**INTRODUCTION**

Although important for all scientific communications, the need for a structured approach in evaluating the credibility* of quantitative models is particularly apparent from a regulatory perspective. The regulator’s role is to ensure that the proposals are scientifically sound and valid given the context of use and the established evidentiary standards, to protect the interest of patients and public health. In order to ensure high scientific quality, objectivity, and consistency in the evaluation of submitted models and consequent decision making, the regulatory systems need to be aligned in their perspectives and approaches across agencies and across assessors. Moreover, given that drug development is becoming a global, multidisciplinary and multistakeholder activity, the regulatory authorities, the drug developers, as well as the academic milieu would benefit greatly from sharing perspectives and preferably a general framework on model evaluation.

Several data and model evaluation frameworks have been proposed over the years, often with the goal to improve clarity, information flow, and decision making. Whereas the learn and confirm paradigm introduced by Sheiner\(^1\) is shaping the thinking on pharmacometric models, the Observe, Orient, Decide, and Act (OODA) loops originally introduced by Colonel Boyd for military use\(^2\).
is often used in other settings to support the iterations from acting to observing the outcomes and determining the next steps. Likewise, the paradigm of model-informed drug development (MIDD) can perhaps be described as these iterative loops of conducting studies, consolidating the knowledge on a medicinal product, the patients and the disease into pharmacometric models, and then determining how to best perform the next study or answer the next question. \(^3\)\(^-\)\(^5\)

However, when it comes to assessing the trust in a model for a specific application (model planning and evaluation), currently available tools/guidance documents are skewed to the technical aspects. There is an unmet need for a holistic tool that provide an end-to-end link from the initial question to the final model-informed decision making, similarly to what is proposed with the estimand framework for statistical assessment of clinical efficacy trials. \(^6\) We suggest the risk-informed credibility framework, \(^7\) originally proposed for model assessment in the medical device domain, could be used for this purpose and also offers strong support for the pharmacometric models. This approach provides a very clear interface between what we want to know and the credibility of the model as a method to provide an answer. The approach aligns well with the framework presented in the European Medicines Agency’s (EMA) Guideline on Physiologically-Based Pharmacokinetic (PBPK) Modeling \(^8\) and both the International Council of Harmonization (ICH) and EMA guidelines on pediatric extrapolation \(^9\)\(^-\)\(^11\) and fits well with our perspectives on model evaluation. We see the application of this framework in the context of regulatory assessment of pharmacometric models as an opportunity that could help to improve not only the transparency in the communication with sponsors and other stakeholders, but also increase the scientific quality, objectivity, and consistency in the evaluations. Most important, we consider this as a valuable mean to address some of the challenges currently faced throughout the EU drug assessment process, as detailed in the section on current challenges.

Although the risk-informed credibility framework \(^7\) appears to have been written from a predominantly mechanistic modeling perspective, we find that the general mindset is well-suited also for empirical (top-down) models. Taking the nonlinear mixed-effects models (NLMEM) in the form of population-pharmacokinetic (PK) models as a general example, it can be postulated that even population-PK models have elements of system knowledge (mechanistic elements) in them that make the key components of the framework relevant. The function of the drug elimination terms, such as clearance or elimination rate, can be viewed as accounting for the sum of the underlying physiological processes involved in eliminating the drug from the body. Similarly, the addition of compartments to a pharmacometric model functions as a means to account for the sum of biological processes that are involved in distributing the drug within the body. This way of looking at the population-PK model structures emphasizes the mirroring of the model structure to the underlying physiological processes. However, there are inherent differences between input and output for mechanistic (bottom-up) and empirical (top-down) models as well as differences in the contexts of drug and device development. Some adaptations of the framework are therefore considered beneficial to aid its use for pharmacometric models. In the sake of clarity, the key features of the empirical pharmacometric models that mandate specific considerations for their assessment will be discussed. Subsequently, the points to consider for these types of models from a credibility perspective will be discussed and examples of implementation provided, including remaining challenges and gaps.

Two tables are proposed for explicit description of key attributes in the process of model evaluation. These tables are proposed as helpful tools to facilitate and streamline the communication between stakeholders.

The adapted framework will be applied to two hypothetical cases’ examples of regulatory submissions using the two tables, similarly to the exercise done by US Food and Drug Administration (FDA) on PBPK models. \(^12\)

**REGULATORY ASSESSMENT OF PHARMACOMETRIC MODELS IN THE CONTEXT OF DRUG DEVELOPMENT**

The pharmacometric methods were revamped in the 1990s, thanks to the work by Sheiner and Beal who applied pharmaco-statistical and computational methods to personalize dosing regimens in postmarketing settings and to optimize drug development. They proposed a new software for nonlinear mixed-effects modeling (NONMEM) of PK and pharmacodynamic (PD) data. \(^13\) Whereas taking its roots in the academic domain, the use is now widespread and the modeling approach a mainstay of drug development. The toolbox for model building and model evaluation is big and has expanded as challenges in understanding the model credibility have emerged. Many of the available model evaluation tools for pharmacometric models are focused on the technical aspects of model evaluation and could be relevant to diverse types of applications. \(^14\)

In the context of drug development and regulatory assessment, models are proposed in well-defined contexts of use. When prediction models are proposed to be used to
inform decision making in drug approval, an important aspect of the context of use may be that the model fully or in part replaces the standard requirements for a fully powered clinical study. An example of this can be replacing a clinical drug-drug interaction or a bioequivalence study. To ensure that evidentiary standards are not sliding, it is important to bear the benchmark of clinical confirmation in mind. Model misspecifications or inadequacies may thus pose risks for the individual patients or at the public health levels and the model acceptability thus needs to account for the risks. On this background, pre-defining the requirements for model acceptability in an application-wise manner becomes a necessity irrespective of the methodology used. This also highlights the need for understanding the strengths and limitations of the model evaluation tools, how they apply to the model used (e.g., population-PK and dose-exposure-response models) and their relevance for the scientific question of interest. In the context of drug development and evaluation, most frequently pharmacometric models are used for the following applications:

- Description of PK and PD data and quantitative characterization of their determinants (e.g., age, bodyweight, organ [liver and kidneys] impairment, comedications, and comorbidities).
- Characterization of the impact of change in formulation on drug efficacy or safety (e.g., modified release, biosimilars, etc.).
- Characterization of the impact of change in dosing regimen on drug efficacy or safety (e.g., change in dosing frequency for more convenience or better compliance).
- Trial design optimization.
- Dose finding/selection.
- Waive a dose finding study for a new indication.
- Waive a clinical drug-drug interaction study.
- Waive a PK, PK/PD, or efficacy and safety trial or parts of such trials in unstudied or limitedly studied (sub) populations (e.g., children, aged patients, and rare disease).

In terms of EMA regulatory procedures, pharmacometric modeling and simulation data are included in scientific advices, protocol assistances, qualification advices, qualification opinions, pediatric investigation plans, marketing authorization applications, postmarketing signals for changes in product information, and postmarketing referrals for changes in marketing authorization conditions. Given the wide range of settings and applications where pharmacometric approaches are used, implementation of the principles presented in the regulatory guidance documents and best practice publications is not always straightforward and several challenges are encountered as detailed in the next section. Having a framework endorsed by all stakeholders could be very helpful in aiding some of the challenges.

**CURRENT CHALLENGES IN REGULATORY EVALUATION OF PHARMACOMETRIC MODELS IN THE CONTEXT OF DRUG DEVELOPMENT**

In our experience, the following challenges are commonly reported by regulators and developers involved in the process of developing and assessing pharmacometric models for drug development.

**From regulators**

- The questions the models are answering are very seldom described.
- Model objectives are not clear or in line with the actual use of the model. Often the model objective is set as describing PK and the model then used for other purposes, such as informing decisions on dosing in subpopulations.
- The adequacy of the input data is seldom discussed. An example could be whether a sufficiently broad range of the covariates that characterize the subpopulation, such as age, weight, etc., are represented in the database to allow generalized conclusions.
- Models are not sufficiently well evaluated/validated. Inconsistencies in the approaches taken by sponsors in the modeling approaches chosen for answering very similar scientific questions.
- Poor reporting, where aspects important for the model evaluation are not included in the report and/or the granularity of the reporting is too low to allow a secondary assessment of the modeling exercise by the regulator.

**From sponsors/developers**

- Lack of guidance for the case of interest.
- Requirements for model acceptability are unclear.
- Inconsistencies in the opinion or issues raised on similar modelling approaches for answering very similar scientific questions.
- Insufficient training/experience of regulators: the assessor does not understand the sponsor’s proposal.
- Poor reporting on assessment: unclear, inadequate, or unnecessary questions from regulators.
APPLICATION OF THE RISK INFORMED CREDIBILITY FRAMEWORK TO PHARMACOMETRIC MODELS

The key concepts of the credibility framework have been described in the American Society of Mechanical Engineers (ASME) standard as well as recapped in the FDAs White Paper on the potential use of the framework in the assessment of PBPK models. We refer to these references for an in-depth description of the concepts but introduce the steps here to guide the examples in the next sections. For each step, a description of expected activities is given, followed by a comment on the sponsor versus regulatory responsibilities, and a comparison to the EMA guideline on reporting the results of population-PK models, which is the reference document for pharmacometric model submissions to the EMA.

Some key terms that are used in the European regulatory network are introduced and discussed in relation to the risk-informed credibility framework, regulatory impact, and extrapolation. At the EMA, the assessment of regulatory impact has played a key role since its introduction by Shepard at the EMA/European Federation of Pharmaceutical Industries and Associations (EFPIA) modeling and simulation workshop. The regulatory impact describes the role played by the model in the regulatory decision making. The regulatory impact is described as low, moderate, or high, and the requirements for the rigor of the credibility evidence increase with the impact. This terminology is now largely understood and widely used in the EU regulatory network. Although the concepts of model influence and decision consequences are key elements of the risk-informed credibility framework, the specific term “regulatory impact” is naturally not included. The model influence will outline the contribution of the model versus other evidence that will influence the overall decision within the concerned drug development program. However, contrarily to the regulatory impact, the model influence does not explicitly compare/contrast the proposed modeling approach to the alternative methods that can be used to address the question of interest. For example, if a fully powered crossover bioequivalence study is proposed to be replaced by a model-based approach with much less data collection, the regulatory impact would likely always be high, whereas the overall model risk could be moderate to high depending on the drug’s therapeutic window and/or other aspects that could influence the decision consequence. As a pragmatic solution, we suggest incorporating the term regulatory impact in the framework, as part of the activities related to establishing risk-informed credibility in regulatory assessments. Extrapolation in a quantitative modeling context can be seen as accepting the model predictions without requirement for confirmatory data generation in a (fully powered) clinical study. As an example, if an adult population-PK model is allometrically scaled to pediatric patients and used to derive dosing recommendations, and these predictions are not confirmed by PK data in the pediatric patients—the PK and related dosing recommendations in children have been obtained based on extrapolation. The scientific basis of the underlying knowledge, as either captured explicitly by the model or assumed not to impact and thus not accounted for, then becomes paramount in order to understand the model credibility.

Steps of the risk-informed credibility assessment framework

The different steps of the framework, as presented in the risk-informed credibility framework, are schematically shown in Figure 1 and the steps 1–4 below.

Step 1 – The specific scientific question(s) of interest

The credibility framework starts by identifying the question of interest, which describes the question, concern, or decision that will be addressed. The question of interest is often broader than the scope and role of the model in answering it, as both modeling data and other in vitro,
preclinical, or clinical data may be used to address it. This is explicitly described in step 2 of the framework.

In general, the adequate definition of the scientific questions of interest is an industry liberty. Whereas the European and other regional regulations and scientific guidelines specify which studies and data are needed for drug development and approval, often the questions to be addressed by these activities are not explicitly presented in the submissions received. As specific questions naturally emerge consequent to the characteristics of the medicinal product under development, some flexibility is clearly necessary.

Although the concept of the question of interest is not described in the EMA guidance on population-PK models, it is recommended that the model objective of the analysis is stated. An example is given in this manner: “an objective be to build a model that describes the data and to test the possible influence of various specified covariates on the parameters of the model.” It clearly is helpful to understand that the model is planned to be used to inform on the impact of a covariate. However, if the conclusion on the model assessment is that the model cannot be trusted, the general nature of such a statement makes it difficult to understand whether other approaches may need to be taken to answer the original concern or interest in the covariate.

Step 2 – Risk-informed model credibility

Explicit descriptions of the context of use, the model risk, and the resultant requirement on the credibility activities are given in this step. The context of use is the specific role and scope of the computational model used to address the question of interest. The description of the context of use could be written out as a short and concise description on what the model will be used for, what data informs the model development, whether the model prediction will be confirmed with generation of new data, and/or what other evidence that supports the decision. In this step, the regulatory impact is assessed. The description of the regulatory impact should outline the influence the model will have on the final decision as well as what the current established methods are for answering the question(s) of interest. If the model is moving into a context of replacing a clinical study or other established methods of answering the question, which often would represent a request for extrapolation, this should be clearly described. The decision consequence in regulatory submission generally relates to the risk to the patient in case the modeling predictions or assumptions lead to erroneous regulatory decisions. Other risks of harm can be relevant depending on the question of interest.

For models that inform regulatory decision making, both sponsors and regulators are responsible for assessing the requirements for risk-informed model credibility.

The context of use, the impact of the model on the decision, the decision consequences, and the resultant requirements on the credibility activities are not explicitly addressed in the EMA guidance on population-PK. Continuing the example of the model objective to investigate the impact of a covariate; this would normally imply that the model would be used to inform the decision of whether the dosing recommendations needs to be revised in a subpopulation. The consequence of not addressing these aspects is that it often is unclear what weight the model will have in deciding whether to adapt the dose, what the risk is to the patients are if the model informed decision is wrong, and whether the credibility evidence is relevant and adequate.

Step 3 – Credibility activities

The process of providing evidence that the model is credible, includes performing verification and validation activities. To reach the required rigor of evidence for the context of use and the model risk, the sponsor chooses the relevant specific activities to be conducted. Credibility factors are used to describe the rigor for each of the activities. The credibility framework in this manner structures the process of model assessment. The rationale for the grading is to support the planning and comparison of the activities that can impact the model credibility. Ideally, the approach, the related activities, and the requirements on the credibility factors should be prespecified in the model analysis protocol and included in the study reports as relevant. The execution of the plan established in the model analysis protocol and study protocols should be performed as relevant.

In general, the sponsors are responsible for performing these steps by writing the model analysis plan, the study protocol, as well as executing the plan. Regulators can be involved by giving early advice on the overall plan as well as specific aspects of modeling/study protocols.

The EMA guidance on population-PK models does not use the terms verification and validation activities, however, it does address both. For the verification, it is requested that the software and version used should be stated, the estimation and simulation methods described, and relevant output files should be presented. For the validation, in line with the risk-informed credibility framework, the EMA guidance on population-PK models describes that the amount and type of model evaluation will depend upon the objective(s) of the model.
development. It is required that a justification is provided for the model evaluation procedures and tools used for the specific evaluation. Although no specific recommendations or examples with methodological choices have been given for low or high impact models, which could have been helpful, several tools and procedures that are available for use are described. The guideline prescribes that the model structure (described as structural, error, and covariate models) should be presented. Further, the guideline recommends describing the quality and range of input data for high impact models. This step, the validation, is the main focus of the guideline and the level of detail on the methodological sections is quite fair. The EMA population-PK guideline recommends that an analyses plan is written before conducting the analysis, however, it further describes that considering the explorative nature of the analysis, the information requested is described as requirement for the model report, rather than the protocol. Whereas it is agreed that this can be a sound approach for early stages of learning and for purely descriptive models, there clearly is a need to step up on the prespecification for prediction models that are used for extrapolation or confirmatory purposes.

Step 4 – Assessment

Upon completion of the credibility activities, an assessment is done to determine the credibility of the model, considering the context of use (COU), model risk, credibility goals, verification, and validation results, as well as other knowledge acquired during the process.

For models that inform regulatory decision making, both sponsors and regulators are responsible for this step.

The scope of the EMA population-PK guideline is on reporting the results of a population-PK model to ensure that the assessor(s) can perform a secondary assessment of the modeling exercise. Accordingly, no clear guidance is given on how to balance the results of the verification and validation activities to the model objective. In practice, for technical assessment of models, most sponsors are following best practice as reflected in the current literature in the field and use state of the art tools for the assessment of the different credibility factors (e.g., for the comparisons of the observed versus simulated/predicted data, numerical, graphical, and statistical tools are used). The main challenge, as seen from the regulatory perspective, is that the choice of methods is seldom linked to the question, COU, and the model risk. The COU differ greatly from a context of describing PK to a context of informing a decision of extrapolation and the model risks will be very different for narrow therapeutic index drugs versus drugs with wide safety margins. If these aspects are not considered, it may result in the choice of a methodological approach that is not credible, which for the pharmacometric models often is related to insufficiencies in the input data.

“Quickstart” tools to support the risk informed credibility assessment – Credibility matrix and Credibility activities tables

As a pragmatic way forward of aligning the current regulatory approach for model evaluation with the risk-informed credibility framework, we suggest using the outline provided in the quickstart tables (Tables 1 and 2). The tables are intended to give an overview of the suggested use of the model (Table 1), the planned goals, and the current status on the rigor of the credibility evidence (Table 2). To be clear, it is by no means the intention that these tables will replace the well-described reports on population-PK/PD or other pharmacometric models. The potential benefit of the tables is providing a quick overview of the modeling exercise, which may aid in focusing the assessment process and preparing concise communication with other domain experts. An additional aspect is that the tools may aid in improving the consistency of decision making, by supporting explicit descriptions of both the model risk for the intended use and the thresholds for acceptability on the credibility activities. The tables are being tested in an EMA pilot for PBPK assessments, and currently explored for population-PK/PD models at some National Competent Authorities. We will also present two hypothetical case examples to inspire the scientific community and get feedback on the value of initiating further pilots with industry.

The elements of the credibility matrix table (Table 1) provide an interface between the question(s) of interest and the credibility of the model as a method in answering the question. For quick overview of the modeling approach, the key information on the investigational product and the type of model should be provided in addition to an explicit request for writing out the model-informed decision.

Whereas Table 1 presents the overall plan for the modeling exercise, we feel an additional tool is needed to give an overview with sufficient granularity on the credibility goals and activities (steps 2 and 3). Table 2 is proposed for this purpose. Table 2 presents a structured approach to the credibility factors related to each activity, and includes columns that allows modelers and assessors as relevant to insert both the planned and the obtained score of the various credibility factors.
| Credibility matrix | Description                                                                 |
|--------------------|-----------------------------------------------------------------------------|
| Investigational product | For a specific application: Describe the drug substance, formulation and route of administration. For qualification of a platform, describe the properties/characteristics of the type of investigational products that are relevant. |
| Type of model | State the general type of model (NLMEM, agent-based, etc.) as well as the popular and/or commercial name of the model as relevant (pop-PK, PBPK, QSP, etc.) |
| Scientific question(s) of interest | State the scientific question(s) the model is intended used to answer. What is the answer we need to inform our concern or (clinical) decision? If the models are used to answer several questions they can be numbered and handled within the same table or split in a separate table. |
| Context of use | Describe the specific role and scope of the computational model to address the question of interest. The context should be outlined as a short and concise description of what the outputs of the model will be used for, what data (type) that is used for building the model as well as what other data or evidence that supports the decision. |
| Regulatory impact | The description of the regulatory impact should outline the influence the model will have on the final decision as well as what the current evidence standard is for answering the question(s) of interest. If the model is moving into a context of replacing a clinical study or other established methods of answering the question, which often would represent a request for extrapolation, this should be clearly described. The regulatory impact is described as low, medium, or high. |
| Risk based analysis of decision consequence | Describe the actual risks for the patients in case of wrong decisions from the model. For drug treatment, risks are often related to patients being over- or underexposed to the drug, however, other risks may also be relevant. The model risk is the composite of the regulatory impact (model influence) and the decision consequence and should be described here. |
| Requirements on the credibility activities | Describe requirements for the credibility activities given the model risk and the current evidentiary standard. Different approaches can be taken to defining acceptance criteria for the credibility activities and individual factors. The approaches must also be seen in light of what is possible: what is the current benchmark on the rigor of the credibility activities? Does this meet the current evidentiary standard in answering the question of interest? To define the required precision/accuracy level of the model predictions, both clinical (pharmacology, exposure-response, therapeutic window, etc.) and quantitative considerations (graphical and numerical tools) needs to be made to ensure the scientific plausibility of the modeling approach. The full outline of the goals for the verification and validation activities can be given in the table on the credibility activities. Critical verification and validation activities can be briefly summarized here. Examples of this may include explicit discussion of key model assumptions relevant for the individual activities and how these are mitigated by sensitivity analysis, uncertainty quantification, further data collection or other approaches. The following outline is suggested for a structured description:  
  - Model verification activities and related acceptancy criteria  
  - Code  
  - Calculations  
  - Model validation activities and related acceptancy criteria  
  - Model structure and key parameters: link with the pathophysiology and pharmacology described by the model. Key assumptions and mitigation such as quantification of sensitivities and uncertainties.  
  - Observed data (external datasets or internal data for model building): What is considered minimum requirement. Key assumptions and mitigation approaches such as quantifying impact of uncertainty, further data collection, etc.  
  - Model assessment: Graphical tools, numerical tools, clinical pharmacology considerations, etc.  
  - Applicability of the verification and validation activities for the context of use.  
  - Relevance of the quantities of interest (such as exposure metrics).  
  - Relevance of the validation activities for the context of use. |
Most of the credibility factors are self-explanatory, or easy to grasp with a detailed reading of the ASME publication7 and we refer the readers to the standard for the in-depth description. References to the section of verification and validation (V&V)40 describing the relevant factor is included in Table 2. The verification activities and factors are relevant for the empirical pharmacometrics models. However, for the validation activities, we find that some adaptations of the V&V40 are needed to better fit with the drug domain. In the device setting, the term comparator is used for the data set(s) that the model predictions are compared with. As the term is not familiar in the drug domain, we use the term observed data. The observed data may be the data that informs the model structure and model input parameters for the empirical pharmacometrics model (internal data set), but may also consist of several data sets, such as external data for specific validation activities. The credibility factors related to the “test samples” describe and grade the extent and quality of the data that are collected to inform on the PK, PD, efficacy, or safety of the drug as relevant for the question of interest. Many of the medical device models are describing physical processes where the conditions of the experiments can be controlled, which means the test samples can be measured at different conditions. Different perspectives could be taken on how to interpret the test condition(s) in the physiological/pharmacological setting and transform it to credibility factors that are informative. For a PK study, it would be possible to perform it at hypothermic conditions in addition to at normal temperatures, and if done this could be seen as investigating two conditions. However, we find that this is not so useful considering all the other intrinsic and extrinsic factors know to affect the pharmacology of a drug and disease status of patients. We thus find it more useful to include an element that reflects the extent of which intrinsic and extrinsic factors that have been explored and included in the data set(s). Although a composite score of these may not be appreciably informative, it still gives an indication of the thoroughness of the investigation of potential sources of variability, and we find it is worth testing whether such credibility factors can provide value for model evaluation.

**CREDIBILITY EVIDENCE BRIEFLY SUMMARIZE THE RIGOR OF THE CREDIBILITY ACTIVITIES THAT WERE IMPLEMENTED AND THE RESULTS. OTHER EVIDENCE INFORMING THE DECISION MAY ALSO BE SUMMARIZED.**

**CASE STUDIES**

With these tools at hand, we will present an assessment of two hypothetical examples, designed based on a look at actual procedures submitted for regulatory review for Marketing Authorization Application and Scientific Advice procedures, respectively. The examples are typical of everyday regulatory assessment of population-PK/PD models and chosen to illustrate the following two scenarios:

1. Example 1. Question of interest 1: waiver for bioequivalence study to bridge across formulations and manufacturing processes, and 2: need for dose adjustment in subpopulations.

2. Example 2. Question of interest 1: dose selection for the confirmatory trial, and 2: dose selection for a pediatric study in 6 months to 12-year-old children.

Although hypothetical examples are used to not disclose the exact models and the sponsors, the main features that we wish to shed light on should still be graspable. The outcomes of the assessment by means of filling the quickstart tools are provided in Tables 3 and 4 for example 1 and in Tables 5 and 6 for example 2. The section on selected rigor of the credibility factors is left empty, as the exercise is an assessment of what was obtained for the two examples without prespecification of the activities.

**GENERAL DISCUSSION**

The current use of population-PK/PD models frequently moves beyond the objective of describing the observed data. The models are often proposed used for decision making on the dosing rationale in subpopulations/ such as pediatric patients and as support for extrapolation. This shift in the model objectives is very welcome. However, using models for higher impact purposes comes with some costs on the requirements for the rigor of the evidence to support that the model is credible. We propose that the risk-informed credibility assessment is a framework fit for supporting the development and assessment of empirical...
models/ such as the population-PK models. To explore and present our perspectives/ we compared the framework to the guidance provided in the EMA guidance on population-PK models and assessed two population-PK models with the two quickstart tools the credibility matrix and the credibility activity tables.

When looking at the EMA guideline on the reporting of population-PK models from this perspective, it becomes

---

**TABLE 2** Credibility activities - template with guidance

| Activity                  | Credibility factor                                      | Rigor         |
|---------------------------|---------------------------------------------------------|---------------|
|                           |                                                          | Selected<sup>a</sup> | Range<sup>b</sup> | Obtained<sup>a</sup> | Credibility<sup>c</sup> |
| Verification              |                                                         |               |                   |                   |                           |
| Code                      | Software quality assurance                              | (a–c; 5.1.1.1) |                   |                   |                           |
|                           | Numerical code verification                             | (a–d; 5.1.1.2) |                   |                   |                           |
| Calculation               | Discretization error                                    | (a–c; 5.1.2.1) |                   |                   |                           |
|                           | Numerical solver error                                  | (a–c; 5.1.2.2) |                   |                   |                           |
|                           | Use error                                               | (a–d; 5.1.2.3) |                   |                   |                           |
| Validation                |                                                         |               |                   |                   |                           |
| Computational model       | Model structure                                         | (a–c; 5.2.1.1) |                   |                   |                           |
|                           | Model input parameters                                  |               |                   |                   |                           |
|                           | Quantification of sensitivities                         | (a–c; 5.2.1.2.1) |                   |                   |                           |
|                           | Quantification of uncertainties                         | (a–c; 5.2.1.2.2) |                   |                   |                           |
| Observed data             | Test samples. Measurement uncertainty                   | (a–c; 5.2.2.1.1) |                   |                   |                           |
|                           | Quantity                                               | (a–c; 5.2.2.1.1) |                   |                   |                           |
|                           | Range of characteristics                                | (a–d; 5.2.2.1.2) |                   |                   |                           |
|                           | Measurements                                           | (a–c; 5.2.2.1.3) |                   |                   |                           |
|                           | Uncertainty                                            | (a–d; 5.2.2.1.4) |                   |                   |                           |
|                           | Test conditions. Intrinsic and extrinsic factors        |               |                   |                   |                           |
|                           | Quantity                                               | (a–c; 5.2.2.2.2) |                   |                   |                           |
|                           | Range                                                  | (a–d; 5.2.2.2.2) |                   |                   |                           |
|                           | Measurements                                           | (a–c; 5.2.2.2.3) |                   |                   |                           |
|                           | Uncertainty of test condition measurements              | (a–d; 5.2.2.4)  |                   |                   |                           |
| Assessment                | Equivalency of input parameters                         | (a–c; 5.2.3.1)  |                   |                   |                           |
|                           | Output comparison                                       |               |                   |                   |                           |
|                           | Quantity                                               | (a–b; 5.2.3.2.1) |                   |                   |                           |
|                           | Equivalence                                            | (a–c; 5.2.3.2.2) |                   |                   |                           |
|                           | Rigor                                                  | (a–d; 5.2.3.2.3) |                   |                   |                           |
|                           | Agreement                                              | (a–c; 5.2.3.2.4) |                   |                   |                           |
| Applicability             | Relevance of quantities of interest                     | (a–c; 5.3.1)   |                   |                   |                           |
|                           | Relevance of the validation activities to the COU       | (a–d; 5.3.2)   |                   |                   |                           |

Abbreviation: COU, context of use.

<sup>a</sup> Insert the goal for the credibility factor at planning stage (Selected) and the obtained score when the modeling exercise have been performed (Obtained).

<sup>b</sup> The scoring range for the individual credibility factors. Please refer to the V&V40 standard for guidance on grading, where the relevant credibility factor is presented under the quoted paragraph. In general, (a) implies little or no activities on the feature, whereas the highest letter (b, c, or d) implies every aspect investigated and impact accounted for and (b) and (c) denoting intermediate activities where relevant.

<sup>c</sup> The overall credibility on the factors can be scored here, with the categories low, medium, and high.
apparent that steps 1 and 2 of the risk-informed credibility assessment are not clearly addressed, whereas steps 3 and 4 are generally covered, albeit suffering from the lack of information on steps 1 and 2. Clarity on the questions of interest would be an important progress, that would make the prespecification of the credibility activities as well as the assessment of the models more straightforward. Although the EMA guidance on population-PK models describes mainly what the sponsor should report from the modeling exercise, this indirectly poses requirements on some of the credibility activities.  

The information included in the tables for the two NLMEMs are high level and illustrative assessments, not including further details as normally presented in the sponsor’s model analysis reports and the regulatory assessment reports. Even so, the elements of the tables aid in providing an overview of the model use and credibility that is far beyond what we normally see. The assessment shows that it is possible to relate each of the credibility activities on verification and validation to an aspect of the NLMEMs in a manner that makes sense

| Credibility matrix | Description |
|-------------------|-------------|
| Investigational product | A new biologic entity (drug X) |
| Type of model | An NLMEM – pop-PK model and two ER regression models |
| Scientific Question(s) of Interest | 1. Is there a clinically relevant impact of the change in manufacturing processes and in the formulation on the PK of the drug? (Q1)  
2. Are there individual or subgroups of patients in need of dose adaptations compared to the target population? If yes, what would be the appropriate dose adjustment? (Q2) |
| COU | The COU are described separately by question of interest  
1. A pop-PK model was built using rich PK data from phase I studies and will be used to inform the characterization of the effects of formulation and manufacturing process on PK through covariate analysis. A dedicated BE study has not been performed, and the data and the covariate analysis is the only evidence that will be generated to inform the decision on the similarity of the new formulation.  
2. A pop-PK and two exposure-response models were built using data from phase II and III studies to describe the PK characteristics of drug X following subcutaneous administration, and to describe the relationships between drug X exposure and two PD response end points. The models will be used to support the decision on whether there are subpopulations that deviate in exposure levels to a degree where dosing adjustments are needed. |
| Regulatory impact | The Regulatory impact is defined separately by question of interest  
1. High (waiver for a dedicated BE study).  
2. Moderate (additional and key evidence will be available from other sources). |
| Risk-informed decision consequence | The decision consequence is medium to high due to the currently known safety profile of drug X with some serious adverse events as well as lower treatment response rates predicted and observed in the subgroups of patients.  
The overall model risk is considered high for both COU given the consequences of inappropriate dosing in subgroups of patients, with risk of therapeutic failure and life-threatening side effects, while there are safer and more effective treatment alternatives for this indication. |
| Requirements on the credibility activities | Key acceptability criteria are described separately by question of interest  
1. The use of the suggested modeling approach for answering the question of interest represents a new method. There are other established methods of high credibility, and the standard validation activities for pop-PK models are not considered adequate for providing a credible answer. In addition to the numerical and graphical analysis, as described in the EMA guideline on reporting pop-PK models, the modeling and simulation needs to be powered to detect the magnitude of effect that would be of concern (based on BE margins).  
2. The standard numerical and graphical analysis are considered appropriate for the internal model validation. |
| Credibility evidence | The final results are not yet available. The credibility activities outlined for Q1 needs to be presented in more detail in order to understand whether the approach could be acceptable. The activities performed for the interim step of investigating the need for dose adjustment in subpopulations (Q2) is considered relevant and adequate. |
| Model informed decision | Pending more details on Q1. The final decision on dosing recommendations in subpopulations (Q2) will only be made after further clinical data is available. |

Abbreviations: BE, bioequivalence; COU, context of use; EMA, European Medicines Agency; ER, exposure-response; NLMEM, nonlinear mixed effect model; PD, pharmacodynamic; PK, pharmacokinetic; pop-PK, population pharmacokinetic; Q, question.
from an assessment perspective. The scoring of the thoroughness of the investigation into each credibility factor can appear tedious and unrewarding. However, in our experience, this systematic approach to describing the characteristics, range, and accordingly the threshold for acceptability is exactly what is missing in most of the pharmacometric model submissions. Whereas the grading of the credibility factors may seem to be open for subjectivity, even today, the individual model developers or assessors will perform their implicit or explicit scoring of the same elements (the structural model, the covariate selection, the adequacy of the database, the comparison of predicted to observed, etc.) to end up with a recommendation on whether we can use/
accept the model for a specific application. We see the structured approach as a tool to make any differences in opinion explicit and thus also a means to derive and harmonize best practices. Although the 2–4 category grading seems adequate to describe the thoroughness of the investigation on most factors, we suggest that the grading of the observed data could benefit from increased resolution. For transparency and to encourage dialogue on this aspect, we have outlined our approach on the scoring of the observed data both for the test samples and the test condition (intrinsic and extrinsic factors) in Table 2. As this is a first attempt to increase the resolution on how we describe and grade the credibility of the observed data and additional independent factors (such as covariates and study design features), we are open to other views and suggest that this needs further discussions to agree on how to best define relevant and informative credibility factors. If these considerations are made when planning the drug development and the clinical studies, it opens for explicit considerations on the sufficiency of the database to be generated. It also facilitates dialogue on how we could systematically add data on expected worst case scenarios to make sure that we have a sufficient database (increase the boundaries of the tested region) for answering the questions of interest.

As seen in the credibility matrix for the examples, both models are suggested to be used to answer two different questions. This use of a model to answer several questions is encountered in most of the regulatory submissions. Sometimes the regulatory impact and/or the decision consequence is higher for one of the questions of interest, which may lead to a conclusion that the model is credible for one of the questions, but not for the other. We frequently experience that the communication of this can be challenging. Outlining the various questions of interest separately, facilitates explicit differentiation between the model applications, which again encourages clear dialogue on potential differentiated requirements on the rigor of the credibility activities.

| Credibility matrix          | Description                                                                 |
|-----------------------------|-----------------------------------------------------------------------------|
| Investigational product     | A medicinal product (drug Y)                                               |
| Type of model               | An NLMEM – pop-PK model and one ER model                                   |
| Scientific question(s) of  | 1. What is the appropriate dose to be used in the confirmatory phase III study in adolescents and adults? (Q1)  
| interest                    | 2. What is the appropriate dose to be used in the different pediatric subgroups in the PK and ER study in children of 6 months to 12 years of age? (Q2) |
| COU                         | A pop-PK model and an exposure-response model were built using data from the adults and adolescents phase I and II studies to describe the PK characteristics of drug Y following subcutaneous administration, and to describe the relationships between drug Y exposure and the PD response measured as 3 clinical endpoints. These models were used to inform dose selection for the phase 3 study and for the PK and E-R study in younger children as part of an (efficacy) extrapolation approach. |
| Regulatory impact           | The Regulatory impact is defined separately by question of interest         |
| Risk-informed decision      | The decision consequence is considered to impose a medium risk for patient harm. The risks are mostly related to trial failure (efficacy) or insufficient data from the phase III trial for informing final dosing recommendations if the observed exposures are not as expected. The overall model risk is considered medium. |
| consequence                 | Key acceptability criteria are described separately by the question of interest |
| Requirements on the         | 1. The standard numerical and graphical analysis as described in the EMA guidance on reporting pop-PK models are considered appropriate.  
| credibility activities       | 2. The standard numerical and graphical analysis as described in the EMA guidance on reporting pop-PK models are considered appropriate. The model should be continuously updated with available data in children and the dosing recommendations adjusted as needed. |
| Credibility evidence        | The final results not yet available. The activities performed for the interim step of planning the doses for phase III is considered relevant and adequate. |
| Model-informed decision     | The suggested approach is considered credible for answering interim questions on doses to be tested in the phase III trials.  
|                             | The final decision on dosing recommendations in the target adult and in the pediatric population will only be made after clinical data is available. |

Abbreviations: COU, context of use; EMA, European Medicines Agency; ER, exposure-response; NLME, nonlinear mixed effect model; PD, pharmacodynamic; PK, pharmacokinetic; pop-PK, population pharmacokinetic; Q, question.
CREDIBILITY ASSESSMENT OF PHARMACOMETRIC MODELS

Overall, the assessment of the NLMEMs show that the credibility activities are scoring at the lower end of the scales, both for the verification and the validation. We think this can be possible for empirical models in a physiological or pharmacological setting. It does not mean we should not use these models, but it makes the dialogue on their strengths and weaknesses easier. We find that the credibility matrix table supports both the communication with peers and other domain experts and makes the assessment process more efficient by providing a clear focus. Direct comparisons across modeling exercises are also possible with the quickstart tools. This opens the option

| TABLE 6 Example 2 – Credibility activities for the population-PK model at current version |
|---------------------------------------------------------------|
| **Activity** | **Credibility factor** | **Rigor** | **Range** | **Obtained** | **Credibility** |
|-----------------|------------------------|-----------|-----------|-------------|----------------|
| Verification   |                        |           |           |             |                |
| Code           | Software quality assurance | -         | (a–c) a   | Low        |
|                | Numerical code verification | -         | (a–d) a   | Low        |
| Calculation    | Discretization error   | -         | (a–c) a   | Low        |
|                | Numerical solver error | -         | (a–c) a   | Low        |
|                | Use error              | -         | (a–d) a   | Low        |
| Validation     |                        |           |           |             |                |
| Computational model | Model structure | -         | (a–c) b   | Medium     |
|                | Model inputs           |           |           |             |                |
|                | Quantification of sensitivities | -       | (a–c) a-b | Low        |
|                | Quantification of uncertainties | -       | (a–c) a   | Low        |
| Observed data  | Test samples. Measurement uncertainty | | | | |
|                | Quantity               | -         | (a–c) b-c | Medium     |
|                | Range of characteristics| -         | (a–d) c   | Medium     |
|                | Measurements           | -         | (a–c) b   | Medium     |
|                | Uncertainty            | -         | (a–d) a   | Low        |
|                | Test conditions. Intrinsic and extrinsic factors | | | | |
|                | Quantity               | -         | - 16      |             |
|                | Range                  | -         | (a–d) b   | Medium to low |
|                | Measurements           | -         | (a–c) b   | Medium     |
|                | Uncertainty of test condition measurements | -       | (a–d) a   | Low        |
| Assessment     | Equivalency of input parameters | -       | (a–c) NA  | Input data equal to observed data |
|                | Output comparison       |           |           |             |                |
|                | Quantity               | -         | (a–b) b   | Medium     |
|                | Equivalence            | -         | (a–c) b   | Medium     |
|                | Rigor                  | -         | (a–d) a   | Low        |
|                | Agreement              | -         | (a–c) b-c based on level 1 comparisons | Medium     |
| Applicability  |                        |           |           |             |                |
| Relevance of quantities of interest | -         | (a–c) a   | Low        |
| Relevance of the validation activities to the COU | -         | (a–d) b   | Low        |

Abbreviations: COU, context of use; PK, pharmacokinetic.
for establishing best practices and for harmonization of requirement for models used for answering very similar questions.

RECOMMENDATIONS

On a general note, we suggest the credibility framework is well-suited as an overarching framework supporting MIDD. We foresee that, in addition to such an overarching credibility framework, preferably agreed at the ICH cross-regional level, there will still be a need for specific guidance on model types or applications frequently used in drug development. One could perhaps say that this need is already apparent by the fact that some of the regions already have guidance on pharmacometric models, and that the guidelines are separated by the model type (population-PK, PB/PK, etc.) and related applications. The role of the general framework should be to ensure consistency across the various domains of device and drug development and assessment, such as the quality, preclinical, clinical, and health technology domains. The general framework would ensure that new modeling approaches not yet covered by specific guidelines or annexes would still be developed according to a sound set of criteria for establishing credibility evidence. Specific guidance would not need to restate the general framework but could provide the details and methodological choices for verification and validation approaches relevant for the specific model type and related applications.

TERMINOLOGY

Context of use

The specific role and scope of the computational model to address the question of interest.7

Credibility

The trust in the predictive ability of the computational model for the COU.7

Decision consequence

Decision consequence is the significance of an adverse outcome resulting from an incorrect decision.7 In drug therapy, this is typically related to over- or underexposure to a drug but may also be related to other aspects.

Empirical models

Set of mathematical rules and algorithms built to match observed data.18,19 This type of model arises from a top-down modeling approach, which consists in starting from observations or data to reconstitute a set of rules explaining those data. This can be done thanks to statistical, mathematical, and/or computational-intelligence methods. In the context of biological hierarchy, top-down modeling amounts to starting from the higher level (e.g., organism and organs) and breaking it down to smaller elementary pieces (e.g., cell, molecules). Data-driven modeling and reverse engineering are synonyms of top-down modeling.

Extrapolation

Extrapolation in a quantitative modeling context can be seen as accepting the model predictions without requirement for confirmatory data generation, such as in a (fully powered) clinical study.

Mechanistic models

In opposition to the top-down approach, bottom-up modeling consists in the definition of a set of theoretical rules based on known mechanisms at lower scale of organization to reconstitute higher order observed behavior.18,19 This type of model is essentially hypothesis-driven and allows to test the validity of the underlying mechanisms, and to explain an observation, hence the term of white box model.

Model risk

The model risk is the possibility that the use of the model will lead to a decision that will result in patient harm and/or other undesirable impacts and it reflects the risk the decision maker incurs when using the model prediction to support a decision. It is a composite of regulatory impact (or model influence) and the decision consequence.7

Regulatory impact

A specific framework to illustrate how regulators weight the importance of models.5,16 The degree of regulatory scrutiny, level of documentation, and the need for early dialogue is proportional to the weight of the M&S exercise
in regulatory decision making. The regulatory impact is described as low, moderate, or high, and benchmarking against the current evidentiary standard is implicitly included in the concept of regulatory impact.

Qualification

Process to establish the regulatory acceptability of a specific use of a methodology for the development of medicinal products.20 Typically, model validation, verification, and uncertainty evaluation are embedded in the qualification process.

Validation

Validation is the process of assessing the degree to which the computational model is an appropriate representation of the reality of interest.7

Verification

A computational model is the numerical implementation of an underlying mathematical model. The objective of verification is to ensure that the mathematical model is implemented correctly and then accurately solved. Verification is composed of two activities; code verification and calculation verification.7

Question of interest

The specific scientific question, concern, or decision that is addressed.7

DISCLAIMER

The views expressed are the personal views of the authors and cannot be quoted as being made on behalf of the affiliated agencies.

CONFLICT OF INTEREST

The authors declared no competing interests for this work.

AUTHOR CONTRIBUTIONS

Both authors wrote the manuscript.

REFERENCES

1. Sheiner LB. Learning versus confirming in clinical drug development. Clin Pharmacol Ther. 1997;61:275-291.
2. OODA loop. Wikipedia. https://en.wikipedia.org/wiki/OODA_loop Accessed January 24, 2021.
3. Marshall SF, Burghaus R, Cosson V, et al. Good practices in model-informed drug discovery and development: Practice, application, and documentation. CPT Pharmacometrics Syst Pharmacol. 2016;5:93-122.
4. Wang Y, Zhu H, Madabushi R, et al. Model-informed drug development: current US regulatory practice and future considerations. Clin Pharmacol Ther. 2019;105:899-911.
5. Manolis E, Rohou S, Hemmings R, Salmonson T, Karlsson M, Milligan P. 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. 2013;2:1-4.
6. ICH. Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials E9(R1). 2019 [cited 2021 18.06] Available from: https://database.ich.org/sites/default/files/E9-R1_Step4_Guide line_2019_1203.pdf
7. The American Society of Mechanical Engineers. Assessing Credibility of Computational Modeling through Verification and Validation: Application to Medical Devices. American Society of Mechanical Engineers; 2018.
8. European Medicines Agency. Guideline on the reporting of physiologically based pharmacokinetic (PBPK) modelling and simulation. EMA/CHMP/458101/2016. (2018). https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-reporting-physiologically-based-pharmacokinetic-pbpk-modeling-simulation_en.pdf Accessed January 28, 2021.
9. European Medicines Agency. Concept paper on extrapolation of efficacy and safety in medicine development. EMA/129698/2012. (2013). http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/04/WC500142358.pdf. Accessed January 28, 2021.
10. ICH. ADDENDUM TO ICH E11: Clinical investigation of medicinal products in the pediatric population E11 (R1). (2017) https://database.ich.org/sites/default/files/E11_R1_Addendum.pdf. Accessed January 28, 2021.
11. European Medicines Agency. Adopted reflection paper on the use of extrapolation in the development of medicines for paediatrics - Revision 1. EMA/189724/2018. (2018). https://www.ema.europa.eu/en/documents/scientific-guideline/adopted-reflection-paper-use-extrapolation-development-medicines-paediatrics-revision-1_en.pdf. Accessed January 28, 2021.
12. Kuemmel C, Yang Y, Zhang X, et al. Consideration of a credibility assessment framework in model-informed drug development: potential application to physiologically-based pharmacokinetic modeling and simulation. CPT Pharmacometrics Syst Pharmacol. 2020;9:21-28.
13. Beal SL, Sheiner LB, Boeckmann AJ, Bauer RJ (Eds). NONMEM 7.4 users guides. https://nonmem.iconplc.com/nonmem743_guides. (1989–2018).
14. Nguyen TH, Moukassi M-S, Holford N, et al. Model evaluation of continuous data pharmacometric models: metrics and graphics. CPT Pharmacometrics Syst Pharmacol. 2017;6:87-109.
15. European Medicines Agency. Guideline on reporting the results of population pharmacokinetic analyses. CHMP/EWP/185990/06. (2007) https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-reporting-results-population-pharmacokinetic-analyses_en.pdf. Accessed January 28, 2021.
16. Shepard T. Role of modelling and simulation in regulatory decision making in Europe. Presentation at the EMA/EFPIA Modeling and Simulation Workshop. (2011). https://www.ema.europa.eu/en/events/european-medicines-agency-european-federation-pharmaceutical-industries-associations-modelling. Accessed January 28, 2021.

17. Harnisch L, Shepard T, Pons G, Della Pasqua O. Modeling and simulation as a tool to bridge efficacy and safety data in special populations. *CPT Pharmacometrics Syst Pharmacol*. 2013;2:e28.

18. Viceconti M, Pappalardo F, Rodriguez B, et al. In silico trials: verification, validation and uncertainty quantification of predictive models used in the regulatory evaluation of biomedical products. *Methods*. 2021;185:120–127.

19. Viceconti M, Pappalardo F, Rodriguez B, Horner M, Bischoff J, Musuamba Tshinanu F. In silico trials: Verification, validation and uncertainty quantification of predictive models used in the regulatory evaluation of biomedical products. *Methods*. 2021;185:120–127.

20. European Medicines Agency. Qualification of novel methodologies for medicine development. https://www.ema.europa.eu/en/human-regulatory/research-development/scientific-advice-protocol-assistance/qualification-novel-methodologies-medicine-development-0. Accessed January 28, 2021.

How to cite this article: Skottheim Rusten I, Musuamba FT. Scientific and regulatory evaluation of empirical pharmacometric models: An application of the risk informed credibility assessment framework. *CPT Pharmacometrics Syst Pharmacol*. 2021;10:1281–1296. doi:10.1002/psp4.12708