EMA
EUROPEAN MEDICINES AGENCY

A regulatory perspective on performance verification of mechanistic models - application to PBPK-based DDI predictions

An EMA qualification team perspective

Pieter J. Colin
Seconded National Expert
Scientific Advice Office
Scientific Evidence Generation Department
EMA

Table of contents

- The Simcyp DDI qualification Matrix
- Predictive performance based on DDI QM
- Simulation study to reproduce results from DDI QM
- Model-based uncertainty quantification
- Reporting uncertainty for upcoming applications in line with the CoU

The Simcyp DDI qualification matrix

University of Washington Drug Interaction Database

- Manually curated collection of human in vitro and clinical (in vivo) data available (>17,000 records).
- Acquired by Certara in 2023

Simcyp DDI qualification matrix (DDI QM)

- Mixture of DDI studies: CI/MBI, different CYPs, weak/moderate/strong inhibitors, sensitive substrates (f_m <0.8) and less sensitive substrates
- Datasets in DDI QM were not used for finetuning drug-specific parameters (compound files)

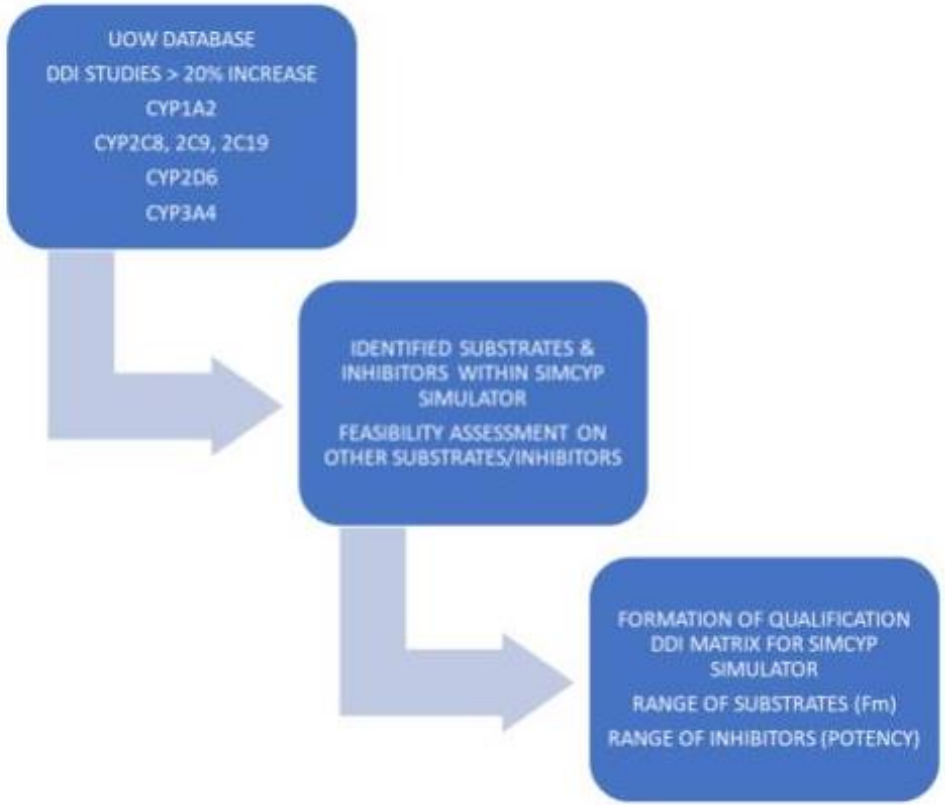
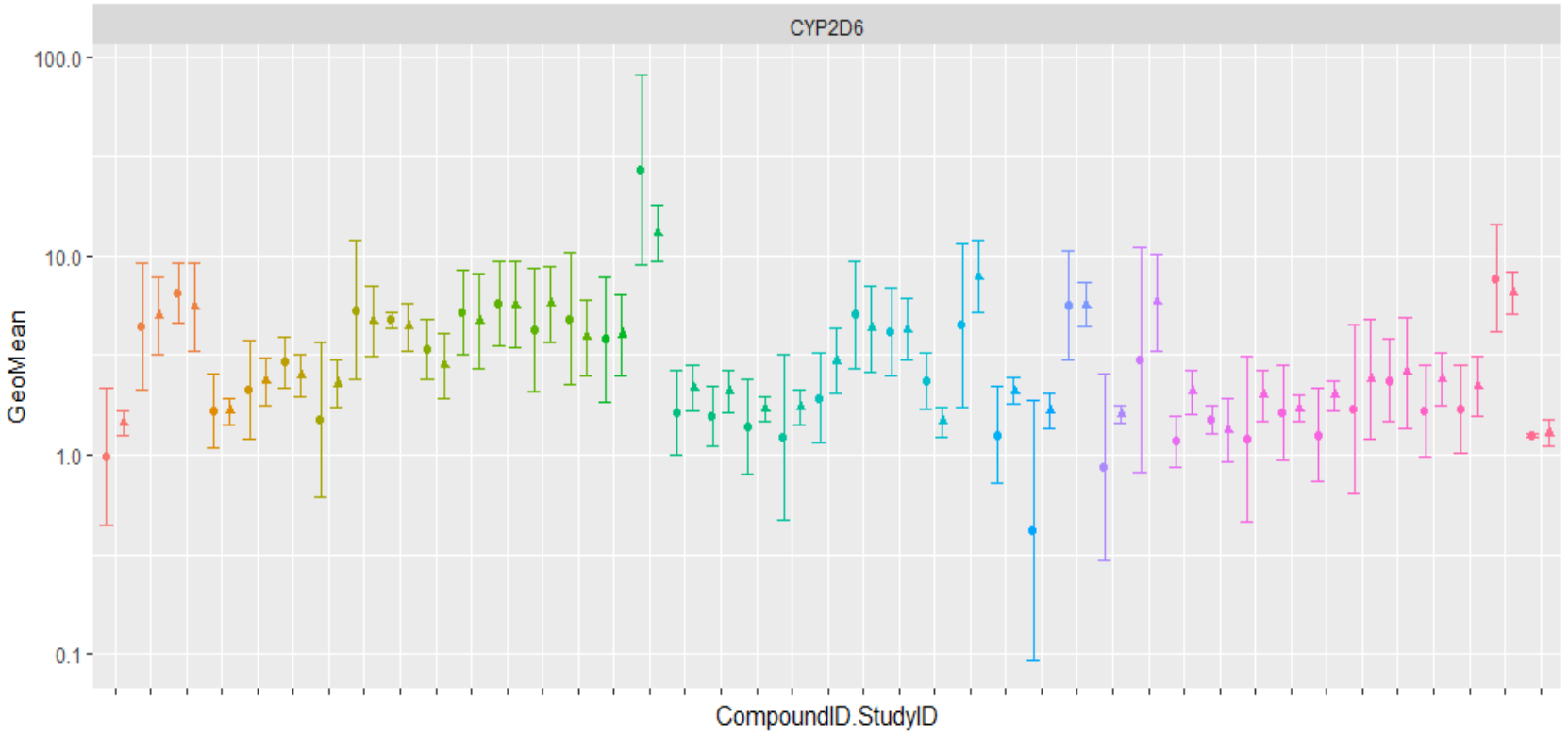


Figure 1. The workflow used to derive the DDI Qualification Matrix.

The Simcyp DDI qualification matrix



Observed data:

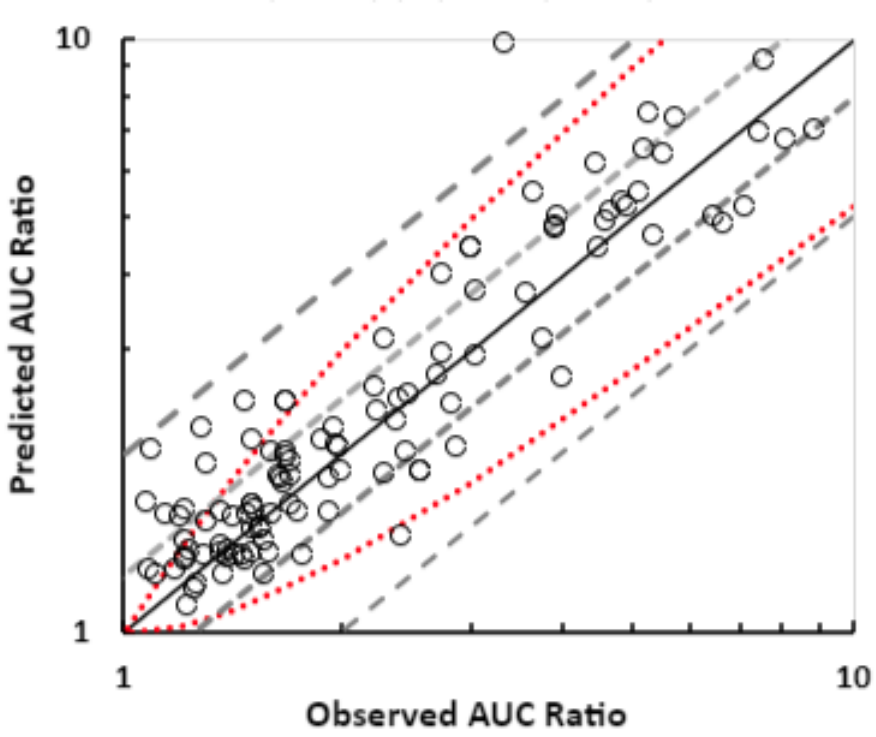
- Geometric mean AUC/C_{max} ratio
- Measure of uncertainty: SD, CI, CV%
- Nsub [range: 2, 28]

Predictions:

- GMR
- BSV in AUC/C_{max} ratio

Predictive performance based on DDI qualification matrix

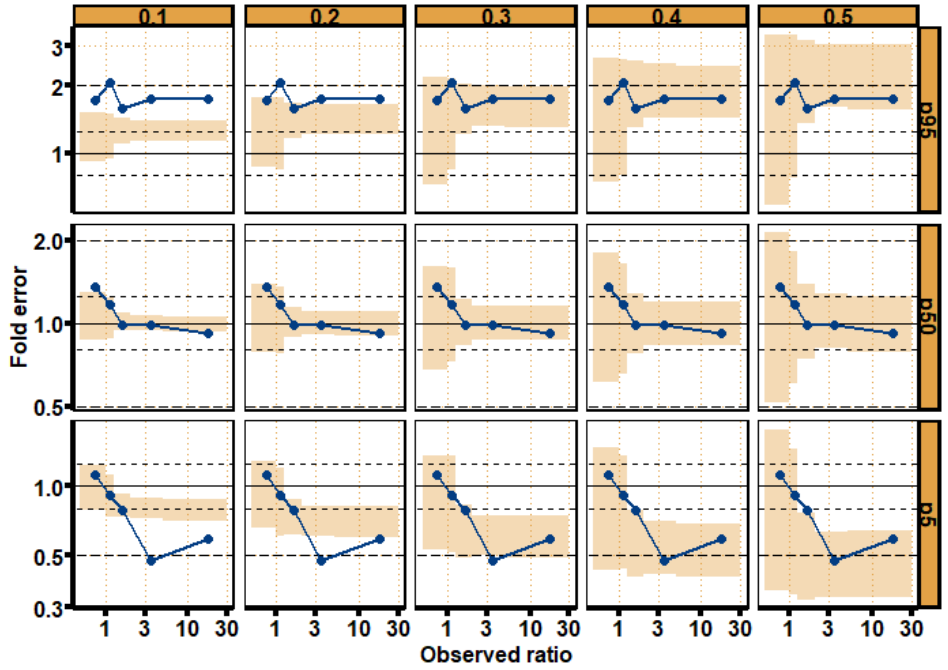
- Performance metrics (for MBI):
  - AFE, AAFE
  - %Within [0.80 - 1.25], %Within Guest criterion, %Within [0.50 – 2.00]
  - % Missclassified



- ❑ No universal acceptance criterion - CoU & QoI define required precision for the predictions
- ❑ Heterogeneity in quantity (number of subjects) and quality (precision of reported point estimates) of the information across CYPs, inhibition mechanisms is not considered
- ❑ Pass/fail type of assessment doesn’t allow thorough understanding of the properties of the underlying system

Simulation study to reproduce results from DDI qualification matrix

- VPC-style graphical evaluation accompanied with NPC to explore different hypotheses about the underlying system (e.g. increasing bias and/or increasing prediction variability)
- Simulate DDI qualification matrix while introducing bias and/or variability in predictions across studies/object-precipitant pairs and contrast with observed results from DDI QM



		Observed DDI				
		Induction	No inhibition	Weak inhibition	Moderate inhibition	Strong inhibition
Predicted DDI	Induction	0 2 [1; 5]	0 2 [1; 4]	0 2 [1; 4]	0 1 [1; 2]	
	No inhibition	3 1 [1; 3]	2 1 [1; 4]	4 2 [1; 6]	0 1 [1; 3]	
	Weak inhibition	0 1 [1; 2]	1 2 [1; 5]	10 8 [3; 13]	7 5 [2; 10]	0 1 [1; 2]
	Moderate inhibition	0 1 [1; 1]	1 1 [1; 2]	3 4 [1; 8]	21 22 [14; 30]	4 2 [1; 6]
	Strong inhibition			0 1 [1; 1]	3 2 [1; 6]	24 27 [19; 35]

- ❑ Results from the DDI QM (MBI) were in line with a scenario assuming >30 % prediction variability and ±15 % bias for moderate inhibition towards weak inhibition

Simulation study to reproduce results from DDI qualification matrix

- VPC-style graphical evaluation accompanied with NPC to explore different hypotheses about the underlying system (e.g. increasing bias and/or increasing prediction variability)
- Simulate DDI qualification matrix while introducing bias and/or variability in predictions across studies/object-precipitant pairs and contrast with observed results from DDI QM

  - Limitations:
    - System properties were evaluated one-by-one
    - For practical reasons different sources of variability and bias were lumped in a single parameter
    - Other (untested) simulation scenarios could have described the data better
    - Not explicitly accounting for uncertainty of points estimates in DDI QM

Model-based uncertainty quantification

Bayesian hierarchical meta-analysis*

- $R_{Obs} = f(R_{Pred}, X, \{\Theta, \Omega, \Sigma\})$
- Joint estimation of multiple sources of variability (CYP, Study, ID)
- More flexibility for investigating factors contributing to variability in DDI QM – performance of the system
  - ❑ size of clinical trial, design characteristics, uncertainty in reported point estimates, …
  - ❑ linear or higher-order relationships between predictive performance and R_{Pred}

➢ Credible intervals/regions derived from the posterior distribution could be used to infer the properties of the system

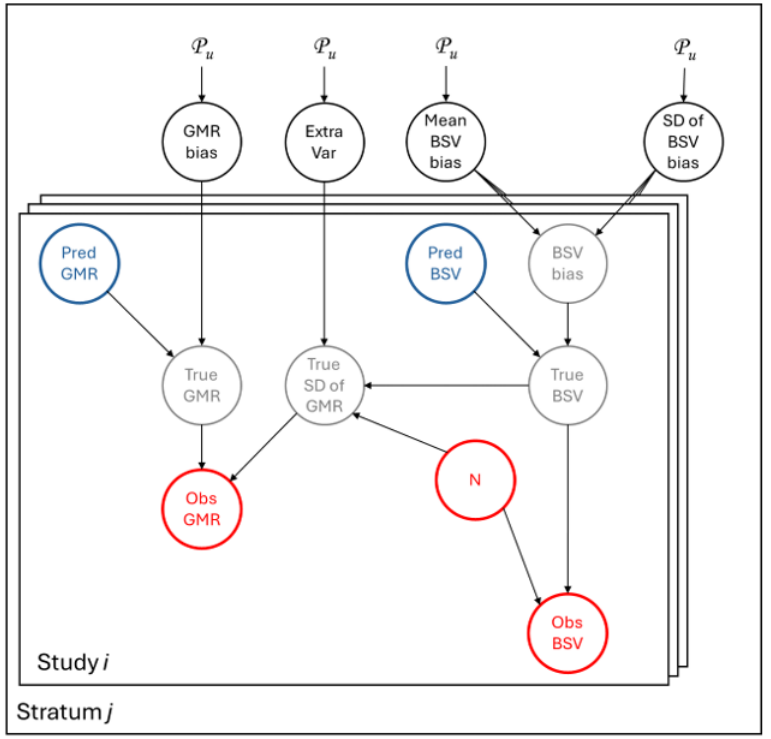
➢ Possibility to propagate what we have learned about the system, incl. the associated uncertainty, in meaningful diagnostics for assessors

* Often referenced in the context of: “Bayesian calibration or bias correction”, “Probabilistic sensitivity analysis”, “Model discrepancy analysis”, etc.

Model-based uncertainty quantification

Bayesian hierarchical meta-analysis

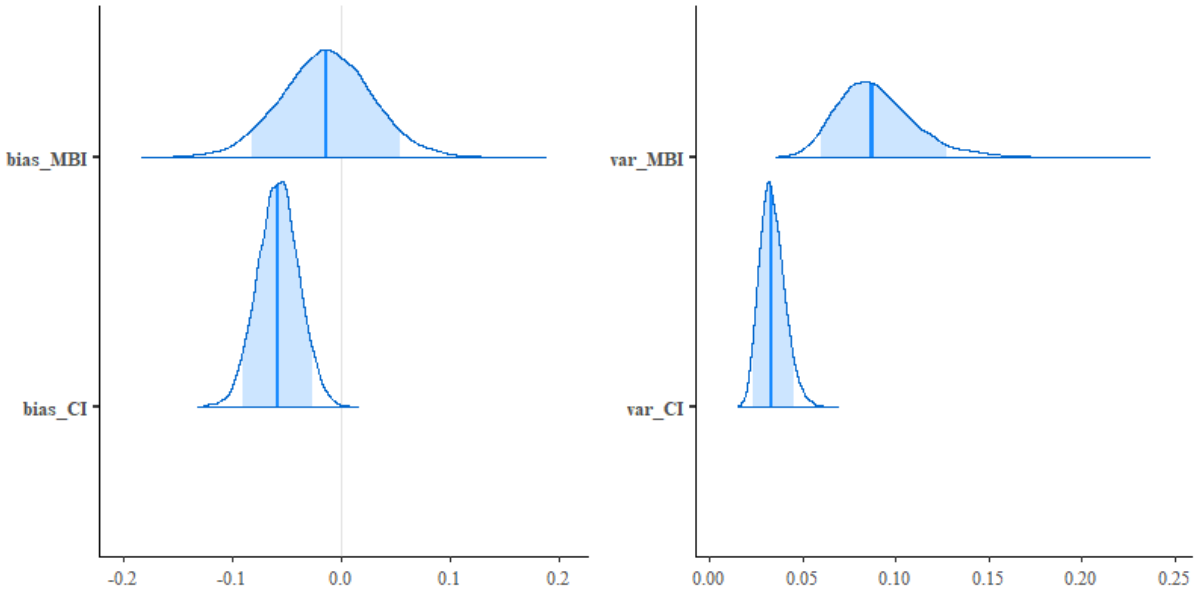
- Simcyp predicted **GMR** and **BSV** were bridged to the **observed GMR** and **observed BSV** by **parameters** estimated from the data (acknowledging that both sampling variability and imprecision* in the predictions may cause the observed GMR to differ from the true GMR)
- GMR_{bias} and imprecision were estimated as fixed effects, BSV_{bias} was estimated as a random effect
- Covariate screening and different stratification and/or hierarchical modelling strategies were explored to understand variability in bias and imprecision across the DDI QM
- In line with CoU, focus was on GMR_{bias} and/or imprecision
- Leave-one-out cross validation & posterior predictive checks drove model selection
- Sensitivity analyses were conducted to understand the influence of model complexity on the results



Model-based uncertainty quantification

Conclusions:

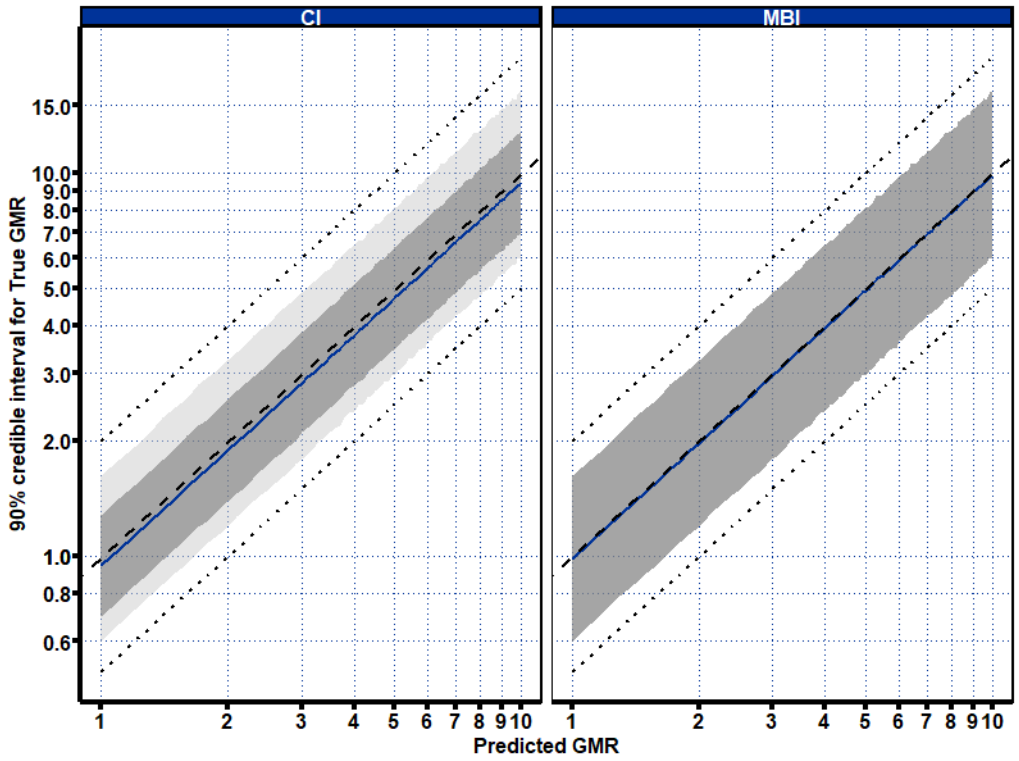
- ✓ Data does not support splitting up GMR_{bias} and imprecision per CYP
- ✓ GMR_{bias} and imprecision depend on type of inhibition (MBI vs. CI)
- ✓ BSV is under-predicted (out-of-scope for this qualification)



Reporting uncertainty for upcoming applications in line with the CoU

Possibility to propagate what we have learned about the system, incl. the associated uncertainty, in meaningful diagnostics for assessors

Qualification opinion proposes different visualizations to contextualize the uncertainty associated with GMR_{AUC} and GMR_{Cmax} predictions (incl. R-code).



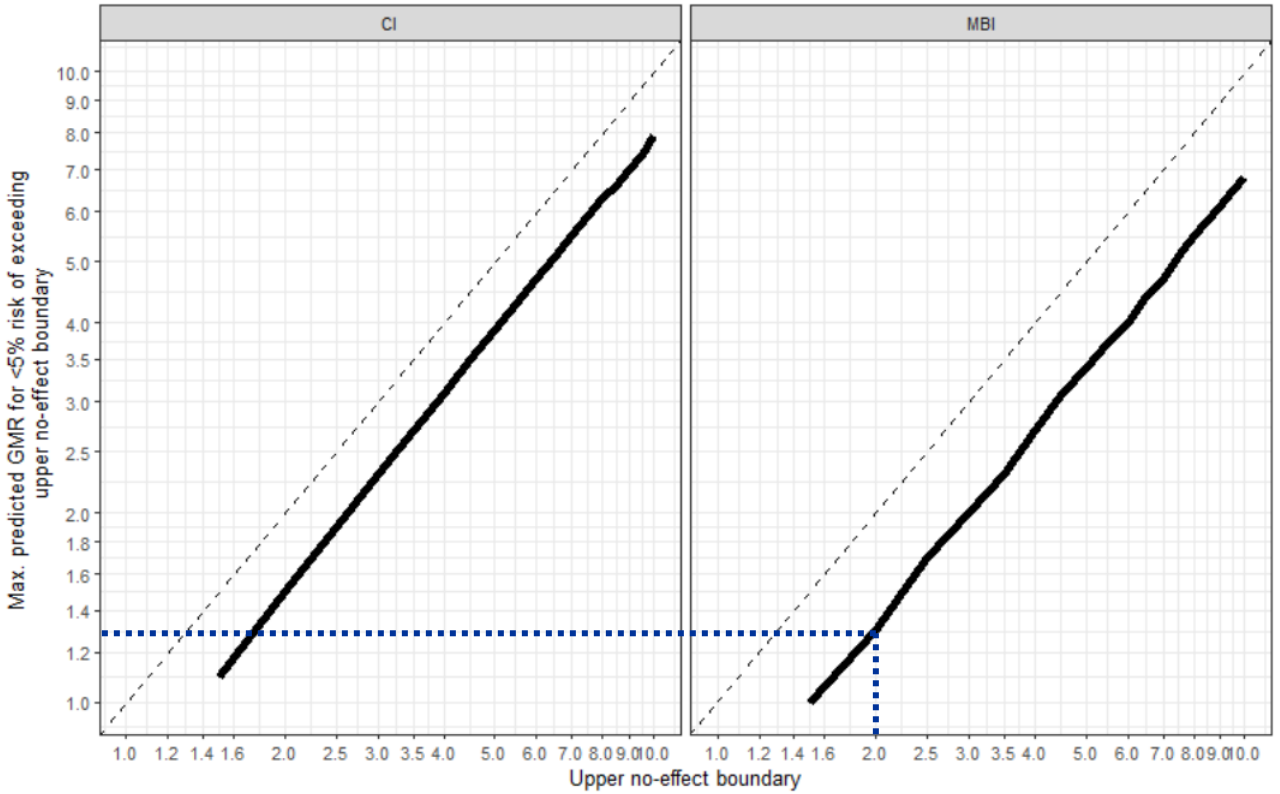
<u>Q1: How accurate are GMR predictions?</u>

GMR_{Pred} = 2 ~ 90% Cred.Int. [1.4; 2.5] for CI

Reporting uncertainty for upcoming applications in line with the CoU

Q2: Assuming a k-fold upper no-effect boundary, what predicted GMR can we tolerate?

Maximum predicted GMR_{AUC} (y-axis) for <5 % risk of exceeding the k-fold relative upper no-effect boundary (x-axis)



<u>Caveat:</u> Risk should not be interpreted as the likelihood of individual DDI PK metrics falling outside the no-effect boundaries (beyond the scope of the qualification)

Conclusions

- **Conventional** performance metrics (AFE, AAFE, etc.) were insufficient for assessing the predictive performance and to contextualize the uncertainty in the predictions for future applications

- **Simulation-based diagnostics** (VPCs, NPCs) were useful to assess the sensitivity for detecting deficiencies in the predictive performance of the system with the validation data at hand

- A **model-based assessment of the discrepancies** between observed and predicted data allowed
  (i) a more granular assessment of the predictive performance (e.g. by testing competing hypotheses around bias/imprecision) and
  (ii) to propagate our findings into diagnostic plots that are useful for assessors when evaluating new model applications (within the CoU)

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

Thank you

Pieter.colin@ema.europa.eu

Follow us