

UPPSALA
UNIVERSITET

CONFIRMS CONSORTIUM · DELIVERABLE 4

Quantifying uncertainty in complex models

A simulation study evaluating uncertainty-quantification methods for PBPK models in regulatory decision-making

Niels Hendrickx[1] · Xiaomei Chen[1] · Gustaf Wellhagen[1] · Robin Svensson[2] · Andrew Hooker[1]

1. Department of Pharmacy, Uppsala University · 2. Swedish Medical Products Agency



UPPSALA
UNIVERSITET

THE CHALLENGE

PBPK models inform regulatory decisions, yet their predictive uncertainty is rarely quantified

Models can substitute for clinical data

PBPK models are used to support waivers of drug–drug-interaction (DDI) studies and to inform dose selection in special populations, including children.

Dependence on many uncertain inputs

Each model combines hundreds of physiological and drug parameters drawn from prior literature, every one carrying its own uncertainty.

Uncertainty is rarely quantified

Deterministic one-at-a-time sensitivity analysis provides no probabilistic interpretation and neglects parameter correlations.

UPPSALA UNIVERSITET

EMA (2018). Guideline on the reporting of physiologically based pharmacokinetic (PBPK) modelling and simulation. EMA/CHMP/458101/2016.

THE CORE IDEA

Two complementary types of uncertainty

PART 1 — Bayesian Hierarchical Meta Analysis (BHMA)

Output-level uncertainty

How well does the platform predict reality across many past studies?

- Compares predicted vs observed DDI outcomes over 219 historical studies
- Estimates the platform's bias and imprecision
- Asks: can this platform be qualified for regulatory use?

Observed prediction performance → model

PART 2 — Probabilistic Sensitivity Analysis (PSA)

Input-level uncertainty

How does uncertainty in a model's parameters flow to its predictions?

- Starts from uncertainty in drug-specific parameters
- Propagates it through the model to the output
- Asks: how reliable is this one model's prediction?

Model parameters → prediction

UPPSALA UNIVERSITET

EMA (2025). Qualification opinion for Simcyp Simulator.
Lash TL et al. (2014). Good practices for quantitative bias analysis. Int J Epidemiol 43(6):1969–85.

OBJECTIVES

Research objectives

01

Evaluate BHMA

for qualifying PBPK platforms

- What data are needed?
- How does it behave operationally?
- Is it robust to misspecification?
- Are there simpler alternatives?

02

Implement PSA

for input-parameter uncertainty

- Build a usable PK-Sim workflow
- Test how the prior distribution shapes the result

03

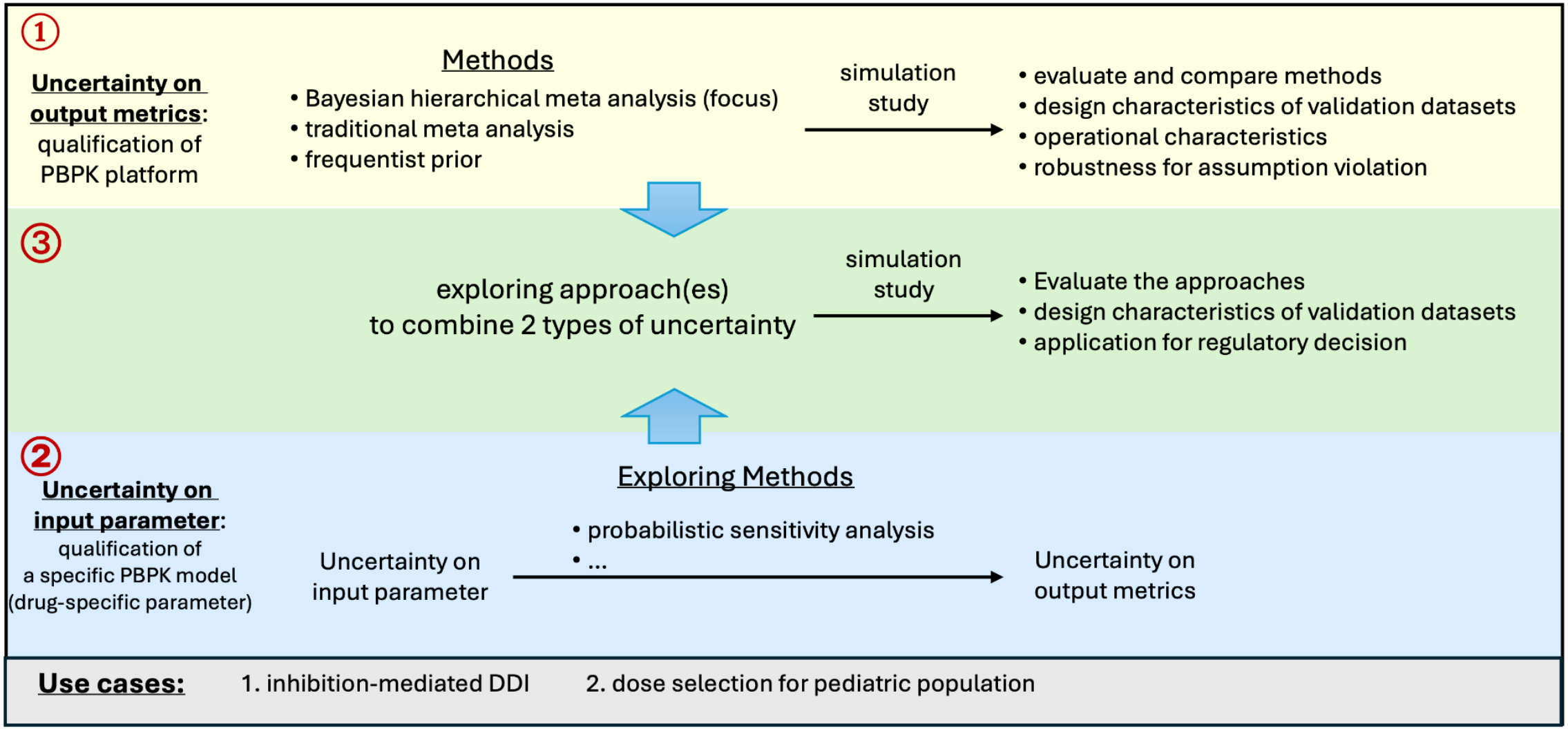
Combine the two

to inform regulatory decisions

- Clarify what each uncertainty means
- Explore how to merge them quantitatively

UPPSALA UNIVERSITET

Overview of simulation study plan



PART 1

Can we trust the platform?

Evaluating Bayesian Hierarchical Meta-Analysis (BHMA) for qualifying PBPK platforms



UPPSALA
UNIVERSITET

Overview of the BHMA model

Approach

For 219 past DDI studies, BHMA compares the platform's predicted exposure ratio (GMR) against the observed one.

What it estimates

- Bias — does the platform systematically over- or under-predict?
- Between-study variance — how much do predictions scatter around the truth?

Why "hierarchical"

Each study informs, and is informed by, a shared platform-level distribution — pooling strength across studies.

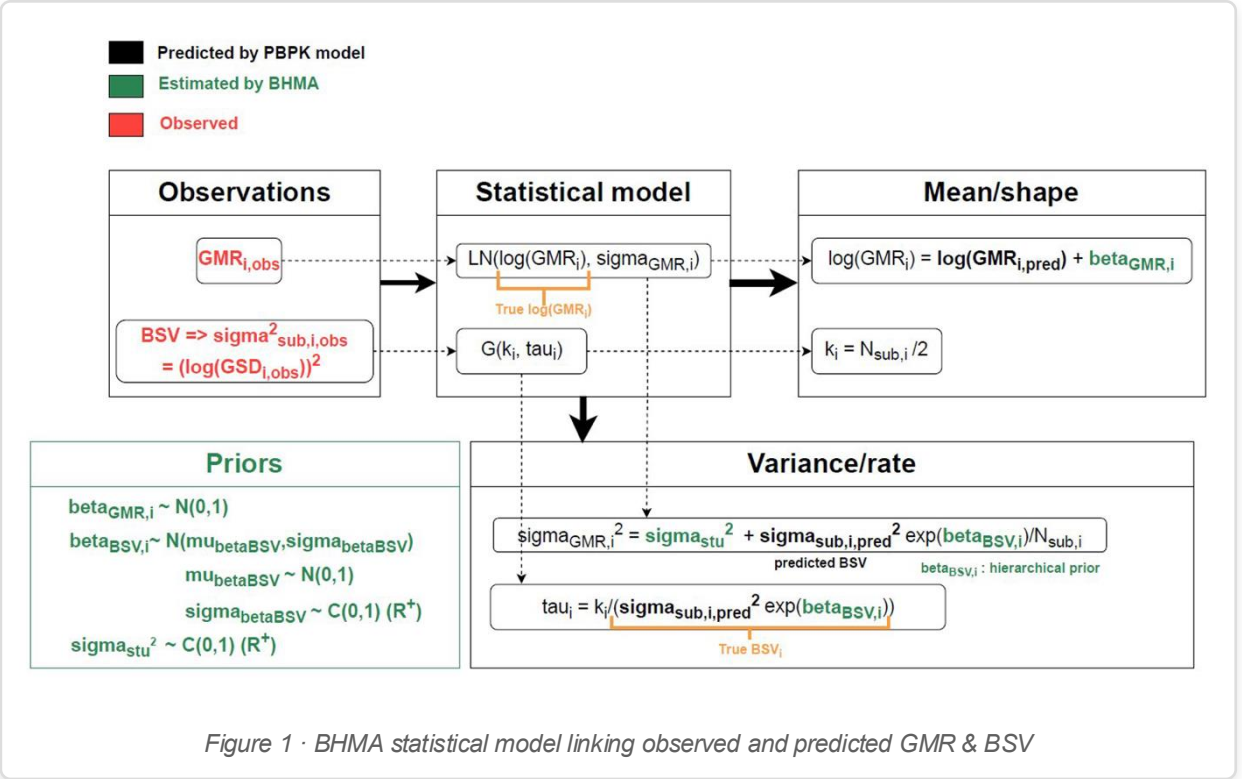


Figure 1 · BHMA statistical model linking observed and predicted GMR & BSV

UPPSALA UNIVERSITET

EMA (2025). Qualification opinion for Simcyp Simulator (EMADOC-1700519818-2006369).

Simulation study design: 80 scenarios, 1000 replicates each

1

Define a scenario

Vary studies/set (4–200), subjects/study (5–30) and platform bias & imprecision → 80 combinations

2

Simulate 1000 datasets

For each scenario, generate 1000 virtual validation sets from the BHMA model

3

Fit BHMA

Re-estimate the model on every replicate (Stan; 4 chains)

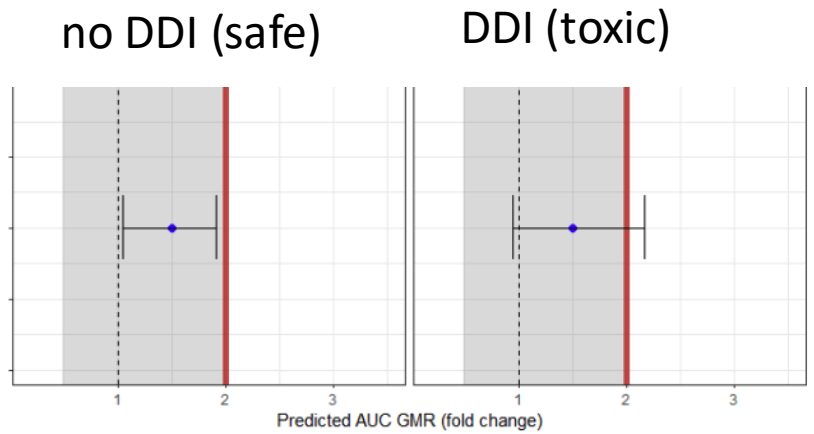
4

One-case prediction

Predict an extra DDI and build its predictive credible interval (PCI)

Evaluated by: parameter bias & RRMSE · convergence (R-hat, ESS) · posterior-predictive p-values · type I error & power for declaring no DDI

Simulation workflow (one simulation case)



true BHMA model

true GMR distribution

Simulate validation datasets
- observed and predicted GMRs
- observed and predicted BSVs

estimating BHMA model

Model results

set true GMR

Simulate a dataset for a hypothetic drug
- predicted GMR

credible interval of the GMR of the hypothetic drug, mapping to a safety margin

At least ~20 DDI studies are required for reliable estimation

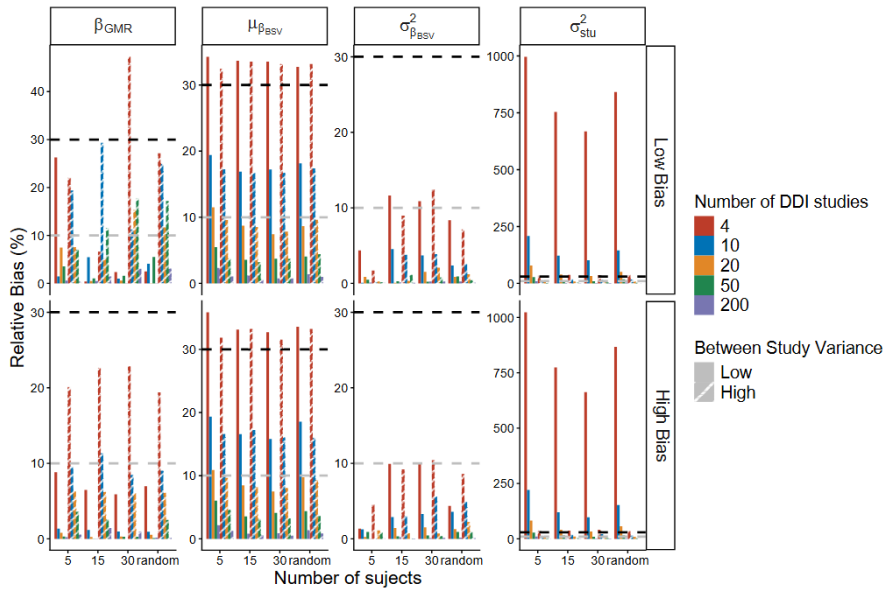


Figure 2 · Relative bias of all BHMA parameters

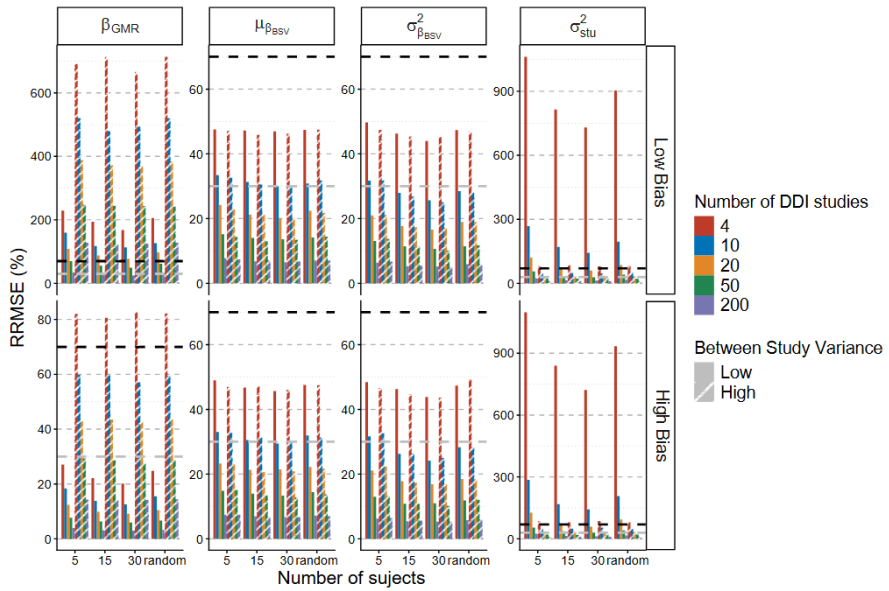


Figure 3 · Relative RMSE of all BHMA parameters

Below 20 studies, bias and RRMSE increase substantially; **at ≥20 studies** both decrease to acceptable levels. The number of subjects per study had limited influence.

UPPSALA UNIVERSITET

The model is well-calibrated to the observed data

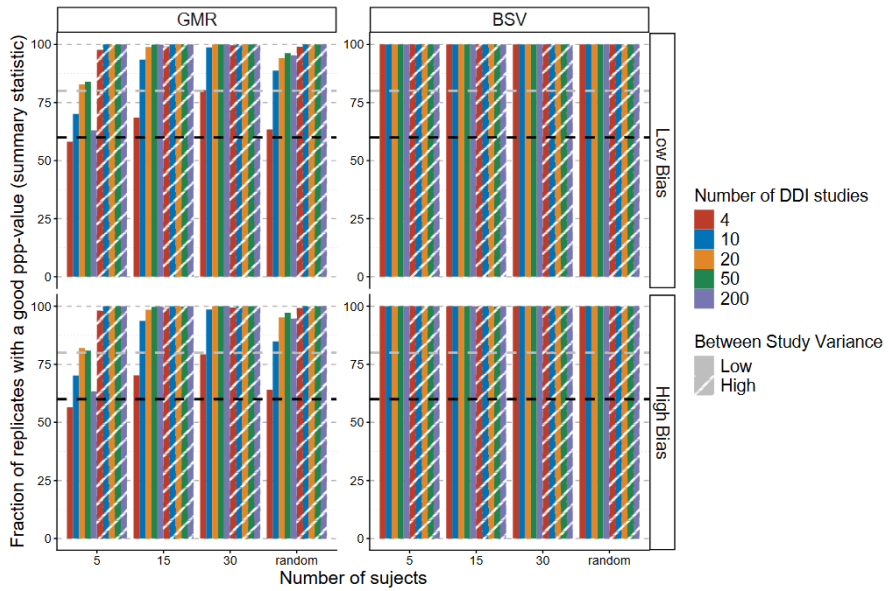


Figure 4 · "Summary-statistic" posterior-predictive p-values

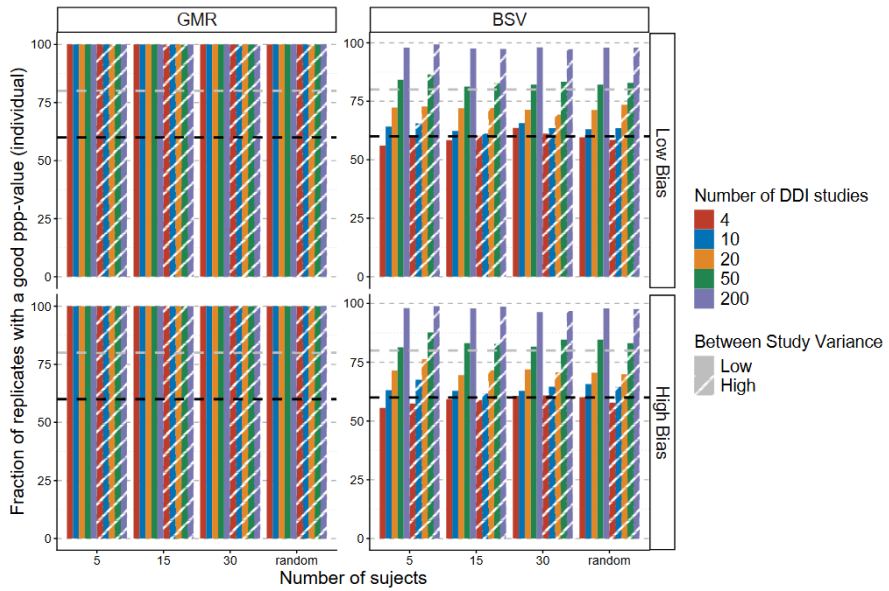


Figure 5 · "Individual" posterior-predictive p-values

Both posterior-predictive checks approach their 0.5 target in most replicates. **Deviations occur only with fewer studies** — consistent with a small-sample effect rather than misspecification.

UPPSALA UNIVERSITET

PART 1 · RESULT

Platform imprecision, rather than bias, is the primary determinant of power

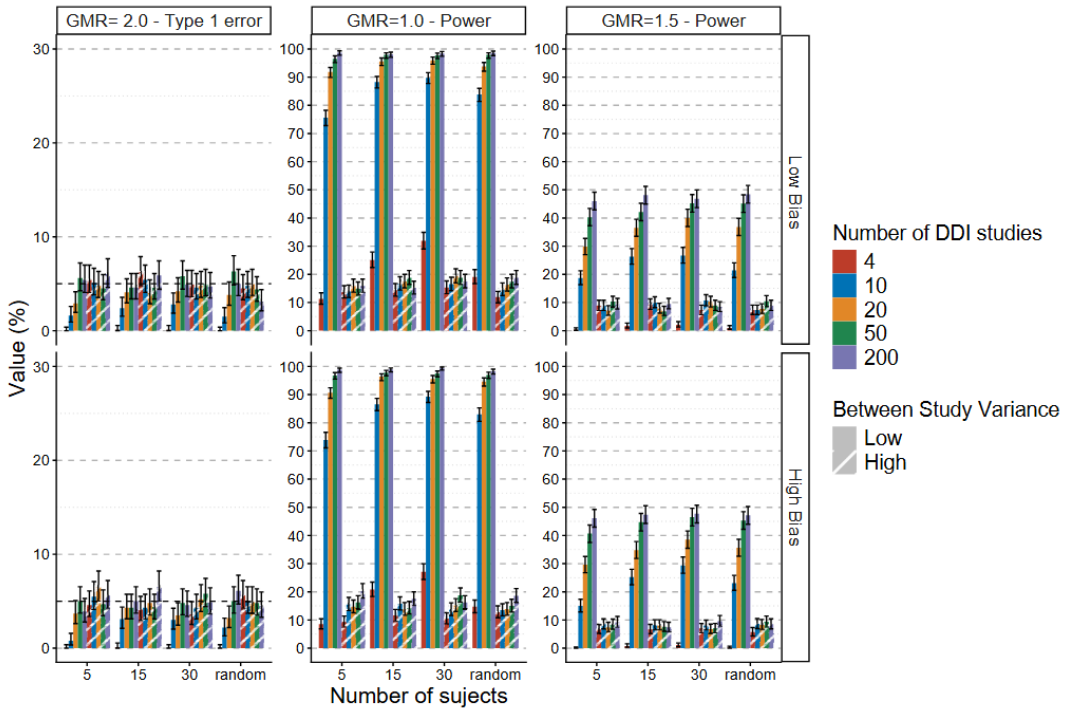


Figure 6 · Type I error & power to declare a DDI (90% PCIs)

Type I error controlled

Maintained at the nominal level across scenarios, irrespective of platform bias.

High variance reduces power

Large between-study variance reduces power across all scenarios.

Power increases with study count

Power increases with the number of studies, plateauing around 20.

UPPSALA UNIVERSITET

BHMA is robust to most misspecification; informative priors are the main exception

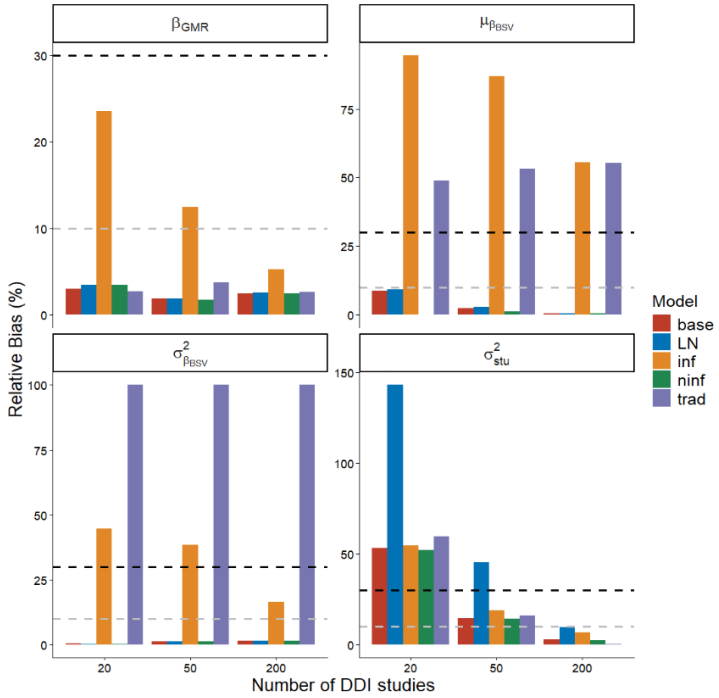


Figure 7 · Relative bias across misspecified models

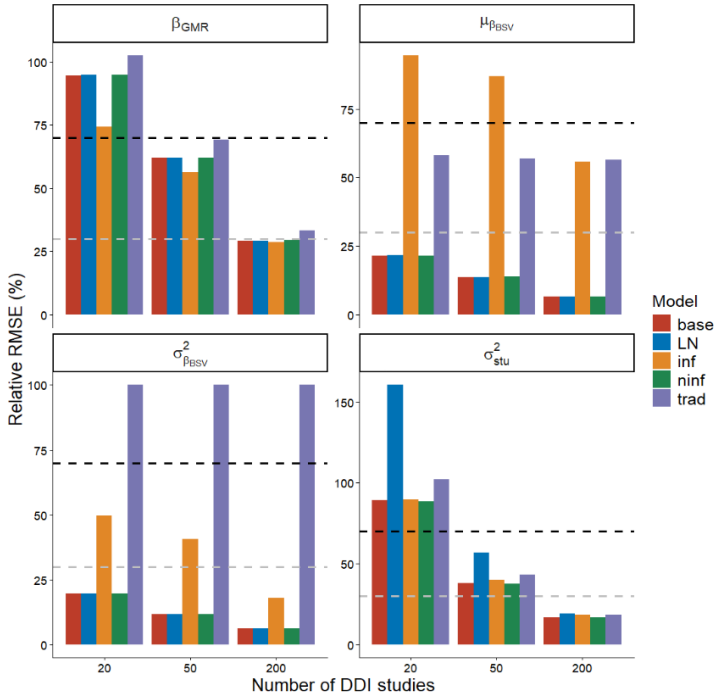


Figure 8 · Relative RMSE across misspecified models

Models tested

ninf – Non-informative prior
behave like the base model

trad – Traditional MA
biases the BSV mean

LN – Log Normal prior
high bias at ≤50 studies

inf – Informative prior
large bias on most parameters

UPPSALA UNIVERSITET

ELPD identifies misspecified models, while type I error remains controlled

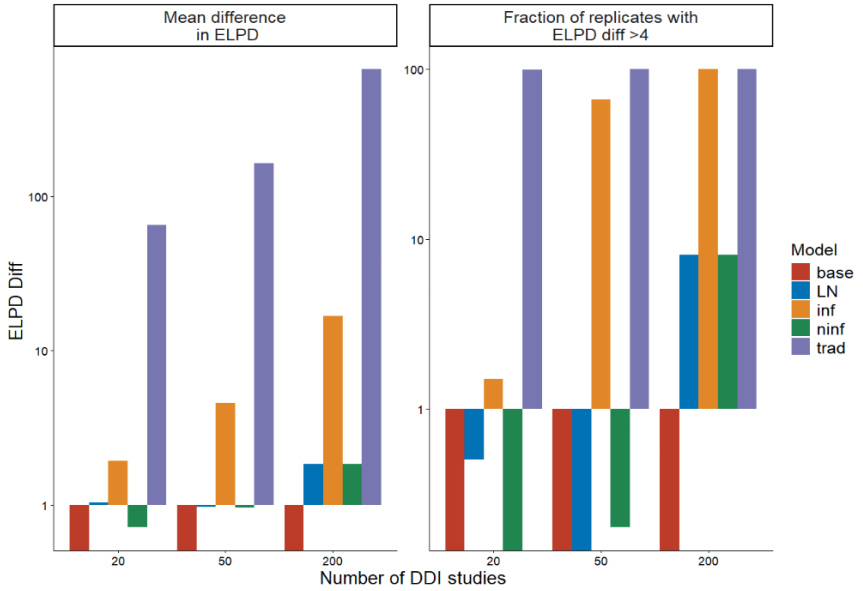


Figure 9 · Difference in ELPD vs the base model

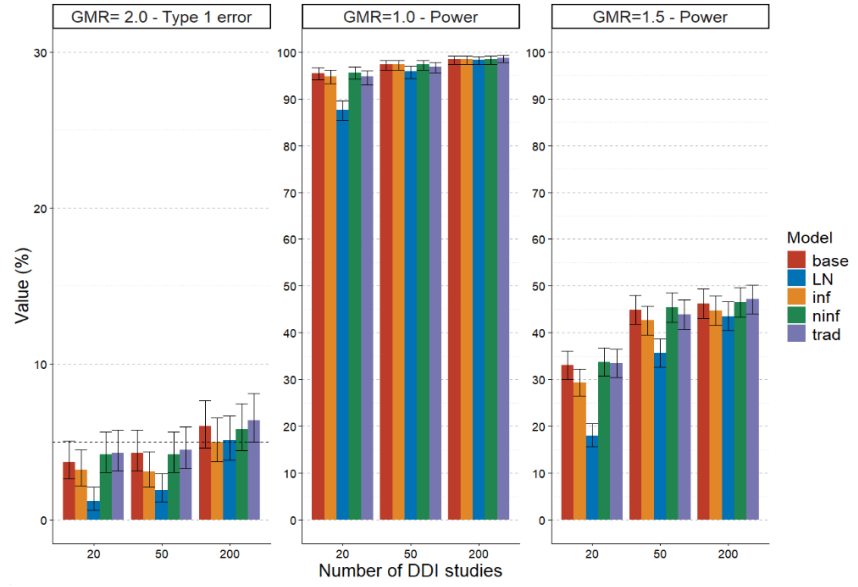


Figure 10 · Type I error & power across misspecified models

ELPD clearly distinguishes the traditional and informative-prior models from the base model. The lognormal model appears acceptable by ELPD yet yields overly conservative intervals, mostly due to over-estimation of between-study variance with a low number of DDI studies

UPPSALA UNIVERSITET

PART 1 · RESULT

Exposure matching for pediatric dosing may require more validation data*

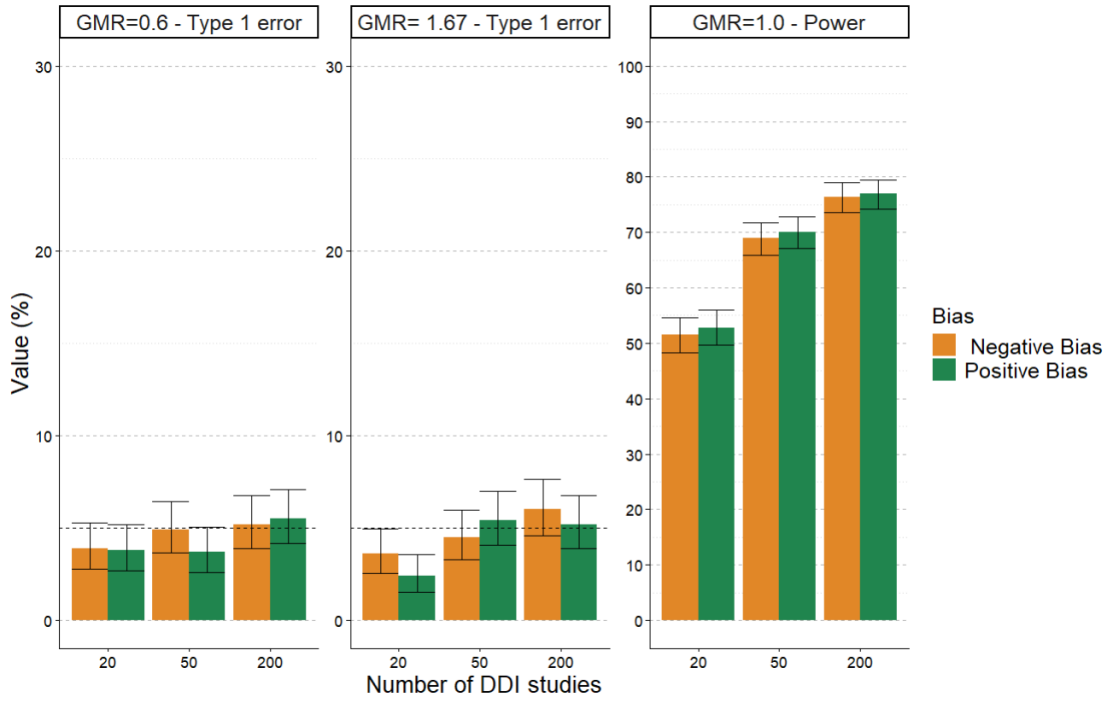


Figure 11 · Type I error & power to achieve exposure matching (0.6–1.67 limits)

The pediatric scenario

Two one-sided tests against tighter 0.6–1.67 exposure limits.

Findings

Power reaches only ~50% at 20 studies and ~78% at 200 studies.

Implication

Where pediatric datasets are limited, PBPK modelling alone may carry insufficient certainty for dose selection*.

* *This exposure-matching exercise is exploratory and of limited applicability in the paediatric setting. The adoption of 0.6–1.67 geometric mean ratio bounds and TOST represents a methodological choice rather than broadly accepted criteria.*
Moreover, adequate exposure matching should address not only central tendency but also variability across populations, which is not fully captured here.
The datasets and BHMA model were developed in a DDI context and, given the limited paediatric data in the public domain and additional uncertainty related to developmental factors, may not adequately reflect the complexities of pediatric extrapolation.

UPPSALA UNIVERSITET

A frequentist approach (NONMEM + SIR) shows comparable performance

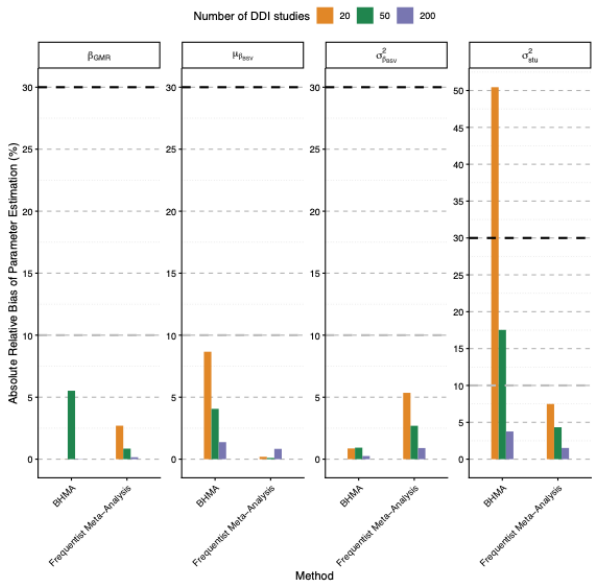


Figure 12 · Absolute relative Bias: NONMEM vs BHMA

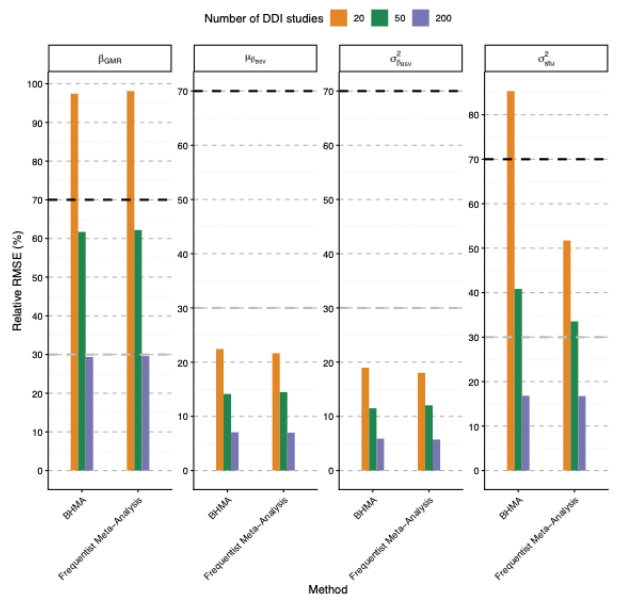


Figure 13 · Relative RMSE: NONMEM vs BHMA

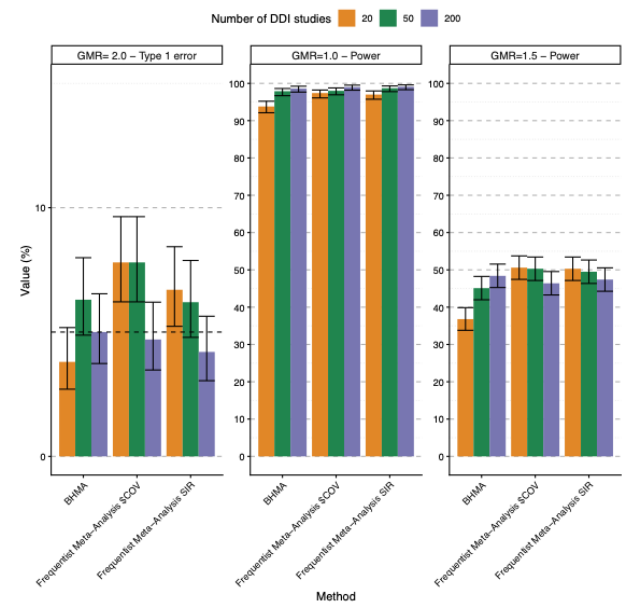


Figure 14 · Type I error & power by method

NONMEM showed bias, error and power comparable to BHMA, and estimated between-study variance more accurately. With 50 validation studies, Sampling Importance Resampling (SIR) retained the nominal type I error, a viable prior-free alternative.

UPPSALA UNIVERSITET

PART 2

How much do the inputs matter?

2

Probabilistic Sensitivity Analysis (PSA): propagating parameter uncertainty through a PBPK model



UPPSALA
UNIVERSITET

PART 2 · METHOD & FINDING

The choice of parameter distribution should reflect the available evidence

Mean + variance known
Normal distribution

Min + max only
Uniform distribution

Min, max + mode
Triangular distribution

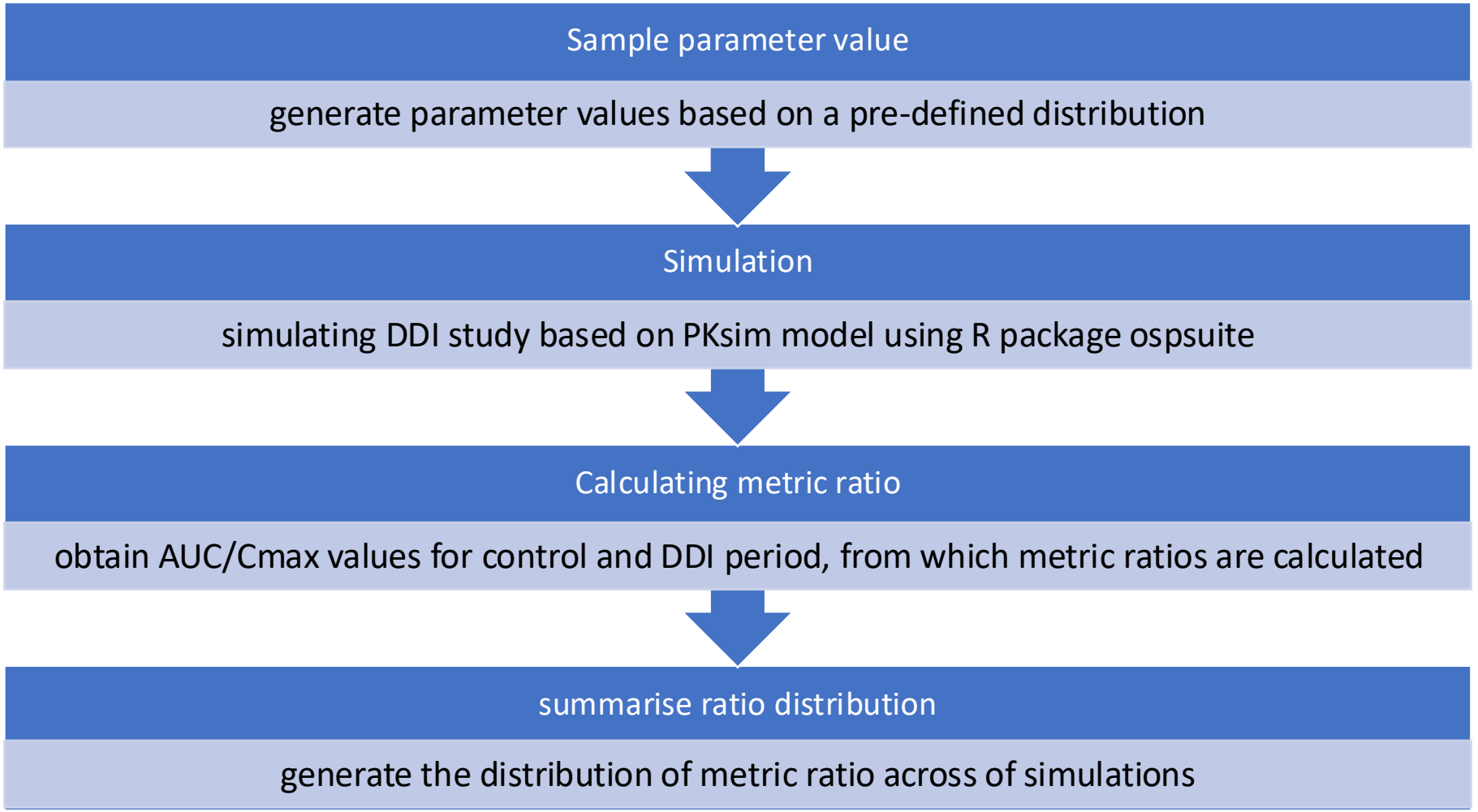
Must stay positive
Log-normal / truncated normal

Maximum-entropy principle: use the least-restrictive distribution consistent with what you actually know.

PSA is feasible in PK-Sim. Implemented via the ospsuite R package. The selection of distribution depends on the source of uncertainty and information of the uncertainty (e.g. min/max or variance)

UPPSALA UNIVERSITET

The workflow of PSA for PKsim



PART 3

Combining the two sources of uncertainty

A realistic DDI case study combining platform-level (BHMA) and parameter-level (PSA) uncertainty



UPPSALA
UNIVERSITET

PART 3 · THE CASE STUDY

Predicting a weak-inhibitor DDI from a strong-inhibitor calibration

Midazolam

Sensitive CYP3A4 substrate (the investigated substance)

Ketoconazole

Strong inhibitor — used to calibrate the model

Cimetidine

Weak inhibitor — to predict midazolam-cimetidine interaction

1

Simulate DDI data

Virtual midazolam + ketoconazole study (9 subjects)

2

Optimize parameters

Estimate CYP3A4 k_cat / K_m in PK-Sim

3

Run PSA

Propagate parameter uncertainty to the midazolam–cimetidine ratio

4

Combine with BHMA

Build predictive intervals under four schemes

All PBPK models from the Open Systems Pharmacology library, run in PK-Sim via ospsuite (R).

UPPSALA UNIVERSITET

The simulated study reproduces the expected CYP3A4 interaction

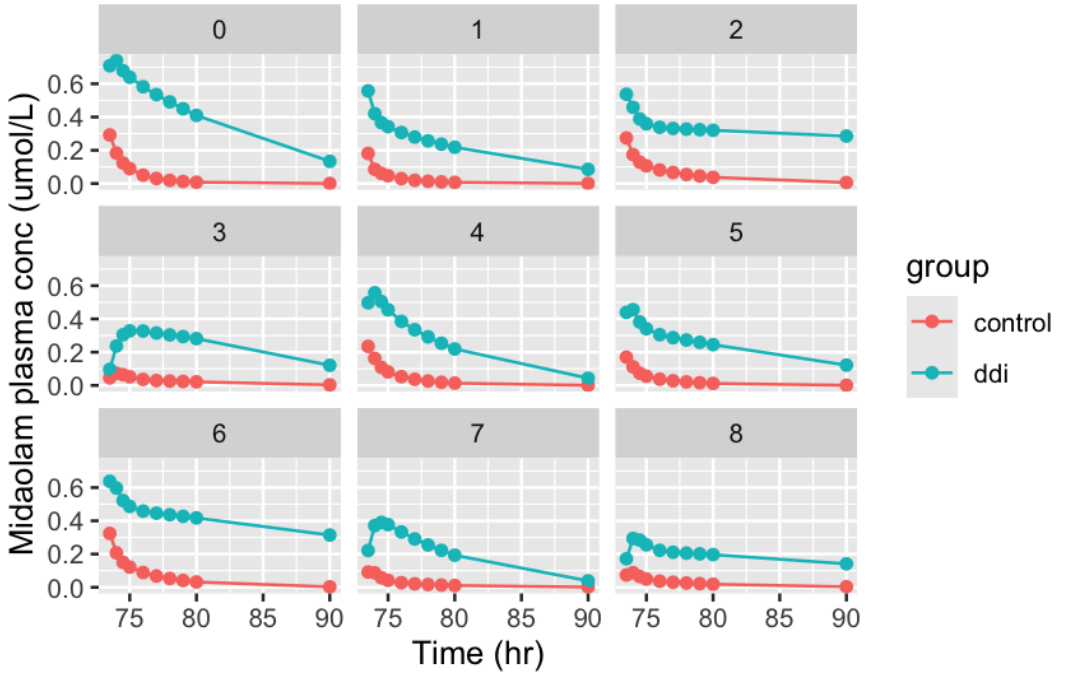


Figure 15 · Midazolam concentration–time profiles, with vs without ketoconazole (9 subjects)

UPPSALA UNIVERSITET

Simultaneous optimization of two parameters compromises identifiability

Data	Optimization	Param	Estimate	SE	RSE %	Corr.
Mean PK	k_cat only	k_cat	10.9	0.36	3.3	—
	K_m only	K_m	3.31	0.10	3.0	—
	k_cat & K_m	k_cat	14.76	14.60	**98.9**	0.999
		K_m	3.80	4.08	**107**	0.999
All individual PK	k_cat only	k_cat	12.10	0.39	3.2	—
	K_m only	K_m	2.95	0.10	3.4	—
	k_cat & K_m	k_cat	6.41	3.79	**59.1**	0.998
		K_m	1.34	0.89	**66.4**	0.998

Table 1 · Parameter-optimization results (PK-Sim)

- Two cases carried into PSA: 1-parameter and 2-parameter.

UPPSALA UNIVERSITET

Single-parameter PSA: a well-constrained input yields a narrow output distribution

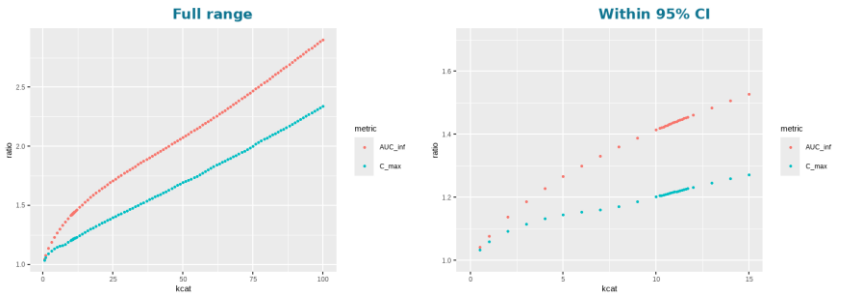


Figure 16 · Deterministic SA — k_cat vs ratios (locally linear, globally saturating)

	Mean	RSE %	5–95% range
k_cat (log-normal)	10.90	3.31	9.64 – 12.23
AUCinf ratio	1.436	0.61	1.405 – 1.468
Cmax ratio	1.216	0.44	1.197 – 1.235

Table 2 · PSA summary, single-parameter case

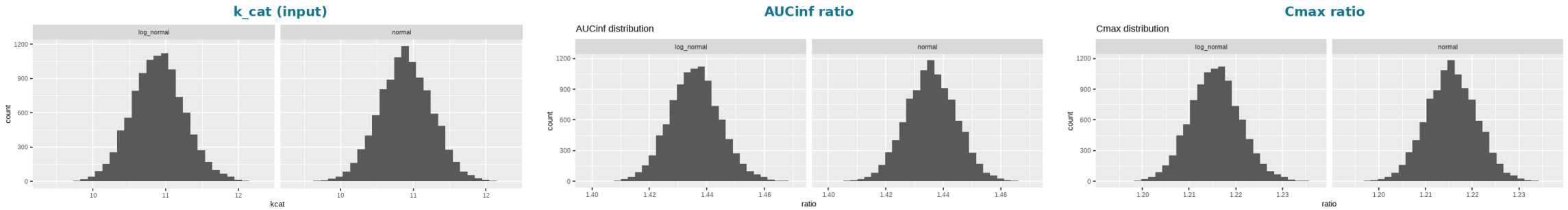


Figure 17 · PSA input/output distributions (log-normal vs normal)

UPPSALA UNIVERSITET

Normal and log-normal assumptions are effectively equivalent here (relative uncertainty ~3%).

2-parameter PSA: Correlated inputs broaden the input distributions, but their effects partly offset due to high correlation

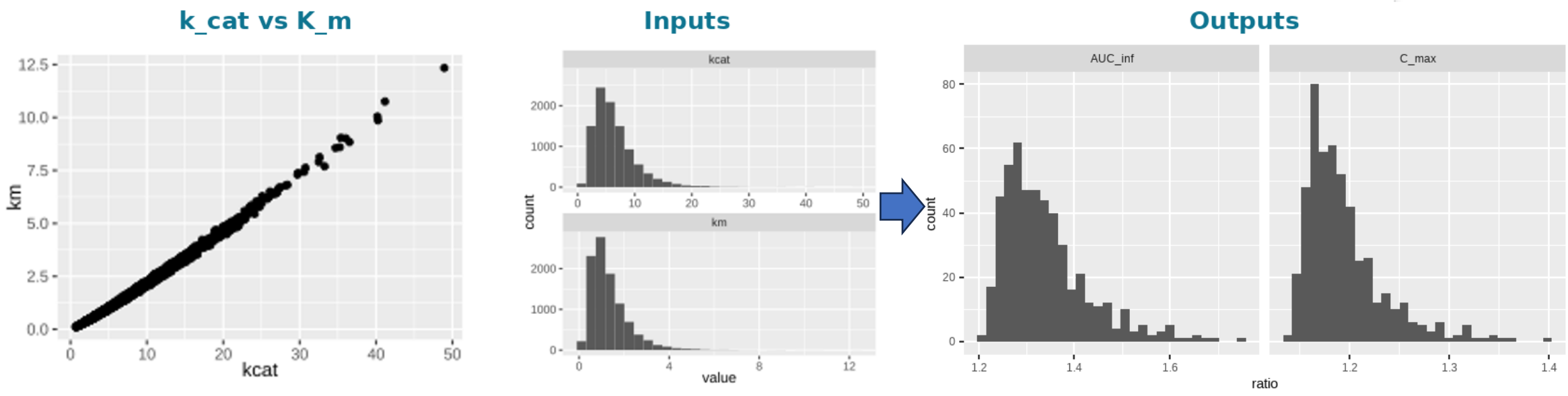


Figure 18 · PSA with bivariate log-normal k_cat & K_m (inputs and outputs)

	Mean	RSE %	5–95% range
k_cat	6.47	**59.2**	2.23 – 14.00
K_m	1.35	**66.7**	0.41 – 3.12
AUCinf ratio	1.339	6.65	1.24 – 1.52
Cmax ratio	1.194	3.35	1.15 – 1.27

UPPSALA UNIVERSITET

4 ways of combining BHMA and PSA uncertainties

(1) BHMA

$$GMR^{(m)} = exp\left(log\,\widehat{GMR} + \beta_{GMR}^{(m)} + Z^{(m)} \cdot \sigma_{stu}^{(m)}\right)$$

(2) PSA

$$90\%\,PCI = \left[Q_{0.05}\left(GMR_{PSA}^{(k)}\right), Q_{0.95}\left(GMR_{PSA}^{(k)}\right)\right]$$

(3) BHMA + PSA

$$GMR^{(m,k)} = exp\left(log\left(GMR_{PSA}^{(k)}\right) + \beta_{GMR}^{(m)} + Z^{(m)} \cdot \sigma_{stu}^{(m)}\right)$$

(4) BHMA without between-study variance

$$GMR^{(m)} = exp\left(log\,\widehat{GMR} + \beta_{GMR}^{(m)}\right)$$

Let $\widehat{GMR}$ denote the fixed PBPK-predicted GMR, $GMR_{PSA}^{(k)}$ the k-th draw (k = 1, …, K) from the PSA-derived AUC_{inf} ratio distribution, $\beta_{GMR}^{(m)}$ and $\sigma_{stu}^{(m)}$ the m-th posterior draw (m = 1, …, M) of the BHMA bias parameter and between-study standard deviation, respectively, and $Z^{(m)} \sim N(0,1)$ drawn independently for each m.

Platform-level variability (BHMA) dominates; drug-specific uncertainty (PSA) is comparatively small

1-parameter estimation

2-parameter estimation

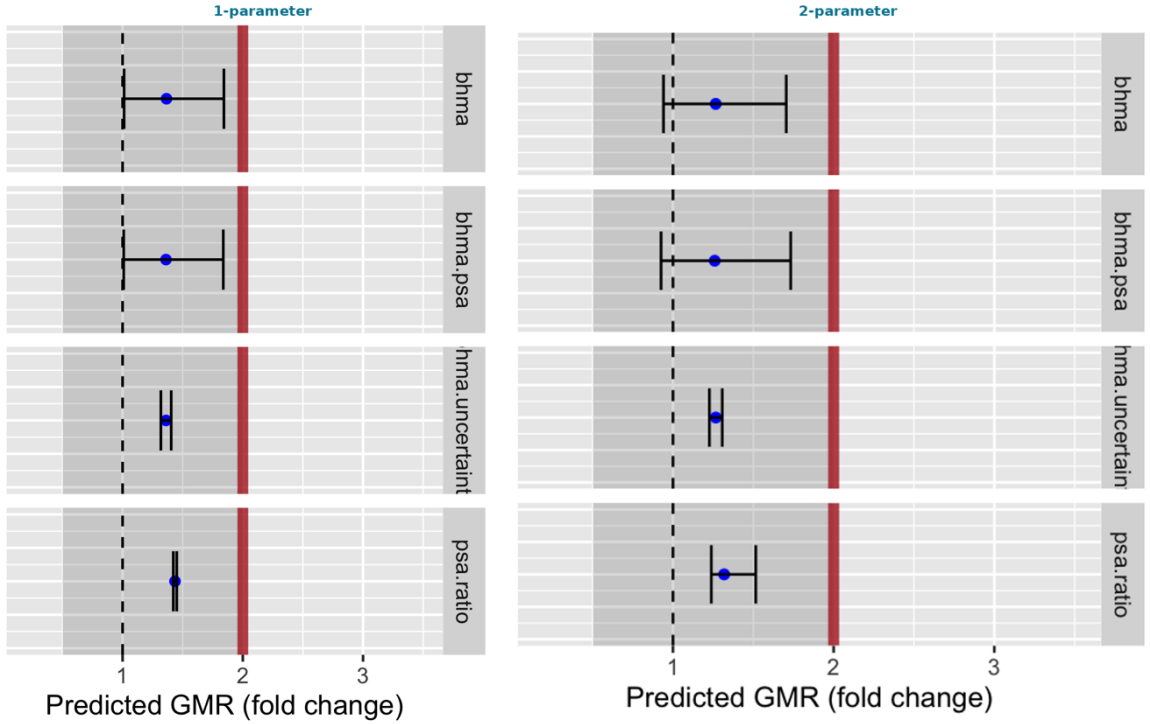


Figure 19 · 90% predictive intervals for the AUCinf ratio under four schemes (1- and 2-parameter)

Four schemes

BHMA —
Platform uncertainty + between-study variance

BHMA + PSA —
Both sources layered together

BHMA, no between-study variance —
Bias term only

PSA —
Drug-specific parameter uncertainty only

UPPSALA UNIVERSITET

SYNTHESIS

Synthesis and implications

4

Key findings and implications for regulatory decision-making

UPPSALA
UNIVERSITET

KEY MESSAGES · PART 1

Summary of findings: BHMA

Data requirements

≥ 20 DDI studies for reliable estimates; subjects-per-study matters only below that.

Operational settings

The Simcyp qualification settings (4 chains, 2500+2500 iterations) converge in >99% of replicates.

Determinant of power

Type I error is consistently controlled; power is governed by platform imprecision rather than bias.

Pediatric matching

May require more studies than the DDI case; PBPK alone may be insufficiently certain.

Robustness

Robust to most misspecification; informative and lognormal-variance priors are the exceptions.

Alternative

A frequentist NONMEM + SIR approach is an equally valid, prior-free option, given enough validation data sets.

UPPSALA UNIVERSITET

Combining the two uncertainties

Two sources of the uncertainties and observation from the case study

What each method captures

BHMA = platform-level performance across historical studies (includes model-quality scatter).

PSA = one model's input-parameter uncertainty, propagated forward. PSA cannot detect structural model error.

Case-study findings

Between-study variance over the validation studies dominated interval width; uncertainty in drug-specific parameters added little variance.

Potential approaches for combining uncertainties

1 **Compare separately**

If PSA uncertainty greatly exceeds BHMA's, the drug-specific model may be of less quality than the validation set.

2 **Sum both sources (used here)**

Layer parameter uncertainty on top of platform uncertainty

3 **Integrate PSA into BHMA**

Future work: build parameter uncertainty into the BHMA structure for one coherent total.



UPPSALA UNIVERSITET