Contains Nonbinding Recommendations

Draft – Not for Implementation

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

---

Draft Guidance for Industry and Food and Drug Administration Staff

DRAFT GUIDANCE

This draft guidance document is being distributed for comment purposes only.

Document issued on December 23, 2021.

You should submit comments and suggestions regarding this draft document within 60 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to http://www.regulations.gov. Submit written comments to the Division of Dockets Management (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. Identify all comments with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions about this document, contact the 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

Additional Copies

Additional copies are available from the Internet. You may also send an e-mail request to CDRH-Guidance@fda.hhs.gov to receive a copy of the guidance. Please use the document number 1500056 to identify the guidance you are requesting.

Table of Contents

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

Draft Guidance for Industry and Food and Drug Administration Staff

This draft guidance, when finalized, will represent the current thinking of the Food and Drug Administration (FDA or Agency) on this topic.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FDA has developed this draft guidance document to assist industry and FDA staff in assessing the *credibility* of computational modeling, defined as trust in the predictive capability of a computational model, used to support medical device premarket submissions (i.e., Premarket Approval (PMA) Applications, [1] Humanitarian Device Exemptions (HDEs),[2] Investigational Device Applications (IDEs),[3] Premarket Notifications (510(k)s),[4] and De Novo requests[5]) or qualification of Medical Device Development Tools (MDDTs); (refer to FDA's guidance titled "Qualification of Medical Device Development Tools"[6]). 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 risk-based framework that can be used in the credibility

[1] 21 CFR part 814
[2] 21 CFR part 814 subpart H
[3] 21 CFR part 812
[4] 21 CFR part 807 subpart E
[5] 21 CFR part 806 subpart D
[6] https://www.fda.gov/regulatory-information/search-fda-guidance-documents/qualification-medical-device-development-tools

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) models. This guidance is intended to help improve the consistency and transparency of the review of CM&S evidence, to increase confidence in the use of CM&S in regulatory submissions, and to facilitate improved interpretation of CM&S 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 standard(s) referenced in this document, see the FDA Recognized Consensus Standards Database.[7]

The contents of this document do not have the force and effect of law and are not meant to bind the public in any way, unless specifically incorporated into a contract. This document is intended only to provide clarity to the public regarding existing requirements under the law. FDA guidance documents, including this guidance, should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word *should* in Agency guidance 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 rapidly increasing.[8] 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. The appropriate and expanded use of CM&S in obtaining accurate and precise results to support regulatory submissions necessitates the development of processes and approaches that promote consistency 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*

[7] Available at https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfStandards/search.cfm

[8] 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.

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 converted into an input to the patient model, and the patient model response is converted into a signal passed back to the controller.[9]

2. **CM&S used within medical device software.** Computational modeling may be implemented as device software functions,[10] which may include software as a medical device (SaMD)[11] that is intended to be used for one or more medical purposes without being part of a hardware medical device, or implemented in software in a medical device (SiMD) that is typically embedded within or part of a hardware device. For example, device software functions that analyze patient data as inputs to a 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 application of CM&S where device performance is evaluated using a 'virtual cohort' of simulated patients with realistic anatomical and physiological variability representing the indicated patient population. *In silico* clinical trials can complement real world clinical trials (e.g., augment or reduce the size of, or provide improved inclusion-exclusion criteria), rather than replace them.[12]
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 and be considered for the Medical Device Development Tools (MDDT) Program[13] 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"[14]).

In all cases, there is a need to demonstrate that the computational model is credible. Methodologies for model credibility assessment have been established in the scientific literature[15] and continue to evolve. Demonstrating model credibility involves various activities

[9] 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.

[10] 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.

[11] See FDA website on "Software as a Medical Device (SaMD)," available at https://www.fda.gov/medical-devices/digital-health-center-excellence/software-medical-device-samd

[12] 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.

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

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

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

that include verification, validation, uncertainty quantification, applicability analysis, as well as adequacy assessment (see the Section IV for definitions). The FDA-recognized standard American Society of Mechanical Engineers (ASME) V&V 40 *Assessing Credibility of Computational Modeling through Verification and Validation: Application to Medical Devices* provides a risk-based framework for assessing verification, validation, and uncertainty quantification (VVUQ) activities for computational modeling of medical devices. However, ASME V&V 40 assumes the ability to perform traditional validation activities, that is, comparison of model predictions against well-controlled validation experiments. For computational models used in regulatory submissions, there are often many different sources of evidence that are available to support model credibility, including results from clinical studies, robust model calibration results, or population-level validation results. This guidance uses key concepts of ASME V&V 40 but provides a more general framework for demonstrating CM&S credibility in medical device regulatory submissions that incorporate such non-traditional evidence.

III. Scope

The purpose of this guidance document is to provide a general framework for assessing CM&S credibility in medical device regulatory submissions that incorporates both traditional V&V evidence and/or other types of supporting data. This guidance document is applicable to physics-based, mechanistic, or other first principles-based 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 statistical or data-driven models such as machine learning or artificial intelligence.

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. 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. We recommend that manufacturers seek feedback on their specific use of CM&S through the Q-submission process (refer to FDA's guidance titled "Requests for Feedback and Meetings for Medical Device Submissions: The Q-Submission Program"[16]).

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.

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

Adequacy assessment: the process of evaluating the evidence in support of credibility of a computational model, for a given context of use, and making a determination on whether the evidence is sufficient

Applicability: the relevance of a credibility assessment activity (e.g., validation activities) to support the use of the computational model for a context of use

Calculation verification (also called solution verification)**:** "the process of determining the solution accuracy of a calculation"[17]

Code verification: "the process of identifying errors in the numerical algorithms of a computer code"[18]

Comparator: the test data that are used for validation, which may be data from bench-testing or *in vivo* studies

Computational model: "the numerical implementation of the mathematical model performed by means of a computer"[19]

Context of use (COU): "a statement that defines the specific role and scope of the computational model used to address the question of interest"[20]

Credibility: "the trust, established through the collection of evidence, in the predictive capability of a computational model for a context of use"[21]

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

Credibility factors: fundamental aspects of the credibility assessment process that break down the analysis of verification, validation, or other sources of credibility evidence

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

Mathematical model: "the mathematical equations, boundary conditions, initial conditions, and modeling data needed to describe a conceptual model"[22]

[17] Reprinted by 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, copywrite 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.

[18] ibid

[19] ibid

[20] ibid

[21] ibid

[22] ibid

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)

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"[23]

Quantity of interest: "the calculated or measured result from a computational model or comparator, respectively"[24]

Question of interest: "the specific question, decision, or concern that is being addressed"[25]

Uncertainty quantification: the process of identifying, characterizing and quantifying those factors that could affect the accuracy of computational results

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"[26]

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"[27] See also *calculation verification* and *code verification.*

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

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

FDA recommends the following process when assessing the credibility of computational modeling used in a medical device regulatory submission. Detailed information on the key

[23] Reprinted by 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, copywrite 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.

[24] ibid

[25] ibid

[26] ibid

[27] ibid

concepts in the framework below are provided in subsequent sections. See Figure 1 for an illustration of an overview of the framework using a hypothetical example.

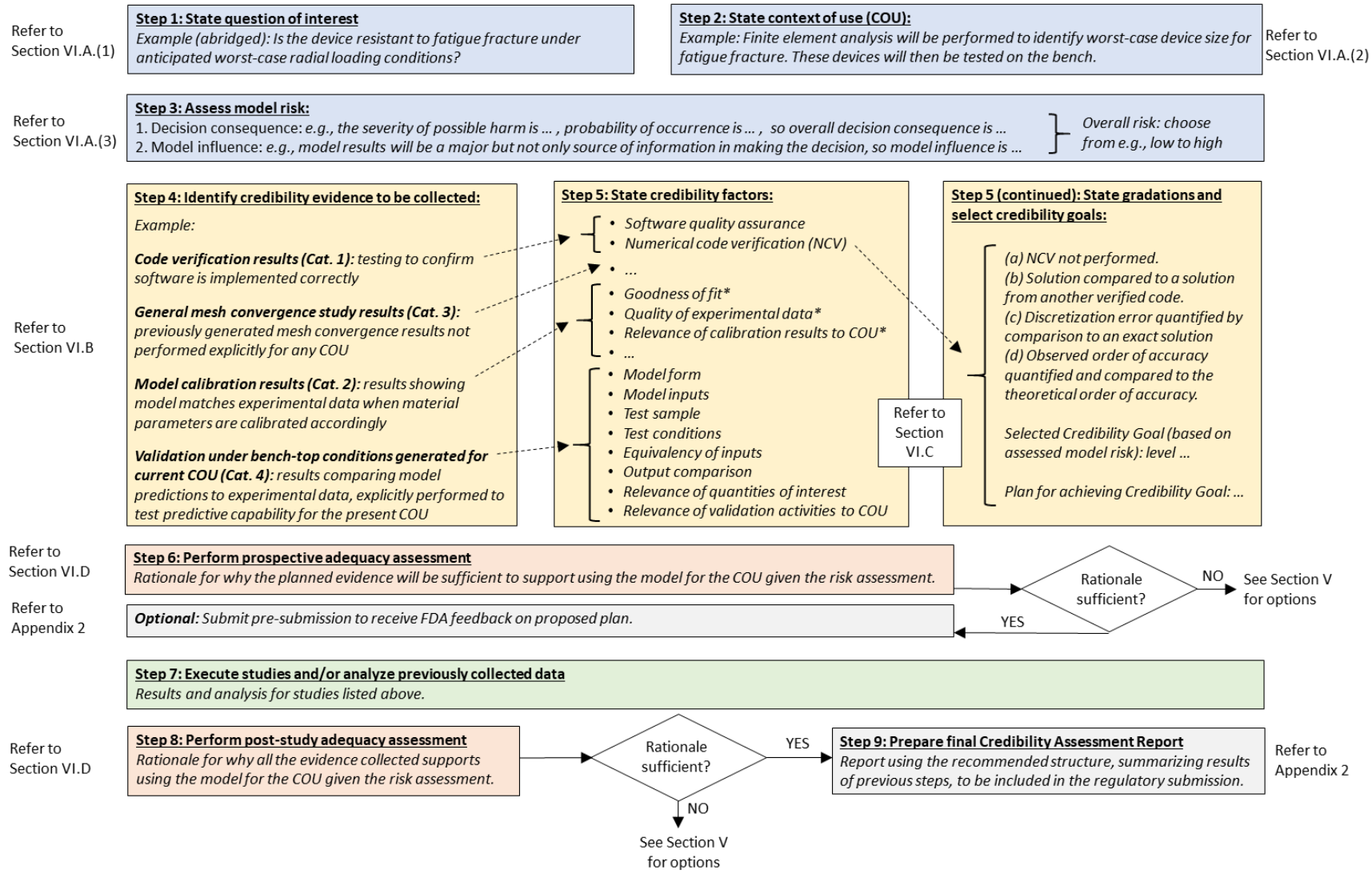
1. Describe the **question(s) of interest** to be addressed in the regulatory submission that will be informed by the computational model. See Section VI.A.(1) for details.
2. Define the **context of use** (COU) of the computational model. See Section VI.A.(2) for details.
3. Determine the **model risk**. See Section VI.A.(3) for details.
4. Identify and categorize the **credibility evidence**, either previously generated or planned, which supports 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 and set prospective **credibility goals** for each credibility factor, with a plan to achieve these goals. See Section VI.C for a discussion of credibility factors 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" [28]) 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 discussion on options. If any changes are made at this stage, go back to Step 2.
9. Prepare a report on the credibility of the CM&S for inclusion in the regulatory submission. See Appendix 2 for **reporting recommendations.**

FDA is recommending 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

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

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.

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.



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 "describe the **question(s) of interest** to be addressed in the regulatory submission that will be informed by the computational model." The question of interest is defined in ASME V&V 40 as "the specific question, decision, or concern that is being addressed." 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.

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 the model predictions are used in. For example, for a device which performs patient-specific simulation 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 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." The COU of the model is defined as the specific role and scope of the computational model used to address the question of interest.[29] The COU should include a detailed description of what will be modeled and how model outputs will be used to answer the question of interest, including a statement on whether other information (e.g., bench testing, animal 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.

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 is completely distinct from the indications for use or intended use of the device.

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 as a non-clinical assessment model (NAM), 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"[30]).

(3) Model risk

Step 3 of the framework is to "determine the **model risk.**" Model risk is defined as "the possibility that the computational model and the simulation results may lead to an incorrect decision that would lead to an adverse outcome."[31] 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. ASME V&V 40 recommends assessing model risk based on two factors, **model influence** and **decision consequence**.

[29] ASME V&V 40 *Assessing Credibility of Computational Modeling through Verification and Validation: Application to Medical Devices*

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

[31] ASME V&V 40 *Assessing Credibility of Computational Modeling through Verification and Validation: Application to Medical Devices*

Model influence is the contribution of the computational model relative to other contributing evidence 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 made relative to the empirical fatigue test observations.

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, although in some cases, impact on the clinician may also be considered. For example, when evaluating decision consequence for the aforementioned device durability COU, you should consider the potential patient harms that could result in the event the implanted device fractures.

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. For example, 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 including any risk mitigations procedures in place.

In general, we recommend assessing decision consequence by considering both the potential severity of harm *and* the probability of occurrence of harm following an appropriate risk management procedure (e.g., see ISO 14971[32] and ISO/TR 24971[33]). 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. When possible, reports of adverse events for the same or similar device types can be helpful in identifying these potential hazards and harms. The overall decision consequence should be assessed by considering all potential harms that may occur due to an incorrect decision, accounting for any risk mitigation procedures in place.

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

- Model influence will be dependent on whether other information (e.g., bench or animal test results) are also provided in the regulatory submission to address the question of interest.

[32] ISO 14971 *Medical devices — Application of risk management to medical devices*
[33] ISO/TR 24971 *Medical devices — Guidance on the application of ISO 14971*

- 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 'controlling' (i.e., 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, but the recommendation is intended to be used in conjunction with other medical information to make a clinical decision, we recommend you consider if there is reasonably foreseeable misuse related to the degree clinicians may rely on the device output without considering additional clinical information that may be available.
- 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 first principles-based computational models used in software as a medical device (SaMD), the risk categorization framework in FDA's guidance titled "Software as a Medical Device (SAMD): Clinical Evaluation"[34] can also be used to inform assessment of decision consequence.

For models submitted for MDDT qualification:

- If the MDDT is a computational model only, model influence is expected to be 'controlling' (i.e., 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 supports 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 VVUQ activities. Because of this, we adopt the more general term of "credibility evidence," which is any evidence that could support the credibility of a computational model. In Table 1 below, ten distinct categories of credibility evidence are provided along with definitions. The objective of defining these categories is to

[34] https://www.fda.gov/regulatory-information/search-fda-guidance-documents/software-medical-device-samd-clinical-evaluation

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 1, 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 1: Ten categories of credibility evidence. Categories 1, 4 and 5 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	General non-COU evidence	Calculation verification and/or validation evidence gathered for the model under conditions that are broad and not specific to the COU.
4	Evidence generated using bench-top conditions to support the current COU	Calculation verification and/or validation evidence using bench-top conditions, that was explicitly planned and generated to support the current COU.
5	Evidence generated using *in vivo* conditions to support the current COU	Same as previous category except using *in vivo* conditions.
6	Evidence generated using bench-top conditions to support a different COU	Calculation verification and/or validation evidence using bench-top conditions, that was planned and generated to support a different COU.
7	Evidence generated using *in vivo* conditions to support a different COU	Same as previous category except using *in vivo* conditions.
8	Population-based evidence	Statistical comparisons of population-level data between model predictions and a clinical data set. (Note: individual-level comparison between model predictions and a clinical dataset falls under Category 5.)
9	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.
10	Model plausibility	Evidence that supports the validity of the governing equations, model assumptions, and input parameters only.

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 availability and the challenge of gathering enough 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, 5, 6 or 7); and
- validation (Categories 3, 4, 5, 6, 7 or 8) or other evidence pertaining to the model's ability to reproduce real-world behavior (Categories 2, 9, 10).

You can also submit multiple types of evidence within each group (e.g., submitting Category 3, 4 and 8 results) 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.

(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. Code verification is important to demonstrate that there are no bugs in the software that affect simulation accuracy. It does not need any comparison of model predictions with real-world data.

Example:

- For solid mechanics, fluid dynamics, 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[35]), including confirmation that the error converges to zero at the expected convergence rate as spatial and temporal discretization size are decreased.

(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

[35] 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

calibrated (whether optimized or manually tuned) to minimize the discrepancy between model results and data. Nevertheless, robust model calibration evidence can still support model credibility. This type of evidence is strongest 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 to match the data, or if many parameters were calibrated.

Examples:

- In solid mechanics, demonstrating that a constitutive model of a material closely matches a test specimen's measured stress-strain behavior, after calibrating constitutive parameters to minimize the discrepancy.
- In physiological modeling, demonstrating that a personalized 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 modeling tissue heating *in vivo*, demonstrating that the first principles-based bioheat transfer model accurately predicts/estimates relevant spatio-temporal *in vivo* tissue heating in appropriate tissue types, after the blood-tissue heat transfer related coefficient has been calibrated based on the heating (i.e., relevant spatio-temporal temperature distribution).

(3) General non-COU evidence

General non-COU evidence is calculation verification and/or validation evidence gathered for the model under conditions that are broad and not specific to the COU. This category refers to evidence that was not generated for any specific COU but could support credibility of the model for the current COU. Typically, the evidence will be **general validation evidence.** This category is especially relevant to general-purpose or multi-application computational models (e.g., some simulation software packages) for which it is common to compare model predictions under a variety of conditions with experimental data, for example, comparison to relevant benchmark data to demonstrate reliability of the model. This category is also especially relevant to computational models of physiological systems, where it is common to demonstrate the ability to reproduce the range of physiological behaviors when publishing or releasing the model. General validation results are also often utilized when complex models are validated in a hierarchical manner, using simple benchmark validation cases before considering potentially more involved COU-specific validation.

Examples:

- 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. These are general validation results because they were not generated for any specific COU. A manufacturer of a physiological closed-loop control (PCLC) device that uses the model in *in silico* testing of the control algorithm could potentially utilize the

previous general validation results to support the model credibility in a PCLC testing COU.

- In fluid dynamics, comparing simulations with classical wind tunnel measurements (e.g., flat-plate boundary layer, lift and drag on objects) or other non-COU benchmark experiments designed for validation (e.g., a benchmark nozzle[36]). As part of this, calculation verification studies are performed to estimate the numerical uncertainty in the simulation predictions.

(4) Evidence generated using bench-top conditions to support the current COU

This category refers to calculation verification and/or validation evidence generated using bench-top conditions explicitly to support the current COU. There are two features of this category:

i. "Bench-top conditions," which means that the verification and/or validation activities were performed using conditions that reflect bench-top testing and not clinical or animal testing (for those see Category 5 below). However, the COU could be either bench-top or *in vivo*; see examples below.

ii. "To support the current COU," which means that the verification and/or validation evidence was explicitly planned and generated to support the credibility of the model for the current COU (as opposed to a different COU; see Category 6).

In many cases, this category of evidence will align closely with the verification and/or validation evidence described in ASME V&V 40.

Examples:

- In the following example, both the COU and the validation simulations correspond to bench-top testing:
  - In solid mechanics, a manufacturer of a new family of peripheral stents plans to perform benchtop durability testing to assess fatigue resistance. A computational model of the stent family is developed, and simulations of the bench test are used to identify worst-case stent sizes to minimize the number of physical experiments. Calculation verification and validation evidence are generated by performing finite element simulations of radial loading for a subset of the stents using multiple mesh resolutions and comparing predicted and measured force-displacement relationships.

- In the following example, the COU corresponds to *in vivo* conditions but the validation simulations correspond to bench-top testing:

[36] 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.

- 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. Energy absorption during MR scanning will be predicted. 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.

(5) Evidence generated using *in vivo* conditions to support the current COU

This category refers to calculation verification and/or validation evidence generated using *in vivo* conditions that is explicitly generated to support the current COU. There are two features of this category.

i. "*In vivo* conditions," which means that the verification and/or validation activities were performed using conditions that reflect representative *in vivo* animal or human use.

ii. "To support the current COU," which means that the verification and/or validation evidence was explicitly planned and generated to support the credibility of the model for the current COU (as opposed to a different COU; see Category 7). This category applies to patient-level validation of a patient-specific computational model. For example, a clinical trial evaluating the performance of SaMD that uses patient-specific computational simulation falls under this category.

Examples:

- In fluid dynamics, a clinical software tool, which uses a patient-specific model of the coronary arteries to predict the fractional flow reserve, is validated by comparing simulations against invasive measurements of fractional flow reserve in the same patient. Also, a calculation verification study is performed to estimate the numerical uncertainty in these simulation predictions.
- A manufacturer develops a computational model-based tool that predicts if a patient will respond positively to proposed therapy, and 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 characteristics (ROC) curve.
- In heat transfer, a first principles-based thermal model is validated to predict relevant spatio-temporal *in vivo* tissue heating using humans and/or animal models for a known spatio-temporal distribution of *in vivo* power density in appropriate tissue.

(6) Evidence generated using bench-top conditions to support a different COU

This category refers to calculation verification and/or validation evidence generated using bench-top conditions that is generated to support a different COU. This category is the same as Category 4 except that the evidence was planned and generated to support a different COU. This category is relevant to situations where model development, verification and validation using bench-top conditions were successfully performed for one COU ('COU1'), and later the same model is used for a new COU ('COU2'). In this case, the verification and validation results for COU1 may be able to support the model for COU2. This would streamline the verification and validation activities for COU2. However, the evidence is expected to have less relevance (i.e., applicability) as compared to the evidence from Category 4.

Examples:

- In solid mechanics, a manufacturer developed a computational model of a family of peripheral stents, estimated the numerical uncertainty by performing a calculation verification study, 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 which will undergo durability testing to assess fatigue safety. Subsequently the manufacturer seeks a new indication for the same stents in different vasculature and a computational model of the stents in new loading conditions is developed. The previously collected calculation verification and validation results may be able to support the credibility of the model in the new loading conditions in the new vasculature.
- In electromagnetics, a computational model of MR-induced thermal heating of an implantable device was 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 be able to support the model for this new COU for known transmit coil configurations.

(7) Evidence generated <u>using *in vivo* conditions</u> to support a <u>different COU</u>

This category refers to calculation verification and/or validation evidence generated using *in vivo* conditions that was generated to support a different COU. This category is the same as Category 5 except the evidence was planned and generated to support a different COU. This category is relevant to situations where model development, verification and validation using *in vivo* conditions were successfully performed for one COU ('COU1'), and later the same model is used for a new COU ('COU2'). In this case, the verification and validation results for COU1 may be able to support the model for COU2. This would save the expense of performing new verification and validation activities for COU2. However, the evidence is expected to have less relevance (i.e., applicability) as compared to the evidence from Category 5.

Examples:

- In the examples for Category 6, the previous and new COUs involved different indications for use of the same device. Alternatively, the COUs could correspond to different versions of similar devices, as in the following examples:
    - In solid mechanics, a manufacturer uses a software platform to compute the device mechanics for one device (e.g., shoulder arthroplasty) under simulated *in vivo* conditions (e.g., rotations), performs a calculation verification study, and validates the predictions against relevant *in vivo* data. Later, the manufacturer wishes to use the same software for a different device (e.g., reverse shoulder arthroplasty). The previous calculation verification and validation evidence may be able to support the credibility of the new device model.
    - In heat transfer, a first principles-based thermal model is validated to predict relevant spatio-temporal *in vivo* tissue heating using humans and/or animal models for a known spatio-temporal distribution of *in vivo* power density in appropriate tissue. If the nature of the spatio-temporal temperature distribution (i.e., magnitude and gradients in space and time) is comparable between two devices for the full range of device specifications, the previous validation evidence may be able to support the credibility of the new device model for comparable indications for use.

(8) Population-based evidence

Population-based evidence consists of statistical comparisons of population-level data between model predictions and a clinical data set. 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 5). 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 5.

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.[37]

- In drug development, a large number of physiologically-based pharmacokinetic models are developed to simulate pharmacokinetic properties (e.g., plasma concentration as a function of time) of a drug across the population. Data from clinical trials can be used to validate the model. Model predictions of average response can be compared with study results for various subject populations (e.g., healthy volunteers, patients, or special populations) and clinical conditions (e.g., different doses, dosing frequencies, or routes of administration).

(9) 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 data (therefore, this is not validation evidence); 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 hemodynamics model correctly predicts the onset of transitional or turbulent flow at conditions where such phenomena are expected. A SaMD manufacturer that uses this 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 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.

(10) Model plausibility

Model plausibility is solely supported by evidence of the validity of the governing equations, model assumptions, and input parameters. A claim of model plausibility is an argument that "the

[37] 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.

model is credible" because the governing equations are expected to hold, assumptions are reasonable, 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 scientific 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 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 epidemiology, a susceptible-infectious-recovered (SIR) model of a novel infectious disease is developed. It is not possible to validate model predictions against data on the actual number of infected individuals since it has not spread widely enough yet. Credibility of the model predictions is then based primarily on belief in the validity of the governing equations (which may be supported by historical validation of SIR models for other outbreaks), together with evidence that the model parameters (e.g., basic reproduction number, infection rate, recovery rate) have been accurately measured for the new disease.

C. Credibility Factors and Credibility Goals

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

Credibility factors are fundamental aspects of the credibility assessment process that break down the analysis of verification, validation, or other sources of non-traditional credibility evidence. For 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 that break down calculation verification, validation and applicability.

To establish creditability factors and credibility goals, we recommend the following process. Refer to Figures 1 and 2 for examples.

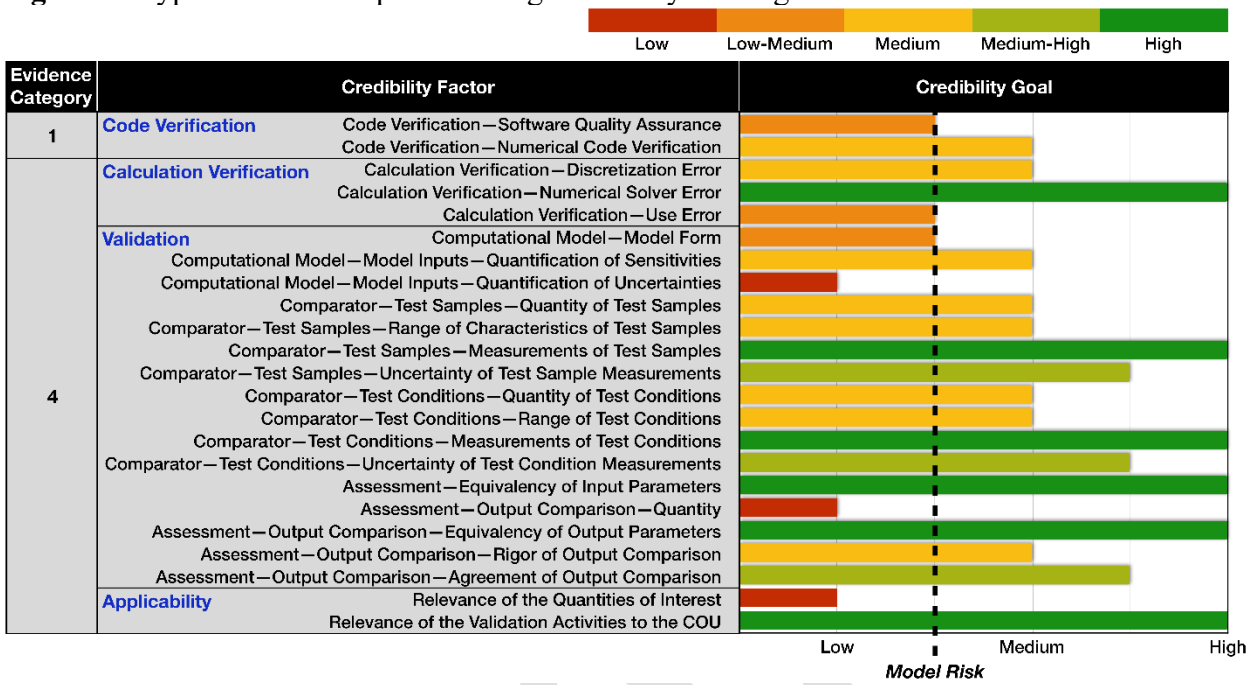
- Step 5.1: State credibility factors relevant to the type of credibility evidence you plan to gather. When relevant, we recommend using ASME V&V 40 credibility factors. For example, if you plan to gather 'validation evidence generated using bench-top conditions to support the current COU' (Category 4), we recommend using ASME V&V 40 credibility factors related to validation and applicability. For non-traditional VVUQ evidence categories that are not explicitly covered by ASME V&V 40 (e.g., model calibration evidence, population-level evidence, or model plausibility – Categories 2, 8, or 10, 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., 'validation evidence generated using bench-top conditions to support the current COU' as primary and 'general non-COU 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. See Figure 1.
- Since the relevance of the evidence to support using the model for the COU is especially important, we recommend defining 'applicability' credibility factors for each set of credibility evidence (as emphasized in Appendix 1 and illustrated in Figure 1).

- Step 5.2: Following ASME V&V 40, for each credibility factor, define a gradation of activities that describes progressively increasing levels of 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) estimated and accounted for in the quantitative error of fit.
- Step 5.3: Following ASME V&V 40, for each credibility factor, select a 'credibility goal' from the gradation, based on the model risk as assessed in Step 3. Higher risk questions of interest 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. For credibility factors for which the goal is less than the level commensurate with model risk (see Figure 2), for example, due to practical constraints, a rationale should be provided to explain why the activities are sufficient for overall model credibility.
- Step 5.4: For each credibility factor, 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.

Figure 2 presents a hypothetical example of this process. In this example, two types of credibility evidence are planned, 'Code Verification Results' (Category 1) and 'Evidence Using Benchtop Conditions to Support the Current COU' (Category 4). In this example, the Category 4 evidence includes both calculation verification and validation 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 2. 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 shown in red), a rationale should be provided for why the activities are sufficient to support overall model credibility.

Figure 2: Hypothetical example of setting credibility factor goals.



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, and 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*? In contrast to *model accuracy,* which is quantifiable through validation, *model adequacy* warrants a careful decision to be made based using engineering and clinical judgement, based on all available information.[38]

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

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

the planned credibility evidence is expected to be sufficient to support using the model for the COU, given the risk assessment.

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. We recommend that you take into consideration the following questions and recommendations in 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 activities such as code verification, calculation verification, sensitivity analysis, uncertainty quantification all considered at some point of the planning of credibility assessment activities? If not, we recommend that you clearly justify not performing these credibility activities based on the model risk (see Section VI.C).
- Were the credibility goals met? If the goal was not met for a factor or multiple factors, this means it was not possible to perform the analysis at the desired level of rigor. In this case, to support the use of the model, we recommend that you provide a justification regarding the impact of the affected credibility factor(s) on the risk associated with using the model to address the question of interest.

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. However, you should still provide a scientific 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 model validation accuracy targets. In this case, we recommend you state 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, as part of the adequacy assessment, we recommend considering the model predictions of the COU quantity(ies) of interest 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, model accuracy from the validation comparison) and any potential 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 far below unacceptable 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 far below unacceptable levels. This information could be used to further justify the adequacy of the model predictions 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, you should modify the model, reduce the model influence, modify the COU, and/or revise the plan to generate credibility evidence (prospective adequacy assessment) or collect 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 credibility factors (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 software that forms part of a medical device, testing performed for software verification will likely encompass code verification of the computational model. 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"[39] 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.

Category 2: Model calibration evidence

- For credibility factors (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;
    - any steps taken to ensure the model is not overfitted; and
    - numerical methods for obtaining the calibrated results.
- If no validation results are available and calibration results are the primary source of evidence for model credibility, consider evaluating the relation between calibration

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

conditions and COU conditions, and between calibration quantities of interest and COU quantities of interest.

Category 3: General non-COU evidence

- For credibility factors (Step 5 of the framework), if the evidence is traditional calculation verification or validation evidence, we recommend using credibility factors defined in ASME V&V 40.
- For general non-COU validation evidence, we recommend paying special attention to the applicability of the general validation evidence used to support the current COU. This should include an assessment of any differences, and impact thereof, between the model used in the general non-COU evidence and the model used in the current COU.

Category 4: Evidence generated using bench-top conditions to support the current COU

- For credibility factors (Step 5 of the framework), if the evidence is traditional calculation verification and/or validation evidence, 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-top validation results to the *in vivo* COU.

Category 5: Evidence generated using in vivo conditions to support the current COU

- For credibility factors (Step 5 of the framework), if the evidence is traditional calculation verification and/or 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., good clinical practices, appropriate statistical techniques, appropriate measures of sensitivity and specificity, positive predictive value), and define appropriate credibility factors for Step 5 of the framework.

Category 6: Evidence generated using bench-top conditions to support a different COU

- For credibility factors (Step 5 of the framework), if the evidence is traditional calculation verification and/or validation evidence, we recommend using credibility factors defined in ASME V&V 40.
- 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 impact thereof, between the model used for the previous COU compared to the model used for the current COU. Also, if your COU will involve making *in vivo* predictions, we recommend paying special attention to the applicability of the bench-top validation results to the *in vivo* COU.
- Consider performing analysis to confirm that the computational model made reliable predictions for the previous COU based on current knowledge of the device performance postmarket. For example, if a computational model was previously validated and used for a device safety COU, but the device was recalled due to safety concerns postmarket

related to that COU, then the computational model may not be appropriate for a new COU involving a new version of the device.

Category 7: Evidence generated using in vivo conditions to support a different COU

- For credibility factors (Step 5 of the framework), if the evidence is traditional calculation verification and/or 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., good clinical practices, appropriate statistical techniques, appropriate measures of sensitivity and specificity, positive predictive value), and define appropriate credibility factors for Step 5 of the framework.
- 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 impact thereof, between the model used for the previous COU compared to the model used for the current COU.
- Consider performing analysis to confirm that the computational model made reliable predictions for the previous COU based on current knowledge of the device performance postmarket. For example, if a computational model was previously validated and used for a device safety COU, but the device was recalled due to safety concerns postmarket related to that COU, then the computational model may not be appropriate for a new COU involving a new version of the device.

Category 8: 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 evaluate and compare the demographics (including sex, age, race and ethnicity), 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 9: Emergent model behavior

- As this is a relatively weak form of demonstrating credibility, we generally do not recommend relying on emergent model behavior as a primary source of evidence for model credibility. In this case, consider strengthening the evidence by quantitatively comparing model predictions to clinical or experimental data rather than comparing model predictions against qualitative knowledge about the system (in which case the evidence would change category and no longer be emergent model behavior 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 credibility factors (Step 5 of the framework), we recommend that you define factors for the relevance of the emergent behavior to the COU, sensitivity of emergent behavior to model input uncertainty, and others.

Category 10: Model plausibility

- As discussed in Section VI.B, model plausibility is a relatively weak argument for model credibility because it does not involve testing the model predictions. Therefore, if model plausibility evidence is the main credibility evidence presented, you should provide a rationale for why validation testing of the model is not possible or warranted, for example, referring to 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 credibility factors (Step 5 of the framework), you should use ASME V&V 40 credibility factors related to model form and model inputs.

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"[40]). 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

[40] 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 silico* device testing or *in silico* clinical trials, refer to FDA's guidance titled "Reporting of Computational Modeling Studies in Medical Device Submissions"[41] (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 other section in the regulatory submission.

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

3. **Device Description:** Include within the report or refer to another section in regulatory submission.
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.
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
            3. Achieved credibility level.
        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. **Limitations**
8. **Conclusions**
9. **References**
10. **Appendices:** Detailed descriptions of credibility assessment study methods and results (if needed).