Workflow for Assessing the Credibility of Patient-Specific Modeling in Medical Device Software

Tool Reference

RST Reference Number: RST24CM01.01

Date of Publication: 03/29/2024

Recommended Citation: U.S. Food and Drug Administration. (2024). *Workflow for Assessing the Credibility of Patient-Specific Modeling in Medical Device Software* (RST24CM01.01). https://cdrh-rst.fda.gov/workflow-assessing-credibility-patient-specific-modeling-medical-device-software

For more information

Catalog of Regulatory Science Tools to Help Assess New Medical Devices

Disclaimer

About the Catalog of Regulatory Science Tools

The enclosed tool is part of the Catalog of Regulatory Science Tools, which provides a peer-reviewed resource for stakeholders to use where standards and qualified Medical Device Development Tools (MDDTs) do not yet exist. These tools do not replace FDA-recognized standards or MDDTs. This catalog collates a variety of regulatory science tools that the FDA's Center for Devices and Radiological Health's (CDRH) Office of Science and Engineering Labs (OSEL) developed. These tools use the most innovative science to support medical device development and patient access to safe and effective medical devices. If you are considering using a tool from this catalog in your marketing submissions, note that these tools have not been qualified as Medical Device Development Tools and the FDA has not evaluated the suitability of these tools within any specific context of use. You may request feedback or meetings for medical device submissions as part of the Q-Submission Program.

For more information about the Catalog of Regulatory Science Tools, email RST_CDRH@fda.hhs.gov.

Workflow for Assessing the Credibility of Patient-Specific Modeling in Medical Device Software

This user manual provides an example workflow for assessing the credibility of patient-specific computational models implemented in medical device software.

Background

Patient-specific models (PSMs) are computational models that have been personalized to a specific patient and can be used to make predictions about that patient. Patient-specific models can be incorporated into medical device software. For example, medical device software tools have been developed that receive patient data as inputs, construct and solve a PSM, and then provide patient-specific predictions to the clinician. Other applications of PSMs, outside the scope of this document, are to predict performance of a medical device that interacts with a patient (e.g., an implant). For an overview of patient-specific modeling for medical devices, see [3].

In 2023, FDA's Center for Devices and Radiological Health (CDRH) released a guidance document, *Assessing the Credibility of Computational Modeling and Simulation in Medical Device Regulatory Submissions - Guidance for Industry and Food and Drug Administration Staff,* that presented a framework for evaluating the credibility of computational models in medical device regulatory submissions [1]. The guidance document builds upon the approach used in ASME V&V40 2018 [2], a consensus standard which describes a risk-informed framework for evaluating the credibility of computational models with medical device applications. ASME V&V40 was developed with a focus on models of medical devices validated against bench test data, and there are challenges in applying ASME V&V40 to PSMs validated using clinical data. To address these challenges, the guidance document [1] provides a more general framework based around eight categories of credibility evidence which could be used to support credibility of the computational models. To follow the framework in [1], users should state what type of credibility evidence will be collected, define 'credibility factors' for each category of evidence, and for each credibility factor, define a 'credibility gradation' of activities of increasing level of rigor. See [1] for discussion of these terms.

ASME V&V40 defines a series of credibility factors and provides example gradations for each, but states that users should define their own gradations as applicable. The framework outlined in the guidance document is more flexible, and does not prescribe credibility evidence, factors, or gradations, allowing the user to define their own as needed. This flexibility ensures the guidance document may be applicable to the wide range of computational models and credibility evidence possible in medical device regulatory submissions. However, there is a need for examples of how to use the guidance document for specific types of models, including lists of relevant credibility factors and potential gradations.

Aim and Scope

The aim of this user manual is to provide an example workflow for assessing the credibility of PSMs in medical device software following the framework outlined in the FDA guidance document, *Assessing the Credibility of Computational Modeling and Simulation in Medical Device Regulatory Submissions - Guidance for Industry and Food and Drug Administration Staff*. The workflow is based on the findings of [3], a journal article focused on credibility assessment of PSMs, which identified unique considerations for evaluating PSMs and discussed how ASME V&V40 could be used for PSMs. The present document extends the findings of [3] to be consistent with the framework outlined in the guidance document.

The scope of this document is PSMs in medical device software. For PSMs used to predict performance of a medical device that interacts with a patient, see the related Regulatory Science Tool "*Example Workflow for Assessing the Credibility of a Medical Device In Silico Clinical Trial*".

The workflow presented below is intended as a useful resource for model developers when evaluating a PSM implemented in medical device software. It is an example workflow only, intended as a starting point for PSM developers following the guidance document. It is not intended to be a comprehensive framework applicable to all PSMs or all conceivable approaches to PSM credibility assessment. Therefore, the workflow may need to be adjusted according to the specific details of the model and planned credibility assessment approach.

Workflow Assumptions

The workflow below assumes the following:

- PSM modeling software has been developed and implemented in medical device software
- Patient medical imaging data or other patient data is used as model input.
- The device may use patient data to construct a three-dimensional representation of the anatomical region of interest. This step may be fully automated (no user decision-making required) or semi-automated (some manual stages required by the user, for example identifying region boundaries in the images or choosing seed points for segmentation).
- The device then solves ordinary differential equations (ODEs) or partial differential equations (PDEs) to compute a quantity of interest, which is output to the user.
- The PSM output is either an established clinical parameter of interest or has been shown to be associated with an established clinical parameter of interest.
- Activities such as parameter sensitivity analysis (for example, to determine which parameters should be personalized) and comparison of different model forms (i.e., comparing different governing equations) were performed during the model development phase rather than credibility assessment phase.
- The patient-specific model will be validated using data derived from a clinical study.

PSM Credibility Assessment Workflow

Given the above assumptions, one possible workflow for PSM credibility assessment, consistent with the general framework in the FDA guidance document [1], is:

1. State the Question of Interest – see FDA guidance document Step 1 for details
2. Describe the Context of Use – see FDA guidance document Step 2 for details
3. Assess Model Risk – see FDA guidance document Step 3 for details
4. Plan to the following activities to generate model credibility evidence:

   Code verification results (Category 1)

   a) **Software quality assurance**. See [2] for discussion.
   b) **Numerical code verification**. See [2] for discussion.

   In vivo validation results with supporting calculation verification and UQ results (Category 4)

   c) **Identify or generate a clinical validation dataset:**
   Identify or plan a clinical study to obtain a dataset for model validation. The dataset will be comprised of data from *N* subjects. Data for each subject will be split into model personalization input data and validation data.
   d) **Generate a set of patient-specific models:**
   Generate *N* patient specific models using the data collected in 4C
   e) **Discretization error analysis:**
   Using a subset (or all) of the patient-specific models generated in Step 4D, perform a mesh convergence analysis. Also perform a time-step convergence analysis if time- dependent governing equations are solved.
   f) **Numerical solver convergence analysis:**
   Using a subset (or all) of the patient-specific models generated in Step 4D, perform a convergence analysis of the numerical solver tolerance(s)
   g) **Use error assessment:**
   See [2] for discussion. The aim of the use error assessment should be to confirm that no errors were made in the validation simulations.
   h) **Inter- and intra-user variability assessment**:
   This step is only applicable to PSMs that are only semi-automated with some manual stages as discussed above. Assess if user variability in the manual stages can lead to different model predictions.
   i) **Uncertainty quantification (UQ):**
   Using a subset (or all) of the patient-specific models generated in Step 4D, assess the impact of uncertainty in model inputs (e.g., uncertainty in personalized inputs due to measurement error) on model outputs. Note that PSM UQ may not be feasible due to lack of data or computational cost in which case a justification for not performing UQ could be provided instead.

j) **Validation**

Compare model output(s) for each subject with that subject's data. The appropriate method of comparison will depend on the type of output. For example, if the output is a binary quantity (such as treatment success or presence of disease), overall accuracy could be quantified by computing sensitivities, specificities, and area under receiver operator characteristics (ROC) curve. If the output is a continuous scalar quantity (such as displacement magnitude or pressure), correlation or Bland- Altmann plots may be appropriate. Other methods may instead be appropriate, depending on the application and data type.

5. Define credibility factors and gradations for the planned credibility evidence, and evaluate the planned activities using the gradations, as discussed in Step 5 of the Guidance. To aid in this task, Tables 1 and 2 lists possible credibility factors and gradations that could be used or adapted with PSMs, for the Categories of evidence collected in Step 4.
6. Assess Adequacy of the planned activities – see FDA guidance document Step 6 for details.
7. Execute studies.
8. Assess Adequacy of the overall results – see FDA guidance document Step 8 for details.

Table 1: Credibility factors and gradations that could be used for code verification results (Category 1) of a patient specific model in medical device software. See Step 5.

Credibility factor	Credibility gradation	Gradation source[1]
Software quality assurance (SQA)	(a) Very little or no SQA procedures were specified or followed. (b) SQA procedures were specified and documented. (c) SQA procedures were specified and documented; the software anomaly list and the software development environment were fully understood, and the impact on the COU was analyzed and documented; quality metrics were tracked.	ASME V&V40 [2]
Numerical code verification (NCV)	(a) NCV was not performed. (b) The numerical solution was compared to an accurate benchmark solution from another verified code. (c) Discretization error was quantified by comparison to an exact solution, and a grid convergence study demonstrated that the numerical solution asymptotically approached the exact solution as the discretization was refined. (d) In addition to the quantification of discretization error and the execution of a grid convergence study as described in (c), the observed order of accuracy was quantified and compared to the theoretical order of accuracy.	ASME V&V40 [2]

[1] Gradations from [2] reprinted from ASME V&V 40-2018, by permission of The American Society of Mechanical Engineers. All rights reserved.

Table 2: Credibility factors and gradations that could be used for in vivo *validation with supporting calculation verification and UQ results (Category 4) of patient specific model in medical device software. See Step 5.*

Sub-category	Credibility factor	Credibility gradation	Gradation source[2]
Calculation verification	Discretization error	(a) No grid or time-step convergence analysis was performed to estimate the discretization error. (b) Applicable grid or time-step convergence analyses were performed, and discretization error was estimated, using one or a small number of patients. (c) Applicable grid or time-step convergence analyses were performed, and discretization error was estimated using a range of representative patients	Adapted from [3]
Calculation verification	Numerical solver error	(a) No solver parameter sensitivity was performed. (b) No solver parameter sensitivity was performed. Solver parameters were established based on values from a previously verified computational model. (c) Problem-specific sensitivity study was performed on solver parameters, confirming that changes in simulation results due to changes in the solver parameters were negligible relative to the model accuracy goal, for one or a small number of patients only. (d) As (c), except with a range of representative patients.	[3]
Calculation verification	Use error	(a) Inputs and outputs were not verified. (b) Key inputs and outputs were verified by the practitioner. (c) Key inputs and outputs were verified by internal peer review. (d) Key inputs and outputs were verified by reproducing simulations as part of an external peer review.	ASME V&V40 [2]
UQ	Model form	(a) Influence of model form assumptions was not explored. (b) Influence of expected key model form assumptions was explored. (c) Comprehensive evaluation of model form assumptions was conducted.	ASME V&V40 [2]
UQ	Input UQ – inputs analyzed	(a) No UQ (b) Expected key personalized inputs analyzed (c) Expected key personalized and non-personalized inputs analyzed (d) Wide range of personalized and non-personalized inputs analyzed	[3]
UQ	Input UQ – rigor of input uncertainty characterization	(a) No UQ (b) Arbitrary choices, e.g., +/- 10% (c) Crude characterization of input uncertainty (d) Precise characterization of input uncertainty but correlation between inputs neglected. (e) Precise characterization of input uncertainty with correlations characterized	[3]

UQ	Input UQ - number of patients	(a) No UQ (b) UQ performed on one patient only (c) UQ performed on small number of patients only (d) UQ performed on large number of patients covering patient population	[3]
UQ	Input UQ – output quantities	(a) No UQ (b) Uncertainty in inputs propagated to compute uncertainty in an output that differs from tool output (c) Uncertainty in inputs propagated to compute uncertainty in the tool output	[3]
UQ	User variability	(a) No inter- or intra-user variability analysis (b) Intra-user variability analysis was performed (c) Inter- or intra-user variability analysis was performed	[3]
Validation	Number of validation subjects	(a) Single subject (b) Multiple subjects, not enough to be statistically relevant (c) Statistically relevant number of subjects	[3]
Validation	Range of characteristics of validation subjects	(a) All validation subjects similar (b) Limited range of characteristics in validation subjects (c) Wide range of characteristics in validation subjects	[3]
Validation	Patient data collected	(a) Key patient data missing *[e.g., because retrospective study]* (b) Most key patient data was obtained. (c) All key patient data was obtained.	[3]
Validation	Patient measurements	(a) Patient measurements were not characterized or were characterized with gross observations, and measurement uncertainty was not addressed. (b) Uncertainty analysis incorporated instrument accuracy only. (c) Uncertainty analysis incorporated instrument accuracy and repeatability (i.e., statistical treatment of repeated measurements). (d) Uncertainty analysis incorporated a comprehensive uncertainty quantification, which included instrument accuracy, repeatability, and other conditions affecting the measurements.	[3]
Validation	(Test condition factors)	Note: also consider whether ASME V&V40 2018 credibility factors related to “Test conditions” should be adapted given the planned validation activities	See discussion in [3]
Validation	Equivalency of Input Parameters	*Note: this factor considers the equivalency of inputs prescribed in validation simulations (e.g., loading conditions) and the conditions used in the clinical study. If there are no such inputs, this factor can be removed* (a) The types of some inputs were dissimilar. (b) The types of all inputs were similar, but the ranges were not equivalent. (c) The types and ranges of all inputs were equivalent.	ASME V&V40 [2]
Validation	Output Comparison – quantity	(a) A single output was compared. (b) Multiple outputs were compared.	ASME V&V40 [2]

Validation	Equivalency of Output Parameters	(a) Types of outputs were dissimilar. (b) Types of outputs were similar. (c) Types of outputs were equivalent.	ASME V&V40 [2]
Validation	Rigor of Output Comparison	*User should define an appropriate gradation, which will be dependent on the specific details of the model (see Step 4J)*	
Validation	Agreement of Output Comparison	(a) The level of agreement of the output comparison was not satisfactory for key comparisons. (b) The level of agreement of the output comparison was satisfactory for key comparisons, but not all comparisons. (c) The level of agreement of the output comparison was satisfactory for all comparisons.	ASME V&V40 [2]
Validation	Relevance of the QOIs	(a) The QOIs from the validation activities were not closely related to those for the COU (b) The QOIs from the validation activities were closely related, though not identical, to those for the COU (c) The QOIs from the validation activities were identical to those for the COU	Adapted from ASME V&V40 [2]
Validation	Relevance of the Validation Activities to the COU	(a) The validation subjects have limited relevance to patient population (b) The validation subjects are representative of the patient population except some sub- populations or extremal cases are not represented. (c) The validation subjects are highly representative of the patient population	New

[2] Gradations from [2] reprinted from ASME V&V 40-2018, by permission of The American Society of Mechanical Engineers. All rights reserved.

Comments

The above workflow assumes Code Verification (Category 1) and *In vivo* validation results (Category 4) will be collected. Other categories of credibility evidence can be relevant to PSMs, as summarized in Table 3. For example, PSM validation could be performed by comparing with bench tests using fabricated (e.g., 3D printed) patient-specific experimental models. In such cases, the above workflow needs to be adapted based on the evidence planned to be collected.

Table 3: Relevance of credibility evidence categories in [1] to PSMs in medical device software

#	Category	Relevance to PSMs in medical device software
1	Code verification	Relevant
2	Model calibration results	May be relevant to some PSMs (e.g., PSMs which compute values of personalized parameters via calibration against observed patient data), generally supporting separate validation results.
3	Bench test validation	Sometimes relevant, for example patient-specific blood flow models can be validated against experiments using fabricated vessels and aneurysms [4]
4	*In vivo* validation	Relevant
5	Population level validation	Unlikely to be relevant
6	Emergent phenomena	May be relevant to some PSMs, generally supporting separate validation results.
7	Model plausibility	Could be used to support credibility of sub-models within the PSM
8	Calculation verification/UQ results under COU conditions	Not relevant to pre-market credibility assessment of PSMs in medical device software, because the COU simulations are run after the device is on the market.

References

[1] FDA, "*Assessing the Credibility of Computational Modeling and Simulation in Medical Device Regulatory Submissions - Guidance for Industry and Food and Drug Administration Staff*", 2023

[2] ASME, *"ASME V&V40 2018, Assessing the Credibility of Assessing Credibility of Computational Modeling Through Verification and Validation: Application to Medical Devices"*, 2018

[3] Galappaththige et al., "*Credibility assessment of patient-specific computational modeling using patient-specific cardiac modeling as an exemplar*", PLOS Computational Biology, 2023

[4] Cito et al., "*Accuracy and Reproducibility of Patient-Specific Hemodynamic Models of Stented Intracranial Aneurysms: Report on the Virtual Intracranial Stenting Challenge 2011*", Annals of Biomedical Engineering, 2011.