decision.

For models used for *in silico* device testing or *in silico* clinical trials:

- The question of interest should describe the specific question, decision or concern being addressed about the device, such as in the device durability example stated in the preceding paragraph and in Figure 1.

For models used within device software:

- The question of interest should cover the specific device functionality(ies) that use the model predictions. For example, for a device that performs a simulation of a patient as part of a diagnostic function, the question of interest may be posed around the clinical decision that is to be made, such as whether or not to treat a patient or diagnose the presence of a disease condition.

For models submitted for MDDT qualification:

- The question of interest should describe the specific question, decision, or concern about the range of devices relevant to the proposed MDDT. For example, "For an active implantable medical device, what is the *in vivo* deposited power during a 1.5T magnetic resonance (MR) scanning procedure and is it below an acceptable threshold?"

(2) Context of use (COU)

Step 2 of the framework is to define the **context of use** (COU) of the computational model. We recommend defining the context of use following the recommendations of ASME V&V 40 together with the clarification points below and specific recommendations for medical device regulatory submissions.

The COU statement should include a detailed description of what will be modeled and how model outputs will be used to answer the question of interest. The COU should also include a statement on whether other information (e.g., bench testing, animal[32] or clinical studies) will be used in conjunction with the model results to answer the question of interest. For example, a possible COU regarding device durability could be summarized as "Combine computational modeling predictions and empirical fatigue testing observations to estimate device fatigue safety factors under anticipated worst-case radial loading conditions," with additional details provided to describe the type of modeling used, key model inputs and outputs, and the specific approach used to combine model predictions with experimental data to answer the question of interest. Since many models have a range of possible uses, it is important to note that the COU describes how the model will be used to answer the question of interest, and may be narrower than the overall model capability.

For models used for *in silico* device testing or *in silico* clinical trials:

- The COU should describe how the model will be used in a simulation study to address the question of interest. Note that in this case, the COU differs from the indications for use or intended use of the device, although the COU may involve addressing a safety or effectiveness question related to the device indications for use or intended use.

For models used within device software:

- The COU should describe how the model will be used within the device. In this case, the COU may be related to the intended use of the device, or a subset thereof, depending on how the device uses the simulation results.

For models submitted for MDDT qualification:

- The model COU is expected to include the MDDT COU information (refer to Section IV.A of FDA's guidance titled "Qualification of Medical Device Development Tools"[33]).

[32] FDA supports the principles of the "3Rs" to replace, reduce, and/or refine animal use in testing, when feasible. We encourage manufacturers to consult with FDA if they wish to use a non-animal testing method that they believe is suitable, adequate, validated, and feasible. We will consider if a proposed alternative method could be assessed for equivalency to an animal test method.

[33] https://www.fda.gov/regulatory-information/search-fda-guidance-documents/qualification-medical-device-development-tools

(3) Model risk

Step 3 of the framework is to determine the model risk. Model risk is assessed because the level of credibility of a model should be commensurate to the risk associated with using the model to address the question of interest. We recommend assessing model risk following ASME V&V 40, which considers model risk as a combination of two factors, **model influence** and **decision consequence**. Below are clarification points and specific recommendations for medical device regulatory submissions.

Model risk should be interpreted as the risk associated with using the model to address the specific question of interest, not risk intrinsic to the model. Decision consequence is generally risk as defined by ISO 14971 *Medical devices — Application of risk management to medical devices*, related to the question of interest. Therefore, model risk can be viewed as risk related to the question of interest, weighted by the influence of the computational model to address the question of interest.

Model influence is the contribution of the computational model relative to other contributing evidence (e.g., bench test results, animal or clinical study results) in addressing the question of interest. For example, evaluating model influence for the aforementioned device durability COU might consider how much influence CM&S results have on the fatigue resistance decision relative to the bench fatigue test results.

Decision consequence is the significance of an adverse outcome resulting from an incorrect decision concerning the question of interest. It is important to note that the decision consequence is the potential outcome of the overall decision that is to be made by answering the question of interest, outside of the scope of the computational model and irrespective of how modeling is used. That is, decision consequence should consider the question of interest, but should not consider the COU of the model. In regulatory submissions, decision consequence will typically involve consideration of potential patient harm. For example, when evaluating decision consequence for the aforementioned device durability example, you should consider the potential patient harm that could result if the implanted device fractures.

In general, we recommend assessing decision consequence by considering both the potential severity of harm and the probability of occurrence of harm, as mentioned in ASME V&V 40. Neglecting probability of occurrence may lead to over-estimating overall model risk and therefore may seem to warrant a higher level of credibility than needed. We recommend following an appropriate risk management procedure (e.g., see ISO 14971 and ISO/TR 24971[34]). The risk management procedure used should consider any specific hazards that are related to the question of interest and then identify any possible hazardous situations and the resultant harm that may occur. Reports of adverse events for the same or similar device types can be helpful in identifying potential hazards and harms, and estimating their associated rates of occurrence. The overall decision consequence should be assessed by considering all potential patient harms that may occur due to an incorrect decision, accounting for any risk mitigation procedures in place.

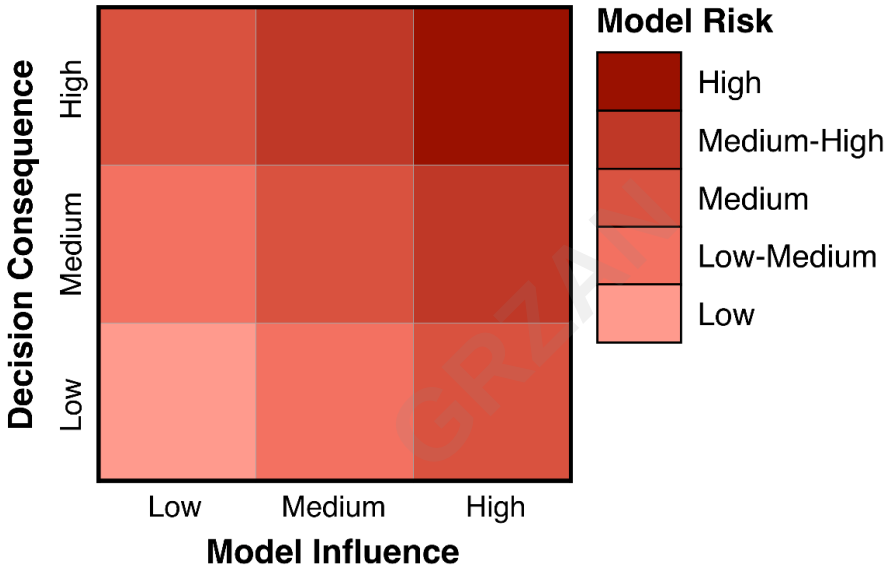
[34] ISO/TR 24971 *Medical devices — Guidance on the application of ISO 14971.*

We acknowledge that for some cases, assessing probability of occurrence may need estimation or subject matter expertise (e.g., for some new devices). See Section V.A.2 of FDA's guidance titled Factors to Consider When Making Benefit-Risk Determinations for Medical Device Investigational Device Exemptions[35] for approaches to estimate probability of occurrence in these situations.

We note that, while the overall risk of a medical device is a major determinant of the device classification, decision consequence should be based on the specific question of interest and not on the specific device class. Therefore, although the overall clinical risk is greater for a class III device than for a class II device, the decision consequence associated with a specific question of interest in a 510(k) submission could be the same or even greater than the decision consequence associated with another question of interest in a PMA application, depending on the specific question of interest. Accordingly, the decision consequence should be solely determined by considering the specific question of interest. For CM&S used to support an IDE application, decision consequence should generally consider the potential harm to trial participants due to making an incorrect decision concerning the question of interest, taking into account the proposed study protocol and including any risk mitigation procedures in place.

Following ASME V&V 40, we recommend using a scheme such as illustrated in Figure 2 to assess model risk considering the combined impact of decision consequence and model influence.

Figure 2: Possible scheme for assessing model risk considering the combined impact of model influence and decision consequence. Alternative schemes may be used instead, for example using a 5x5 or 5x4 grid instead of 3x3.



For models used for *in silico* device testing or *in silico* clinical trials:

[35] https://www.fda.gov/regulatory-information/search-fda-guidance-documents/factors-consider-when-making-benefit-risk-determinations-medical-device-investigational-device

- Model influence will be dependent on whether other information (e.g., bench or animal test results) is also provided in the regulatory submission to address the question of interest.
- When assessing decision consequence, you should consider device hazards that are related to the specific device safety or effectiveness concern that is being addressed, as stated in the question of interest.

For models used within device software:

- Model influence will be dependent on whether other information (e.g., additional direct patient measurements, clinical assessments) will be used in answering the question of interest. If the device takes action based solely on simulation results, model influence will be the highest level. If the simulation results are provided to the clinician to inform a decision, model influence will be dependent on other information available and on the specific language proposed in the labeling for the device. When determining model influence for a device that provides a simulation-based recommendation to a clinician, which is intended to be used in conjunction with other medical information to make a clinical decision, we recommend you examine if there is reasonably foreseeable misuse[36] related to the degree clinicians may rely on the device output without considering additional clinical information that may be available. For example, for a device that provides a simulation-based recommendation to a clinician for adjunctive use, model influence should account for possible misuse where the clinician relies on the model information to a greater degree than intended in the labeling. A model influence of 'zero' or 'negligible' should be well-justified when proposed.
- When assessing decision consequence, device hazards to be considered should be those related to the specific device functionality that the model is used for, as stated in the question of interest.

For models submitted for MDDT qualification:

- If the MDDT is a computational model only, model influence is expected to be the highest level.
- Decision consequence should be assessed based on the potential risk to patients should the tool, when used as specified in the MDDT COU, provide inaccurate information for the question of interest.

B. Credibility Evidence

Step 4 of the framework is to identify and categorize the **credibility evidence,** either previously generated or planned, which would support credibility of the computational model for the COU.

Not all evidence that could potentially support the use of a computational model in medical device regulatory submissions comes from traditional validation activities. Because of this, we

[36] See ISO 14971 *Medical devices — Application of risk management to medical devices* for definition of reasonably foreseeable misuse.

adopt the more general term of "credibility evidence," which is any evidence that could support the credibility of a computational model. The evidence categories defined below represent results from different VVUQ activities. Definitions for each of these activities were provided in Section IV; some clarification points are provided below.

Verification is focused on the software implementation of a numerical algorithm to solve the underlying mathematical model. It can be broken down into code verification and calculation verification. **Code verification** is performed to confirm that numerical algorithms and associated code have been correctly implemented without errors that affect numerical accuracy and involves activities such as software quality assurance and numerical code verification; see ASME V&V 40 for details. The aim of **calculation verification** is to estimate the specific numerical error in quantities of interest arising from, for example, the chosen spatial discretization. Calculation verification may be performed any time a simulation is run. For example, calculation verification can be performed using the validation simulations, that is, using model input values corresponding to the validation experiment(s). Alternatively, calculation verification can be performed using the COU simulations, that is, using the COU model inputs.

Validation involves comparison of model predictions with real world observations, referred to as the comparator. In this guidance, validation is interpreted as comparison against data that is independent of the data used to create the model. Therefore, model calibration, where parameters are tuned or optimized so that the model output matches the real-world observations, is not considered validation. Additionally, comparison of model predictions against predictions from a different model is not considered validation. A related activity to validation is **applicability** assessment**,** which is assessment of the relevance of the validation activities to the COU. Differences between how the model is validated and how it is used in the COU may limit the relevance of the validation activities to the COU.

Uncertainty quantification (UQ) involves estimating the uncertainty in model outputs. Model output uncertainty can arise from a range of factors, including uncertainties in model inputs or uncertainty in model form (see ASME V&V 40 for more information on model inputs and model form). Input UQ is related to sensitivity analysis (SA), which aims to estimate and potentially rank the influence of model inputs on model outputs and can be performed locally around fixed input values or globally using input ranges or distributions. SA can support UQ, for example by reducing the number of inputs with which to perform UQ. However, ultimately it is UQ results – that is, estimation of the uncertainty in model outputs – that support model credibility. As with code verification, UQ and SA can be performed using validation simulations, COU simulations, or both.

In Table 2 below, eight distinct categories of credibility evidence are provided. The objective of defining these categories is to provide a common framework to characterize the available evidence to support a computational model. It is not to characterize the quality or level of rigor of the evidence; the ordering of the categories does not reflect the strength of the evidence. This categorization is not intended to be exhaustive. In some cases, there may be a need to define new categories if the credibility evidence does not fit into any of the following categories. For many computational models, there will likely be evidence from multiple categories that support model credibility, all of which can be included in a regulatory submission.

Following Table 2, each category is discussed in more detail, with key distinguishing features and examples. Specific considerations for each category are also provided in Appendix 1.

Table 2: Eight categories of credibility evidence. Categories 1, 3 and 4 are explicitly within the scope of ASME V&V 40.

	Category	**Definition**
1	Code verification results	Results showing that a computational model implemented in software is an accurate implementation of the underlying mathematical model.
2	Model calibration evidence	Comparison of model results with the same data used to calibrate model parameters.
3	Bench test validation results	Validation results using a bench test comparator. May be supported by calculation verification and/or UQ results using the validation conditions.
4	*In vivo* validation results	Same as previous category except using *in vivo* data as the comparator.
5	Population-based validation results	Comparison of population-level data between model predictions and a clinical data set. No individual-level comparisons are made.
6	Emergent model behavior	Evidence showing that the model reproduces phenomena that are known to occur in the system at the specified conditions but were not pre-specified or explicitly modeled by the governing equations.
7	Model plausibility evidence	Rationale supporting the choice of governing equations, model assumptions, and/or input parameters only.
8	Calculation verification /UQ results using COU simulations	Calculation verification and/or UQ results obtained using the COU simulations, that is, the simulations performed to answer the question of interest

What types of credibility evidence should be included in a regulatory submission? In accordance with ASME V&V 40, the demonstrated credibility of a computational model should be commensurate with the risk associated with using the model. We recognize that the ability to generate credibility evidence may depend upon multiple factors including but not limited to the type of the model, the maturity of the modeling field, and the ability to perform validation. Therefore, this guidance document does not prescribe the specific types of credibility evidence that should be included in a regulatory submission. However, you should consider providing evidence for each of the following general groups since these evaluate different aspects of the model:

- code verification (Category 1);
- calculation verification (Categories 3, 4 or 8); and
- validation (Categories 3, 4 or 5) or other evidence pertaining to the model's ability to reproduce real-world behavior (Categories 2, 6 or 7).

You can also submit multiple types of evidence within each group (e.g., submitting both bench and *in vivo* validation (Categories 3 and 4) if it is appropriate for overall testing of the model and/or it increases the overall credibility in the model. If you have questions on your planned

credibility evidence for your specific model, we recommend that you use the Q-submission process to obtain feedback.

Examples:

- *In silico* device testing:
  - A model of a device that will be used to reproduce a bench test could be supported by: code verification results (Category 1), bench test validation results (Category 3), and calculation verification results (Category 3 or 8).
- Models used within device software:
  - A patient-specific modeling algorithm implemented in a medical device could be supported by: code verification results (Category 1), *in vivo* validation results (Category 4), and calculation verification results (Category 4).
- *In silico* clinical trials:
  - An *in silico* clinical trial where a device safety/effectiveness question is addressed using a virtual cohort of patient models, generated by sampling parameter values across the patient population, could be supported by: code verification results (Category 1), bench test validation results (to validate the device model; Category 3), *in vivo* validation results (to validate the baseline patient model; Category 4), calculation verification results (Category 3, 4 or 8), model plausibility evidence (to support the sampled parameters; Category 7); and population-based validation results (Category 5).

We emphasize that these are examples only, and appropriate evidence will depend on multiple factors as discussed in the preceding paragraph.

(1) Code verification results

Code verification results provide evidence demonstrating that a computational model implemented in software is an accurate implementation of the underlying mathematical model and associated numerical algorithms. Code verification is important to demonstrate that there are no bugs in the software that affect simulation numerical accuracy.[37] Comparison of model predictions against real-world observations is not part of code verification and is addressed separately by validation activities.

Example:

- For solid mechanics, fluid dynamics, heat transfer, electromagnetism, and other domains involving partial differential equations: results comparing the computational model against analytical solutions (e.g., generated using the method of manufactured solutions[38]), including confirmation that the error converges to zero at the expected convergence rate as spatial and temporal discretization size are decreased.

[37] Salari K and Knupp P. Code verification by the method of manufactured solutions (No. SAND2000-1444), Sandia National Lab, 2000.

[38] Aycock KI, Rebelo N and Craven BA. Method of manufactured solutions code verification of elastostatic solid mechanics problems in a commercial finite element solver. *Computers & Structures,* vol. 229, p. 106175, 2020.

(2) Model calibration evidence

Model calibration evidence is the comparison of model results with the same data used to calibrate model parameters. The evidence is an assessment of the "goodness of fit" of simulation results using calibrated model parameters. This is *not* validation evidence because it is not testing of the final model against data independent of model development; instead, model parameters are calibrated (whether optimized or manually tuned) to minimize the discrepancy between model results and data. Calibration evidence is weak in comparison to validation evidence. Nevertheless, robust model calibration evidence can still support model credibility. When the same amount of data are used, this type of evidence is stronger if complex behavior is reproduced after calibrating a small number of parameters in a first principles model. This type of evidence is weaker if the governing equations were chosen solely based on the data rather than underlying principles, or if many parameters were calibrated.

Calibration evidence could be generated for the overall model or for sub-models within the overall model; examples of both are provided below. When the overall model needs calibration of some of its parameters, the calibration results could provide relevant credibility evidence, generally supplementary to separate validation of the overall model. When a sub-model needs calibration to determine the value of sub-model parameters, the calibration results can be important for justifying use of those parameter values, and to provide confidence in the predictions when sub-model dependent variables (e.g., strains) will be extrapolated past values used in validation simulations.

Examples of overall model calibration:

- In physiological modeling, demonstrating that a patient-specific model of a patient's heart closely matches the patient's clinically measured pressure-volume (P-V) loop, after tissue parameters have been calibrated based on the same P-V loop data.
- In heat transfer modeling, demonstrating that the first principles-based model accurately reproduces spatio-temporal *in vivo* tissue heating in different tissues after calibrating the blood-tissue heat transfer coefficient to match the same thermal measurements.

Example of sub-model calibration:

- In solid mechanics, demonstrating that a constitutive model of a material closely matches a test specimen's measured stress-strain behavior across a wide range of strains, after calibrating constitutive parameters to minimize the discrepancy.

(3) Bench test validation results

This category refers to validation results using experimental data from bench testing, not clinical or animal testing (for the latter see Category 4 below). Bench tests are typically performed under well-controlled laboratory conditions, making them advantageous for simulation validation. Note

that 'bench testing' is a broad term that encompasses *in vitro*, cadaveric and other types of non-clinical testing.[39]

Bench test validation results could be supported by calculation verification and/or UQ results using the validation simulations (as opposed to calculation verification and/or UQ results using the COU simulations; see Category 8).

For this type of evidence, either the validation simulations or the bench tests can be prospectively planned or previously generated. As shown in Table 3. This leads to three common cases: prospectively planned validation activities, validation against retrospective experimental datasets, or previously generated validation results. In addition, although the validation involves bench testing, the COU itself could be either bench or *in vivo*. Examples of potential combinations are provided below.

Examples using prospectively planned validation activities:

- In the following example, both the COU and the validation simulations correspond to bench testing:
    - In solid mechanics, a manufacturer of a new family of cardiovascular implants plans to perform bench durability testing to assess fatigue resistance. A computational model of the device family is developed, and simulations of the bench test are used to select worst-case device sizes to minimize the number of bench test articles needed. Validation with supporting calculation verification evidence is generated by performing finite element simulations of loading for a subset of the devices using multiple finite element mesh resolutions and comparing model-predicted and bench-measured quantities of interest.
- In the following example, the COU corresponds to *in vivo* conditions, but the validation simulations correspond to bench testing:
    - In electromagnetics, a manufacturer of a new implantable device plans to assess induced power density during MR imaging using a computational model of the device implanted in anatomical models of a set of virtual patients. The computational model predicts energy absorption during MR scanning. For validation, physical experiments using the same device in a gel phantom tank are compared to simulation results using an *in silico* model of the device in a simulated gel phantom tank.

Example using validation against retrospective datasets:

- In fluid dynamics, a manufacturer uses computational fluid dynamics to assess the performance of a blood-contacting device. The manufacturer compares simulations with classical hydrodynamic laboratory measurements (e.g., flat-plate boundary layer, lift and drag on objects) or other benchmark experiments designed for validation

[39] See also FDA guidance, "Recommended Content and Format of Non-Clinical Bench Performance Testing Information in Premarket Submissions" available at https://www.fda.gov/regulatory-information/search-fda-guidance-documents/recommended-content-and-format-non-clinical-bench-performance-testing-information-premarket

(e.g., a benchmark nozzle or blood pump[40]). Although the validation dataset is not specific to the COU, the validation exercise provides evidence that the model accurately predicts hydrodynamic behavior that are generally relevant to the COU.

Examples using previously generated validation results:

- In solid mechanics, a manufacturer previously developed a computational model of a family of peripheral stents, validated the model by comparing predicted and measured force-displacement relationships under radial loading on the bench, and then used the model to identify worst-case stent sizes to reduce the number of samples that underwent durability testing. Subsequently the manufacturer seeks a new indication for the same stents in different vasculature. A computational model of the stents in the new loading conditions is developed. The previously collected validation results may support the credibility of the model under the new loading conditions associated with the new indications.
- In electromagnetics, a computational model of MR-induced heating near an implantable device was previously developed, validated, and used to generate evidence to support conditions of safe use of the device for 3T MR machines. Subsequently, the same model is used to support conditions of safe use of the device for 7T MR machines. The previous validation results may support the model for this new COU for known transmit coil configurations.

[40] Malinauskas RA, Hariharan P, Day SW, Herbertson LH, Buesen M, Steinseifer U, Aycock KI, Good BC, Deutsch S, Manning KB and Craven B. FDA Benchmark Medical Device Flow Models for CFD Validation. *ASAIO J,* vol. 63(2), pp. 150-160, 2017.

Table 3: Comparison of three common validation cases based on whether the validation simulations are prospectively planned or previously generated (rows) and whether the comparator data are prospectively planned or previously generated (columns).

	Prospectively planned comparator data	**Previously generated comparator data**
Prospectively planned validation simulations	Corresponds to prospectively planned validation activities. • possible to select experiments and simulations to maximize relevance to COU (applicability) • possible to quantify uncertainties in simulation results • possible to quantify comparator measurement error and uncertainty • method of comparison can be chosen	Corresponds to validation against retrospective data. Validation simulations need to be planned to match the comparator. Examples include comparison against literature experimental data or benchmark datasets. • limited control over relevance of the validation activities to the COU; applicability may be low • limited control over ASME V&V 40 comparator credibility factors • possible to quantify uncertainty in simulation results • comparator measurement error and uncertainty may not be available • method of comparison can be chosen
Previously generated validation simulations	Very uncommon	Usually corresponds to previously generated validation results; for example, for a previous COU with a similar model (e.g., in a previous regulatory submission), or general model validation results published in the literature. • limited ability to select experiments and simulations to maximize relevance to COU; applicability may be low • no/limited control over ASME V&V 40 validation credibility factors • uncertainties in simulation results may not be available • comparator measurement error and uncertainty may not be available • no ability to choose method of comparison unless raw data are available

(4) *In vivo* validation results

This category refers to validation results using *in vivo* data as the comparator, in the form of either clinical or animal data. This category assumes subject-level comparison between simulation and comparator when data from one or multiple subjects are available (population-level comparison falls under Category 5). Therefore, this category applies to patient-level validation of a patient-specific computational model, for example, a clinical trial evaluating the performance of a medical device that uses patient-specific computational simulation.

The validation results could be supported by calculation verification and/or UQ results using the validation simulations (as opposed to calculation verification and/or UQ results using the COU simulations; see Category 8).

In this type of evidence, either the validation simulations or the *in vivo* comparator data can be prospectively planned or previously generated. As shown in Table 3, this leads to three common cases: prospectively planned validation activities, validation against retrospective datasets, or previously generated validation results. Some examples are provided below.

Examples using prospectively planned validation activities:

- In fluid dynamics, a clinical software tool, which uses a physics-based patient-specific model of the coronary arteries to predict the fractional flow reserve (FFR), is validated by comparing simulations against invasive clinical FFR measurements in the same patient. A calculation verification study may also be performed to estimate the numerical uncertainty in these simulations.
- A manufacturer develops a computational model-based tool that predicts a quantitative clinical metric with a known correlation to patient outcomes. The manufacturer validates the predictive capability of the tool by performing a clinical trial and computing sensitivity, specificity, positive/negative predictive value, and area under receiver operating characteristic (ROC) curve.
- In bioheat transfer, a first principles-based thermal model is developed to predict *in vivo* tissue heating due to a device (e.g., devices based upon delivering ultrasound, laser, radiofrequency (RF) energy). The model is validated using humans and/or animal models in relevant tissues for appropriate spatio-temporal distribution of *in vivo* power density, by making direct measurements using temperature probes.

Examples using previously generated validation results:

- In solid mechanics, a manufacturer uses a computational model to compute displacements for one device (e.g., shoulder arthroplasty) under simulated *in vivo* conditions (e.g., rotations), performs a supporting calculation verification study, and validates the predictions against relevant *in vivo* data. Later, the manufacturer wishes to use a similar model for a different device (e.g., reverse shoulder arthroplasty). The previous validation and calculation verification may support the credibility of the new device model.

- In bioheat transfer, a first principles-based thermal model is developed to predict *in vivo* tissue heating due to a device (e.g., devices based upon delivering ultrasound, laser, RF energy). The model was validated using humans and/or animal models in relevant tissues for appropriate spatio-temporal distribution of *in vivo* power density, by making direct measurements using temperature probes. Later, the manufacturer wishes to use the same model for a different device. If the nature of the spatio-temporal temperature distribution (i.e., magnitude and gradients in space and time) is expected to be comparable between two devices for the full range of device specifications in the tissue of interest, the previous validation evidence may be able to support the credibility of the model for predicting *in vivo* tissue heating due to the second device.

Another possibility for previously generated validation results occurs for general-purpose or multi-application computational models for which it is common to compare model predictions under a variety of conditions with experimental data. With computational models of physiological systems, it is common to show the model can reproduce the range of physiological behaviors when publishing or releasing the model. Those validation results could be relevant if the physiological model is later used in a medical device COU. For example:

- In physiological modeling, a model of the cardiovascular system is developed and then validated by comparing model predictions of various hemodynamic variables (e.g., mean arterial blood pressure, cardiac output) against recordings from patients throughout a range of normal and pathological conditions. A manufacturer of a physiological closed loop control (PCLC) device that uses the model for *in silico* testing of the control algorithm could potentially utilize the previous validation results to support the model credibility in a PCLC testing COU.

(5) Population-based validation results

Population-based evidence consists of comparisons of population-level data between model predictions and a clinical data set, or potentially other data such as animal or cadaveric data. A distinguishing feature of this evidence is that multiple subjects are involved, but comparison of simulation results and experimental data for the same subject is not performed (i.e., no comparison is made on a patient-level basis; such evidence falls under Category 4). For example, this type of evidence is relevant to validation of 'virtual populations' or 'virtual cohorts,' that is, multiple patient models representing a patient population. Population-based evidence for credibility of the virtual population/cohort could be generated by comparing the mean and standard deviation of a model output across the virtual population/cohort with the mean and standard deviation from a clinical dataset. Population-level clinical trial results would be a part of this category, whereas patient-level clinical trial results fall in Category 4.

Examples:

- In medical imaging, a set of virtual patients is generated by taking an anthropomorphic model of a breast and of lesions and varying key parameters across expected ranges. Comparison of model predictions to individual patient data is not possible because none of the virtual patients correspond to any one actual patient.

Instead, the results of the computer-simulated trial are statistically compared to clinical outcomes to demonstrate that the predictions are consistent with the comparative trial using human subjects and human image interpreters.[41]

(6) Emergent model behavior

Emergent model behavior is evidence that demonstrates that the finalized computational model reproduces phenomena that are known to occur in the system at the specified conditions but were not pre-specified or explicitly modeled by the governing equations. A distinguishing feature of this type of evidence is that simulation results are not directly compared to specific data. Instead, simulation results are assessed using scientific knowledge about the system, possibly based on qualitative experimental observations. This type of evidence is especially relevant to models of physiological systems, because physiological systems often exhibit emergent behavior that is not predictable from knowledge on sub-systems.

Examples:

- In fluid dynamics, a computational model of blood flow through a stenotic vessel is developed, and evidence is collected to confirm the model correctly predicts the onset of transitional or turbulent flow at conditions where such phenomena are expected. A manufacturer that uses the model to predict clinical metrics related to stenosis severity and ischemia could include this information as credibility evidence.
- In cardiac electrophysiology, a model of electrical activity in the heart and torso is developed. It is demonstrated that each simulated electrocardiogram (ECG) in the standard 12-lead ECG has the same morphology as clinical ECGs, in terms of relative size and direction of the P-wave, QRS-complex and T-wave. A cardiac device manufacturer that uses this model for *in silico* testing of their device could include this information as credibility evidence for the cardiac model.

(7) Model plausibility

Model plausibility evidence is the rationale supporting the choice(s) of governing equations, model assumptions, and/or input parameters. A claim of model plausibility is an argument that the model is credible because the governing equations are expected to hold, assumptions are justified, and parameters and other quantities that are input into the model have been justified. A distinguishing feature of this category is that simulations do not need to be run to generate this kind of evidence, because the evidence is based on knowledge about the model, and not on a comparison of model results to data. Since this evidence does not involve testing or assessing the finalized model (i.e., no verification or validation), model plausibility might be the first step in supporting model credibility, but it is generally a weak form of credibility evidence. In some

[41] Badano A, Graff CG, Badal A, Sharma D, Zeng R, Samuelson FW, Glick SJ and Myers KJ. Evaluation of Digital Breast Tomosynthesis as Replacement of Full-Field Digital Mammography Using an In Silico Imaging Trial. *JAMA Netw Open,* vol. 7(1), 2018.

cases where it is very difficult to obtain any experimental data from the system of interest for validation, this may be a primary form of evidence to support model credibility.

Example:

- In solid mechanics, a finite element model of a simple joint arthroplasty device is developed. For the particular combination of implant design, implant material, and loading conditions considered, deformations are anticipated to be well within the linear-elastic regime. The mechanical behavior of the implant material is also well characterized and has been shown to be approximately isotropic at the length scales of interest. Accordingly, plausibility evidence could support the credibility of the implant material model, i.e., a linear elastic model with an isotropic constitutive law, supported by justification for the specific material parameters used. The credibility of the whole model could be supported using plausibility evidence if valid rationales for the governing equations, model assumptions, and input parameters can be made.

(8) Calculation verification/UQ results using COU simulations

This category refers to standalone calculation verification and/or UQ results performed using the COU simulations, which are the simulations performed to answer the question of interest using the COU conditions. Direct *validation* of the COU simulations is not possible, because if comparator data was available for the COU there would be no need for the model. However, calculation verification or UQ analyses are possible using these simulations.

This type of evidence applies to *in silico* device testing or *in silico* clinical trials, but not models in device software or in MDDTs, for which the COU simulations are run after the device is on the market or MDDT qualified.

Examples:

- A finite element model of a medical device is developed to identify worst-case configurations related to a device safety concern. For validation, model predictions were compared to bench test data. A mesh convergence study was performed to confirm the numerical error due to spatial discretization is acceptably small (Category 3 evidence). However, for the COU a different quantity of interest will be analyzed than that considered in the validation study. There is reason to believe a finer computational mesh is needed to resolve this quantity of interest. Therefore, a new mesh convergence study is performed for this quantity of interest using the COU conditions.
- In fluid dynamics, a computational model of blood flow through a ventricular assist device is used to assess the influence of a planned change in manufacturing tolerances on hemolysis. Simulations were previously validated using a single well-characterized device at multiple operating conditions, by comparing with measurements of the velocity field and the corresponding flow-induced stress from particle image velocimetry. To address the question of interest, simulations are performed with accompanying UQ to analyze the influence of the planned change in manufacturing tolerances. In the UQ study, the device dimensions are varied within

the range of the manufacturing tolerances. This input geometric uncertainty is propagated through the model using Monte Carlo sampling to perform a large number of simulations to quantify the influence of geometric variances on the predicted flow-induced stress and blood exposure time in the device, which are closely related to hemolysis. Two separate UQ studies are performed for the original and the proposed manufacturing tolerances to justify that the planned change has a negligible influence on the hemolytic potential of the device.

C. Credibility Factors and Credibility Goals

Step 5 in the framework is to define **credibility factors** for the planned credibility evidence and set credibility goals for each credibility factor, with a plan to achieve these goals.

See ASME V&V 40 for an introduction to credibility factors. As an example, ASME V&V 40 defines two credibility factors for code verification: 'Software quality assurance' and 'Numerical code verification.' Other credibility factors are similarly defined in ASME V&V 40 for calculation verification, validation and applicability.

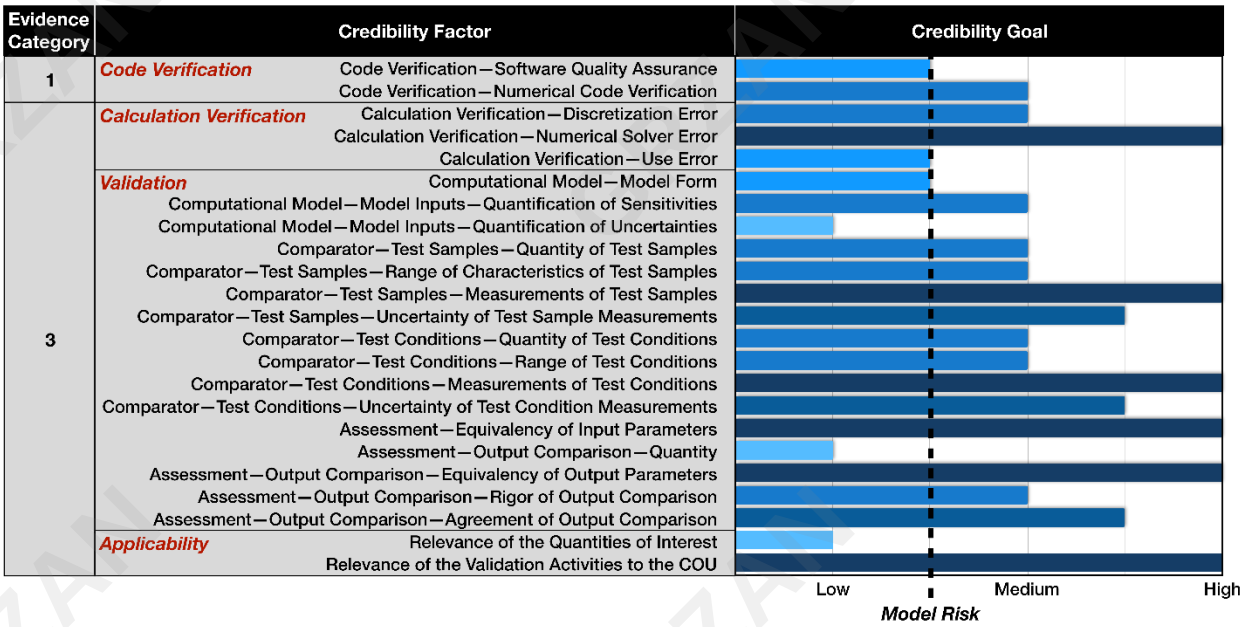
To establish credibility factors and credibility goals, we recommend the following sub-steps for Step 5. Refer to Figures 1 and 3 for examples.

- Step 5.1: State credibility factors relevant to the type of credibility evidence you plan to gather. When relevant, we recommend using ASME V&V 40 credibility factors. For example, if you plan to gather bench test validation results (Category 3), we recommend using ASME V&V 40 credibility factors related to validation and applicability. For evidence categories that are not explicitly covered by ASME V&V 40 (e.g., model calibration evidence, population-based evidence, or model plausibility – Categories 2, 5 or 7, respectively), we recommend defining new credibility factors. For example, if model calibration results will be used in support of model credibility, you could define a 'Goodness of fit' credibility factor, among others.
    - See also Appendix 1 for specific considerations for each category of credibility evidence, including suggested credibility factors.
    - If there are multiple forms of credibility evidence from different categories, with one set being used as the 'primary' source of evidence and other sets as 'secondary' or 'supporting' evidence (e.g., *in vivo* validation results as primary and bench test validation results as secondary), we recommend using ASME V&V 40 credibility factors when possible for the primary evidence and an appropriately limited set of credibility factors for the supporting evidence. This is to avoid an excessive total number of credibility factors when results from multiple categories are used to support the overall credibility of the model.
    - Since the relevance of the evidence to support using the model for the COU is especially important, we recommend defining a 'Relevance to the COU' credibility factor(s) for each set of credibility evidence (as emphasized in Appendix 1). For validation evidence, this is termed 'applicability' (see Definitions section).

- Step 5.2: Following ASME V&V 40, for each credibility factor, define a gradation of activities that describes progressively increasing levels of rigor in investigation. For example, for a 'Goodness of fit' credibility factor for model calibration evidence (Category 2), a possible gradation is:
  a) Qualitative comparison of fit performed.
  b) Quantitative error of fit computed without accounting for any uncertainty.
  c) Uncertainty in fitted parameters (e.g., due to experimental noise) is estimated and accounted for in the quantitative error of fit.
- Step 5.3:
  - Following ASME V&V 40, for each credibility factor corresponding to <u>prospectively planned</u> activities:
    - Select a 'credibility goal' from the gradation, considering the model risk as assessed in Step 3. Higher risk questions of interest generally warrant higher-level credibility goals. It is important to note that in this step, a level of credibility is being proposed for each factor that will contribute to the *overall* credibility of the model. See ASME V&V 40 for examples.
    - If the goal is less than the level commensurate with model risk (see Figure 3), for example due to practical constraints, you should provide a rationale for why the activities are sufficient.
    - Describe a high-level plan to achieve the proposed credibility goal. This should be included in the prospective credibility assessment to justify the level of credibility that is being proposed.
  - For each credibility factor corresponding to <u>previously generated</u> data (e.g., ASME V&V 40 'comparator' credibility factors in the case of validation using a retrospective dataset):
    - Identify which level from the gradation the previously performed activities correspond to.
    - If the assessed credibility level is less than the level commensurate with model risk, you should provide a justification for why the activities are sufficient.

Figure 3 presents a hypothetical example of this process. In this example, two types of credibility evidence are planned, code verification results (Category 1) and prospectively planned bench test validation results (Category 3). In this example, the Category 3 evidence includes both validation and supporting calculation verification results. Model risk was assessed to be Low-Medium. ASME V&V 40 credibility factors are used, and a five-level gradation was defined to grade each credibility factor. Credibility goals were chosen for each factor as indicated in Figure 3. For credibility factors for which the goal corresponds to a credibility level that is not commensurate with model risk (i.e., the three credibility factors with level 'low'), a rationale should be provided for why the activities are sufficient.

Figure 3: Hypothetical example of setting credibility factor goals. In this example all activities are assumed to be prospectively planned.



D. Adequacy Assessment

Steps 6 and 8 of the framework assess the adequacy of the credibility-related activities and results. Step 6 is a prospective adequacy assessment, which asks the question*: if the credibility goals are achieved, will the credibility evidence be sufficient to support using the model for the COU given the risk assessment*? Step 8 is a post-study adequacy assessment and asks the question: *does the available credibility evidence support using the model for the COU given the risk assessment*? Note that adequacy assessment is different from applicability: as per Section V, applicability refers to the relevance of validation activities to the COU but adequacy assessment considers the totality of the credibility evidence. Also, in contrast to *model accuracy,* which is quantifiable through validation, *model adequacy* warrants a careful decision to be made using engineering and clinical judgement, based on all available information.[42]

Performing the prospective adequacy assessment (Step 6) is recommended if you plan to request FDA feedback on planned activities via a pre-submission (as described in Step 6 in Section V) to facilitate the evaluation of your proposed rationale for credibility of the computational model. If performing prospective adequacy assessment, we recommend that you consider the planned credibility evidence, the proposed credibility goals for each credibility factor, and any other relevant information. The prospective adequacy assessment should include a rationale for why the planned credibility evidence is expected to be sufficient to support using the model for the COU, given the risk assessment.

[42] Oberkampf WL and Roy CJ. *Verification and Validation in Scientific Computing*. Cambridge University Press, 2010.

When performing post-study adequacy assessment (Step 8), we recommend that you first re-evaluate the credibility level that was achieved for each credibility factor and whether the credibility goal was met. The post-study adequacy assessment should also include a rationale for why the credibility evidence is sufficient to support using the model for the COU, given the risk assessment. Post-study adequacy assessment can also use the COU simulation results, if available, and related information such as the difference between COU model predictions and safety thresholds (see example below). We recommend that you take into consideration the following questions and recommendations in the post-study adequacy assessment:

Questions:

- Have all relevant features of the model been adequately tested? That is, do the verification, validation and any other credibility evidence sources cover all features of the model relevant to the COU? For example:
    - For models used within device software, have all model-derived device outputs been evaluated as part of the credibility assessment process?
- Were the credibility goals met? If the goal was not met for a factor(s), we recommend that you provide a justification for why the impact of the unmet credibility factor(s) on the risk (associated with using the model to address the question of interest) is acceptable.

Recommendations:

- You may wish to pre-specify quantitative *accuracy targets* for the model validation comparison, such that the model will be considered adequate if the accuracy targets are met. Since quantitative accuracy targets will be application-specific, you should still provide a rationale explaining why this level of accuracy is sufficient to support using the model for the COU. Note that even if pre-specified quantitative accuracy targets for model validation were not met, it may still be possible to use the model for the COU if a valid rationale can be provided, such as based on further analysis. We also recognize that it is not always possible and/or meaningful to pre-specify precise quantitative accuracy targets. In this case, we recommend you pre-specify how you intend to assess the level of agreement between the model results and the validation data.
- When the question of interest includes information concerning a decision or safety threshold, then as part of the adequacy assessment we recommend considering the proximity of model predictions relative to such thresholds. That is, how close is the model prediction to the decision or safety threshold? As part of this assessment, it may also be useful to consider estimates of uncertainty in the COU predictions (e.g., based on uncertainty quantification, calculation verification results, and/or model accuracy from the validation comparison) and, if applicable, uncertainty in the value of the decision or safety threshold. Such considerations could be used to further support the adequacy of the model for addressing the question of interest. For example:
    - For a computational model of MR-induced energy absorption of an implantable metallic device, suppose the COU simulations predict that the power deposited into the surrounding tissue is well within acceptable levels,

and moreover, the uncertainty in predicted power, based on uncertainty quantification and validation, is small. Overall, the 99% confidence interval for power deposited into the surrounding tissue is well within acceptable levels. This information could be used to further justify the adequacy of the model credibility assessment activities for addressing the question of interest.

- It is important to explicitly state any limitations of the model and provide a rationale for why they do not reduce confidence in using the model for the COU, referring to the credibility evidence or other scientific knowledge as appropriate.

If you determine the evidence to be insufficient in either the prospective or post-study adequacy assessment, we recommend that you consider modifying the model, reducing the model influence, modifying the COU, and/or revising the plan to generate credibility evidence (prospective adequacy assessment) or collecting additional evidence (post-study adequacy assessment). See ASME V&V 40 for a discussion on these different options.

Appendix 1. Considerations for Each Credibility Evidence Category

Below are considerations regarding the generation and/or evaluation of credibility evidence, for each category of evidence in Section VI.B. Some of the following considerations may not be applicable depending on specific details of the modeling performed.

Category 1: Code verification results

- For Step 5 of the framework, we recommend using the credibility factors for code verification defined in ASME V&V 40.
- For computational models implemented within medical device software, note that software and model verification and validation are both important but differ in scope and definition. Testing performed for software verification may include code verification of the computational model, although the latter is typically addressed separately and may need consideration of the specific COU. See software verification and validation reporting recommendations in FDA's guidance titled "Guidance for the Content of Premarket Submissions for Software Contained in Medical Devices"[43] and refer to the appropriate tests when describing model code verification activities.
- For computational models that are not part of the device (e.g., *in silico* device testing, *in silico* clinical trials), code verification for the model is unrelated to the device software verification and/or validation and is therefore performed separately from device software verification and validation.
- For computational models that are not part of the device (e.g., *in silico* device testing, *in silico* clinical trials), if a commercial software package was used to develop the computational model, we recommend referring to any information provided by the software manufacturer on software quality assurance and code verification, as relevant.

Category 2: Model calibration evidence

- For Step 5 of the framework, consider defining credibility factors related to goodness of fit, quality of the comparator data, and relevance of calibration activities to the COU.
- Be cautious not to present or confuse calibration evidence as/with validation evidence and ensure that data for calibration is separate or not inclusive of data used for validation.
- Consider evaluating whether final values of all calibrated parameters that have a physical/physiological meaning are within expected physical/physiological ranges.
- Consider quantifying the 'goodness of fit.'
- When reporting calibration results, we recommend that you provide details on the following (if applicable):
    - calibration procedure, including which parameters were calibrated;
    - prior distributions for these parameters if a Bayesian calibration approach was used;
    - details of the simulations run, source and details of experimental/comparator data;

[43] https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-content-premarket-submissions-software-contained-medical-devices

    - any steps taken to ensure the model is not overfitted; and
    - numerical methods for obtaining the calibrated results.
- As discussed in Section VI.B, model calibration evidence is weaker than validation evidence. Therefore, if model calibration evidence is provided as the primary source of credibility evidence, you should provide a rationale for why validation testing of the model is not possible or warranted, for example, referring to the assessed model risk.
- If no validation results are available and calibration results are the primary source of evidence for model credibility, consider evaluating the relation between calibration conditions and COU conditions, and between calibration quantities of interest and COU quantities of interest.

Category 3: Bench test validation results

- For Step 5 of the framework, we recommend using credibility factors defined in ASME V&V 40.
- If the COU will involve making *in vivo* predictions, we recommend paying special attention to the applicability of the bench test validation results to the *in vivo* COU.
- For prospectively planned validation:
    - If possible, we recommend considering that the computational analyst(s) performing the simulations be blinded to the bench test validation data to prevent the potential for bias.[44]
- For validation against retrospective datasets:
    - We recommend that you pay special attention to the applicability of validation results to the COU, since the comparator data were not designed for validating the model for the current COU.
- For previously generated validation results:
    - We recommend that you pay special attention to the applicability of previously generated validation results to the COU, since the previous validation results were not designed to support the model for the current COU. This should include an assessment of any differences, and the impact thereof, between the model used in the previous validation results and the current model.

Category 4: In vivo validation results

- For Step 5 of the framework, if the evidence is traditional validation evidence, we recommend using credibility factors defined in ASME V&V 40.
- If the evidence takes another form (e.g., clinical trial results), we recommend that you generate and evaluate the evidence using the appropriate best practices and methods (e.g., appropriate statistical techniques, appropriate measures of sensitivity and specificity, positive predictive value), applicable regulatory requirements (e.g., good clinical practice regulations[45]), and define appropriate credibility factors for Step 5 of the framework.
- For prospectively planned validation:

[44] See Section 2.5.1 and Section 11.1.4, Oberkampf WL and Roy CJ. *Verification and Validation in Scientific Computing*. Cambridge University Press, 2010.

[45] See Regulations: Good Clinical Practice and Clinical Trials at https://www.fda.gov/science-research/clinical-trials-and-human-subject-protection/regulations-good-clinical-practice-and-clinical-trials

    - If possible, we recommend considering that the computational analyst(s) performing the simulations be blinded to the validation data to prevent the potential for bias.[46]
- For validation against retrospective datasets:
    - We recommend that you pay special attention to the applicability of validation results to the COU, since the comparator data were not designed for validating the model for the current COU.
- For previously generated validation results:
    - We recommend that you pay special attention to the applicability of previously generated validation results to the COU, since the previous validation results were not designed to support the model for the current COU. This should include an assessment of any differences, and the impact thereof, between the model used in the previous validation results and the current model.

Category 5: Population-based evidence

- Consider quantitatively assessing the closeness of the two populations by comparing means, variances, full distributions or using other appropriate statistical methods.
- We recommend that you provide and compare relevant demographic information, anatomy, pathologies, and co-morbidities of the subjects used in: (i) the patient data used to generate the virtual cohort; (ii) the clinical dataset used for validation; and (iii) the intended patient population.
- If the evidence comes from a clinical study without subject-level data, we recommend that you generate and evaluate the evidence using the appropriate best practices and methods (e.g., good clinical practices, appropriate statistical techniques), and define appropriate credibility factors for Step 5 of the framework.

Category 6: Emergent model behavior

- As discussed in Section VI.B, compared to model validation, emergent model behavior is generally relatively weak evidence for model credibility because it does not involve direct comparison with experimental data. Therefore, we generally do not recommend relying on emergent model behavior as a primary source of evidence for model credibility, although it may serve as useful secondary evidence.
- Consider evaluating how important or relevant the emergent behavior is to the COU and explaining why the model reproducing the emergent behavior provides confidence in the model for the COU.
- For Step 5 of the framework, we recommend that you define credibility factors for the relevance of the emergent behavior to the COU, sensitivity of emergent behavior to model input uncertainty, and others.

Category 7: Model plausibility

- As discussed in Section VI.B, compared to model validation, model plausibility is generally a relatively weak argument for model credibility because it does not involve testing the model predictions. Therefore, if model plausibility evidence is the primary

[46] See Section 2.5.1 and Section 11.1.4, Oberkampf WL and Roy CJ. *Verification and Validation in Scientific Computing*. Cambridge University Press, 2010.

credibility evidence presented, you should provide a rationale for why validation testing of the model is not possible or warranted, for example, referring to the assessed model risk.

- Consider evaluating how any assumptions impact predictions by comparing results using alternative model forms, preferably from higher-fidelity models if possible.
- Consider performing uncertainty quantification and sensitivity analysis for the model parameters.
- For Step 5 of the framework, we recommend using ASME V&V 40 credibility factors related to model form and model inputs, as appropriate.

Category 8: Calculation verification/UQ results using COU simulations

- For calculation verification results: for Step 5 of the framework, we recommend using the three calculation verification credibility factors defined in ASME V&V 40.
- For UQ results: for Step 5 of the framework, we recommend using the model input credibility factors defined in ASME V&V 40.
- If you generate this type of evidence, we recommend incorporating the calculation verification and/or UQ results when comparing COU predictions with any decision thresholds (as discussed in Section VI.D, 'Adequacy Assessment'), taking into account the estimated numerical uncertainty and/or output uncertainty from UQ.

Appendix 2. Reporting Recommendations for CM&S Credibility Assessment in Medical Device Submissions

In this Appendix, we provide: (a) recommended information to include when requesting feedback on a CM&S credibility assessment plan in a Q-submission, and (b) recommendations for reporting of CM&S credibility assessment in medical device regulatory submissions.

Requesting FDA Feedback on a Credibility Assessment Plan

We recognize that the generalized framework for assessing model credibility may necessitate interactive feedback from FDA, in particular concerning the model risk assessment and the prospective adequacy assessment (Steps 3 and 6 in Section V, respectively). Manufacturers who wish to receive feedback from FDA can receive feedback on any aspect of their computational modeling and/or credibility assessment using the Q-submission pathway (refer to FDA's guidance titled "Requests for Feedback and Meetings for Medical Device Submissions: The Q-Submission Program"[47]). If requesting feedback on a plan for credibility assessment, we recommend that you provide information on the preliminary and prospective steps in the framework outlined in Section V (Steps 1-6). The following provides an example of how the Q-submission could be organized:

Possible Content to include in a Q-submission on a Credibility Assessment Plan:

1. **Purpose:** The overall purpose of the Q-Submission including goals for the outcome of the interaction with FDA.
2. **Background:** e.g., clinical context or other relevant background information for the device.
3. **Device Description**
4. **Proposed Indications for Use**
5. **Regulatory History**
6. **Description of Computational Model**
7. **Credibility Assessment Plan**
    a. Summary of overall approach
    b. Question of Interest (see Section VI.A.(1))
    c. COU (see Section VI.A.(2))
    d. Model Risk Assessment (see Section VI.A.(3))
    e. Planned Credibility Evidence. For each type of credibility evidence planned, provide the following:
        i. Categorization of evidence per Section VI.B
        ii. Description of evidence to be collected
        iii. Chosen credibility factors (see Section VI.C). For each factor, provide:
            1. Credibility gradation
            2. Proposed credibility goal (or assessed credibility level for previously generated data)

[47] https://www.fda.gov/regulatory-information/search-fda-guidance-documents/requests-feedback-and-meetings-medical-device-submissions-q-submission-program

3. Brief plans for achieving credibility goal

f. Prospective Adequacy Assessment (see Section VI.D).

8. **Specific Questions for FDA**

Recommendations for a Credibility Assessment Report

A Credibility Assessment Report is a self-contained document that can be included as part of a regulatory submission. The report is intended to provide evidence and the rationale for the credibility of CM&S used in a medical device regulatory submission.

Below, we provide an example of how a Credibility Assessment Report could be organized. The outline below only applies to CM&S credibility information and does not provide a recommended format for information pertaining to the model itself. Moreover, for CM&S used in *in silico* device testing or *in silico* clinical trials (see Section II) the outline does not provide recommendations for providing the results of the simulation study. For CM&S used for in sil*ico devic*e testing or in sil*ico cli*nical trials, refer to FDA's guidance titled "Reporting of Computational Modeling Studies in Medical Device Submissions"[48] (hereafter referred to as "Computational Modeling Reporting Guidance") for reporting model details and study results. In this situation, we recommend that you provide two reports: one report describing the model and study results using the Computational Modeling Reporting Guidance, and a separate "Credibility Assessment Report" using the outline described below. In the first report, we recommend you reference your Credibility Assessment Report as appropriate to provide any credibility-related information recommended by the Computational Modeling Reporting Guidance (i.e., Section III: Code Verification, Section VIII: System Discretization—Calculation Verification, and Section X: Validation).

FDA recognizes that the level of detail included in a Credibility Assessment Report will vary and will depend on the specific discipline, type of computational modeling, and the COU of the model, among other factors. Because we expect the level of detail to vary for different types of CM&S, we recommend that your Credibility Assessment Report provide an emphasis on the rationale/justification used when generating and assessing your credibility evidence. The following outline may be helpful to organize the content of your Credibility Assessment Report:

Recommended Content for a Credibility Assessment Report:

1. **Executive Summary:** Include a brief description of the device, the model, the question of interest that the model is used to address, the model COU, the assessed model risk, a summary of the categories of the credibility evidence provided, and a summary of the adequacy assessment with a brief rationale.
2. **Background:** e.g., clinical context or other relevant background for the device. Either provide here or refer to another section in the regulatory submission.
3. **Device Description:** Include within the report or refer to another section in regulatory submission.

[48] https://www.fda.gov/regulatory-information/search-fda-guidance-documents/reporting-computational-modeling-studies-medical-device-submissions

4. **Proposed Indications for Use:** Include within the report or refer to another section in regulatory submission.
5. **Description of Computational Model:** If model details are included elsewhere in the regulatory submission, we recommend referencing accordingly. We recommend providing details on governing equations, model parameter values, methods used to determine parameter values, numerical methods used for solving the governing equations, and other information that could be relevant in evaluating model credibility.
6. **Model Credibility Assessment**
    a. Summary of overall approach
    b. Question of Interest (see Section VI.A.(1))
    c. COU (see Section VI.A.(2))
    d. Model Risk Assessment (see Section VI.A.(3))
    e. Credibility Evidence. For each type of credibility evidence provided, provide the following:
        i. Categorization of evidence per Section VI.B
        ii. Description of evidence
        iii. Chosen credibility factors (see Section VI.C). For each factor, provide:
            1. Credibility gradation;
            2. Prospective credibility goal (if prospectively planned activities) or assessed credibility level (if previously generated data); and
            3. Achieved credibility level (if prospectively planned activities).
        iv. Methods. Full methods may be provided here, or provided elsewhere (e.g., in an Appendix to the Credibility Assessment Report or published in a journal article) and referenced here.
        v. Results. As with the methods, full results may be provided here, or provided elsewhere and referenced here.
    f. Post-study Adequacy Assessment (see Section VI.D).
7. **Credibility Assessment Limitations**
8. **Conclusions**
9. **References**
10. **Appendices:** Detailed descriptions of credibility assessment study methods and results (if needed).