WHITE PAPER

Consideration of a Credibility Assessment Framework in Model-Informed Drug Development: Potential Application to Physiologically-Based Pharmacokinetic Modeling and Simulation

Colleen Kuemmel[1,*], Yuching Yang[1], Xinyuan Zhang[1], Jeffry Florian[1], Hao Zhu[1], Million Tegenge[2], Shiew-Mei Huang[1], Yaning Wang[1], Tina Morrison[3] and Issam Zineh[1]

The use of computational models in drug development has grown during the past decade. These model-informed drug development (MIDD) approaches can inform a variety of drug development and regulatory decisions. When used for regulatory decision making, it is important to establish that the model is credible for its intended use. Currently, there is no consensus on how to establish and assess model credibility, including the selection of appropriate verification and validation activities. In this article, we apply a risk-informed credibility assessment framework to physiologically-based pharmacokinetic modeling and simulation and hypothesize this evidentiary framework may also be useful for evaluating other MIDD approaches. We seek to stimulate a scientific discussion around this framework as a potential starting point for uniform assessment of model credibility across MIDD. Ultimately, an overarching framework may help to standardize regulatory evaluation across therapeutic products (i.e., drugs and medical devices).

Quantitative modeling and simulation methods have become increasingly applied to facilitate drug development and regulatory decision making. These model-informed drug development (MIDD) approaches enable the prediction and understanding of drug pharmacokinetics (PK) and pharmacodynamics. In regulatory applications, they have been routinely used to optimize dosing, provide supportive evidence for efficacy, inform clinical design, and guide regulatory policy.[1] In some instances, quantitative models have served as primary evidence, proving especially useful when clinical trials are not feasible or ethical.[1] Overall, MIDD can be used to expand and accelerate patient access to safe and effective treatments.

A key aspect to the appropriate application of modeling and simulation in drug development and regulatory evaluation is ensuring model credibility. The term *credibility* refers to trust in the predictive capability of a computational model (hereafter referred to as "model") for a particular context of use. Several best practices for establishing confidence in specific quantitative models have been recommended.[2–6] Although best practices and regulatory experience have been used to develop guidance for long-standing MIDD approaches[7–11], other emerging approaches lack guidance. There is, however, no consensus among modeling and simulation approaches or regulatory authorities on how to establish or assess the credibility of a model for regulatory purposes.

An overarching framework for modeling and simulation in regulatory decision making was proposed for drug development.[12] However, there is additional need to consider an expanded framework that provides steps for establishing and assessing model credibility for regulatory decisions, perhaps irrespective of the therapeutic product (i.e., drugs or medical devices) being evaluated. Lack of a consistent evidentiary framework and variable interpretation and use of terminology in describing credibility assessments may hinder clear communication and understanding of regulatory expectations.

The American Society of Mechanical Engineers (ASME) published a standard that could be used by industry and regulatory agencies to assess the credibility of computational models used for medical device applications.[13] The standard does not prescribe specific activities or define criteria required to establish model credibility for a particular context or application. Acknowledging the range of potential applications of modeling and simulation, the standard instead provides a risk-based evidentiary framework for determining the rigor of evidence needed to rely on a model to inform decisions, assessing the adequacy of activities used to establish credibility, and evaluating overall model credibility. Although this framework was developed to assess medical device models, including physiological, engineering, and physics-based models, we hypothesize the framework could be translated to MIDD.

[1]Office of Clinical Pharmacology, Office of Translational Sciences, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, Maryland, USA; [2]Office of Biostatistics and Epidemiology, Center for Biologics Evaluation and Research, US Food and Drug Administration, Silver Spring, Maryland, USA; [3]Office of Science and Engineering Laboratories, Center for Devices and Radiological Health, US Food and Drug Administration, Silver Spring, Maryland, USA. *Correspondence: Colleen Kuemmel (Colleen.Kuemmel@fda.hhs.gov)

To this end, we applied the ASME framework to physiologically- based PK (PBPK) modeling and simulation as an illustrative example with the goal of stimulating a discussion about the utility of such an approach (or an alternative overarching framework) to standardize regulatory evaluation of a variety of models used in drug development.

BASIC PRINCIPLES OF THE RISK-INFORMED CREDIBILITY ASSESSMENT FRAMEWORK

Key concepts of the risk-informed credibility assessment framework are presented in this section. To ensure clarity, terminology defined in the ASME framework are used herein. A list of these definitions is provided in **Table 1**. A conceptual representation of the framework is presented in **Figure 1**.

Concept 1: State question of interest
The first step in applying modeling and simulation to support regulatory decisions is defining the question of interest. The question of interest presents the key question, concern, or decision of the study or development program. As such, the question of interest may be broader than the intended use of model.

Table 1 Key terminology in the risk-informed credibility assessment framework

Term	Definition
Applicability	Relevance of the validation activities to support use the computational model for a specific context of use
Comparator	Test data that are used for validation; may be data from *in vitro* or *in vivo* studies. Selection should be based on context of use
Context of use	Statement that defines the specific role and scope of the computational model used to address the question of interest
Credibility	Trust, established through the collection of evidence, in the predictive capability of a computational model for a context of use
Credibility factors	Elements of the verification and validation process, including applicability, used to establish credibility (listed in **Table 2**)
Decision consequence	Significance of an adverse outcome resulting from an incorrect decision
Model influence	Contribution of the computational model relative to other contributing evidence in making a decision
Model risk	Possibility that the computational model and the simulation results may lead to an incorrect decision and adverse outcome
Question of interest	The specific question, decision, or concern that is being addressed
Validation	Process of determining the degree to which a model or simulation is an accurate representation of the real world
Verification	Process of determining a model or simulation represents the underlying mathematical model and its solution from the perspective of the intended uses of modeling and simulation

Terms and definitions are specified from the American Society of Mechanical Engineers verification and validation 40.[13]

Concept 2: Define context of use
The context of use (COU) describes how the model will be used to address the question of interest, i.e., the specific role and scope of the model. The COU should include a description of additional data sources that will also be used to inform the question of interest (e.g., clinical data). In our experience, ambiguity in the question of interest and COU can result in (i) reluctance to accept modeling and simulation in a given drug development or regulatory review scenario or (ii) undesirably protracted dialogue between drug developers and regulators on the data requirements needed to establish model credibility. It is, therefore, critical to unambiguously and explicitly state the question of interest and how the proposed modeling and simulation approach will address it.

Concept 3: Assess model risk
Model risk is decided by (i) the weight of the model in the totality of evidence for a given decision, i.e., model influence; and (ii) the potential consequences of a wrong decision, i.e., decision consequence. Model influence and decision consequence are shaped by the COU, thus enabling model risk to be case specific.

Concept 4: Establish model credibility
Model credibility should be commensurate with model risk. As such, model risk drives the selection of credibility goals and activities. Credibility goals include desired qualitative or quantitative outcomes (e.g., prespecified acceptance criteria) based on scientific rationale. Credibility activities include verification of the software code and calculations, validation of the model using comparator studies, and evaluation of the applicability of validation assessments to the COU. Verification and validation (V&V) activities, including applicability, are divided into a total of 13 credibility factors (**Table 2**). The description and the rigor of the assessments described for each credibility factor are case specific as they depend on the COU and model risk, respectively. V&V activities and goals (i.e., acceptable outcomes) can be defined in a plan and executed to establish credibility.

Concept 5: Assess model credibility
Upon completion of credibility activities, an assessment can be made to determine if the model is sufficiently credible, considering the COU, risk, credibility goals, V&V results, and other knowledge acquired during the process. Based on the credibility assessment, a model may or may not be accepted for a given regulatory purpose. If, for example, the level of model credibility passes a bar for acceptance for a given COU, the model may be used for the proposed regulatory purpose (e.g., waiving a clinical trial; informing prescription drug labeling). If, however, the model credibility is not sufficiently established for the level of model risk, several outcomes are possible: (i) the model could be downgraded in terms of model influence, necessitating additional lines of evidence to support a regulatory decision; (ii) more data may be needed to increase the rigor of credibility activities or augment the model's output; (iii) the COU can be deemed
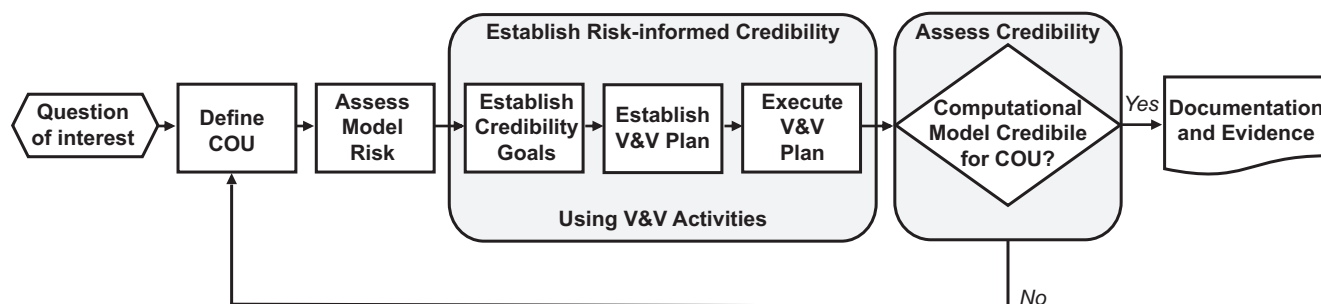
Figure 1 Overview of the ASME V&V 40 risk-informed credibility assessment framework. Modified from ASME V&V 40-2018, by permission of the ASME.[13] All rights reserved. ASME, American Society of Mechanical Engineers; COU, context of use; V&V, verification and validation.

Table 2 Credibility activities and factors

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

List of activities and corresponding factors are specified from the American Society of Mechanical Engineers verification and validation 40.[13]

unacceptable relative to the model's credibility and would therefore be rejected or need to be revised.

APPLICATION OF THE FRAMEWORK TO PBPK MODELING AND SIMULATION

In the following sections, a PBPK model will be used to illustrate the application of the risk-informed credibility framework. Key aspects of the framework, including defining the COU, assessing the model risk, and establishing credibility, will be highlighted using a hypothetical example. In practice, however, all steps in the framework would apply to the credibility assessment of a model.

Hypothetical example
A small molecule drug is in clinical development for the treatment of a chronic, non-life-threatening symptomatic condition that affects people of all ages. Planned clinical studies include assessment of PK and long-term safety and efficacy in adults, adolescents, and children. The drug is primarily eliminated by cytochrome P450 (CYP) 3A4 and has a broad therapeutic window. Clinical drug–drug interaction (DDI) studies in adults demonstrate that drug PK are affected by strong CYP3A4 modulators such that patients require altered dosing. The PBPK model will be developed, refined, and modified throughout the clinical development program to predict (i) PK changes that result from DDIs with CYP3A4 modulators and (ii) PK profiles in children (6–11 years of age) and adolescent (12–17 years of age) patients. As the drug model will serve multiple purposes, there are two questions of interest, each with a different COU.

Concepts 1 and 2: Defining the question of interest and context of use (COU)
To begin, the COUs are defined for each question of interest to outline how the model will be used to inform the question.

Question of interest 1: How should the investigational drug be dosed when coadministered with CYP3A4 modulators?

COU 1: The PBPK model will be used to predict the effects of weak and moderate CYP3A4 inhibitors and inducers on the PK of the investigational drug in adult patients. Simulated peak plasma concentration (C_{max}) and area under the plasma concentration-time curve (AUC) ratios of the investigational drug after a single dose and at steady state will be used to provide dosing recommendations for adults in labeling without the need for additional clinical DDI studies.

Question of interest 2: What is the optimal labeled dose for pediatric patients?

COU 2: Relevant physiological parameters will be changed in the adult PBPK model to predict plasma concentration-time course and exposure metrics in adolescents and children. Predictions at steady state will be used to inform the starting dose for pediatric patients in a clinical trial assessing the PK, efficacy, and safety of the investigational drug. The results of the trial will determine the final labeled dose.

Concept 3: Assessing model risk
Model risk is assessed for each COU. This evaluation considers both the model influence and decision consequence, both of which can be characterized according to a graded scale from low to high risk, defined by the specific purpose of the model.

Model influence. Model influence is described as the role of the model considering all available evidence in addressing the question of interest.

A single scale is proposed for assessment of model influence in both COUs:

Model influence	Description
Low	Model provides minor evidence; substantial nonclinical and clinical data are available to inform the decision
Medium	Model provides supportive evidence; some clinical trial data are available to inform the decision
High	Model provides substantial evidence; no clinical trial data relevant to the context of use or limited clinical trial data from similar scenarios are available to inform the decision

Based on this classification, the model in COU 1 has high influence as the PBPK analyses are used in lieu of conducting clinical DDI studies with weak and moderate CYP3A4 modulators. Dedicated clinical studies in healthy volunteers are used to support dosing with strong CYP3A4 modulators. The model in COU 2 has low influence with respect to the totality of evidence. The PK, safety, and efficacy results from the clinical trial are used to determine final labeled dose; PBPK analyses are used to select the starting dose for pediatric patients in the trial.

Decision consequence. Decision consequence describes the significance of an incorrect decision based on all available evidence. Adverse outcomes resulting from wrong decisions, for example, could include (but may not be limited to) risk of therapeutic failure or risk to patient safety. The significance can be driven by the number of patients likely to be impacted by the wrong decision, the severity of the potential harm, and/or the likelihood of occurrence.

Varying degrees of decision consequence are reflected in the proposed assessment scale:

Decision consequence	Description
Low	Incorrect decision would not result in adverse outcomes in patient safety or efficacy
Medium	Incorrect decision could result in minor to moderate adverse outcomes in patient safety or efficacy
High	Incorrect decision could result in severe adverse outcomes in patient safety or efficacy

In both questions of interest, the decisions relate to the final labeled dose of the drug, thus making an incorrect decision has the potential to result in adverse outcomes in the general patient population postapproval. The likelihood that a wrong decision will lead to an unwanted safety or efficacy outcome may be justified by relevant clinical evidence.

According to the proposed scale, the decision consequence in COU 1 is medium. Although use of CYP3A4 modulators is common in this population, suboptimal dosing of the investigational drug upon comedication is unlikely to

result in severe patient harm based on clinical DDI studies with strong CYP3A4 modulators. The anticipated exposure changes with a moderate inhibitor (or inducer) should be less than that from a strong inhibitor (or inducer). The recommended dosage adjustment for a moderate inhibitor (inducer) can be capped by the dosage adjustment recommended for a strong inhibitor (or inducer). In addition, the labeled dosage adjustment is always considered as a recommended starting dose in patients receiving these comedications. The dose may be adjusted by physicians based on individual patient responses (i.e., efficacy and safety outcomes).

In COU 2, the decision consequence is low. A suboptimal labeled dose for pediatric patients is very unlikely to result in patient harm because the model recommended dose will be assessed in pediatric clinical trials. Clinical data will be generated to support that labeled dose is safe and effective in pediatric patients.
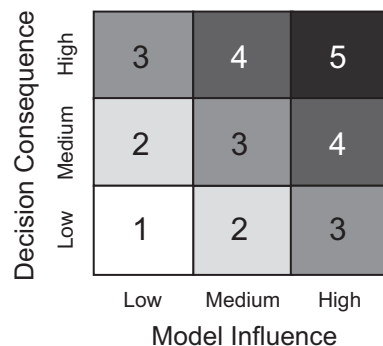
Model risk. With model influence and decision consequence characterized, model risk can be determined. Model influence and decision consequence ratings are mapped to a matrix to assess the model risk for each COU (**Figure 2**).

Model influence is plotted along the *x*-axis and decision consequence on the *y*-axis of the matrix, where an increase in either independent factor leads to an increase in model risk. Accordingly, as the PBPK model in COU 1 has high model influence and medium decision consequence, the model risk level is high (level 4). In COU 2, model influence and decision consequence are both low, resulting in low model risk (level 1).

A comparison of COUs and model risk assessments is provided in **Table 3**.

Concept 4: Establishing credibility
The model risk levels can then be used to select V&V activities and define outcomes that will provide evidence to demonstrate credibility for a COU. The V&V activities proposed should be described according to the model's COU. Potential activities can be graded on a scale from least to most rigorous to align with level of credibility needed. More rigorous activities may be selected for models that have

Figure 2 Model risk matrix for the hypothetical physiologically-based pharmacokinetic model. Model risk moves from low (levels 1–2) then medium (level 3) to high (levels 4–5) as model influence or decision consequence increases. The ratings for model influence and decision consequence are determined independently.

Table 3 Overview of the context of use, model risk assessment, and validation plan for the hypothetical example

	Context of use 1	Context of use 2
Question of interest	How should the investigational drug be dosed when coadministered with CYP3A4 modulators?	What is the optimal labeled dose for pediatric patients?
Context of use	–Simulation to predict effects of weak and moderate CYP3A4 modulators on investigational drug PK –Predictions will be used for dosing recommendations in label –No DDI studies proposed with weak and moderate CYP3A4 modulators; have clinical data with strong CYP3A4 modulators	–Simulation to predict investigational drug PK in children and adolescents –Prediction will be used to inform starting dose for clinical trial –Final labeled dose will be based on clinical trial data in pediatric patients
Model risk	High	Low
Model influence	High: –Model provides substantial evidence –Limited clinical data from similar scenarios to support the decision	Low: –Model provides minor evidence –Primary evidence for labeled dose is pediatric clinical trial
Decision consequence	Medium: –Incorrect decision could result in minor to moderate adverse patient outcomes	Low: –Incorrect decision would not result in adverse outcomes in patient safety or efficacy
Validation plan	For both: ensure model reproduces clinical PK data at different doses from healthy volunteers	
	Ensure model also reproduces: –Clinical PK data when dosed with strong CYP3A4 modulators –Effects observed for other CYP3A4 substrates with weak and moderate modulators	Ensure physiological parameters changed from adult to pediatric model are appropriate and sufficient using clinical PK data with other drugs metabolized by the same pathway (CYP3A) to confirm predictions in similar age populations

CYP, cytochrome P450; DDI, drug–drug interactions; PK, pharmacokinetics.

greater risk and thus require more evidence to demonstrate credibility. The level of evidence collected should be proportional to the level of model risk.

Some examples of how to map model risk to credibility goals for a specific COU has been provided in medical device applications.[13–15] This process requires a team of experts to decide on the appropriate level of rigor and involvement for each V&V activity. To demonstrate the concept at a high level in drug development, some credibility goals and activities of varying rigor are described. Examples from verification, validation, and applicability are presented. Note that simply stated, as described in the ASME standard[13], verification is the process of demonstrating that the equations are solved correctly in a mathematical sense, validation is the process of demonstrating that the correct equations (for the question of interest) are being solved, and applicability is the process of demonstrating that the validation efforts are relevant to the COU.

Although some verification, validation, and applicability steps presented in the ASME standard[13] are not discussed herein, in practice, a complete assessment would address all credibility factors.

Verification. Verification is the first step in establishing the credibility of a model; it ensures the accuracy and reliability of the underlying mathematical code and calculations. PBPK models are developed using ready-made software platforms or user-developed software. Software platforms may provide predefined mathematical representations of tissues and organs and be linked to databases with physiological data or compound files. For all software, code and numerical solutions should be evaluated for error and algorithms should be checked for correct implementation and function.

Additional verification steps may be warranted depending on whether the user or software vendor performed verification

and the level of credibility evidence needed. A vendor typically verifies a software platform for a variety of purposes. However, it is the responsibility of the user to ensure software is sufficiently verified for the intended use. For example, if the model incorporated many stochastic differential equations, a user may check that verification included assessing the stability in the numerical integration. If user-developed software is employed, then it is the user's responsibility to perform verification.

Example: Code verification—software quality assurance. The model is developed using a commercially available PBPK platform. For both COUs, the user confirmed that sufficient software quality assurance (SQA) was performed by the software vendor. The user repeated simulations with test cases provided by the vendor to confirm the results are reproducible on the user's computer. In addition, based on the high level of model risk for COU 1, the potential impact of unresolved software anomalies on the COU was understood by the user. If user-developed code was used in this example, then the user would conduct SQA and specify his or her procedures as appropriate in each COU.

Validation. Following verification of the code and calculations, validation of the model is performed. The purpose of validation is to determine the accuracy of the model to predict observed data and assess the correctness of model assumptions. Validation activities include assessment of the model form. For PBPK models, this relates to evaluating the underlying assumptions in the model structure, including mechanistic equations, and their relevance to the COU. For example, key PBPK model structure uncertainties may be explored through testing of alternative mechanistic equations.

Other validation activities include assessment of model input; this is subdivided into quantification of sensitivities

and uncertainties. In PBPK models, model input relates to physiological system-dependent and drug-dependent parameters. For example, uncertainties in key PBPK model inputs may be explored through the testing of *in vitro in vivo* correlation of drug-dependent parameters. However, as with verification activities, the description of validation activities should be tailored by the COU, and the rigor of the activities selected should be driven by the overall model risk to ensure applicability and sufficient credibility.

Example: Model input validation—quantification of sensitivities. For both COUs, local (i.e., one at a time) and multivariate global sensitivity analyses are performed for uncertain system-dependent and drug-dependent parameters. Considering the high risk for COU 1, additional sensitivity analyses are conducted to characterize the parameters (and underlying processes) that contribute the most to the variability of the model output.

During validation, model predictions are compared with observed data. The ASME standard[13] describes the data that are used for comparison to the simulations as the "comparator." In drug development, comparators can be data from clinical trials. Comparator assessments are divided into the credibility factors: test samples and test conditions (**Table 2**). In PBPK models, this could relate to clinical trial subjects and clinical trial scenarios. For example, predicted PBPK data and observed clinical data may be compared 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). The selection of relevant comparators should be guided by the COU and the availability of data. Various types of credibility evidence may be considered when data may be limited (e.g., clinical trial not feasible). For example, depending on the COU, clinical data from another disease or another population may be used as comparators to build credibility. The number and range of comparators may be chosen to balance model risk. Further discussion on the selection of credibility evidence for validation is provided by Pathmanathan and Gray.[16]

Example: Validation with comparator—test samples/conditions. For both COUs, the adult PBPK model predictions are compared with clinical PK studies evaluating the investigational drug at different doses in healthy volunteers. For COU 1, clinical DDI studies of the investigational drug with strong CYP3A4 modulators serve as comparators. The effects of weak and moderate modulators are validated by comparing model predictions to historical data from DDI studies with other CY3A4 substrates. In COU 2, the adult model is modified with relevant changes in physiological parameters (such as CYP3A enzyme ontogeny, tissue/organ composition, and blood flow rate) to predict the starting dose for the pediatric trial. The changes in relevant physiological parameters are validated with other drugs metabolized by the same pathway (i.e., CYP3A), comparing model predictions to observed clinical PK data in similar aged populations.

Once validation activities are completed, the degree to which the predictions match the observed can be assessed.

For model validation, key activities include assessing the rigor of the comparison method, agreement of the predicted and observed data, and relevance of the validation activities to the COU.

Example: Validation assessment—rigor of output comparison. As the model risk in COU 2 is low, a visual comparison of the steady-state plasma concentration-time profiles for the predicted and observed pediatric PK is sufficient. The model risk for COU 1 is high, thus a more rigorous comparison is performed. Simulations at steady state and after single-dose administration are compared with observed. The predicted and observed mean plasma concentration-time profiles for patients are overlaid on log and linear scales to ensure the model accurately describes baseline PK profiles. Also, the AUC and C_{max} ratios are compared between the predicted and observed data. A twofold difference between the predicted and observed C_{max} is considered an acceptable range of error as the model reproduces the overall plasma concentration-time profiles at various doses and accurately predicts the AUC and C_{max} ratio (i.e., within 25%) in the clinical DDI trial with strong CYP3A4 modulators. Stringent acceptance criteria were applied to the AUC and C_{max} ratio as these were considered the most relevant PK parameters for providing dosing recommendations.

Applicability. Model credibility increases when there is increased overlap between the validation activities and the COU. However, in drug development, there will be differences between how the model is validated and how the model will be used. For example, comparator studies used for validation may not exactly match the conditions of a simulation. Thus, the relevance of the proposed validation activities to the COU should be evaluated and justified. Lack of relevance to the COU can diminish potential credibility gained through validation activities. An example of how applicability can be assessed has been presented for medical device applications.[17]

Example: Relevance of the validation activities to the COU. For both COUs, the selection of validation activities, including comparators (described in the "Example: Validation with Comparator—Test Samples/Conditions" section) is based on the purpose of the model and the availability of clinical data. Comparing PBPK model predictions to clinical PK data where the investigational drug was evaluated at different doses in healthy volunteers is relevant to both COUs as the data set validates the base model. In COU 1, clinical studies of the investigational drug with a strong CYP3A4 inducer and inhibitor are relevant comparators as these data will validate the contribution of the CYP3A pathway and likely represent the worst case DDI scenario. In addition, clinical studies of sensitive CYP3A substrates with weak and moderate modulators provide relevant comparators to validate the clinical DDI potential of these modulators on the CYP3A pathway. For COU 2, clinical studies of other drugs metabolized by CYP3A4 in a pediatric population are relevant comparators to validate

physiological parameters for this pathway in similar-aged subjects.

Concept 5: Assessing credibility

The remaining steps of the framework include defining an appropriate plan for V&V activities and acceptable results for each credibility factor (e.g., acceptable range of fold error between the predicted and observed, level of acceptable uncertainty). Upon execution of the plan, the completed activities and outcomes are reviewed. If both are considered sufficient and acceptable for establishing model credibility based on the COU, model risk, and credibility goals, the results can be documented and used to demonstrate evidence of credibility. If not, the model itself or the COU may be modified, additional V&V activities can be conducted, or model influence can be reduced. For example, if the acceptance criteria prespecified as part of the credibility goals are not met, the model can be refined and validation repeated. This is consistent with the "learn-confirm paradigm."[18] Although this paradigm is germane to the process of establishing and assessing credibility, the framework does not describe model building and refinement and instead focuses on regulatory applications. Further explanation and examples of credibility assessment are provided by Morrison *et al.*[14] and the ASME standard.[13]

IMPACT

In the current regulatory landscape for long-standing MIDD approaches, credibility assessments are specific to each modeling and simulation approach.[7,8] Adoption of such a framework may shift the community toward a uniform approach for future assessments in drug development, irrespective of the modeling and simulation method used or the intended application (e.g., optimize dosing, inform risk/benefit, or trial design, etc.). In doing so, this would standardize what constitutes a credible model and provide a common language to describe risk-informed credibility assessments across modeling and simulation approaches. This may also help to enable more consistent regulatory decision making, minimize the risk of erroneous decisions in the acceptance of modeling and simulation to inform drug labels, and help to align regulator and sponsor expectations.

The use of this framework may facilitate alignment by providing a starting point for high-level discussions regarding the model and how it may be used to address a question of regulatory interest. For example, discussions with regulators on the COU, model risk, and appropriateness and acceptability of the V&V plan may be of value during drug development. A potential mechanism for these discussions is the MIDD pilot program, where early engagement aims to help accelerate drug development through discussion of modeling and simulation strategies, with the potential to reduce late-stage drug development failures.[19] Consideration of this framework also invites discussion of potential changes to regulatory documentation including knowledge management of a model throughout a regulatory lifecycle where the knowledge, COU, and software versions among other things are potentially evolving.

CHALLENGES

There is a simultaneous desire for granularity and flexibility when it comes to the regulatory assessment of computational models. This commonly encountered viewpoint presents regulators with a challenge to provide recommendations for assessment of model credibility in MIDD. Although the framework offers a potential solution, providing discretized, tailorable steps to establish and justify model credibility, alternative approaches may be debated.

Other hurdles may present themselves in accepting this framework or an alternative overarching approach. For example, it may be challenging to relate the terminology used herein to those currently used by the PBPK and broader MIDD community. Various terminology has been used to describe model credibility and V&V activities across MIDD approaches, including PBPK and quantitative systems pharmacology.[20] To add complexity, there is also mismatch in terms and definitions used between the computational science and MIDD communities (including "qualification," "validation," and "verification").[21] Regardless of the vocabulary used across the MIDD community currently, if the goal is to assure a common starting point for dialogue on the evidentiary standards needed for acceptance of a model for a given regulatory purpose, then the terminology used for model credibility assessments needs to be standardized.

Another challenge is shifting the mind-set of how regulators and sponsors currently assess models. The use of the ASME framework would necessitate understanding novel (although we would argue intuitive) concepts, including that model risk drives the selection of V&V activities and that model credibility can increase when V&V activities are rigorous and applicable to the intended use. To adopt this framework and mind-set shift, potential users must move past concerns that different people may arrive at different conclusions regarding assessment of model risk and credibility and selection of V&V activities. These concerns exist even without the framework. The adoption of this framework, we believe, could provide greater transparency and thus enable drug developers and regulators to deliberate and align (even if through iteration) on what V&V activities would be needed for a given COU. In essence, this framework then could derisk the use of MIDD across the continuum of drug discovery, development, and regulatory evaluation.

RECOMMENDATIONS

In the short term, we recommend public discussion on the potential use of the framework using PBPK modeling and simulation as a strawman. This can be accomplished through workshops that include multiple stakeholders. In addition, the potential utility and challenges of using this framework or where and how the framework could be adapted to better serve in regulatory decision making for drug development can be opined on in peer-reviewed literature. Alternative, but overarching, frameworks can also be proposed.

In the long term we see value in internal harmonization (e.g., across all US Food and Drug Administration centers) on the assessments of model credibility, regardless of whether

an alternative framework is ultimately adopted. Beyond this, we advocate for harmonization across regulatory agencies in the future.

1. Wang, Y. *et al.* Model-informed drug development: current US regulatory practice and future considerations. *Clin. Pharmacol. Ther.* **105**, 899–911 (2019).
2. Bai, J.P.F., Earp, J.C. & Pillai, V.C. Translational quantitative systems pharmacology in drug development: from current landscape to good practices. *AAPS J.* **21**, 72 (2019).
3. Friedrich, C.M. A model qualification method for mechanistic physiological QSP models to support model-informed drug development. *CPT Pharmacometrics Syst. Pharmacol.* **5**, 43–53 (2016).
4. Peters, S.A. & Dolgos, H. Requirements to establishing confidence in physiologically based pharmacokinetic (PBPK) models and overcoming some of the challenges to meeting them. *Clin. Pharmacokinet.* (2019). https://doi.org/10.1007/s40262-019-00790-0.
5. Overgaard, R.V., Ingwersen, S.H. & Tornoe, C.W. Establishing good practices for exposure-response analysis of clinical endpoints in drug development. *CPT Pharmacometrics Syst. Pharmacol.* **4**, 565–575 (2015).
6. Byon, W. *et al.* Establishing best practices and guidance in population modeling: an experience with an internal population pharmacokinetic analysis guidance. *CPT Pharmacometrics Syst. Pharmacol.* **2**, e51 (2013).
7. United States Food and Drug Administration. Guidance for industry: population pharmacokinetics <https://www.fda.gov/media/128793/download> (2019). Accessed August 26, 2019.
8. United States Food and Drug Administration. Exposure-response relationships—study design, data analysis, and regulatory applications <https://www.fda.gov/media/71277/download> (2003). Accessed August 26, 2019.
9. United States Food and Drug Administration. Guidance for industry: physiologically based pharmacokinetic analyses—format and content <https://www.fda.gov/media/101469/download> (2018). Accessed August 26, 2019.
10. European Medicine Agency. Guideline on the reporting of physiologically based pharmacokinetic (PBPK) modelling and simulation <https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-reporting-physiologically-based-pharmacokinetic-pbpk-modelling-simulation_en.pdf> (2018). Accessed August 26, 2019.
11. European Medicine Agency. Guideline on reporting the results of population pharmacokinetic analyses <https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-reporting-results-population-pharmacokinetic-analyses_en.pdf> (2007). Accessed August 26, 2019.
12. Manolis, E. *et al.* The role of modeling and simulation in development and registration of medicinal products: output from the EFPIA/EMA modeling and simulation workshop. *CPT Pharmacometrics Syst. Pharmacol.* **2**, e31 (2013).
13. The American Society of Mechanical Engineers. Assessing Credibility of Computational Modeling through Verification and Validation: Application to Medical Devices (American Society of Mechanical Engineers, New York, 2018).
14. Morrison, T.M. *et al.* Assessing computational model credibility using a risk-based framework: application to hemolysis in centrifugal blood pumps. *ASAIO J.* **65**, 349–360 (2019).
15. Parvinian, B. *et al.* Credibility evidence for computational patient models used in the development of physiological closed-loop controlled devices for critical care medicine. *Front. Physiol.* **10**, 220 (2019).
16. Pathmanathan, P. & Gray, R.A. Validation and trustworthiness of multiscale models of cardiac electrophysiology. *Front. Physiol.* **9**, 106 (2018).
17. Pathmanathan, P., Gray, R.A., Romero, V.J. & Morrison, T.M. Applicability analysis of validation evidence for biomedical computational models. *J. Verif. Valid. Uncertain. Quantif.* **2**, (2017). https://doi.org/10.1115/1.4037671
18. Sheiner, L.B. Learning versus confirming in clinical drug development. *Clin. Pharmacol. Ther.* **61**, 275–291 (1997).
19. Madabushi, R. *et al.* The US Food and Drug Administration's model-informed drug development paired meeting pilot program: early experience and impact. *Clin. Pharmacol. Ther.* **106**, 74–78 (2019).
20. Rostami-Hodjegan, A. Reverse translation in PBPK and QSP: going backwards in order to go forward with confidence. *Clin. Pharmacol. Ther.* **103**, 224–232 (2018).
21. Shepard, T., Scott, G., Cole, S., Nordmark, A. & Bouzom, F. Physiologically based models in regulatory submissions: output from the ABPI/MHRA forum on physiologically based modeling and simulation. *CPT Pharmacometrics Syst. Pharmacol.* **4**, 221–225 (2015).