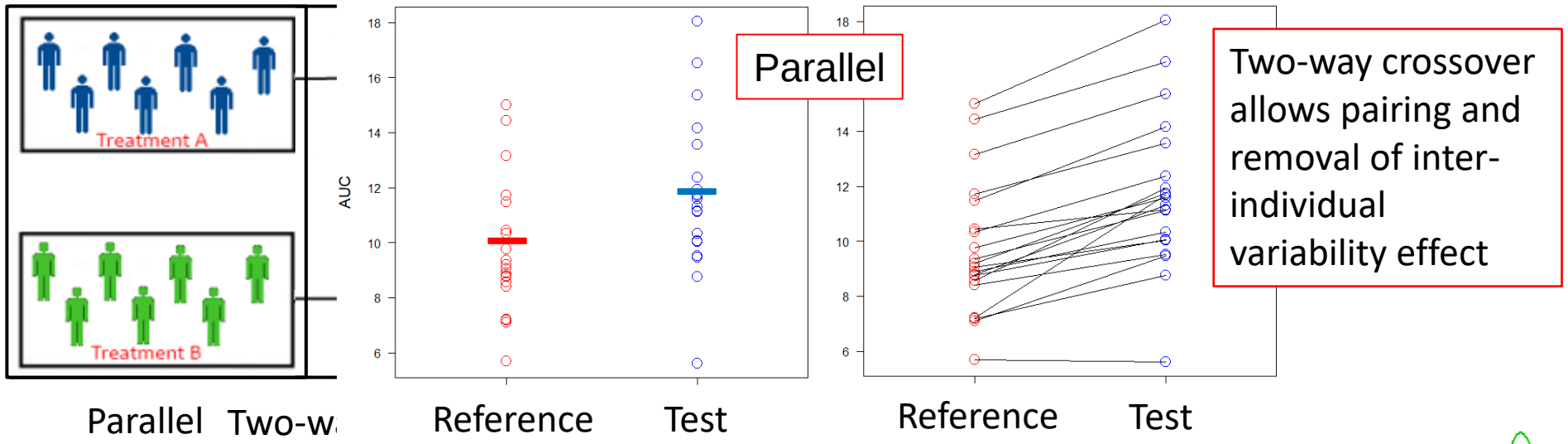
CERTARA®

A Bayesian Virtual Bioequivalence Workflow

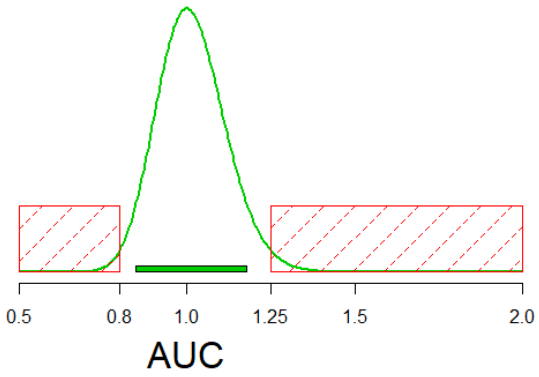
Frederic Bois, Celine Brochot

Standard BE assessment

- BE focus on absorption. Tested using the two one-sided t-test (TOST). Both *AUC* and C_{max} (obtained from observed individual concentration-time profiles) must pass. *AUC* measures extent of absorption, C_{max} rate.
- Typical PK BE studies use two-way crossover study designs, rarely parallel.

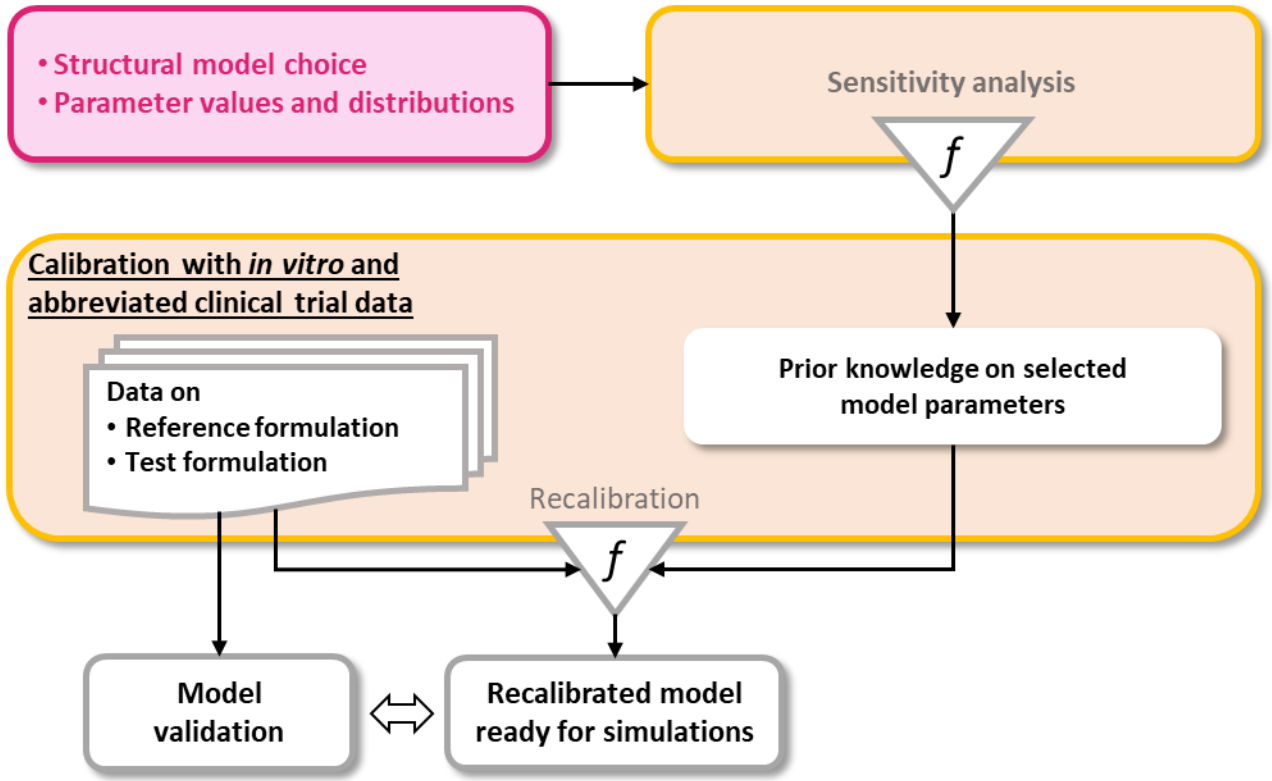


- Standard test of mean differences not usable.
- In TOST, the 90% confidence interval around the geometric mean ratio test/reference of *AUC* or C_{max} has to be between 80 and 125%. Hard to meet if difference or uncertainty or variability are large.



Virtual BE (VBE)

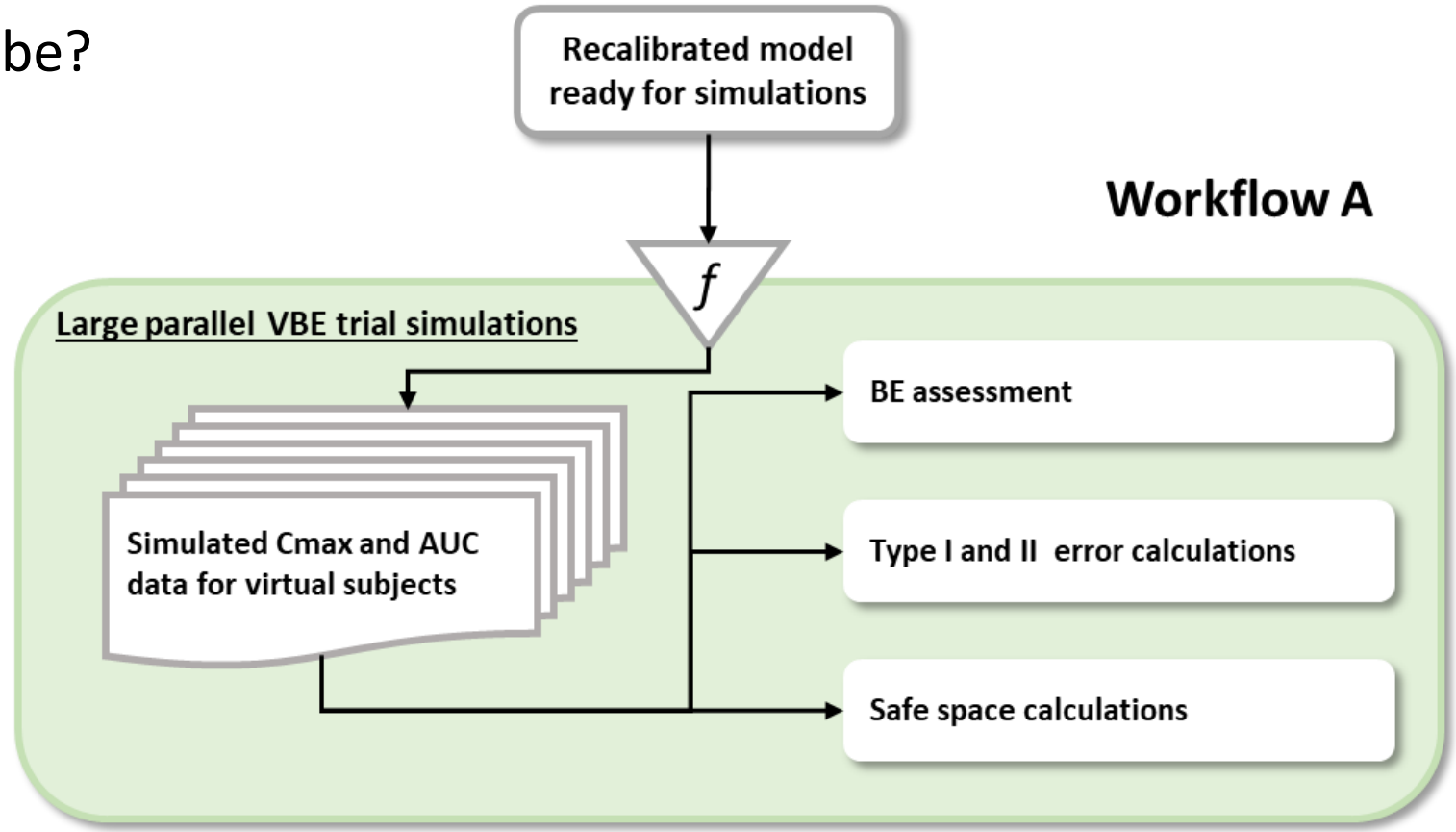
- In *virtual* BE, *no* actual trial to help decide. You simulate virtual subjects in a virtual trial with a model and decide with that.
- Model should be excellent, or you introduce uncontrollable modelling error in your decision.
- PBPK preferable, with *in vitro* data on the difference between test and reference formulations. You can use a small "abbreviated" clinical trial to inform/check your model.
- Data integration is easy in a Bayesian framework.



Calibrating a model with *in vitro* data and an abbreviated trial. SA helps choosing parameters to calibrate. In a Bayesian framework, all information, uncertainties and variability are integrated in a balanced way.

Current state of the art VBE workflow, half-baked

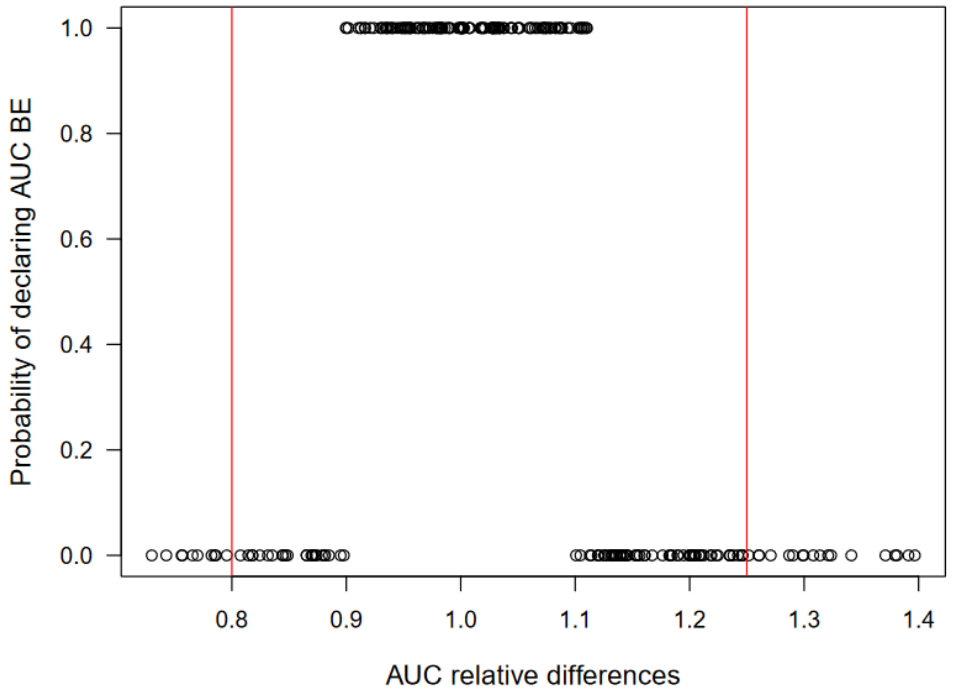
- You take the best model you have and simulate a nice big, otherwise unfeasible, clinical trial and do standard BE analyses of it, *as if it were real*.
- Question:
  How big should my “virtual trial” be?



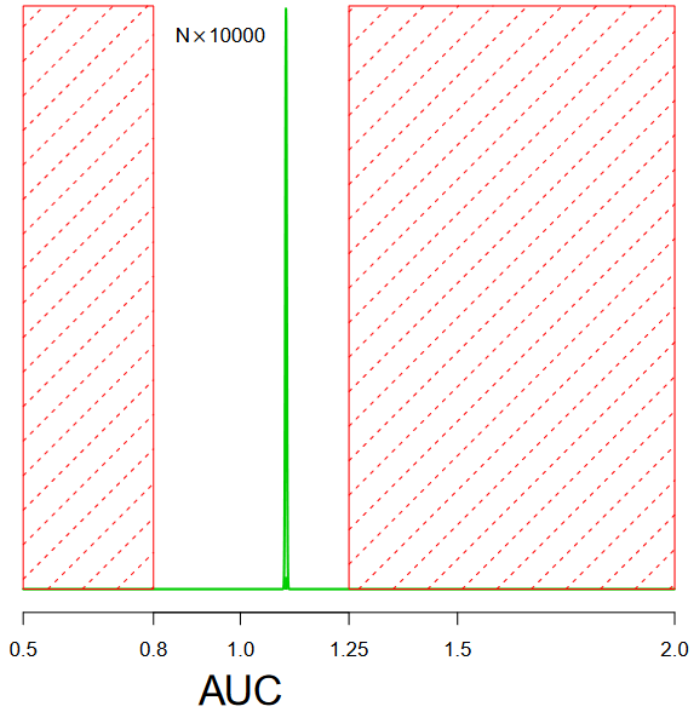
CERTARA

What happens if we limit sample size, like we do in real life for cost reasons?

Power becomes very bad when differences are close to the decision limits



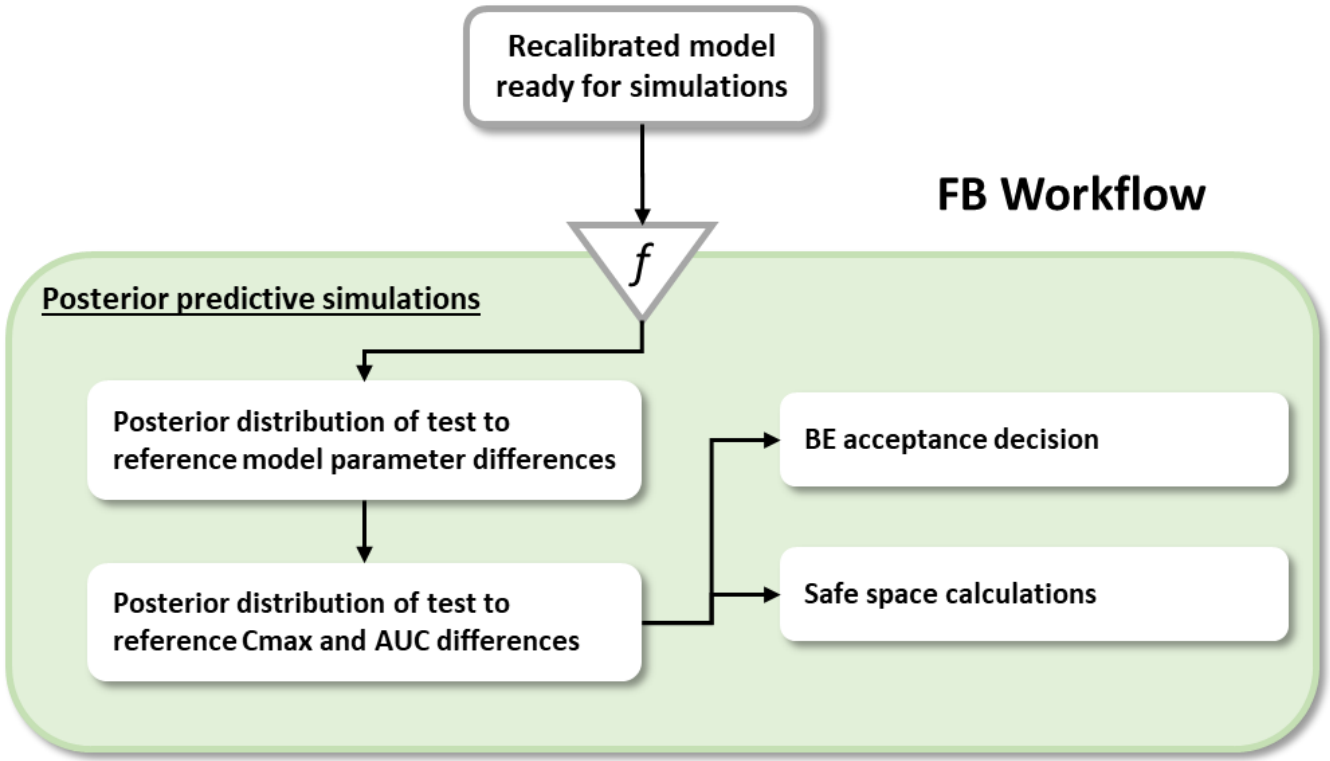
But if we increase trial size, TOST starts breaking down



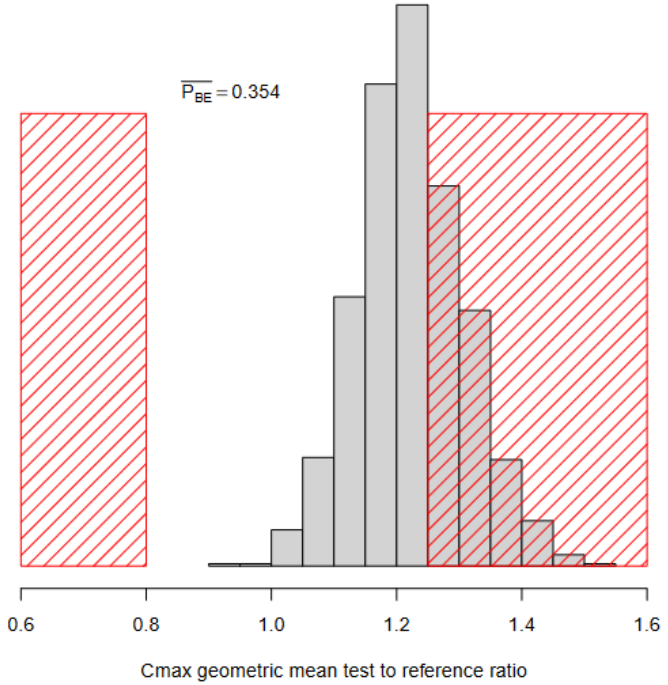
1000 virtual parallel BE trials, 500 subjects per arm. Each trial has different random values of a CQA parameter. BE assessed by TOST.

Beyond the state of the art: a fully Bayesian (FB) workflow for VBE

- All that matters is our confidence that the true test/reference difference in C_{max} (and *AUC*) is between 0.8 and 1.25. That's fair.
- Bayesian calculations can easily estimate that distribution, given all the prior, *in vitro*, abbreviated trial *etc*. information we have.



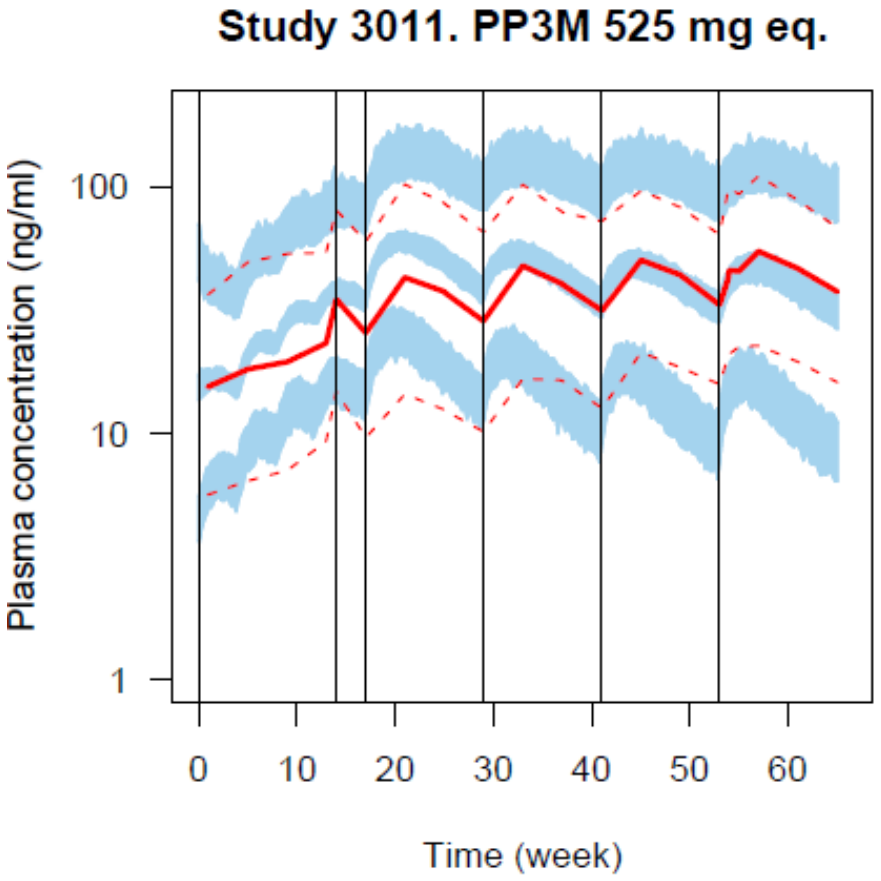
Just get the area outside the 0.8 to 1.25 interval. If > 5% do not declare BE.



CERTARA

Case study: paliperidone palmitate long-acting injectable suspensions

- PP is a treatment of schizophrenia. LAI formulations improve treatment adherence.
- Janssen has two LAI products: PP1M and PP3M. There is no generic for PP3M. High variability and 2-year parallel trial or 5-year cross-over trial would be needed.



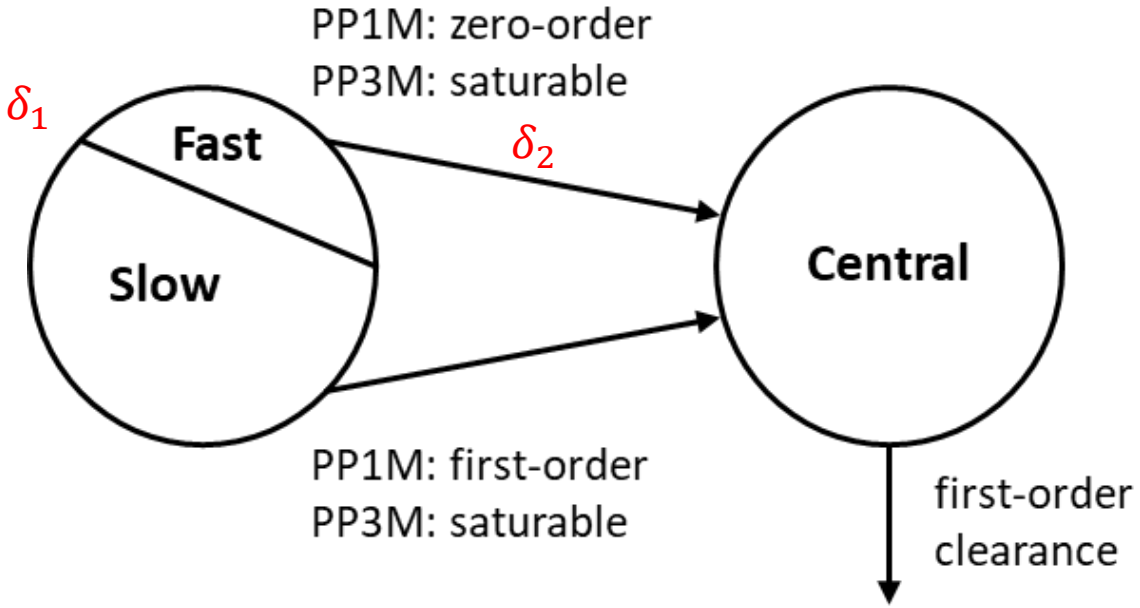
Typical treatment
PP1M: once per month
PP3M: once every 3 months

- Solid and dashed red lines are median, 5th and 95th percentiles reported in Janssen validation plots at various doses.
- Blue areas are 90 % confidence intervals of the median, 5th and 95th percentiles predicted by our implementation.
- 100 simulated clinical trials with 130 subjects per arm.

CERTARA

The Janssen PP1M + PP3M model

- We don’t have a PBPK model.
- But Janssen developed population PK models for PP1M and PP3M. The models were accepted by FDA and fit the data on hundreds of subjects at different doses.
- We cannot use *in vitro* data on a generic alternative, no *in vitro* data available anyway.



We implemented the structural model in C, the population model in BUGS language called by Nimble in R.

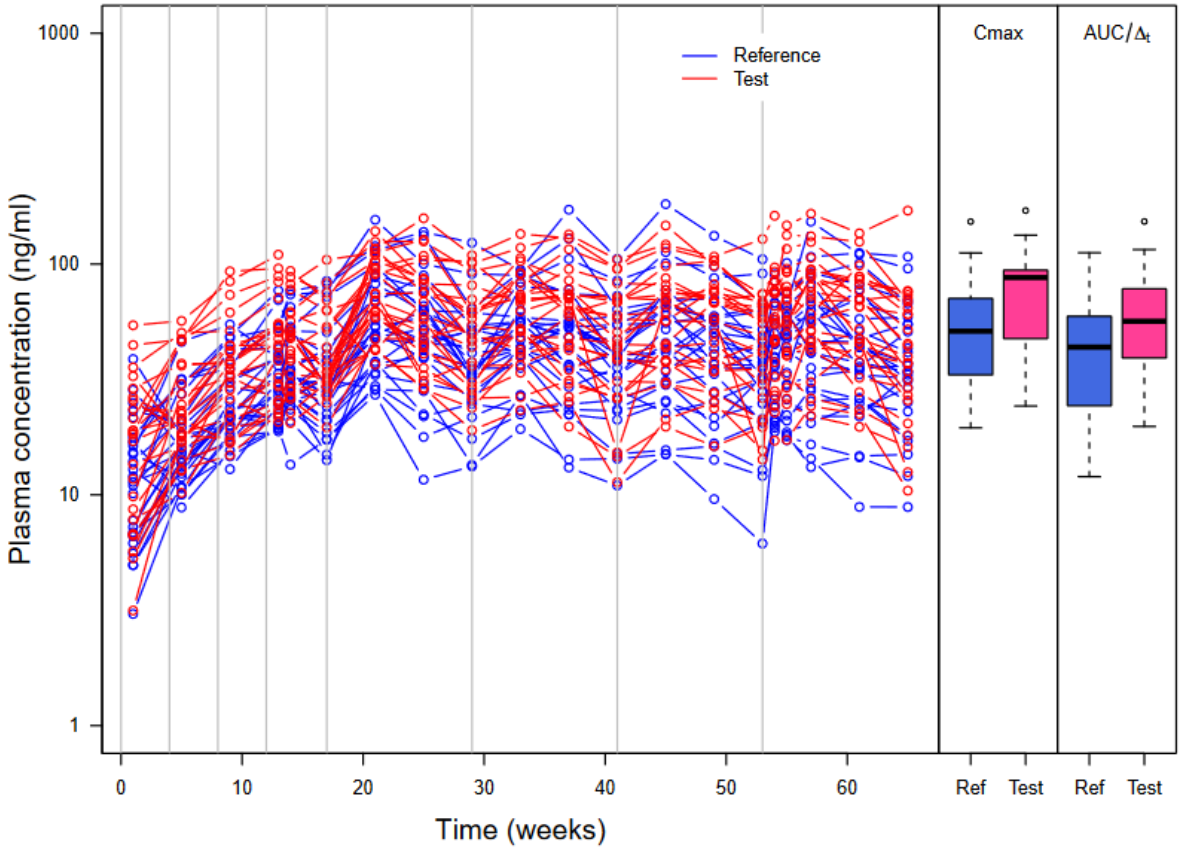
All model simulations and PE done with Nimble.

Rowland *et al.*, Expert Review, 2013

CERTARA

Virtual abbreviated trial data

- Simulated parallel abbreviated virtual BE trial for test and reference, 25 subjects per arm.
- Parameter δ_2 increased by 5% in test formulation.
- Given sensitivity, 5% difference should be bioequivalent, but trial fails to prove BE (as expected).

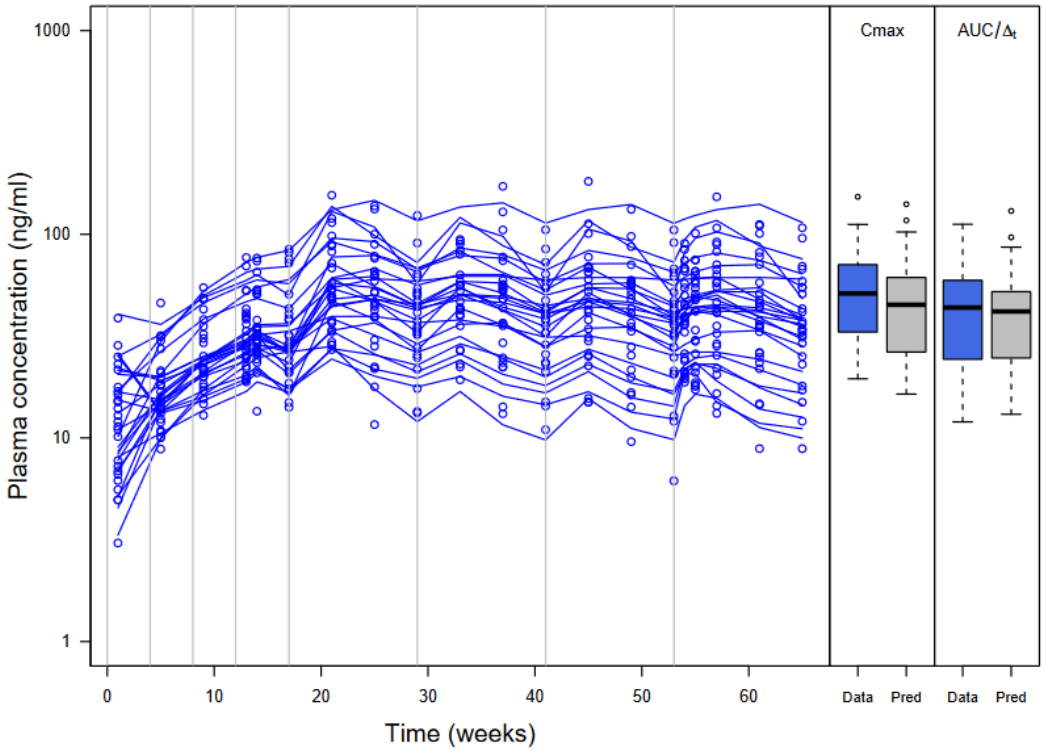


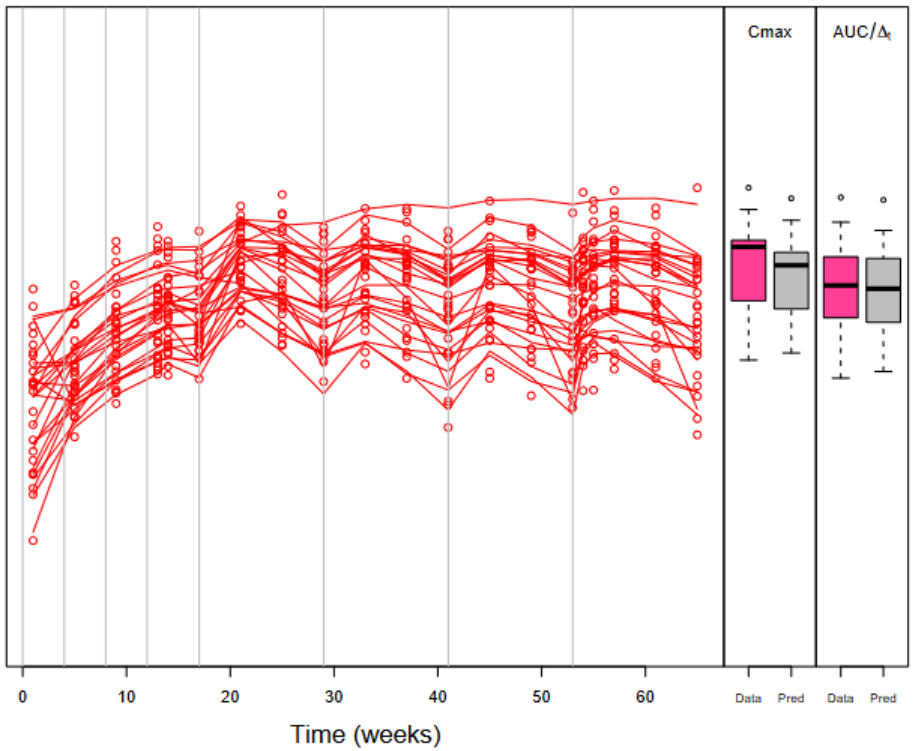
Simulated plasma PP concentrations, C_{max}, and AUC / Δ_t.
Four injections of PP1M 150 mg, four of PP3M 525 mg.

CERTARA

Bayesian calibration of the model to learn about test *vs*. reference differences

- Fit of the model to reference formulation data and test formulation data.
- C_{max}, and AUC / Δ_t are noticeably higher for the test formulation.

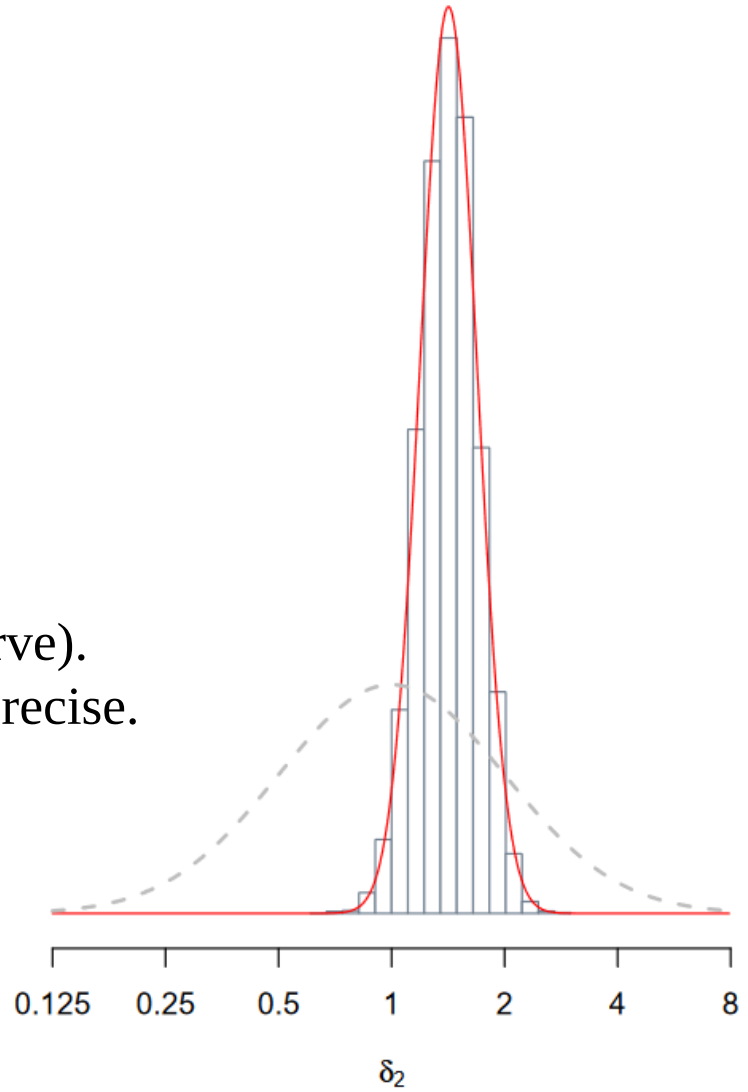




What we conclude about test *vs.* reference difference at the parameter level

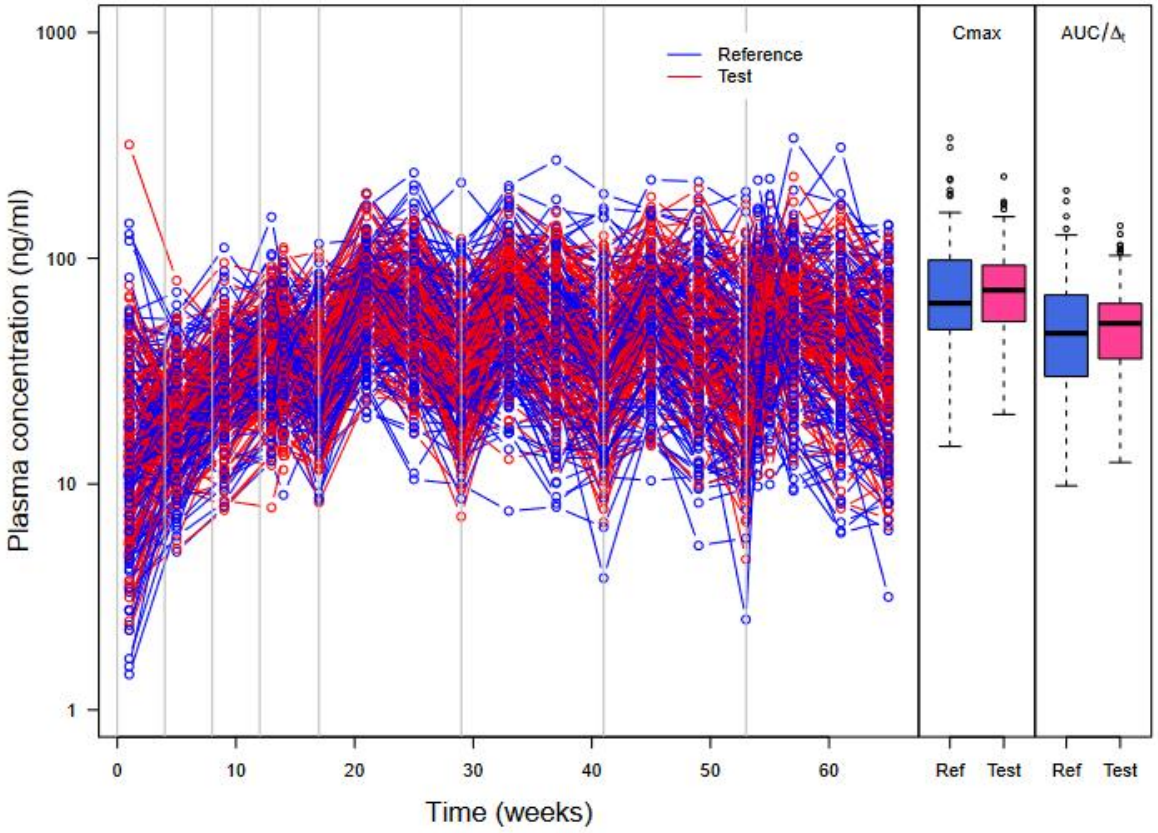
- If we fit the model it is only to estimate the value of parameter δ_2 (we are not supposed to know that it was 1.05).
- We find a value of 1.4 on average (the trial was a bit “weird”).
- Generic producer would have to live with that.

Posterior distribution of δ_2 (histogram and smooth density curve).
Dotted line shows prior distribution. Posterior is much more precise.



Simulation of a “large” virtual trial and BE assessment with workflow A

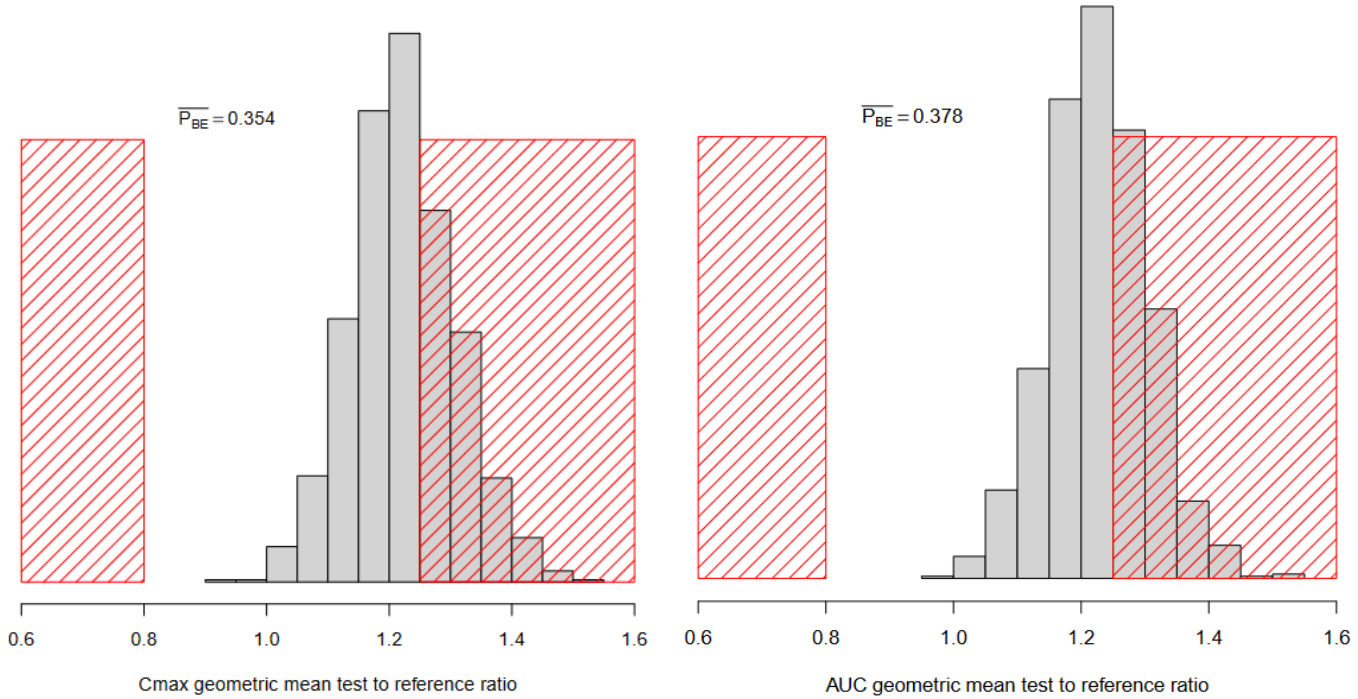
- Simulate a large trial (130 subjects per arm, like original Janssen trials)
- It’s a BE pass! Unlikely, because the abbreviated trial was indicating a large difference.
- Sounds like a scam… even though it’s the right decision!



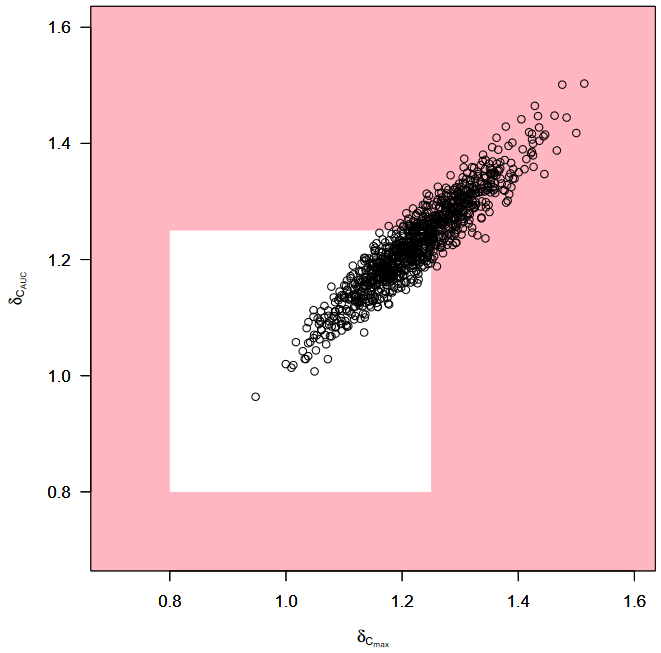
Simulated plasma PP concentrations, C_{max}, and AUC / Δ_t. PP1M 150 mg, PP3M 525 mg reference or test.

Assessment of BE with the FB workflow

- Generate samples from the posterior distributions of C_{max} and *AUC* test to reference ratios.
- Count the fraction off BE bounds. If > 5% do *not* declare BE.
- It's a fail, consistent with the abbreviated trial. FDA should like that… even though it's the wrong decision! Remember we could not use *in vitro* data because it's not a PBPK model…



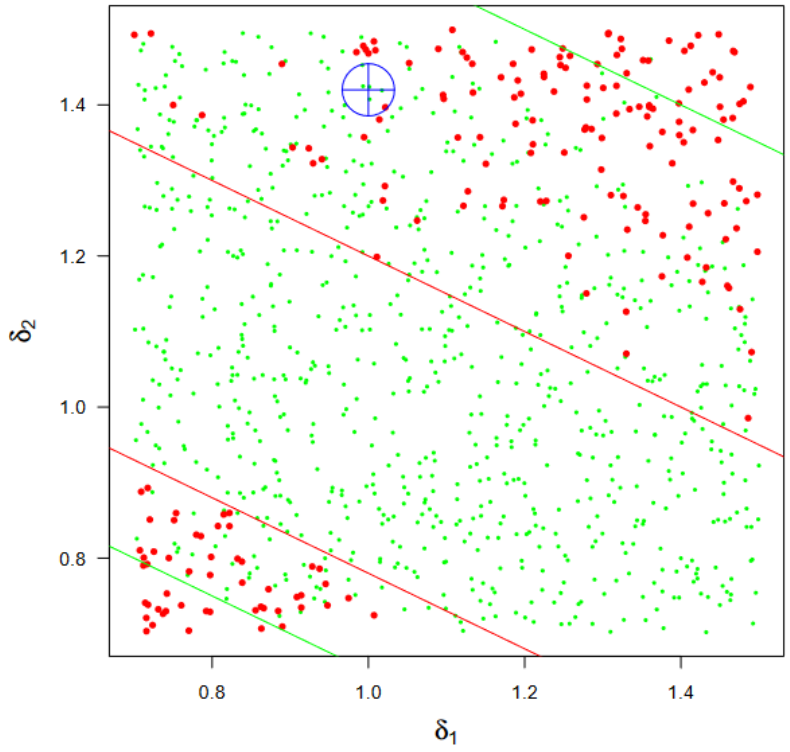
Posterior distributions of C_{max} and *AUC* test/reference ratios (1000 samples). $\overline{P_{BE}}$ is probability of *non*-BE.

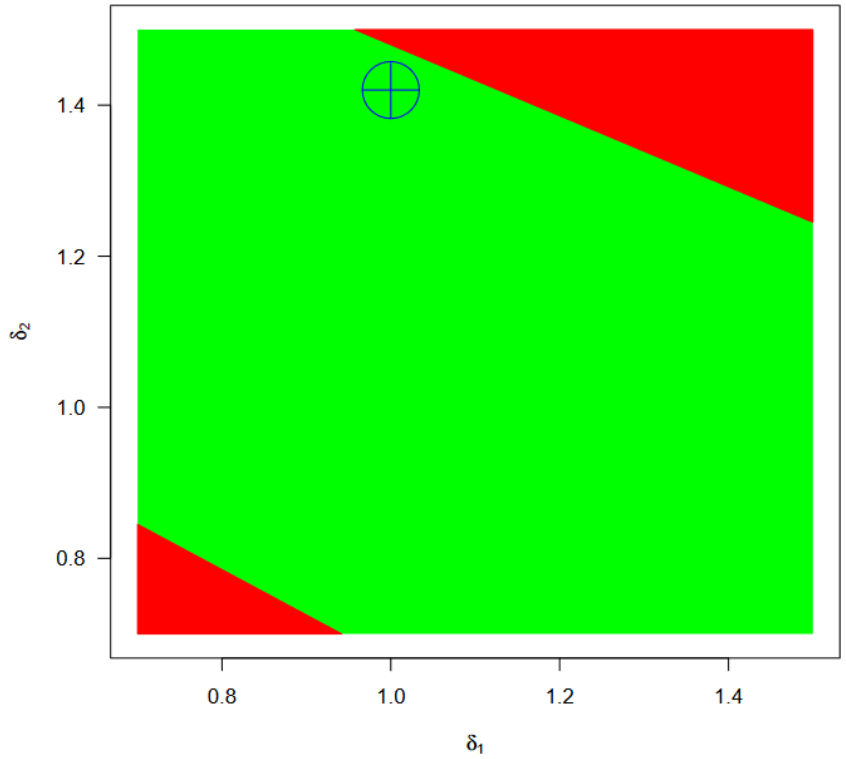


2D correlation between the C_{max} and *AUC* test/reference ratios.

Safe space analyses results from workflows A and FB

- BE safe-space regions for two most influential absorption parameters of the PP population PK model.
- Left: data-based workflow A estimate (1000 trials, 500 subjects per arm); BE trials and *non*-BE trials. The mixed region stems from bad statistical power.
- Right: fully Bayesian workflow, model-based regions are much crisper and consistent with individual-level SA (BE limit reached if δ_2 changed by about 40%.)
- Blue cross marks the location of the simulated full parallel trial we simulated for BE assessment.





CERTARA

Wrap-up

A fully Bayesian VBE workflow is much more powerful than the current practice.

You make better decisions for consumers and producers.