Deliverable 1:
Preliminary study plan

Title of project	***Uncertainty quantification for complex models supporting regulatory decision making***
Tender ID	*EMA/2020/46/L3.02-ROC22*
Consortium	*CONFIRMS* *(CONsortium For Innovation in Regulatory Medical Statistics)*
Objectives	(1) An assessment of the published literature on uncertainty quantification methods for complex models used in drug development and/or regulatory approval. (2) Simulation studies to investigate the operational characteristics of the methods identified in (1) and any relevant unpublished methods *(3)* Toolkit to derive *a priori* recommendations of minimally sufficient validation datasets and to use the best performing methods on available validation datasets.
Scientific contact person	*Andrew C. Hooker, Uppsala University*
Administrative contact person(s)	*Karin Brauneis* *Medical University of Vienna* *Spitalgasse 23, 1090 Vienna, Austria* *Tel: +43 1 40400 74880* *Email: medstat@meduniwien.ac.at*
Date	*2025-06-19*
Version	*1.0*

Role of different organisations participating in the study

Organisation	Key person	Role(s)	Status	Estimated budget allocation (e.g. % of budget allocation)
Medical University of Vienna (MUW), Spitalgasse 23, 1090 Vienna, Austria	Prof. Dr. Martin Posch Assoc. Prof. Dr. Franz König	Consortium Lead, Scientific advisor, statistical expertise, writing and review.	consortium leader	10.7%
Uppsala Universitet (UU), Von Kraemers Allé 4, 751 05 Uppsala, Sweden	Prof. Dr. Andrew Hooker Prof. Dr. Mats Karlsson Dr. Xiaomei Chen, Dr. Yevgen Ryexnik Zhe Huang, MSc.	Scientific advisor, pharmacometrics expertise, complex model expertise, Project management, project implementation, research, writing and review.	Scientific lead, consortium member	57.9%
Universitätsmedizin Göttingen (UMG), Robert-Koch-Str. 40, 37075 Göttingen, Germany	Prof. Dr. Tim Friede Dr. Norbert Benda Paula Starke Lucia Schneider	Scientific advisor, research, statistical expertise, writing and review.	consortium member	13%
Österreichische Agentur für Gesundheit und Ernährungssicherheit GmbH (AGES), Spargelfeldstraße 191, 1220 Wien, Austria	Dr. Florian Klinglmüller Tobias Fellinger, MSc. Elham Yousefi Lynette Caitlin Mikula	Scientific advisor, regulatory science, statistical expertise, writing and review.	consortium member	16.9%
Swedish medical products agency (MPA), Uppsala, Sweden	Dr. Robin Svensson	Regulatory expertise within pharmacometrics and assessing results from complex models	Sub-contractor of UU	1.5%

Table of contents

Title of the study

Uncertainty quantification for complex models supporting regulatory decision making

List of abbreviations

AGES - Österreichische Agentur für Gesundheit und Ernährungssicherheit GmbH

EMA - European Medicines Agency

EPAR - European Public Assessment Report

MAA - Centralised Marketing Authorisation Application

MUW - Medical University of Vienna

SA - Scientific Advice

UMG - Universitätsmedizin Göttingen

UU - Uppsala Universitet

MIDD - Model-informed drug development

PBPK - Physiologically Based Pharmacokinetic

QSP - Quantitative Systems Pharmacology

PK - pharmacokinetic

PKPD - pharmacokinetic and pharmacodynamic

IVIVE - *in vitro-in vivo* extrapolation

ML - machine learning

PD - pharmacodynamic

UQ - Uncertainty quantification

PART I PROPOSED METHODOLOGY FOR CONDUCTING THE WORK

1.1. Abstract

Model-informed drug development (MIDD) plays a crucial role in drug development, and evaluation by regulatory agencies. A key component of MIDD is the use of complex mechanistic models, built on an understanding of physiology, drug characteristics, and pharmacology, to make extrapolations. However, these models often result in inadequate assessment of uncertainty due to their complexity and the sources of parameters. The current proposal aims to enhance drug development and regulatory decision-making by improving uncertainty quantification (UQ) methods for complex models. The primary objectives are to assess existing UQ methods, conduct simulation studies to evaluate these methods, and develop a toolkit and tutorial for applying the best methods. The methodology involves a comprehensive literature review to identify and summarize UQ methods, followed by simulation studies to test selected methods on use-cases such as drug-drug interactions and pediatric dosing. The project will culminate in the development of user-friendly tools and tutorials for implementing UQ methods. The CONFIRMS consortium, with partners from the Medical University of Vienna, Uppsala University, Universitätsmedizin Göttingen, and the Austrian Medicines and Medical Devices Agency, will perform the project. It is structured into five work packages, ensuring a comprehensive approach to improving UQ of complex models in regulatory science.

1.2. Background on the research question

Model-informed drug development (MIDD) plays a crucial role not only in drug development[1] but also in drug evaluation by regulatory agencies[2,3]. A key component of MIDD is the use of complex mechanistic models, such as Physiologically Based Pharmacokinetic (PBPK)[4] and Quantitative Systems Pharmacology (QSP)[5] models. Unlike statistical models, which are based on data, PBPK and QSP models are built on an understanding of physiology, drug characteristics, and pharmacology. These models are typically intricate, with numerous parameters derived from previous preclinical and clinical studies.

By leveraging prior knowledge of the system and the drugs being studied, PBPK and QSP models are often used for extrapolations. This includes determining the first-in-human dose based on preclinical data, predicting pharmacokinetic and pharmacodynamic (PKPD) profiles in special situations (such as drug-drug interactions) and special populations (such as pediatric patients), aiding in dose adjustment decisions, and optimizing clinical trial designs[6,7].

Despite the theoretical and predictive appeal of complex mechanistic models, their practical applications are hindered by several limitations. One major issue is the lack of specific information on mechanisms and physiological parameter values, especially from special populations during model building[8]. This gap, along with other weaknesses, reduces the credibility of these models.

For example, unlike data-based statistical models, current approaches for PBPK and QSP models do not adequately assess uncertainty due to their complexity and the sources of set parameters. Most physiological parameters are fixed at average values from literature, without accounting for parameter uncertainty and variability. Drug-specific parameters, derived from observed data (in vitro, in vivo, and/or clinical), are also fixed without considering potential bias and uncertainty from experimental setups and *in vitro-in vivo* extrapolation (IVIVE)[4,8]. Further, although some parameters are estimated based on clinical data, the uncertainty of those parameters is usually not evaluated or reported.

Additionally, the adage "all models are wrong" applies here; no matter how complex, these models cannot fully represent biological systems, contributing further to model uncertainty. This impairs the predictability of complex models. Discrepancies between simulated data and observed data are typically described using concise metrics, such as the average (absolute) fold prediction error[9]. Models are sometimes validated by comparing the prediction error to an arbitrary value (e.g., 2-fold or 3-fold), which may not provide a clear understanding of the pharmacological significance of these

differences. Moreover, fold-change metrics can oversimplify the diversity in a model's predictive performance and lack the sensitivity needed to detect variations among specific subgroups.

Although sensitivity analysis is regularly conducted, as suggested by regulatory agencies, to evaluate the robustness of PBPK models[10,11], the output is often fold-based uncertainty and/or graphs, focusing on a small portion of key and/or uncertain parameters. Validation data are not directly involved in sensitivity analyses. The above challenges create difficulties for regulators to assess a model's predictive capabilities and its applicability beyond the data used for its development. Further research is needed to develop better methods for characterizing and reporting uncertainty in a model's predictive performance.

In recent years, machine learning (ML) has garnered significant interest in drug discovery and development due to its ability to address several critical challenges in the field[12,13]. During the drug discovery stage, ML algorithms can identify new drug targets, predict potential adverse effects, and optimize drug candidates. In preclinical studies, ML can predict the pharmacokinetic (PK), pharmacodynamic (PD), and safety profiles of new drug candidates. Across different phases of clinical trials, ML approaches can enhance trial management and treatment decisions. For example, ML can predict dose-limiting toxicities[14] and improve dose-finding in phase I studies[15] making more efficient recruitment of trial participants and reducing the required sample size of the phase II studies[16], and build reliable disease progression models by combining ML and pharmacometrics approaches in phase III clinical trials[17].

Similar to mechanism-based models, uncertainty quantification is essential for advancing the reliability and robustness of ML models. However, a widely adopted framework for uncertainty quantification in machine learning is still lacking[18], underscoring the need for further research and development in this area.

Uncertainty quantification (UQ) is a scientific discipline that provides a computational framework for quantifying uncertainties in both inputs and outputs, thereby facilitating predictions with quantified and reduced uncertainties[19]. Typically, UQ involves one or more mathematical models for a quantity of interest, with some uncertainty regarding the correct form of the model or models. These uncertainties are often treated probabilistically[20]. An appropriate UQ method is essential for regulatory decision making which may depend on type I error and power related to hypothesis tests or on confidence intervals / credible intervals related to parameter estimation or prediction. UQ becomes more important when using complicated analysis methods and/or complex models. For example, in previous work, members of our consortium (UU) have demonstrated that in some model-integrated methods based on population pharmacometric models, UQ methods based on information matrices showed inflated type I error, while type 1 error could be controlled with a more appropriate UQ method (sampling importance resampling)[21–23].

Another group in our consortium (UMG) contributed to the literature on another UQ method, namely Bayesian random-effects meta-analysis, in particular in the setting of few studies[24]. Shrinkage estimation in these models can be applied to implement dynamic borrowing across different types of studies[25], regions[26] or dose regimens (combinations of doses and frequencies)[27]. More generally, Bayesian hierarchical models are a flexible framework for evidence synthesis, borrowing of information and prediction, as well as a powerful UQ method, especially for applications where variability and uncertainty arise from multiple sources. That is specifically of interest for this project that focuses on complex mechanistic models, the parameters of which are from literature and data from multiple studies.

Other UQ methods are also of interest for application in complex models used in drug development and evaluation. These include Bayesian calibration[28], model discrepancy analysis[29], and surrogate modeling (e.g., based on polynomial chaos)[30,31]. Various approaches of sensitivity analysis can also be further explored, separately or in combination: (1) probabilistic sensitivity analysis[32] assesses uncertainty by assuming input parameter uncertainty distributions, in contrast to the current practice in PBPK modeling, which often uses fixed values (i.e., deterministic sensitivity analysis). (2) Global

sensitivity analysis[33,34] considers parametric correlations to provide a more comprehensive assessment of model sensitivity. An alternative approach involves using prior information, such as the frequentist prior, to link preclinical and clinical data through PBPK models. This method, employed by Mats Karlsson[35,36] (a member of our consortium), estimates model parameters while assessing uncertainty.

In summary, there is a significant lack of well-established and widely recognized methods for assessing the uncertainty of complex models in drug development and evaluation. Recent advancements in uncertainty quantification (UQ) offer an opportunity to enhance the methodologies for evaluating and validating PBPK, QSP, and ML models, thereby increasing the credibility of these modeling tools for regulatory decision-making. Our consortium, comprising experts in pharmacometrics and statistics, is ideally positioned to undertake this interdisciplinary project, advancing the application of complex models in drug development and regulatory decision-making.

1.3. Objectives

1. To identify available statistical methods for UQ that are potentially applicable to complex models used in drug development and/or regulatory evaluation and decision making.

   This will provide a summary of the current state of the art in UQ in the context of complex models used in drug development and/or regulatory evaluation and decision making (e.g., PBPK, QSP, ML models). Furthermore, it will give a comprehensive overview on the current knowledge on the properties and interpretability of these methods, as well as applicability for PBPK, QSP models, and ML. Moreover, it will provide a sound basis for the selection of methods to be studied further in this project.

2. To assess the operational characteristics of selected UQ methods

   The comprehensive simulation study and summary of theoretical results will provide a detailed assessment of the selected approaches, covering different method settings (such as selections of priors for bayesian approaches), trial designs, and scenarios for the data generating process. This will be the basis for an appraisal of the methods with regard to their performance in the simulation and to derive recommendations.

3. To promote the application of selected methods

   This will be achieved by publishing the study results, developing a toolkit and tutorials. The resultant deliverables will provide users an easy-to-use toolkit and instructions on how to derive a priori recommendations for validation datasets and to carry out UQ methods for complex modeling.

1.4. Methodological approach

As proposed in the call, the methodological approaches consist of the following deliverables:

1. Definition of a preliminary study plan to define the literature review as well as selection criteria to select methods and use-cases to be studied in the simulation study; technical details on simulation study; and a preliminary outline of the tutorial.
2. Literature Review
3. Writing of the study protocol and selection of methods, use-cases and associated design features of the validation data
4. Simulation study report (theoretical assessment, software development, simulation study, definition of recommendations and writing of the report)

5. Writing of a manuscript
6. Developing a toolkit and tutorials

Below we describe each element in more detail.

Deliverable 1: Preliminary study plan

The preliminary study plan will include a protocol for the literature review and an outline of the simulation study including anticipated operational characteristics and preliminary technical details. It will also include strategies for toolkit development and an outline of the tutorial. In general, the preliminary study plan will provide guidance on coordination and execution of the project, allowing flexibility for adjustment based on the findings from the literature review and discussion with the EMA.

Deliverable 2: Literature review

A comprehensive literature search will be conducted in electronic literature databases to identify (a) methodological articles on methods for quantifying uncertainty in QSP and PBPK models, (b) validation approaches of QSP and PBPK models, and (c) approaches of uncertainty quantification in other fields dealing with large/complex models. In a first step, titles and abstracts will be screened and papers that clearly do not meet the predefined inclusion criteria will be excluded. In a second step, full texts of the potentially relevant publications will be reviewed according to pre-specified inclusion and exclusion criteria. In an additional hand search, reference lists of included publications and relevant reviews will be checked. From the full texts, relevant data will be extracted to perform a critical appraisal of the methods. These will serve as a basis for the selection of methods for the simulation study. Details on the systematic literature review are provided in the Subprotocol A: Literature review (Section 5.2).

Additionally, a search of regulatory documents (e.g. EPARs, EMA-SA letters, regulatory guidance) using AGES' EMA Scientific Advice repository, EMA's Scientific Explorer and MPH's M-RECON will be performed to identify regulatory texts, marketing authorization procedures and scientific advice where uncertainty quantification methods for QSP or PBPK models are discussed. This search will be performed using a search strategy similar to the literature review. Identified documents will be screened for the assessment or discussion of methods to compare large complex models with validation data or with uncertainty quantification of these models. The review will be conducted by AGES employees with appropriate access credentials and who are bound by confidentiality agreements to ensure the protection of sensitive information. All accessed documents and procedures will be handled in compliance with applicable confidentiality and security policies. The findings shared outside the agency - i.e. within the consortium or via eventual publications - will be limited to high-level summaries, broad categories, and aggregated data that will not reveal or compromise any specific confidential information. Examples of shared results may include the frequency of use of specific methods or categories of methods and generalized parameter estimates for disease-related models, without disclosing proprietary details or sensitive procedural content. This data will then serve to identify potential candidate methods, define parameter ranges for distributional scenarios in the simulation study, and to derive case studies to illustrate the findings of the simulation study. For a detailed description of the review of regulatory procedures please see the corresponding Subprotocol B: EMA EPAR review (Section 5.3).

In order to identify software implementations of methods identified in the systematic review, articles included will be checked for references to software implementations (either in main text or appendices). In addition, for each method identified in the literature review an internet search will be performed and articles citing the parent article will be scanned with the aim to identify potential software implementations. In addition, widely used software repositories such as CRAN will be screened. The software review will be performed similarly as the recent software review on platform trials (Meyer et al. 2021). The primary objective of this review will be to identify software implementations suitable to be

used in the simulation study. Therefore, we will focus on open-source and free software implementations primarily in R (R Core Team 2018). Consequently, this review part may not be comprehensive with respect to the entire software landscape. However, corresponding findings will be reported in the review and thereby complement the theoretical discussion.

Deliverable 3: Simulation study protocol

The study protocol for the simulation study will be based on the findings of the literature review as well as review of regulatory documents. In particular, the protocol will specify the objectives of the simulation study, list UQ methods and related assumptions and/or settings to be investigated, and representative use-cases. In addition, a series of design features of validation datasets will be listed for the simulation study to evaluate the studied methods. For more information see the next section on "assessment of methods". Criteria for the selection of methods are: reported favourable operating characteristics (e.g. type I error rate and power, coverage and width of confidence and/or credible intervals, as well as bias and imprecision of model-generated metrics of interest such as the ability to detect the least significant difference). All aspects of the simulation study will be specified in a detailed protocol (Deliverable 3) and will be discussed and agreed to with EMA before finalization.

Deliverable 4: Simulation study

A simulation study will be conducted following the simulation study protocol (Deliverable 3). Currently, we plan to execute the study with the following key aspects, with details subject to change based on the literature review and discussions with the EMA.

1. **Use-cases**: Two to three use cases will be included, focusing on complex models used during drug development. These use cases will represent common challenges in the application of PBPK/QSP/AI modeling and will be selected based on a review of marketing authorization procedures. Specifically, we are interested in:
    a. **PBPK models for predicting drug-drug interactions**. The PBPK modeling approach is a valuable tool for predicting drug-drug interactions (DDIs) during regulatory approval processes. DDIs for small molecules typically occur through drug metabolism or transporters[37]. During drug development, DDIs can be initially explored through in vitro studies and subsequently confirmed in clinical trials. However, it is impractical to conduct clinical trials for all possible DDI combinations. Consequently, PBPK modeling has become an essential tool for predicting DDIs. A use-case for a simulation study could be investigating enzyme-mediated inhibition of CYP3A4[38], for example, or predicting transporter-mediated DDIs. The latter presents more challenges including the limited abundance of transporters in elimination organs and the physiological limitations of in vitro systems in mimicking in vivo conditions, often leading to inconsistencies between simulated and observed DDI clinical studies[4,8] quantification of uncertainty even more crucial for interpreting PBPK-predicted DDI clinical trial results. The use case could be similar to the study reported by Posada et.al[39], where researchers applied a PBPK model to predict the OAT3 (organic anion transporter 3)-mediated DDI between baricitinib, a drug investigated for the treatment of inflammatory diseases and OAT3 perpetrators.
    b. PBPK models for extrapolating drug doses and computing sample sizes for pediatric studies. Conducting clinical trials in pediatric populations presents significant challenges due to ethical concerns, limited patient availability, and potential risks. These constraints often lead to insufficient clinical data for determining appropriate dosing in children. However, pediatric studies are required for new drug applications as part of market authorization. Consequently, PBPK modeling becomes an invaluable tool for predicting adequate dosing strategies for pediatric patients. Nevertheless, the lack of information in children, such as maturation profiles of protein expressions, increases the bias and uncertainty of predictions from PBPK models, which may be poorly characterized.
    c. **ML-based prediction of the concentration and the exposure-response modeling.** Population pharmacokinetic (PopPK) modeling is essential for predicting drug

concentration to understand drug behavior in different subjects. However, there are scenarios when existing PopPK models result in poor prediction accuracy and, thus, limit their effectiveness in adjusting dosages for individual patients. Additionally, these models often lack uncertainty quantification, making it hard to assess their reliability in clinical settings. Using machine learning for predicting concentrations combined with uncertainty quantification techniques may enhance the practical application of PopPK models in real-world clinical practice. This use-case could look like that presented by Verhaeghe *et al.* [40,41]. The ML-predicted concentration can be then used in building exposure-response models and decision making.

2. **UQ methods**
    a. **Bayesian hierarchical meta-regression** methods integrate variability and uncertainty from multiple sources, making them valuable for complex mechanistic models. These models often rely on parameters derived from literature and data from multiple studies, including preclinical in vitro and in vivo studies, as well as clinical studies. Different prior distributions and weights will be investigated
    b. **"Frequentist" prior** approaches can integrate prior information into model estimation based on observed data, serving as an efficient and powerful tool to link preclinical and clinical data. Unlike the Bayesian approach, a penalty term is introduced in the objective function during model estimation when parameters deviate from prior values. In previous studies[35,36], this method showed similar results to the Bayesian approach while reducing the run-time for a whole-body PBPK model. It can be a useful tool for analyzing population PBPK models and is easily implemented in NONMEM, a software used for nonlinear mixed modeling for population PK.
    c. **Sensitivity analyses** are routinely performed as required by regulatory agencies for uncertainty evaluation. Therefore, they will be included in the simulation study as a reference method for method comparison. The parameters and their corresponding investigated values will be based on the original study in the use cases.
    d. **Other UQ methods** will be chosen after literature review and discussion with EMA. Potentially, they will include Bayesian calibration, Probabilistic sensitivity analysis, model discrepancy analysis, and Surrogate-Based uncertainty, as suggested in the technical specification document.

3. **A simulation framework** will be created to streamline the simulation study procedures and will include the following components (see section 1.6 for additional information). The transporter-mediated DDI use-case of baricitinib[39] will be used here to illustrate the simulation study when necessary.
    a. **Generating simulated datasets** using:
        i. **simulation models**, ideally, available from selected use-cases. If not available, we will establish the models based on reported information. It is noted that the models will be modified for simplification and/or meeting the purpose of the simulation study. For example, drug-specific parameters for transporter kinetics will be adjusted:
            1. for evaluating type I error. The datasets will be simulated under the null hypothesis, which is the existence of a minimally clinically relevant DDI. This can be achieved by setting related parameters, such as the OAT3-mediated baricitinib uptake parameter Vmax, so that the true geometric mean ratio of AUC or Cmax between with and without drug co-administration is located at the edge of concluding DDI (i.e., 0.8 or 1.25). To find appropriate sets of parameters, a simulation similar to a sensitivity analysis will be performed.
            2. for power. The datasets will be simulated under the alternative hypothesis, which is the absence of a clinically relevant DDI (simulation made within the range of clinically irrelevant DDI). This will be done by setting related parameters so that the geometric mean ratio of AUC or

Cmax is within the range of 0.8 and 1.25, e.g. AUCratio=1. We may simulate datasets using different PK models to investigate the impact of different sources of uncertainty, such as

- PBPK models that consider uncertainty or variability on certain parameters, i.e., population PBPK.
- population PK models for data simulation while using reported PBPK models for analysis will provide an opportunity to explore UQ methods with model misspecification

ii. **Study designs** for validation data will be based on use-cases. For example, clinical DDI studies commonly use crossover designs. In addition, a series of study designs will be included to evaluate the robustness of the methodological approaches with respect to certain features of the validation dataset: total sample size, sample size of subgroups, number of observations per subject/experiment, missing data due to below limitation of quantification, (in)balance between subgroups, etc. In addition, we are interested in investigating the application of optimal design theory[42,43] for required study design features of the validation dataset in the context of different UQ methods.

b. **performing UQ analysis:** the investigated methods (e.g., Bayesian hierarchical meta-analysis, "frequentist priors", and others) will be performed on each simulated validation dataset and prior information from in vitro update and inhibition studies. As a result, the model parameter and secondary PK exposure metrics (such as AUC, Cmax and their ratio between with and without perpetrators) will be obtained together with their uncertainty. The 90% confidence interval for AUC and Cmax ratios will be used for the conclusion of DDI. We plan to perform the simulation study in R. Ideally, there are available R packages to perform investigated UQ methods. If this is not the case, and implementation is feasible within time and cost constraints, then we will create code for the method.

c. **Simulation summary:** Each simulation study will be summarised according to the simulation study protocol, and performance measures such as the type I error rate, power, bias, root mean squared error, variance, coverage and width of confidence/credible intervals, etc. will be reported. Comparisons will be performed among UQ methods as well as different method settings, study designs, based on which a general recommendation of UQ methods will be provided. The summary of the simulation study design will be similar to our previous work that investigated the performance of model-integrated bioequivalence methods[21,22].

Deliverable 5: Writing of a manuscript.

Results will be submitted to international, peer-reviewed journals as open access publications, and presented at national and international conferences. See section 1.7 for more information.

Deliverable 6: Toolkit and tutorial

The theoretical appraisal of methods, the case studies and the simulation results will be used to derive *a priori* recommendations of minimally sufficient validation datasets and recommend the best performing methods on available validation datasets for uncertainty quantification of complex models. This will, for example, include which supportive analysis should be provided to assess the robustness of the results and how to evaluate the underlying assumptions. The best practice guide (tutorial)
shall give clear recommendations on how to present the results, e.g., which statistical measures and figures should be included in the study report for evaluation by regulators. For example, we expect that confidence/credible intervals will be an important part of the output of any study report for assessment by regulators. The toolkit will include a set of R-scripts tailored to the use-cases investigated in deliverable 4.

The tutorial and toolkit are currently envisioned as an R package accompanied by a web-based tutorial or vignette, similar to this example. Preliminary Outline of the Tutorial:

1. **Overview of Findings**: A summary of the literature review and key results from the simulation study.
2. **Recommendations**. Guidance on:
    a. Minimally sufficient validation datasets.
    b. Best-performing methods for uncertainty quantification of complex models, based on available validation data.
3. **Supportive Analyses**: Suggestions on additional analyses to assess the robustness of results.
4. **Assumption Evaluation**:How to critically evaluate the underlying assumptions of the applied methods.
5. **Results Presentation:** Best practices for presenting results, including recommended statistical measures and visualizations to be included in study reports for regulatory review.
6. **Toolkit Integration**: Direct links and instructions for using the toolkit to implement the recommended methods.

1.5. Use of empirical data

We will rely on data from completed marketing authorization procedures or from published literature to inform key aspects of the simulation study. In line with the clinical scenario evaluation framework this will include disease specific features concerning assumptions related to treatment effect on specific model parameters, variability and underlying models, as well as options concerning clinical trial design, recruitment process and relevant estimand definitions. Historical data will be limited to aggregate data available in publications, study reports, EPARs or other publicly available sources. Use of patient level data is not anticipated. Aggregate study data may be extracted (e.g. point estimates) to inform the choice of relevant parameter space ranges for the simulation study. In addition these procedures will serve as examples for a number of case studies that will be developed to illustrate each of the distributional scenarios considered in the simulation study and derive recommendations in the form of estimand templates for each case.

It is anticipated that individual patient level data will be simulated from the relevant complex models investigated in the simulation study, or from other models describing the same clinical situation. For example, simulations from population PK models may be used to generate data that is similar to that generated from a PBPK or QSP model, but with different underlying assumptions.

Additional case studies may be identified by means of a review of EPARs (see section 1.4 above) and from examples discussed in the literature.

1.6. Statistical software and programming

A simulation framework will be implemented as a dedicated software package using the statistical programming language GNU R[44]. The framework will facilitate design, execution, summary and visualisation of results. Implementation as an R-package also aids collaborative development, code-sharing between partners and EMA, as well as execution of specific simulation studies on the high-performance computing infrastructure available at MUW, UMG and UU. If software is needed that is not available in R/Python, then the tools may be used with consortium member licences. In this case interface functions (e.g. to facilitate import of data generated in other software, or export of simulated data for analysis in other software) will be implemented. Design and implementation will take into account recommendations in Hallgren[45], Sigal and Chalmers[46], and Lee *et al.*[47], where suitable development of the core simulation framework and organization of the code will use and expand on existing implementations such as SimDesign[48]. This package provides utilities to structure, implement and run Monte Carlo Simulation Designs. SimDesign nicely accommodates the intended modular approach to simulation design and provides features to facilitate parallelization, quality assurance, and integration of software written and executed outside R (e.g. SAS).

The simulation software will be implemented in a modular way. It is expected that the components will comprise:

- a core simulation module that implements dispatch, execution and concatenation of results;
- several modules implementing algorithms to fit and evaluate different models;
- one or more modules to generate pseudo-random numbers according to various distributional scenarios;
- an output and presentation module that implements tabulation and plotting of simulation results.

Development of the core simulation model will be led by the team at UU. High performance cluster computing facilities are available at MUW, UMG, and UU.

Such a modular simulation framework has the attractive property as individual components can be implemented by different partners, permits extensive use of existing libraries and promises high reusability for potential follow-up investigations. Seamless interaction between modules is guaranteed by comprehensive specification and documentation of module interfaces.

Simulation designs (i.e. parameter scenarios) will be specified according to scenarios derived from the reviews of the scientific literature and regulatory procedures. Optimization of validation data designs will rely on the R package PopED developed by UU.

Data generating functions will largely rely on existing packages such as PK-sim[49], MoBi[49], nlmixr[50], and NONMEM[51]. The module shall be implemented to facilitate specification of distributional scenarios by different partners in the consortium.

Analysis functions to fit and evaluate individual models will be implemented by different partners according to their specific expertise. High re-use of existing tools (e.g. PK-sim, MoBi, nlmixr, NONMEM, pharmpy/pharmr[49], xpose[52]/xpose4[53], PsN[52]), the last three of which are developed by UU, are anticipated. Not only does this reduce the amount of development work but also ensures use of high quality code that adheres to quality specifications as required by CRAN and that has been validated through peer review.

Summary functions will be implemented to compute various measures to quantify summary model metrics as well as performance characteristics for methods comparison (bias, Type I error, etc.).

Compartmentalization of output and presentation into a separate module permits that corresponding software development can be deferred to a later stage in the project. For testing purposes and preliminary communication of early results between partners and EMA a rough prototype can suffice. This will free up resources in the initial stages to focus on implementation of the core functionality and permit flexible adaptation of the output to meet publication and presentation needs.

1.7. Publication and communication of results

Results will be submitted to international, peer-reviewed journals as open access publications, and presented at national and international conferences. The draft manuscripts will also be put on public repositories such as arxiv.org. Furthermore, we will organize webinar sessions to present and discuss the result of the review and simulation studies with EMA (and leave it to the discretion of EMA which stakeholders should be invited, e.g., assessors from EMA and NCAs). Relevant reporting standards will be considered for publications, e.g. the adaptive design CONSORT extension (ACE) or for the systematic literature review the PRISMA-Statement ('Preferred Reporting Items for Systematic Reviews and Meta-Analyses'). A draft manuscript will be delivered to EMA in month 15 of the proposed work (see timelines below). Our consortium is committed to open science and reproducible research such that storage of data and code for long-term availability is planned for the project. We actively encourage and support timely publication of data and code wherever possible. Software will be implemented as an r-package. This permits reuse of code from existing packages via dependencies.

Development code will be maintained using a distributed version control system in order to manage and track changes to the software code. Using a distributed system facilitates collaboration between development staff, as well as distribution of preliminary working versions of the package among users (researchers, EMA). For example R-packages can be directly installed from github circumventing the need for frequent submissions to CRAN or the exchange of compiled binary packages. It is foreseen that the final code will be made available under a free software licence (to be agreed with EMA) and, if applicable, published on CRAN. All deliverables will be delivered as PDF files. In addition, software code will be delivered as text files. Generated data sets will be provided as csv files.

1.8. Limitations of the research methods

The proposed literature review will use filters to limit the number of papers to be screened manually. Therefore, papers will be missed. To address this risk, we investigate the impact of each filter on the number of identified manuscripts and inspect in random samples excluded papers.

In general, an important limitation of simulation studies is the challenge to cover all relevant scenarios as the number of scenarios to be considered increases exponentially with the dimension of the parameter space. To maximize the robustness of conclusions based on simulation studies, the scenarios to be investigated will be selected based on (i) scenarios that have been discussed in the literature or observed in marketing authorisation procedures and (ii) where theoretical considerations suggest that they may give upper bounds for operating characteristics. Case studies and simulation studies will be mainly based on published data and models. While simulations from models may not provide the same distributional information as patient level data, they should provide relatively good descriptions of the overall population.

Risk No	Risk	Mitigation Strategy
1	Key personnel leaves post	Backups will be provided either internally or by other consortium members. All consortium members have qualified staff who can step in for the specific tasks .
2	Delay in completion of individual tasks	Coordination and monitoring through the executive board. Reallocation of tasks between key staff and consortium members if necessary to avoid delay.
3	Pandemic situation, travel restrictions or access to offices	All consortium members are well equipped for remote working. Simulation servers can be accessed remotely.
4	Unforeseen interim research results make a change of the study protocol necessary	Potential changes to the agreed study protocol will be discussed in the executive board and EMA and the protocol accordingly amended.

1.9. Ethical aspects

As part of the work on this project, statistical methods for quantification of complex models will be assessed. These assessments will be based on theoretical analysis as well as simulation studies. Therefore, the assessment of these methods in itself does not raise ethical concerns. At this point, we do not anticipate the use of patient level data. In case patient level data would be used, appropriate approval by local ethics committees will be obtained.

1.11. References

1. Milligan PA, Brown MJ, Marchant B, et al. Model-based drug development: a rational approach to efficiently accelerate drug development. *Clin Pharmacol Ther*. 2013;93(6):502-514. doi:10.1038/CLPT.2013.54
2. Wang YN, Zhu H, Madabushi R, Liu Q, Huang SM, Zineh I. Model-Informed Drug Development: Current US Regulatory Practice and Future Considerations. *Clin Pharmacol Ther*. 2019;105(4):899-911. doi:10.1002/cpt.1363
3. Madabushi R, Seo P, Zhao L, Tegenge M, Zhu H. Review: Role of Model-Informed Drug Development Approaches in the Lifecycle of Drug Development and Regulatory Decision-Making. *Pharm Res*. 2022;39(8):1669-1680. doi:10.1007/S11095-022-03288-W/FIGURES/3
4. Jones H, Rowland-Yeo K. Basic Concepts in Physiologically Based Pharmacokinetic Modeling in Drug Discovery and Development. *CPT Pharmacomet Syst Pharmacol*. 2013;2(8):63. doi:10.1038/psp.2013.41
5. Braakman S, Pathmanathan P, Moore H. Evaluation framework for systems models. *CPT Pharmacomet Syst Pharmacol*. 2022;11(3):264-289. doi:10.1002/psp4.12755
6. Lin W, Chen Y, Unadkat JD, Zhang X, Wu D, Heimbach T. Applications, Challenges, and Outlook for PBPK Modeling and Simulation: A Regulatory, Industrial and Academic Perspective. *Pharm Res*. 2022;39(8):1701-1731. doi:10.1007/s11095-022-03274-2
7. Zhao P, Zhang L, Grillo J, et al. Applications of Physiologically Based Pharmacokinetic (PBPK) Modeling and Simulation During Regulatory Review. *Clin Pharmacol Ther*. 2011;89(2):259-267. doi:10.1038/clpt.2010.298
8. Tsamandouras N, Rostami-Hodjegan A, Aarons L. Combining the 'bottom up' and 'top down' approaches in pharmacokinetic modelling: fitting PBPK models to observed clinical data. *Br J Clin Pharmacol*. 2015;79(1):48-55. doi:10.1111/bcp.12234
9. Jones HM, Parrott N, Jorga K, Lavé T. A novel strategy for physiologically based predictions of human pharmacokinetics. *Clin Pharmacokinet*. 2006;45(5):511-542. doi:10.2165/00003088-200645050-00006
10. Food and Drug Administration. Physiologically based pharmacokinetic analyses — format and content guidance for industry. *Cent Drug Eval Res*. Published online 2018.
11. Committee for Medicinal Products for Human Use (CHMP). Guideline on the reporting of physiologically based pharmacokinetic (PBPK) modelling and simulation. *Eur Med Agency*. Published online 2018.
12. Vamathevan J, Clark D, Czodrowski P, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov*. 2019;18(6):463-477. doi:10.1038/s41573-019-0024-5
13. Ryeznik Y, Sverdlov O. Pharmacometrics and machine learning in drug development. In: Pais A, Vitorino C, Nunes S, Cova T, eds. *Artificial Intelligence for Drug Product Lifecycle Applications*. Academic Press; 2025:99-108. doi:10.1016/B978-0-323-91819-0.00004-X
14. Bedon L, Cecchin E, Fabbiani E, et al. Machine Learning Application in a Phase I Clinical Trial Allows for the Identification of Clinical-Biomolecular Markers Significantly Associated With Toxicity. *Clin Pharmacol Ther*. 2022;111(3):686-696. doi:10.1002/cpt.2511
15. Fagerholm U, Hellberg S, Spjuth O. Advances in Predictions of Oral Bioavailability of Candidate Drugs in Man with New Machine Learning Methodology. *Mol Basel Switz*. 2021;26(9):2572. doi:10.3390/molecules26092572
16. Réda C, Kaufmann E, Delahaye-Duriez A. Machine learning applications in drug development. *Comput Struct Biotechnol J*. 2020;18:241-252. doi:10.1016/j.csbj.2019.12.006
17. Bertolini D, Loukianov AD, Smith A, et al. Forecasting progression of mild cognitive impairment (MCI) and Alzheimer's disease (AD) with digital twins. *Alzheimers Dement*. 2021;17(S9):e054414. doi:10.1002/alz.054414
18. Friedrich S, Antes G, Behr S, et al. Is there a role for statistics in artificial intelligence? *Adv Data Anal Classif*. 2022;16(4):823-846. doi:10.1007/s11634-021-00455-6
19. Smith RC. *Uncertainty Quantification: Theory, Implementation, and Applications*. Society for Industrial and Applied Mathematics; 2013. doi:10.1137/1.9781611973228
20. Sullivan TJ. *Introduction to Uncertainty Quantification*. Springer Cham; 2015.

21. Chen X, Nyberg HB, Donnelly M, et al. Development and comparison of model-integrated evidence approaches for bioequivalence studies with pharmacokinetic end points. *CPT Pharmacomet Syst Pharmacol*. 2024;13(10):1734-1747. doi:10.1002/psp4.13216
22. Nyberg HB, Chen X, Donnelly M, et al. Evaluation of model-integrated evidence approaches for pharmacokinetic bioequivalence studies using model averaging methods. *CPT Pharmacomet Syst Pharmacol*. 2024;13(10):1748-1761. doi:10.1002/psp4.13217
23. Dosne AG, Bergstrand M, Karlsson MO. An automated sampling importance resampling procedure for estimating parameter uncertainty. *J Pharmacokinet Pharmacodyn*. 2017;44(6):509-520. doi:10.1007/s10928-017-9542-0
24. Friede T, Röver C, Wandel S, Neuenschwander B. Meta-analysis of few small studies in orphan diseases. *Res Synth Methods*. 2017;8(1):79-91. doi:10.1002/jrsm.1217
25. Röver C, Friede T. Dynamically borrowing strength from another study through shrinkage estimation. *Stat Methods Med Res*. 2020;29(1):293-308. doi:10.1177/0962280219833079
26. Röver C, Ursino M, Friede T, Zohar S. A straightforward meta-analysis approach for oncology phase I dose-finding studies. *Stat Med*. 2022;41(20):3915-3940. doi:10.1002/sim.9484
27. Günhan BK, Meyvisch P, Friede T. Shrinkage estimation for dose-response modeling in phase II trials with multiple schedules. *Stat Biopharm Res*. 2022;14(2):249-261. doi:10.1080/19466315.2020.1850519
28. Kennedy MC, O'Hagan A. Bayesian Calibration of Computer Models. *J R Stat Soc Ser B Stat Methodol*. 2001;63(3):425-464. doi:10.1111/1467-9868.00294
29. Strong M, Oakley JE. When Is a Model Good Enough? Deriving the Expected Value of Model Improvement via Specifying Internal Model Discrepancies. *SIAMASA J Uncertain Quantif*. 2014;2(1):106-125. doi:10.1137/120889563
30. Owen NE, Challenor P, Menon PP, Bennani S. Comparison of Surrogate-Based Uncertainty Quantification Methods for Computationally Expensive Simulators. *SIAMASA J Uncertain Quantif*. 2017;5(1):403-435. doi:10.1137/15M1046812
31. Sudret B, Marelli S, Wiart J. Surrogate models for uncertainty quantification: An overview. In: *2017 11th European Conference on Antennas and Propagation (EUCAP)*. ; 2017:793-797. doi:10.23919/EuCAP.2017.7928679
32. Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. *Int J Epidemiol*. 2014;43(6):1969-1985. doi:10.1093/ije/dyu149
33. Najjar A, Hamadeh A, Krause S, Schepky A, Edginton A. Global sensitivity analysis of Open Systems Pharmacology Suite physiologically based pharmacokinetic models. *CPT Pharmacomet Syst Pharmacol*. 2024;13(12):2052-2067. doi:10.1002/psp4.13256
34. McNally K, Cotton R, Loizou GD. A Workflow for Global Sensitivity Analysis of PBPK Models. *Front Pharmacol*. 2011;2. doi:10.3389/fphar.2011.00031
35. Langdon G, Gueorguieva I, Aarons L, Karlsson M. Linking preclinical and clinical whole-body physiologically based pharmacokinetic models with prior distributions in NONMEM. *Eur J Clin Pharmacol*. 2007;63(5):485-498. doi:10.1007/s00228-007-0264-x
36. Gisleskog PO, Karlsson MO, Beal SL. Use of prior information to stabilize a population data analysis. *J Pharmacokinet Pharmacodyn*. 2002;29(5-6):473-505. doi:Doi 10.1023/A:1022972420004
37. Food and Drug Administration. M12 Drug Interaction Studies Guidance for industry. *Cent Drug Eval Res Cent Biol Eval Res*. Published online 2024.
38. Posada MM, Morse BL, Turner PK, Kulanthaivel P, Hall SD, Dickinson GL. Predicting Clinical Effects of CYP3A4 Modulators on Abemaciclib and Active Metabolites Exposure Using Physiologically Based Pharmacokinetic Modeling. *J Clin Pharmacol*. 2020;60(7):915-930. doi:10.1002/jcph.1584
39. Posada MM, Cannady EA, Payne CD, et al. Prediction of Transporter-Mediated Drug-Drug Interactions for Baricitinib. *Clin Transl Sci*. 2017;10(6):509-519. doi:10.1111/cts.12486
40. Verhaeghe J, Van Der Donckt J, Ongenae F, Van Hoecke S. Powershap: A Power-full Shapley Feature Selection Method. Published online 2022. doi:10.48550/ARXIV.2206.08394
41. Verhaeghe J, Dhaese SAM, De Corte T, et al. Development and evaluation of uncertainty quantifying machine learning models to predict piperacillin plasma concentrations in critically ill patients. *BMC Med Inform Decis Mak*. 2022;22(1):224. doi:10.1186/s12911-022-01970-y
42. Nyberg J, Bazzoli C, Ogungbenro K, et al. Methods and software tools for design evaluation in

population pharmacokinetics-pharmacodynamics studies. *Br J Clin Pharmacol*. 2015;79(1):6-17. doi:10.1111/bcp.12352
43. Nyberg J, Ueckert S, Strömberg EA, et al. PopED: An extended, parallelized, nonlinear mixed effects models optimal design tool. *Comput Methods Programs Biomed*. 2012;108(2):789-805. doi:10.1016/J.CMPB.2012.05.005
44. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2024. https://www.R-project.org/
45. Hallgren KA. Conducting Simulation Studies in the R Programming Environment. *Tutor Quant Methods Psychol*. 2013;9(2):43-60. doi:10.20982/tqmp.09.2.p043
46. Sigal MJ, Chalmers RP. Play It Again: Teaching Statistics With Monte Carlo Simulation. *J Stat Educ*. 2016;24(3):136-156. doi:10.1080/10691898.2016.1246953
47. Lee S, Sriutaisuk S, Kim H. Using the Tidyverse Package in R for Simulation Studies in SEM. *Struct Equ Model Multidiscip J*. 2020;27(3):468-482. doi:10.1080/10705511.2019.1644515
48. Chalmers RP, Adkins MC. Writing Effective and Reliable Monte Carlo Simulations with the SimDesign Package. *Quant Methods Psychol*. 2020;16(4):248-280. doi:10.20982/tqmp.16.4.p248
49. Nordgren R, Belin S, Chen X, et al. Pharmpy: a versatile open-source software package for pharmacometrics. In: *PAGE. Abstracts of the Annual Meeting of the Population Approach Group in Europe.* Abstr 11039; 2024. https://www.page-meeting.org/default.asp?abstract=11039
50. Fidler M, Wilkins JJ, Hooijmaijers R, et al. Nonlinear Mixed-Effects Model Development and Simulation Using nlmixr and Related R Open-Source Packages. *CPT Pharmacomet Syst Pharmacol*. 2019;8(9):621-633. doi:10.1002/psp4.12445
51. Beal SL, Sheiner LB, Boeckmann AJ, Bauer RJ. NONMEM User's Guides. (1989-2023).
52. Nordgren R, Hooker AC, Belin S, et al. Open source software from the Uppsala University Pharmacometrics group. In: *PAGE. Abstracts of the Annual Meeting of the Population Approach Group in Europe.* Abstr 11037; 2024. https://www.page-meeting.org/default.asp?abstract=11037
53. Keizer RJ, Karlsson MO, Hooker A. Modeling and Simulation Workbench for NONMEM: Tutorial on Pirana, PsN, and Xpose. *CPT Pharmacomet Syst Pharmacol*. 2013;2(6):e50. doi:10.1038/psp.2013.24

PART II. PROPOSED ORGANISATION OF WORK

The **CON**sortium **F**or **I**nnovation in **R**egulatory **M**edical **S**tatistics **(CONFIRMS)** brings together statistical and regulatory expertise from different institutions, sectors and countries as well as from different quantitative methods areas in drug development. This includes expertise on innovative clinical trial designs (including adaptive group-sequential designs and master protocols, optimal designs for dose-finding experiments), research synthesis (including meta-analytic approaches with Bayesian methods, causal inference), pharmacometrics (including modelling & simulation, PK/PD modelling with complex longitudinal models) and regulatory statistics (e.g., assessment of estimands for regulatory decision making; assessment of type I error rate control and multiplicity adjustments). Such methods are of interest for common therapeutic areas as well as rare diseases. All academic partners are not only experienced in the design, conduct, analysis, interpretation and reporting of clinical trials but have also a broad experience in the development and assessment of novel statistical methods as well as planning and conduct of simulation studies. The partners included in the CONFIRMS consortium are

- the Center for Medical Statistics, Informatics, and Intelligent Systems (CeMSIIS) at the Medical University of Vienna (MUW), Austria,
- the Department of Medical Statistics at the University Medical Center Göttingen (UMG), Germany,
- Department of Pharmacy at Uppsala University (UPP), Sweden and
- the Austrian Medicines and Medical Devices Agency (MEA), Austria.

The **Pharmacometrics Research Group at Uppsala University (UU)** is home to the Pharmacometrics research group at the Department of Pharmacy. This group, with ~40 scientists, has research that focuses on drug development and involves methodological aspects in the area of PKPD, longitudinal modelling, optimal and adaptive experimental design, as well as applications in the areas of oncology, infectious diseases, cardiovascular diseases, neurodegenerative and autoimmune disorders. It has research-related interactions with European agencies as well as with the FDA and PMDA. UU will be the lead for this tender if awarded to CONFIRMS.

The **Medical University of Vienna (MUW)** is one of the largest medical schools, and the biggest health institution in Austria. The **Center for Medical Statistics, Informatics, and Intelligent Systems** (CeMSIIS) of the MUW has a long tradition in research on innovative statistical methodology for clinical trials. The research areas include group sequential trials, adaptive design, platform trials, multiple testing, regulatory statistics and many collaborative medical research projects. Members of the CeMSIIS have been involved in regulatory activities both on national and EMA level. MUW provides an excellent research infrastructure, large patient registries, a statistics library, electronic access to all relevant journals, statistical software and simulation servers

The **Department of Medical Statistics, University Medical Center Göttingen (UMG),** collaborates with a range of clinical partners providing statistical support and data management solutions for clinical trials and bioinformatics projects. Furthermore, the Department conducts a program of methodological research with funding from national and international sources. The Department has a research focus on evidence synthesis, Bayesian statistics, adaptive designs, causal inference and rare diseases and, through appropriate external appointments, in regulatory sciences, estimands, multiple comparisons procedures, missing data. The Göttingen team will support and guide the clinical scenario evaluation of different design and analysis options and provide regulatory input with respect to the requirements on the validity metrics and the estimand strategy.

The **Austrian Medicines and Medical Devices Agency** (MEA), a business division of the Austrian Agency for Health and Food Safety **(AGES)** is among the most active national agencies within the European regulatory system with respect to the assessment of applications for European marketing authorizations and EMA scientific advice procedures. With ten statistical experts the methodology group at AGES is among the largest in the European regulatory system. The team has ample experience with the evaluation of state-of-the-art statistical approaches in drug development and has unique access to

the leading edge of international drug development. The AGES maintains a well curated repository of past scientific advice (SA) procedures. The AGES team will provide the regulatory context and support the development and review of study protocols and relevant case studies.

2.1. General approach for the organisation of the study

The work on this project is organised into five work packages as seen in the table below. Each of the work packages is divided into tasks and deliverables to further structure the work.

WP Title	**PM**	**Lead (contributors)**	**Tasks**	**Deliverables**
WP1 Project Management and initial study planning	2.5	UU (All)	Task 1.1 Organisation of consortium meetings, monitoring of project progress Communication with EMA, Submission of deliverables. Task 1.2 Planning of the literature review and key methodological aspects	D1 Preliminary study plan
WP2 Literature Review	3	UU (All)	Task 2.1 Definition of algorithmic search strategy including keywords, and selection of journals Task 2.2 Reviewing of identified manuscripts and information extraction and summarising the key findings Task 2.3 Review EPAR to identify use-cases	D2 Report on literature review
WP3 Study Planning	2	UU(All)	Task 3.1 Selection of use-cases and related models, UQ methods, design characteristics and operating characteristics to be studied.	D3 Simulation study protocol
WP4 Evaluation of UQ methods	3	UU	Task 4.1 preparing models (PBPK, QSP, and/or ML) for selected use-cases	D4 Final Study Report including recommendations on best practice D5 Manuscript
	3	UMG, UU (All)	Task 4.2 Preparing R code to carry out selected UQ methods	
	2	UU (All))	Task 4.3 Establishing simulation study workflow	
	5	UU, AGES (All)	Task 4.4 Performing simulation study	
	2.5	UU (All)	Task 4.5 Preparation of Manuscript	

WP5 Developing a toolkit and tutorial	4	UU (All)	Task 5.1 Developing a toolkit Task 5.2 preparing tutorial	D6 Toolkit and Tutorial

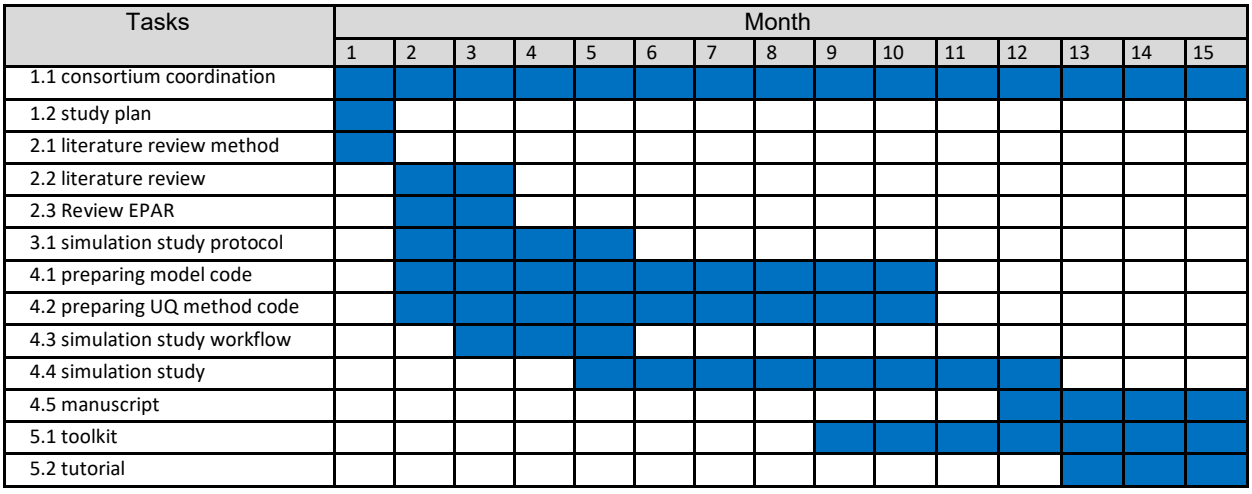
2.2. Roles and responsibilities

Person name (if applicable)	**Organisation**	**Function in the study**	**Description of the function**
Andrew Hooker	*UU*	*Pharmacometrics expert, **Project lead***	*Overall coordination and oversight of the project, responsible for the submission of deliverables, contribution to the various tasks with pharmacometrics expertise, methods, and examples.*
Martin Posch	*MUW*	*Senior Statistician, Regulatory expert, **Consortium lead***	*Overall coordination and oversight of the consortium, support of the literature review, contribution to study plan and protocol, aid in the development of study design and scenarios for simulations studies*
Mats Karlsson, Xiaomei Chen, Other group members	*UU*	*Pharmacometrics experts, lead software development.*	*Contributions to the various tasks with pharmacometrics expertise, methods, and examples. PBPK expertise. Development of software code.*
Franz König	*MUW*	*Senior Statistician, Regulatory Expert*	*Support of the literature review, Contribution to study plan and protocol, aid in the development of study design and scenarios for simulations studies*
Tim Friede *Norbert Benda* *Other group members*	*UMG*	*Senior statisticians and regulatory experts*	*Literature review, support of the development of the study plan and protocol, simulation studies and case studies. Writing of the report.*
Robin Svensson	*SMPA*	*Regulatory expert in pharmacometrics, practical regulatory knowledge in the evaluation of simulations from large complex models*	*Aid in development of the literature review, aid in simulation study development, methods, examples.*
Florian Klinglmüller *Tobias Fellinger*	*AGES*	*Senior statisticians and regulatory experts*	*Review of regulatory documents, development of software code*

Other group members			

2.3. Plan and timelines for deliverables

Month (M)	Due Date	Activity including important milestones and deliverables (D)
M1	2025/06/19	Preliminary study plan (D1)
M3	2025/08/18	Literature review (D2)
M5	2025/10/17	Simulation study Protocol (D3)
M12	2026/05/18	Simulation study report (D4)
M15	2026/08/17	manuscript (D5)
M15	2026/08/17	Toolkit and Tutorial (D6)



2.4. Communication with EMA and third parties

The consortium will regularly communicate with EMA to inform about the progress of the project and align next steps. Steering committee meetings are planned to be held every month and will typically take place virtually (see the table below). For each deliverable at least one draft deliverable will be provided. A point to point answer to comments received by EMA will be provided. If EMA requests an additional meeting, a TC will be organised within 5 working days involving the relevant members of the consortium. Communication with third parties will be achieved through the two stakeholder webinar sessions where CONFIRMS will present and discuss the result of the review and simulation studies with EMA (and leave it to the discretion of EMA which stakeholders should be invited, e.g., assessors from EMA and NCAs). Furthermore, presentations at scientific conferences and workshops will be planned together with EMA.

A working prototype software package will be developed before completion of D3 (Study Protocol) for the purpose of planning and evaluation of feasibility. Software will be hosted at Github and/or CRAN, to provide timely access to the developed software.

Month	**Date**	**Objective**
M0	2025/05/15	Kick off Meeting to discussproject.
M1	2025/06/16	Discuss the preliminary study plan including literature and EPARs review plan.
M2	2025/07/17	Discuss progress of literature and EPARs review.
M3	2025/08/20	Discussion of the literature review results and discussion of the study protocol, planning for next steps
M4	2025/09/18	Discuss progress of current tasks in project
M5	2025/10/16	Presentation and discussion of the Study Protocol
M6	2025/11/13	Discuss progress of current tasks in project
M7	2025/12/11	Discussion of simulation study progress
M8	TBD	Discuss progress of current tasks in project
M9	TBD	Discussion of simulation study progress
M10	TBD	Discuss progress of current tasks in project
M11	TBD	Discuss progress of current tasks in project
M12	TBD	Meeting to present the simulation study report and to discuss the development of the manuscript, toolkit and tutorials
M13	TBD	Discuss progress of current tasks in project
M14	TBD	Discuss progress of current tasks in project
M15	TBD	Meeting to present the manuscript, toolkit and tutorials

PART III. QUALITY CONTROL

3.1. General approach to quality management and control

The consortium will use a common structure and template for deliverables. For quality assurance, a system with internal reviewers within the consortium will be installed (e.g., statistical reviewers from partner institutions not involved in a specific task). In addition, in accordance with EMA, regular meetings with EMA representatives will be held to discuss project progress, interim results and obtain guidance for further work. Relevant regulatory guidance documents (e.g., ICH E9, ICH E9 (R1) addendum and methodological and therapeutic EMA guidance documents) will be considered. Reviewers from the partner organizations will be involved to discuss underlying assumptions. The study plan will be publicly pre-registered in the (EU PAS Register). Similarly, for clinical trial simulations a simulation plan will be pre-agreed. To ensure high-quality for tasks where statistical programming is involved (e.g., for clinical trial simulations or analysis of data) the CONFIRMS consortium will follow the reproducible research principle of the Biometrical Journal, e.g., by using automated reports and providing access to software code on repositories such as R-packages on CRAN or GitHub. To ensure high-quality output a predefined list of checks will be implemented, e.g., comparisons of operating characteristics of standard designs with published results. Critical parts will be validated by another consortium member, e.g., by re-programming or assessment by another partner institution.

3.2. Specific aspects of quality management and control

Systematic literature review

Relevant steps of the systematic review of the literature will be replicated independently. The search strategy for the electronic databases will be checked by a second reviewer to ensure that all relevant terms are included. The search results will be screened by two reviewers independently and disagreements will be resolved by a third reviewer. For the data extraction a pre-specified data extraction form will be used. The extraction will be done by two independent reviewers. Disagreements will be resolved by a third reviewer if necessary.

Software

To ensure that software code will be intuitive to read and debug comprehensive naming and coding conventions will be agreed between involved partners. In addition complete interface specifications and common object and data-type models will be defined at the design stage. Software code will be extensively documented. For all high-level functions manual pages will be written (facilitated by packages roxygen and devtools). Usage of the overall package will be described in a vignette.

In order to produce code that is flexible and extensible a functional programming type of approach - that prioritizes mapping over looping - will be used. Such an approach (e.g. relying on packages provided within the tidyverse) facilitates the development of computationally efficient code that can be easily scaled on the parallel computing infrastructure available within the consortium.

To ensure timely detection and correction of implementation errors, a comprehensive unit testing framework will be implemented (e.g. using r-package testthat). Test cases will be prospectively planned and implemented independently from corresponding software modules. At each development iteration, results from data generating processes and analysis methods will be automatically checked against predefined test cases with known outcomes. In addition outputs from data generating procedures and analysis results will be routinely checked visually and using summary statistics.

Simulation study

Several measures to ensure safe and reliable execution of simulation studies will be

implemented. Errors and warning conditions (e.g. to detect convergence failures) will be implemented and tracked. In addition failsafe conditions (e.g. to terminate execution in case of overly long runtimes) will be implemented. Filename conventions will be specified to avoid accidental overwriting of existing files. To minimize the impact of network errors, power outages or other hardware failures. Intermediate results will be saved to harddisk and functions implemented that permit continuation of computation once the hardware failure has been resolved.

Reproducibility of simulations will be ensured by thorough tracking of seeds, software versions and hardware configurations. Simulation study results will be checked against results from previous related studies previously conducted by members of the Consortium. Results from novel scenarios will be checked for plausibility visually and using summary statistics. Simulation study reports will be reviewed by internal reviewers preferably from a partner in the Consortium not involved in the implementation and execution of the study.

PART IV. ANY OTHER BUSINESS

4.1. Declarations of interest

Tim Friede reports personal fees from Actimed, Apellis, argenx, Aslan, AstraZeneca, Bayer, Biogen, BiosenseWebster, Boehringer Ingelheim, Bristol Myers Squibb, Cardior, CSL Behring, CVRx, Daiichi Sankyo, Enanta, Galapagos, Novartis, Pfizer, PRInnovation, Priothera, RECARDIO, Relaxera, Roche, Upstream Bio, Viatris/ Mylan and VICO Therapeutics for consultancies including data monitoring committees, steering committees and advisory boards. Tim Friede does not hold any shares. The Department of Medical Statistics at UMG did not receive any grants from industry. Norbert Benda does not report any relevant financial interests.

Martin Posch is a registered EMA European Expert with an up-to-date Declaration of Interest. Franz König has participated in DSMBs.

Mats Karlsson and Andrew Hooker report current grants to the UU PMX group from Roche, and GSK. Both are also advisors to, and own shares of, the pharmaceutical consulting company Pharmetheus AB. Mats Karlsson owns shares in the pharmacometrics educational company Wellhagen & Karlsson AB. Andrew Hooker owns shares in the pharmacometrics educational company Hooksson AB. Xioamei Chen, Yevgen Ryexnik and Zhe Huang report no relevant financial interests.

Florian Klinglmueller, Elham Yousefi, Lynette Caitlin Mikula and Tobias Fellinger are all registered EMA European Experts with up-to-date Declarations of Interest. They do not report any relevant financial interests.

4.2 Funding

This work is funded by the EMA Project “Uncertainty quantification for complex models supporting regulatory decision making” - Re-opening of competition EMA/2020/46/TDA/L3.02-ROC22.

PART V. SUPPLEMENTARY DOCUMENTS

5.1. List of Supplementary documents

5.2. Subprotocol A: Literature review

Objective

The main objective of the literature review is to identify available uncertainty quantification (UQ) methods and their applicability to quantify uncertainties in complex models such as physiologically based pharmacokinetics (PBPK), quantitative system pharmacology (QSP), and machine learning (ML) in the context of regulatory decision making in drug development. Based on the review, the performance of the identified methods will be assessed and compared under a wide range of scenarios.

Searches

A comprehensive literature search that will be conducted in relevant electronic databases. This could include the following databases: ***PubMed***, ***Web of Science***, ***Scopus, MathSciNet, MEDLINE***, ***EMBASE***, ***Google scholar***. AI-powered search tool(s) will be investigated to improve/update search results obtained from the electronic databases, this could include, but is not limited to:

- Elicit (https://elicit.com)
- Scopus AI (https://elsevier.com/products/scopus/scopus-ai)
- Undermine (https://undermind.ai)

An adaptive literature search may be applied based on the volume and relevance of results at each stage. If an initial search results in too many publications (e.g., 100,000), the strategy will be adapted by making the search more specific by adjusting keywords, applying stricter filters, or focusing on certain publication types. Another option can be using the large language models to process the search results. If, on the contrary, the number of papers found is too small (e.g., 10), the search will be extended by relaxing criteria, adding synonyms, or including additional databases. Also, the search scope can be expanded by adding extra types of complex models such as quantitative system toxicology (QST) and physiologically based biopharmaceutics models (PBBM). The search adjustment steps will be decided based on the discussion within the research team.

To explicitly compare AI-assisted and manual search methods, both approaches can be piloted: a small sample of papers will be identified using each method, and their relevance to the research question will be assessed. The AI-assisted search is expected to expedite the scanning and analysis phases, while manual searching may offer deeper contextual understanding and nuanced selection. The pilot study will inform which method—or combination—yields the most relevant and high-quality literature for the review.

The search terms are broadly divided into two categories:

1. Uncertainty Quantification Methods
2. Complex Modeling Approaches

For each category, search strategies will be developed to comprehensively capture all known and relevant methods and models (see Table 1). The final set of records will consist of those that meet criteria from both categories.

The search on UQ methods for machine learning models will be done separately from that for PBPK and QSP models. To reduce the number of publications that are focused on UQ for ML models, ***ML model-***

related search terms will be combined with specific terms related to modeling in drug development (e.g., “interval prediction” AND “Machine learning” AND “drug development”; “uncertainty” AND “deep learning” AND “pharmacometrics”, etc.).

The main UQ methods to be searched will be related to

- Frequentist approaches;
- Bayesian approaches;
- Sensitivity analysis;
- Surrogate models;
- Stochastic spectral methods;
- Interval analysis and fuzzy (soft) computations.

Additionally, other UQ methods can be included such as

- Uncertainty propagation;
- Model discrepancy;
- Distributional uncertainties;
- Filtering.

Table 1. Details of the literature search terms. Note that the final set of search terms may change based on the total number of identified articles and to make sure that articles known to be relevant to the current investigation are included.

#	Search terms	
1	“uncertainty” OR “uncertainty quantification” OR “UQ” OR “uq” OR “quantifying uncertainty” OR “uncertainty evaluation” OR “evaluating uncertainty” OR “parameter uncertainty” OR “model uncertainty” OR “structural uncertainty”	***uncertainty***-related terms
2	“Physiologically based pharmacokinetics” OR “PBPK ”quantitative system pharmacology” OR “QSP” OR “machine learning” OR “ML” OR “modeling” OR “quantitative system toxicology” OR “QST” or “physiologically based biopharmaceutics modeling” OR “PBBM”	***model***-related terms
3	“machine learning” AND “neural network" AND “deep learning” AND “supervised learning” AND “unsupervised learning” AND “reinforcement learning” AND “classification” AND “regression” AND “clustering” AND “random forest” AND “XGBOOST” AND “transformer	***ML models***-related terms

	model" AND "large language model" AND "explainable AI" AND "pattern recognition" AND "active learning" AND "anomaly detection" AND "generative adversarial networks" AND "scientific machine learning" AND "SciML" AND "neural ODE" AND "universal ODE"	
4	**Frequentist approach**: "frequentist uncertainty quantification" OR "frequentist confidence intervals" OR "frequentist prediction intervals" OR "maximum likelihood estimation uncertainty" OR "frequentist hypothesis testing uncertainty" OR "frequentist error propagation" OR "confidence interval estimation" OR "frequentist parameter estimation" OR "bootstrap uncertainty quantification" OR "profile likelihood uncertainty" OR "Fisher information uncertainty" OR "frequentist coverage probability" OR "frequentist prior"	***UQ methods***-related terms
5	**Bayesian approach**: "Bayesian hierarchical model" OR "Bayesian uncertainty quantification" OR "Bayesian inference for uncertainty quantification" OR "Bayesian methods in uncertainty quantification" OR "Bayesian inverse uncertainty quantification" OR "Bayesian updating for uncertainty" OR "Markov chain Monte Carlo for Bayesian uncertainty" OR "Variational inference Bayesian uncertainty" OR "Bayesian deep learning uncertainty quantification" OR "Bayesian hierarchical modeling for uncertainty" OR "Gaussian process Bayesian uncertainty" OR "Bayesian posterior predictive uncertainty" OR "Bayesian model evidence (Bayes factors) for uncertainty" OR "Modular Bayesian approach uncertainty quantification" OR "Bayesian calibration"	
6	**Sensitivity analysis (local and global)**: "local sensitivity analysis uncertainty quantification" OR "global sensitivity analysis uncertainty quantification" OR "variance-based sensitivity analysis" OR "Morris method global sensitivity" OR "input-output sensitivity analysis" OR "non-intrusive sensitivity analysis" OR "model output sensitivity to input uncertainty"	
7	**Surrogate models**: "surrogate models uncertainty quantification" OR "surrogate modeling for uncertainty quantification" OR "surrogate-based uncertainty quantification" OR "surrogate model uncertainty	

	propagation" OR "Kriging surrogate models uncertainty" OR "polynomial chaos surrogate models" OR "ensemble surrogate models uncertainty" OR "surrogate model error quantification"	
8	**Stochastic spectral methods**: "stochastic spectral methods uncertainty quantification" OR "polynomial chaos expansion uncertainty quantification" OR "PCE uncertainty quantification" OR "generalized polynomial chaos" OR "gPC" OR "stochastic Galerkin method" OR "stochastic collocation method" OR "Karhunen–Loève expansion" OR "non-intrusive spectral projection" OR "spectral methods for uncertainty propagation" OR "regression-based polynomial chaos"	
9	**Interval analysis**: "interval analysis uncertainty quantification" OR "interval uncertainty quantification" OR "interval methods uncertainty quantification" OR "interval arithmetic uncertainty analysis" OR "non-probabilistic uncertainty quantification" OR "interval propagation methods" OR "set-theoretical uncertainty quantification" OR "interval-based uncertainty modeling" OR "interval contractors"	
10	**Fuzzy computing**: "fuzzy uncertainty quantification" OR "fuzzy computing uncertainty quantification" OR "fuzzy logic uncertainty quantification" OR "fuzzy analysis uncertainty quantification" OR "fuzzy sets for uncertainty modeling" OR "fuzzy interval analysis" OR "fuzzy uncertainty propagation" OR "fuzzy rule-based systems uncertainty" OR "fuzzy decision making under uncertainty" OR "fuzzy sensitivity analysis" OR "fuzzy–interval uncertainty analysis" OR "fuzzy modeling for imprecise data" OR "fuzzy numerical methods uncertainty"	
11	**Uncertainty propagation**: "uncertainty propagation methods" OR "forward uncertainty propagation" OR "probabilistic uncertainty propagation" OR "Monte Carlo uncertainty propagation" OR "simulation-based uncertainty propagation"	
12	**Model discrepancy**: "model discrepancy uncertainty quantification" OR "predictive uncertainty model	

	discrepancy" OR "model error quantification" OR "discrepancy function uncertainty" OR "prediction intervals model discrepancy" OR "model misspecification uncertainty" OR "identifiability model discrepancy" OR "model updating with discrepancy" OR "maximum mean discrepancy for uncertainty" OR "calibration and model discrepancy" OR "discrepancy analysis"	
13	**Distributional uncertainty**: "distributional uncertainty quantification" OR "distribution shift uncertainty quantification" OR "distributional robustness uncertainty" OR "robust uncertainty quantification" OR "model uncertainty under data distribution changes" OR "distributional robustness methods" OR "stable distribution propagation uncertainty" OR "calibration under distributional uncertainty" OR "ensemble methods distributional uncertainty"	
14	**Filtering**: "filtering uncertainty quantification" OR "Bayesian filtering uncertainty quantification" OR "Kalman filter uncertainty quantification" OR "Ensemble Kalman filter uncertainty" OR "particle filter uncertainty quantification" OR "sequential Monte Carlo filtering uncertainty" OR "optimal filtering for uncertainty quantification" OR "posterior uncertainty filtering" OR "filtering for parameter estimation uncertainty" OR "filtering and uncertainty propagation"	

Following the database search, the resulting records will be exported and duplicants merged using systematic review software (e.g., Rayyan). The set of articles will then be screened for relevance based on predefined inclusion and exclusion criteria. Screening will be conducted independently by two reviewers. Titles and abstracts will be assessed, and any records that clearly do not meet the inclusion criteria will be excluded at this stage.

Type of studies to be included

Methodological research articles and investigations where uncertainty quantification methods are presented and/or applied. ***Due to time constraints and relevance of the results, the search will be restricted to a set of pre-specified journals***:

- journals, focusing on modeling and its applications in drug discovery and development (3D);

- some search systems allow restricting search by selecting a journal category (***3D-related categories*** will be set).

The following inclusion and exclusion criteria will be used:

Inclusion criteria

- Time frame (e.g., from [YEAR] to present) may be used as a filter to support adaptive search strategies.
- Research and methodological publications (peer-reviewed articles, conference proceedings, reviews, reports), discussing UQ methods for complex models.
- Publications, presenting case studies, where UQ methods for complex models have been applied (clinical trials, simulation study, modeling).

Exclusion criteria

- Investigations/publications that are not relevant to the main objective:
    - Not focusing on UQ for modeling problems considered.

Condition or domain being studied

UQ methods for PBPK, QSP, ML models.

Participants/population

not applicable

Intervention(s), exposure(s)

not applicable

Comparator(s)/control

not applicable

Main outcome(s)

- The review focuses on existing UQ methods for complex models (with a focus on PBPK, QSP, and ML).
- We will investigate which methods are available in the literature and applied in such a context.
- We briefly review the available software mentioned in the publications.

Measures of effect

not applicable

Additional outcome(s)

not applicable

Data extraction (selection and coding)

For all relevant records after the screening step, abstracts and full texts will be obtained and reviewed according to the pre-specified inclusion criteria.

Search results will be organized in a table that may contain the following information:

- ID
- Title
- Authors
- Source
- Year
- PUI/DOI/FullText link
- Type of paper
- Inclusion? (yes/no)
- If no inclusion, reason
- Other reason
- Simulation? (yes/no)
- Software? (yes/no)
- Type of model (PBPK/QSP/ML)
- UQ method
- Case study
- Comments

Risk of bias (quality) assessment

not applicable

Strategy for data synthesis

not applicable

Analysis of subgroups or subsets

not applicable

References

1) Sullivan TJ (2015) “Introduction to Uncertainty Quantification.” Springer. doi: 0.1007/978-3-319-23395-6

2) Smith RC (2024) “Uncertainty Quantification. Theory, Implementation, and Applications.” 2nd ed. *SIAM*. doi:10.1137/1.9781611977844

5.3. Subprotocol B: EMA EPAR and SA review

Title: Review of regulatory procedures where uncertainty quantification of mechanistic models were of concern – scoping review of EMA Public Assessment Reports and Scientific Advice procedures

Protocol Version 1.0

Date: 2025-06-18

Abstract

Objective: Identify EMA marketing authorization procedures and scientific advice final letters where uncertainty quantification of mechanistic models were of concern in order to identify methods used in a drug development context, relevant parameter ranges for a simulation study, derive case studies, and regulatory recommendations.

Introduction: Modelling and simulation approaches play a crucial role in drug development, supporting decisions such as dose selection, extrapolation between populations, and predicting clinical outcomes. However, assessment of these models and interpretation of their results depends on the clear quantification and communication of uncertainty. This scoping review aims to systematically identify and map existing methods proposed, suggested, or discussed in regulatory procedures for quantifying uncertainties in mechanistic modelling within medicines development.

Inclusion criteria: We will include marketing authorization procedures granted and EMA Scientific Advice Final Letters issued before June 1st 2025. Procedures will include initial authorizations as well as variations e.g. extending the use to other therapeutic areas. We will include procedures for which the European Public Assessment Report (EPAR) and EMA Scientific Advice Final Letter discusses uncertainty quantification in relation to mechanistic models used to address scientific questions in the context of a drug development programme. The scope of the review will be limited to assessments of mechanistic models (including PBPK and QSP) excluding procedures discussing model uncertainty in relation to e.g. statistical models for estimation of dose response from clinical studies and population PK analyses. Procedures – especially scientific advice letters - where uncertainty quantification are noted by the Applicant only, but not addressed in the regulatory assessment may be excluded.

Methods: EPARs available in the database at paediatricdata.eu and Scientific Advice Final Letters (EMA-SA FAL) available in AGES' in-house regulatory search system will be searched for paragraphs matching a predefined list of keywords related to the uncertainty quantification (UQ) for mechanistic models. Results will be screened to exclude matches unrelated to UQ for mechanistic models. For selected matches full-text of EPARs and EMA-SA FALs will be obtained and information items related to UQ for mechanistic models, the methods used to quantify uncertainty, model specifications and parameter estimates relevant to inform parameters for simulation scenarios may be extracted. The final extraction process will be determined in a two-step process, starting with a pilot review of five procedures to obtain an overview of the detail of model and method description, the depth of discussion and identify common items suitable for systematic extraction.

Introduction

This protocol concerns the review of regulatory procedures: EMA centralised marketing authorization procedures and EMA scientific advice procedures. The main objective of this review is to identify EMA marketing authorization procedures where uncertainty quantification of mechanistic models were of concern. Based on the review, we will define relevant parameter ranges for distributional scenarios in the simulation study, to derive case studies in order to illustrate the findings of the simulation study and derive regulatory recommendations.

The review focuses on marketing procedures where uncertainty quantification of mechanistic models was of concern. Thus, we will investigate which methods were used or proposed to evaluate the uncertainty related to estimates and conclusions from mechanistic models intended to inform scientific questions relevant to the development of medicinal products. In addition, we intend to extract aggregate data on specific models, estimates of model parameter estimates and related uncertainty measures in order to define relevant parameter ranges for a simulation study and derive case studies for illustration.

Review question

The main question of this review is to identify documents, where UQ methods were used to address uncertainty challenges in mechanistic models like PBPK and QSP models. For selected documents we will investigate:

- what UQ methods were used;

- what were the key findings: prediction uncertainties, parameters uncertainties, etc.

In addition, we intend to extract aggregate data on specific models, estimates of model parameter estimates and related uncertainty measures in order to define relevant parameter ranges for a simulation study and derive case studies for illustration.

Eligibility criteria

Inclusion criteria

1. Regulatory procedures with a positive opinion (EPAR) or final advice letter (EMA-SA) (issued before June 1st 2025) where uncertainty quantification of mechanistic models was identified as an issue during regulatory assessment.
2. Initial authorization, variation e.g. extending the authorised use to another therapeutic area, or Scientific Advice Final Letter.
3. Procedures for which the EPAR or EMA-SA FAL discusses at least one method for uncertainty quantification of a mechanistic model intended to inform regulatory conclusions on aspects of the drug development programme.
4. Procedures that contain a regulatory discussion of the UQ method.

Exclusion criteria

1. Marketing authorization procedures currently under review.
2. EMA-SA procedures currently under assessment (i.e. EMASA FAL not issued before June 1st 2025).
3. Marketing authorization procedures withdrawn by the Applicant or with a negative opinion.
4. Procedures that do not discuss at least one mechanistic model.
5. Procedures that do not discuss UQ.
6. Procedures where the discussion of UQ does not refer to a mechanistic model (e.g. confidence intervals for parameters estimated in a dose-finding clinical trial or population pharmacokinetic analyses).
7. Procedures where UQ is mentioned by the Applicant but not discussed in the regulatory assessment.

Methods

Search strategy

A comprehensive search of EMA Public Assessment Reports (EPARs) will be conducted in electronic databases (e.g. paediatricdata.eu). In addition, the AGES Database of EMA Scientific Advice Final Letters will be searched to identify related Scientific Advice procedures. Search terms will include terms such as

"uncertainty quantification" in different types of spelling and related prespecified search terms. The searches will be restricted to procedures with a positive opinion and EMA SA were the FAL has been issued by CHMP.

Search Strategy: The following structured search terms will be used:

Category	Terms Included
Modelling-related Terms	"in silico", "in-silico", "modelling", "PBPK", "pharmacometric", "simulation", "dose-response", "extrapolation", "quantitative systems pharmacology", "QSP"
Uncertainty-related Terms	"uncertainty quantification", "uncertainty evaluation", "model uncertainty", "parameter uncertainty", "structural uncertainty", "evaluation of uncertainty", "confidence interval", "credible interval", "sensitivity analysis", "probabilistic sensitivity analysis", "prediction interval"

Searches will combine both categories using the AND operator to increase the likelihood that potential matches meet inclusion criterion 3 and exclusion criteria 4 and 5.

An initial limited search of paediatricdata.eu and AGES Database was undertaken to evaluate the feasibility of a search strategy based on full-text queries (see below). The final search strategy, including all search terms and conditions for selection may be adapted following the pilot review (see below) and in response to the systematic review of the modelling literature. Only rules applicable to items provided by the respective databases will be included (e.g. excluding criteria for index terms, title, keywords).

In case an insufficient number of relevant documents can be identified using the above search strategy, terms to identify modelling-related and uncertainty-related discussions may be expanded to cover broader modelling or model evaluation contexts like goodness of fit or model validation.

Procedure identified in the EPAR search may be followed up by a secondary search in the EMA-SA FAL database to identify corresponding EMA-SA procedures and investigate to what extent UQ concerns noted in the EPAR were discussed in related scientific advice.

Search results for individual search terms will be exported as delimited text files, including for each result at least identifiers for procedure, active substance, matching paragraph text, and source file. Results will be collated for further processing. Duplicates will be removed and EPARs and Final Advice letters will be filtered to match predefined combination rules of search terms.

The databases to be searched include:

- paediatricdata.eu – Full-Text search of EMA EPARs: https://paediatricdata.eu/shiny/users/ralfherold/emaepars/ (accessed 2025-06-13)
- AGES internal database of EMA Scientific Advice letters

Study/Source of Evidence selection

Matching paragraph text from EPARs will be screened for assessment against the inclusion criteria for the review. Screening will be performed using a screening form developed by the reviewers. For this purpose, a small number of search results will be screened in a pilot test by AGES reviewers to evaluate the usability of the form. Subsequently, screening will be performed by two or more independent reviewers. Potential disagreements between reviewers results will be resolved based on discussion between involved reviewers and potentially a third additional reviewer.

For potentially relevant procedures the EPAR will be retrieved in full. The full text of selected procedures will be assessed in detail against the inclusion criteria by two or more independent reviewers. Reasons for exclusion of procedures will be recorded and reported in the review. Any disagreements that arise between the reviewers at each stage of the selection process will be resolved through discussion, or with an additional reviewer. The results of the search and the study inclusion process will be reported in full

in the final review and presented in a Preferred Reporting Items for Systematic Reviews and Meta-analyses extension for scoping review (PRISMA-ScR) flow diagram (Trico et al. 2018).

In case the number of initial matches exceeds 100, screening by single reviewers only will be considered. In this case random selection of procedures for which a second reviewer will be asked to confirm screening results, as well as, an option to request peer review in case of unclear eligibility will be implemented.

Data Extraction

Data will be extracted from EPARs and Final Advice Letters included in the review by two or more independent reviewers using a data extraction form developed by the reviewers. The data extracted will include specific details about the procedure, product, indication, modelling context, mechanistic model, as well as key aggregate model parameter and uncertainty measure estimates.

In the first step only data from 5 EMA-SA FALs and 3 EPARs will be reviewed with the aim to obtain information on items including:

- Drug development question addressed
- Type of model and specification
- Uncertainty quantification methods proposed or applied
- Recommendations provided by the EMA regarding uncertainty quantification
- Estimates of model parameters
- Estimates of uncertainty measures

Depending on the level of detail, breadth of discussions encountered in the regulatory documents the final set of items to extract for the review will be finalized and a standardized extraction form and review instructions will be developed.

Data extraction will be performed by two or more independent experts from AGES and MPA which have appropriate access credentials. Any disagreements that arise between the reviewers will be resolved through discussion, or with an additional reviewer.

In case the number of included procedures exceeds 30, data extraction by single reviewers only will be considered.

Data Analysis and Presentation

In general, we will use descriptive statistics to summarize data extracted from EPARs and Scientific Advice Final Letters. We will provide an overview of the number of procedures, therapeutic areas, and specific indications involved. We intend to qualitatively and quantitatively summarize the types of modelling approaches utilized, the specific drug development questions addressed (e.g., dose selection, extrapolation), and the methods proposed, discussed, or recommended by EMA for quantifying uncertainty. We will offer a summary of relevant outcome measures, including model parameters, confidence or credible intervals, sensitivity analyses, and probabilistic approaches to uncertainty, where applicable. Findings will be summarized according to the type of uncertainty discussed, therapeutic area, and product class.

Acknowledgements

This protocol is based on the JBI scoping review template (accessed 2025-06-18).

Funding

This work is funded by the EMA Project "Uncertainty quantification for complex models supporting regulatory decision making" - Re-opening of competition EMA/2020/46/TDA/L3.02-ROC22.

Conflicts of interest

See Section 4.1. Declarations of interest.

References

Tricco AC, Lillie E, Zarin W, O'Brien KK, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med. 2018 Oct 2;169(7):467-473. doi: 10.7326/M18-0850. Epub 2018 Sep 4. PMID: 30178033.