An Introduction to Bayesian Estimation in NONMEM

Tim Waterhouse
February 22, 2023

METRUM
RESEARCH GROUP

Acknowledgements

- Curtis K. Johnston
- Matthew Wiens
- John Mondick
- Jonathan French
- Bill Gillespie

- “Tutorial: Bayesian Estimation in NONMEM” (manuscript submitted for publication)
- Introduction to Bayesian pharmacometric data analysis with NONMEM
  - ACoP 2019 workshop by Bill Gillespie and Curtis Johnston

METRUM RESEARCH GROUP

Key messages

- Bayesian estimation can and should be used for pharmacometric models
- NONMEM speaks Bayes
- Some thought is required
  - Selecting priors
  - Processing output

nature reviews drug discovery https://doi.org/10.1038/s41573-023-00638-0

Perspective Check for updates

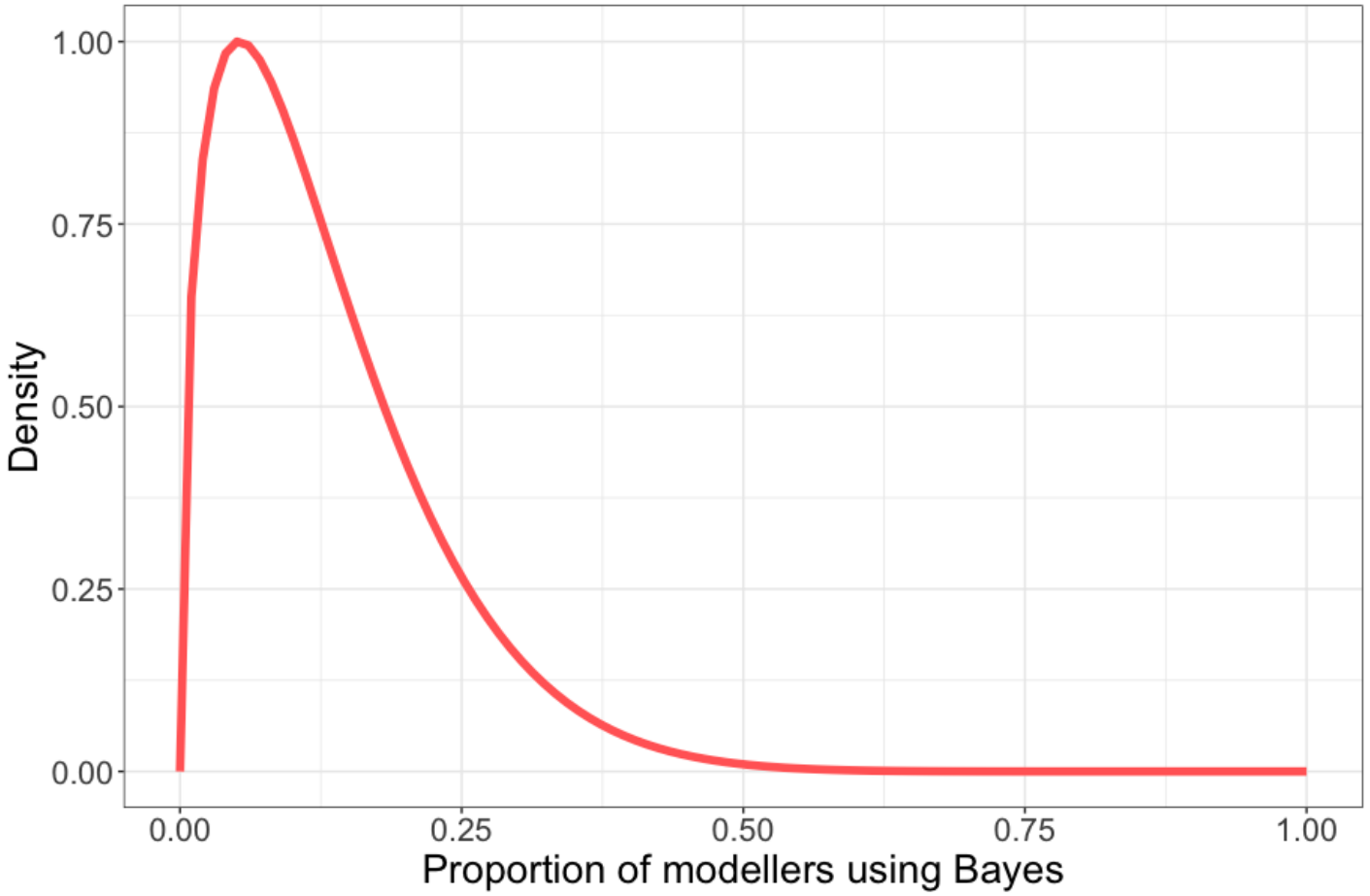
Application of Bayesian approaches in drug development: starting a virtuous cycle

Stephen J. Ruberg[1] ✉, Francois Beckers[2], Rob Hemmings[3], Peter Honig[4], Telba Irony[5], Lisa LaVange[6], Grazyna Lieberman[7], James Mayne[8] & Richard Moscicki[8]

"… despite advances in Bayesian methodology, the availability of the necessary computational power and growing amounts of relevant existing data that could be used, Bayesian methods remain underused in the clinical development and regulatory review of new therapies."

METRUM RESEARCH GROUP

My prior belief in PMx use of Bayes



$p \sim Beta(1.5, 10)$

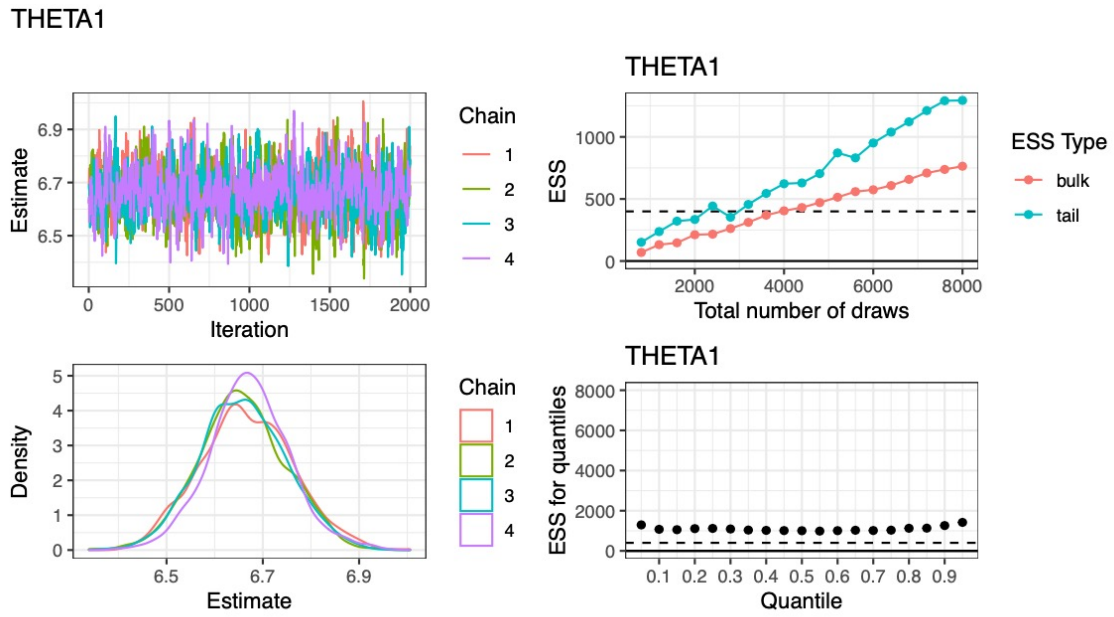
https://bit.ly/ctsi-bayes

METRUM
RESEARCH GROUP

Overview

- Why Bayes?
- What is Bayesian analysis?
- Bayesian estimation in NONMEM
- Bayesian diagnostics
  - MCMC
  - Model

Why Bayes?

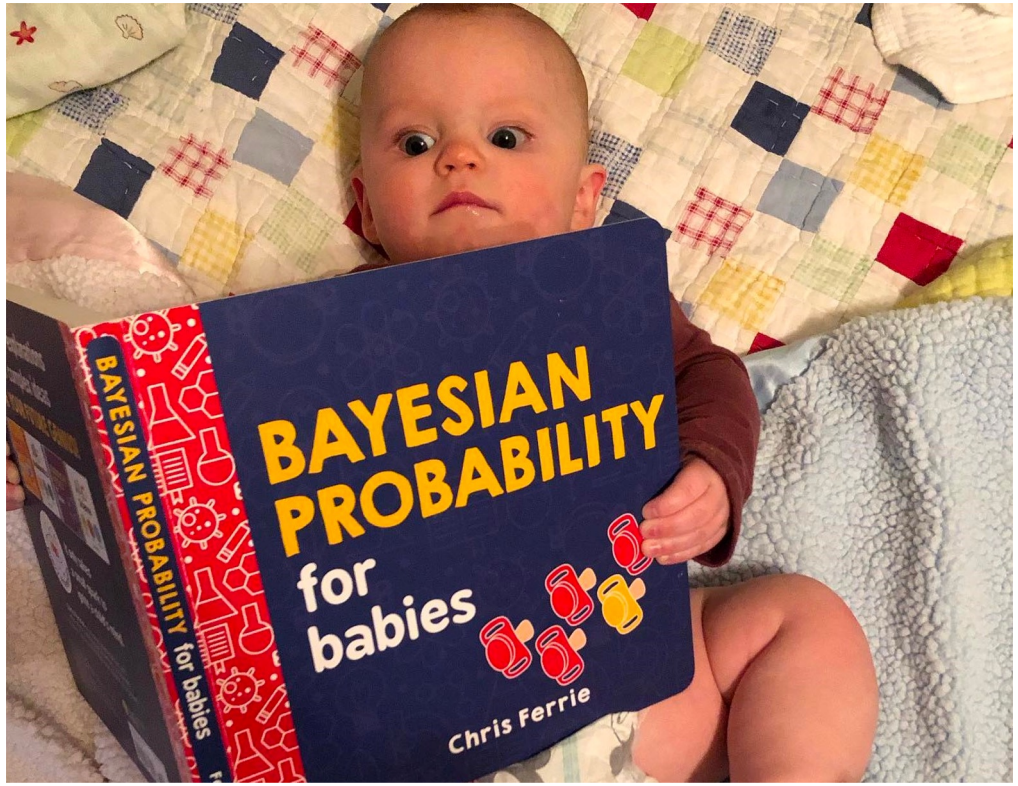


METRUM
RESEARCH GROUP

Why Bayes?

- Incorporate prior information
- Complexity in terms of random effects and hierarchies
- Analysis of data from heterogeneous sources
- Full posterior gives the best estimate of uncertainty
- Probabilistic inference for decision making

What is Bayesian analysis?



https://twitter.com/beth_fossen/status/1227244290763563008

METRUM
RESEARCH GROUP

Bayes' Rule

Posterior distribution

Likelihood

Prior distribution

$$P(\theta, \Omega, \Sigma|y) = \frac{P(y|\theta, \Omega, \Sigma)P(\theta, \Omega, \Sigma)}{P(y)}$$

Data

(Purportedly) The Reverend Thomas Bayes

Bayes' Rule

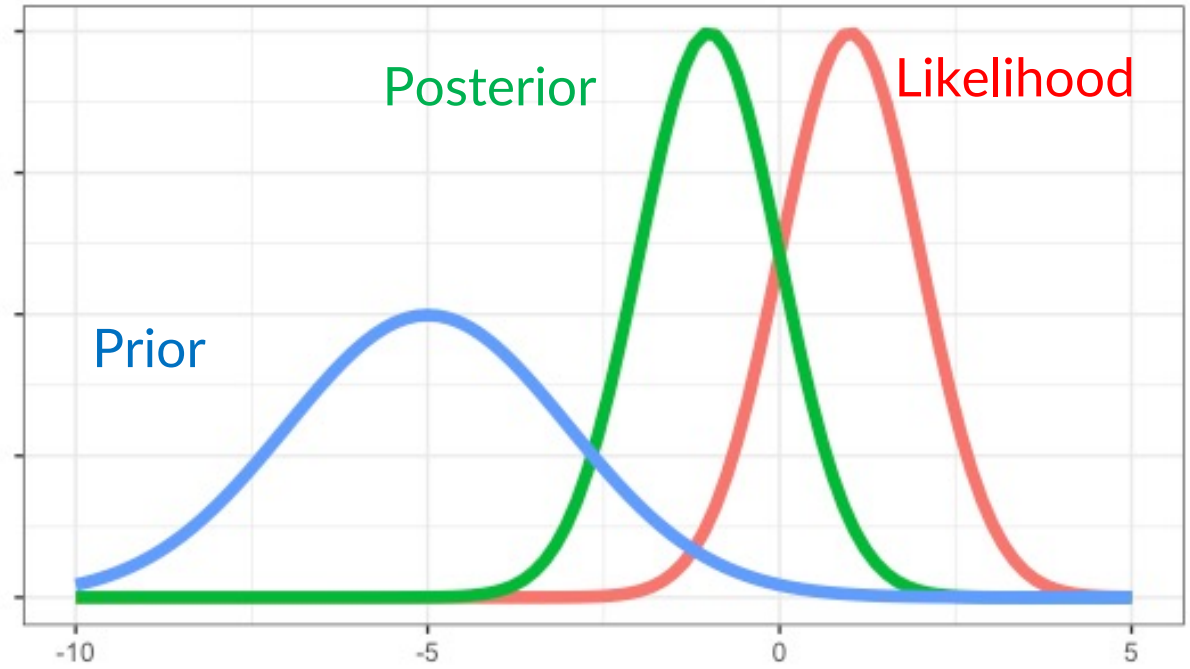
Posterior distribution

Likelihood

Prior distribution

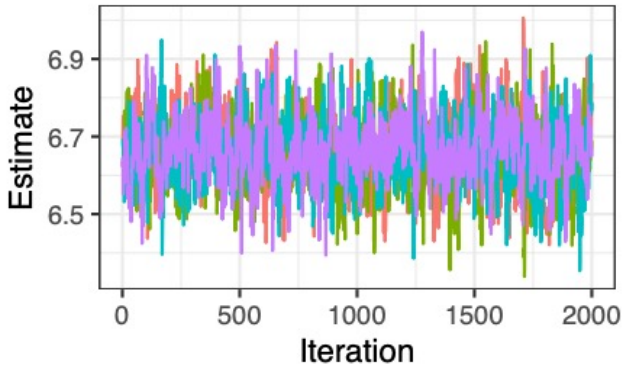
$$P(\theta, \Omega, \Sigma | y) = \frac{P(y|\theta, \Omega, \Sigma)P(\theta, \Omega, \Sigma)}{P(y)}$$

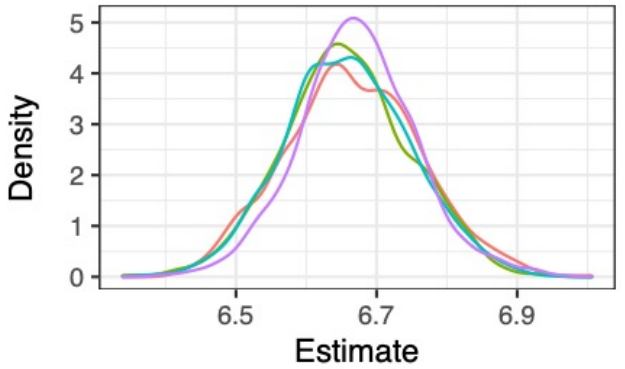
Data



Bayesian Modeling Computation

- Typically, no closed form posterior distribution
- Markov chain Monte Carlo (MCMC) used to sample from posterior
  - Metropolis-Hastings (MH)
  - Gibbs sampling
  - Hamiltonian Monte Carlo (HMC)





METRUM
RESEARCH GROUP

Prior distributions

- Represents prior knowledge or belief about model parameters
- Degrees of prior informativeness:
  - Informative
  - Weakly informative
  - Uninformative (e.g., uniform over positive real numbers)
- Explore with **prior predictive simulation**

Bayesian estimation in NONMEM

Control stream

- MU reference when possible
  - Allow Gibbs sampling (vs MH) for `METHOD=BAYES`
  - Analytic derivatives for `METHOD=NUTS`
- Prefer unbounded THETAs
  - Log or logit transform where possible
- Specify as many priors as possible

Control stream: Priors for THETAs

- Normal distribution
  - Mean **$THETAP**
  - Variance **$THETAPV**
    - Shorthand:
      $THETAPV BLOCK(5) FIXED VALUES(10,0)
- t-distribution (METHOD=NUTS)
  - Set degrees of freedom in **$EST TTDF** or **$TTDF**

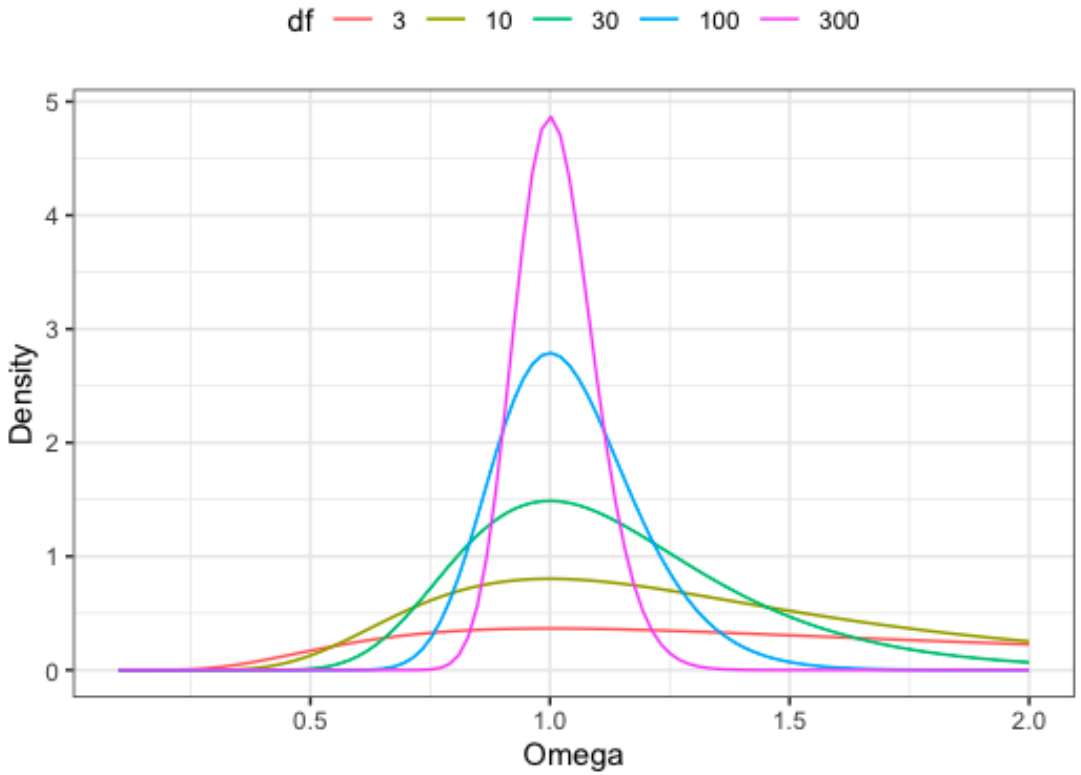
Control stream: Priors for OMEGAs

- Inverse Wishart distribution

$$f_{\mathbf{X}}(\mathbf{X};\mathbf{\Psi},\nu) = \frac{|\mathbf{\Psi}|^{\nu/2}}{2^{\nu p/2}\Gamma_p(\frac{\nu}{2})}|\mathbf{X}|^{-(\nu+p+1)/2}e^{-\frac{1}{2}\operatorname{tr}(\mathbf{\Psi}\mathbf{X}^{-1})}$$

where $\mathbf{X}$ and $\mathbf{\Psi}$ are $p \times p$ positive definite matrices, $|\cdot|$ is the determinant, and $\Gamma_p(\cdot)$ is the multivariate gamma function.

https://en.wikipedia.org/wiki/Inverse-Wishart_distribution



METRUM RESEARCH GROUP

Control stream: Priors for OMEGAs

- Inverse Wishart distribution
  - Mode **$OMEGAP**
  - Degrees of freedom **$OMEGAPD**
- Additional options for **METHOD=NUTS**:
  - Lognormal or half-t-distribution for SDs (**$EST OVARF**)
  - Lewandowski-Kurowicka-Joe (LKJ) distribution for correlation matrix (**$EST OLKJDF**)

$$f_{\mathbf{X}}(\mathbf{X};\mathbf{\Psi},\nu) = \frac{|\mathbf{\Psi}|^{\nu/2}}{2^{\nu p/2}\Gamma_p(\frac{\nu}{2})}|\mathbf{X}|^{-(\nu+p+1)/2}e^{-\frac{1}{2}\operatorname{tr}(\mathbf{\Psi}\mathbf{X}^{-1})}$$

where $\mathbf{X}$ and $\mathbf{\Psi}$ are $p \times p$ positive definite matrices, $|\cdot|$ is the determinant, and $\Gamma_p(\cdot)$ is the multivariate gamma function.

https://en.wikipedia.org/wiki/Inverse-Wishart_distribution

METRUM RESEARCH GROUP

Inverse Wishart OMEGA prior guidance

$$\mathrm{df}_i = CV(\Omega)^{-2} + n + 3$$
$$\mathrm{df} = \min(\mathrm{df}_i)$$
$$\Omega_{\mathrm{prior}} = \frac{\mathrm{df} - n - 1}{\mathrm{df}} E(\Omega)$$

where

- df_i = degrees of freedom for i*th* OMEGA diagonal
- E(Ω) = expected value of OMEGA diagonal
- CV(Ω) = desired coefficient of variation for OMEGA diagonal
- n = number of diagonal elements in the OMEGA block

to set Ω_{prior} (`$OMEGAP`) and df (`$OMEGAPD`)

Control stream: Priors for SIGMAs

- Inverse Wishart distribution
  - Mode `$SIGMAP`
  - Degrees of freedom `$SIGMAPD`
- Options for `METHOD=NUTS`:
  - Lognormal or half-t-distribution for SDs (`$EST SVARF`)
  - Lewandowski-Kurowicka-Joe (LKJ) distribution for correlation matrix (`$EST SLKJDF`)

Estimation options: initial estimates

- Multiple (e.g., 4) chains using **METHOD=CHAIN**.
  - Generate 4 sets of initial estimates with **METHOD=CHAIN NSAMPLE=4 FILE=1000.chn**
  - Use **CTYPE** option to sample initial THETAs from
    - uniform (% above and below **$THETA**), or
    - bounds in **$THETA** (not recommended!), or
    - normal distribution defined by **$THETAP** and **$THETAPV**
  - OMEGA and SIGMA initial estimates from inverse Wishart distributions
    - Degrees of freedom from **DF** and **DFS**

Estimation options: Sampling algorithm

- Metropolis-Hastings (MH) (`METHOD=BAYES`)
- Gibbs sampling (`METHOD=BAYES` with MU referencing)
- Hamiltonian Monte Carlo (HMC) (`METHOD=NUTS`)

MH is Meh
Gibbs is Good
HMC is How Maestros Compute

METRUM
RESEARCH GROUP

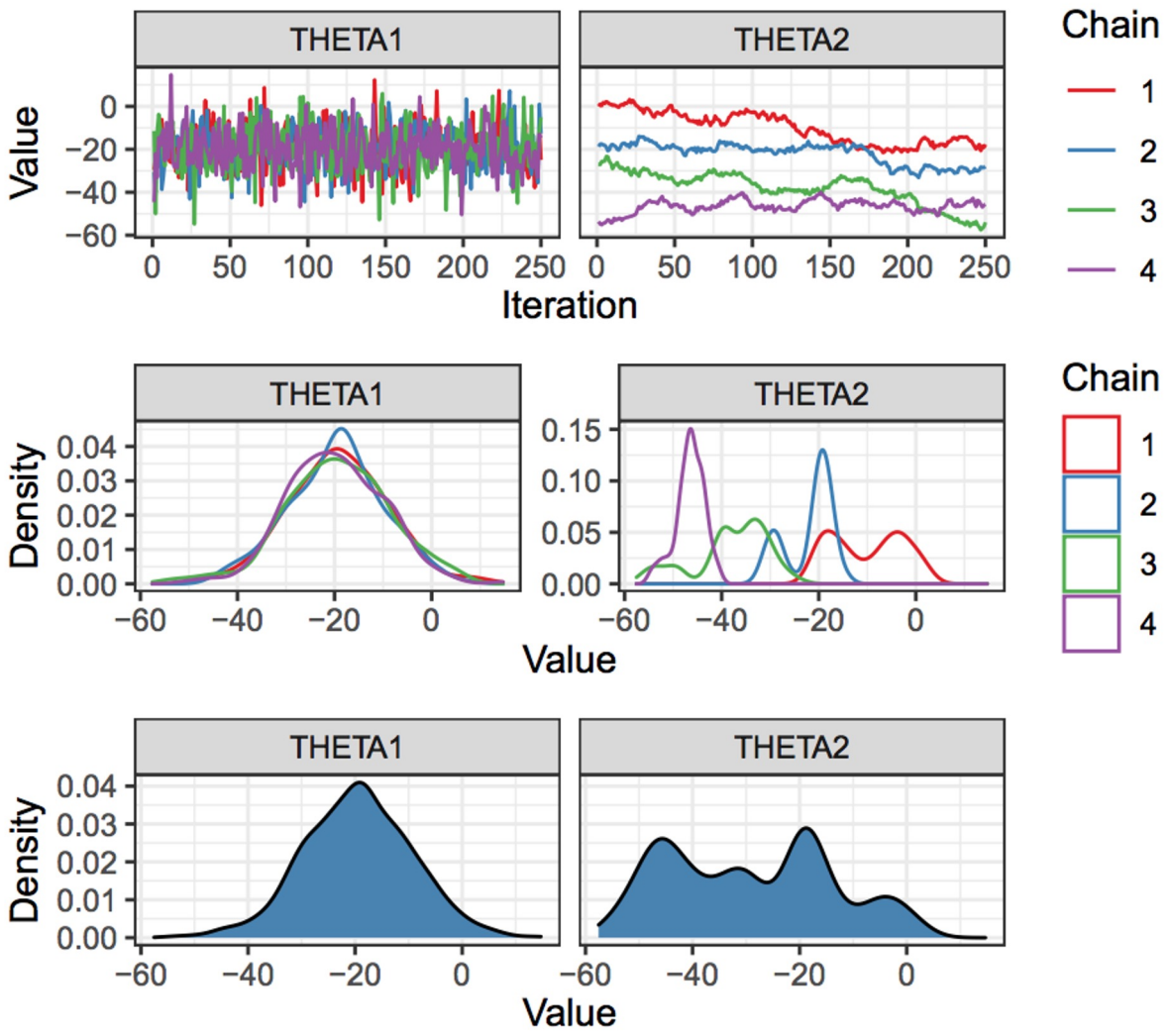
Estimation options: Individual posteriors

- **`BAYES_PHI_STORE=1`**
- Set of ETA samples for each draw from posterior
  - Provide individual-level summaries of uncertainty
  - Diagnostics (e.g., shrinkage, IPRED over full posterior)

Estimation options: Convergence testing

- `CTYPE=0`: no termination test (default, recommended)
- Tests based on changes in parameter estimates and/or objective function does not ensure convergence

Bayesian diagnostics



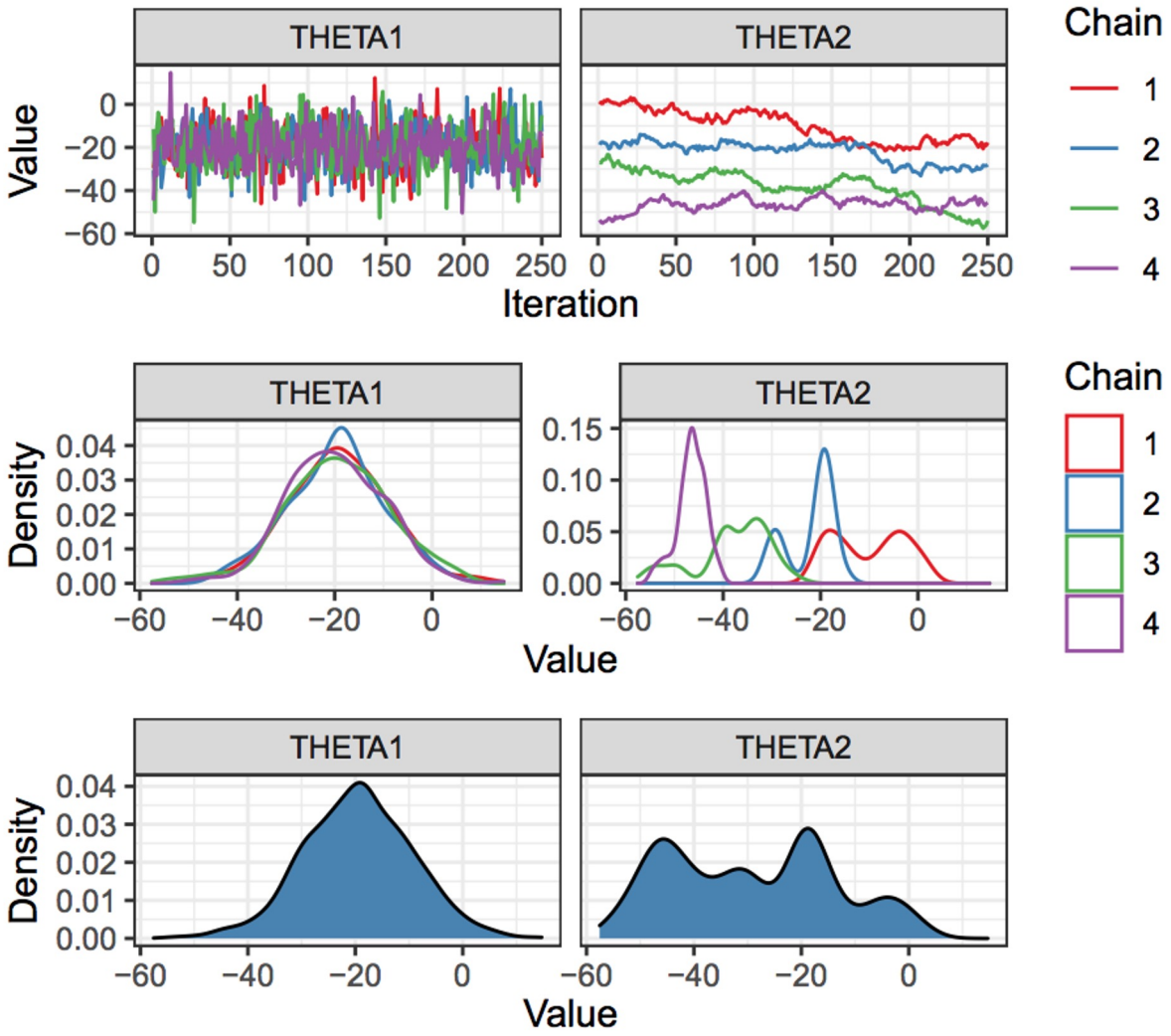
METRUM
RESEARCH GROUP

Bayesian diagnostics in NONMEM

- Diagnostics should consider full posterior (across all chains)
- NONMEM generates summaries (means, standard errors, shrinkages, etc.) within each chain
- Further post-processing is required to summarize and diagnose models across all chains

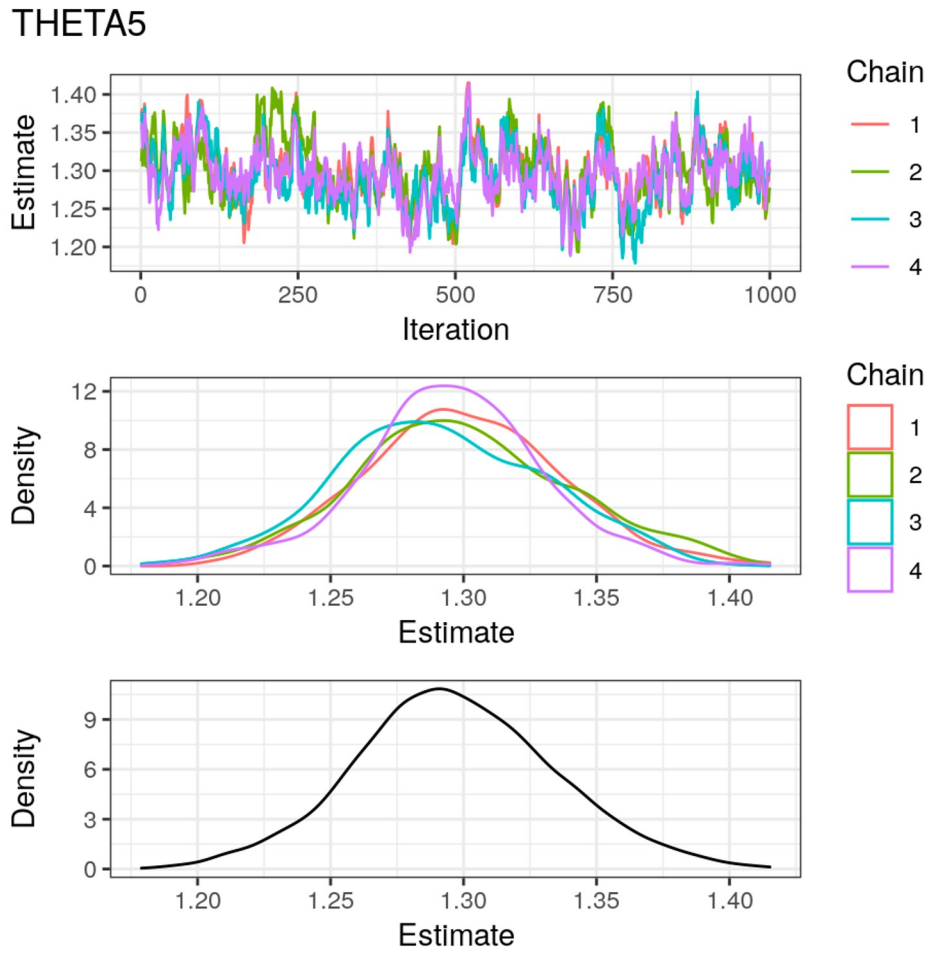
MCMC convergence diagnostics: graphical

- Trace plots
  - Check for stationary distribution with reasonable autocorrelation
  - ✅ fuzzy caterpillar
  - ❌ wiggly snakes
- Density plots
  - Common distribution between chains



MCMC convergence diagnostics: numerical

- Rhat ($\hat{R}$)
  - Measure of between-chain variance vs within-chain variance
  - Desire Rhat close to 1
- Effective sample size (ESS)
  - Measure of sampling efficiency
  - Bulk (location of distribution)
  - Tail (5th and 95th percentiles of distribution)
  - Desire ESS > ≈400

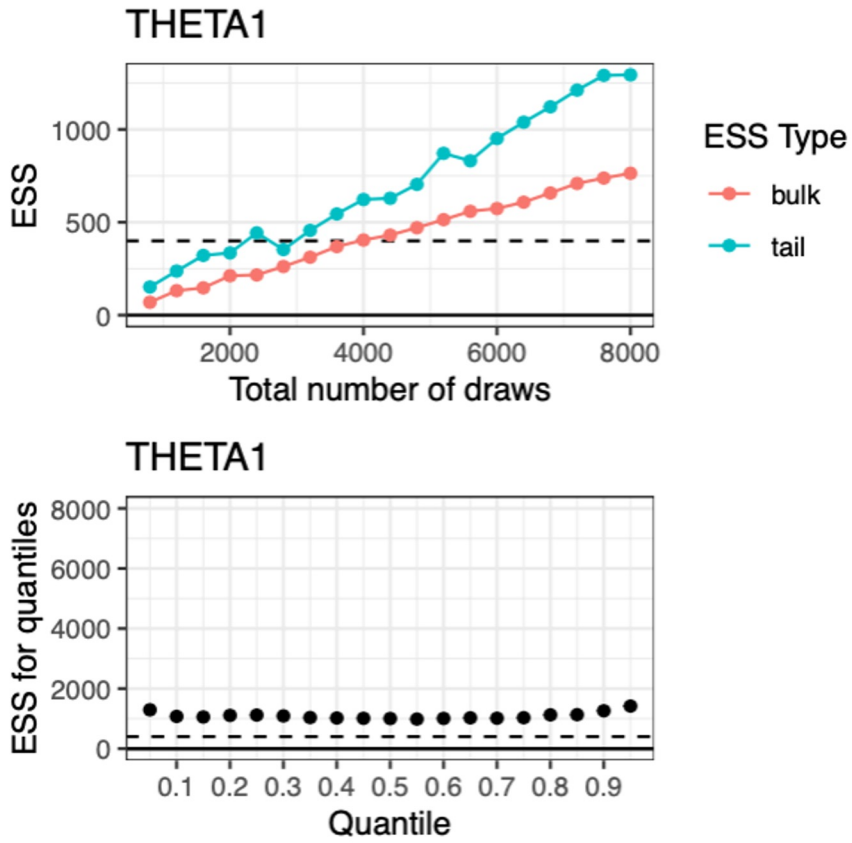


Bulk ESS = 188
Tail ESS = 354
Rhat = 1.01

METRUM RESEARCH GROUP

MCMC convergence diagnostics: more graphical

- ESS vs draw
  - Will longer chains solve convergence issues?
- ESS vs quantile
  - Ensure convergence across all quantities of interest



Addressing convergence issues

- Constrain parameter space to plausible region using tighter priors
- For IIV parameters, sampling can also be improved by setting initial estimates for individual ETA values
  - E.g., single iteration of ITS first
- Reparameterize
  - Simplification: Consider identifiability
    - Non-Bayesian estimation can cover a multitude of sins
  - Non-centered parameterization
  - Truncated Emax
- Sampling algorithm: HMC (NUTS) > Gibbs > MH

Bachman, W. J., & Gillespie, W. R. (1998, February). " Truncated sigmoid E-max models": A reparameterization of the sigmoid E-max model for use with truncated PK/PD data. In *CLINICAL PHARMACOLOGY & THERAPEUTICS* (Vol. 63, No. 2, pp. 199-199)

Addressing convergence issues: Non-centered parameterization

- "Devil's funnel" common in hierarchical (mixed effects) models
  - Sampler cannot explore sharp "neck"
- "Matt trick":

  $x' \sim N(0, 1)$

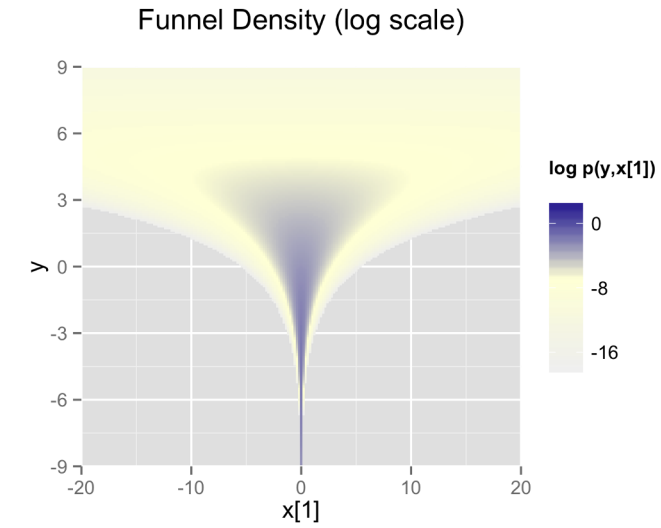
  $y' \sim N(0, 1)$

  $x = \exp(y/2) \cdot x'$

  $y = 3 \cdot y'$
  - `NUTS_EPARAM=2 NUTS_MASS=BD`
  - Set when `AUTO=2`

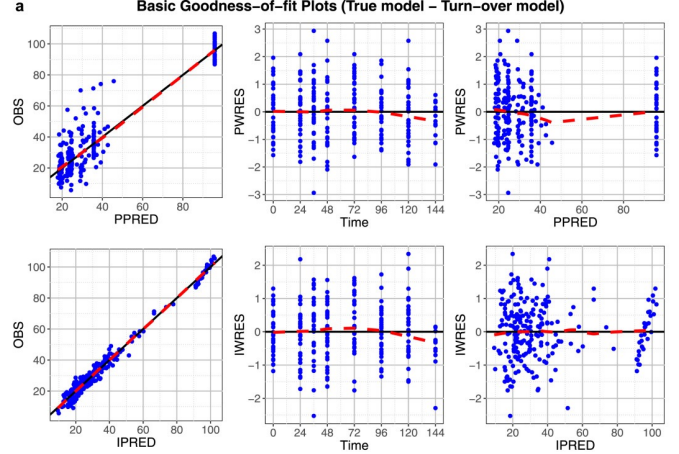
$y \sim N(0, 3)$

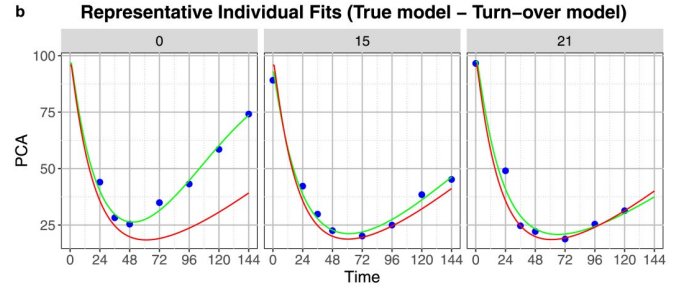
$x \sim N(0, \exp(y/2))$

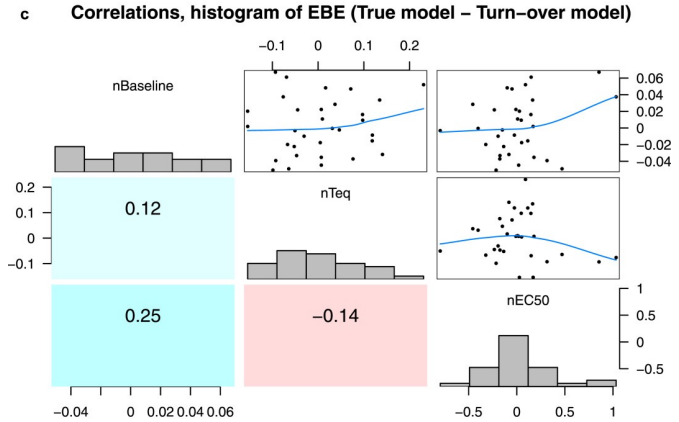


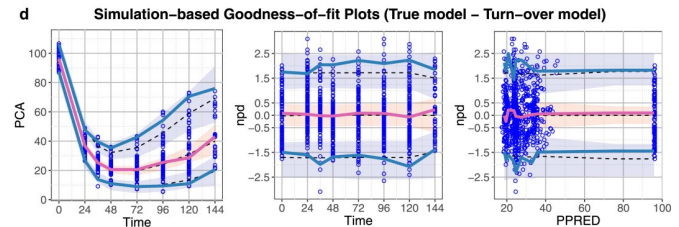
https://mc-stan.org/docs/stan-users-guide/reparameterization.html

Model diagnostics









METRUM RESEARCH GROUP

All typical PMx diagnostics apply

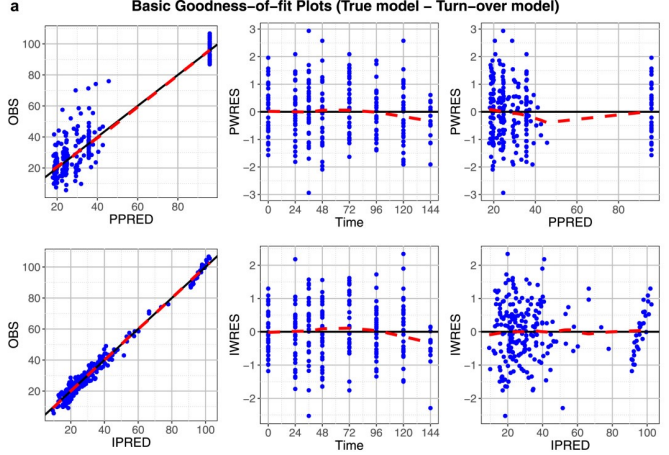
TUTORIAL

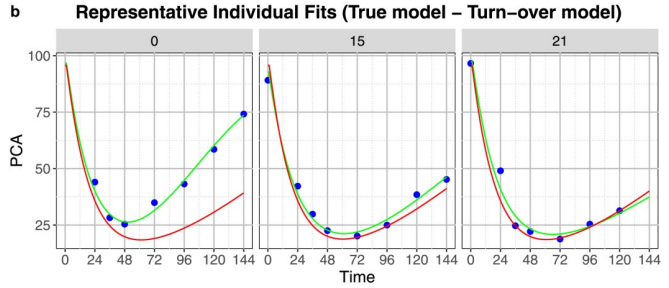
Model Evaluation of Continuous Data Pharmacometric Models: Metrics and Graphics

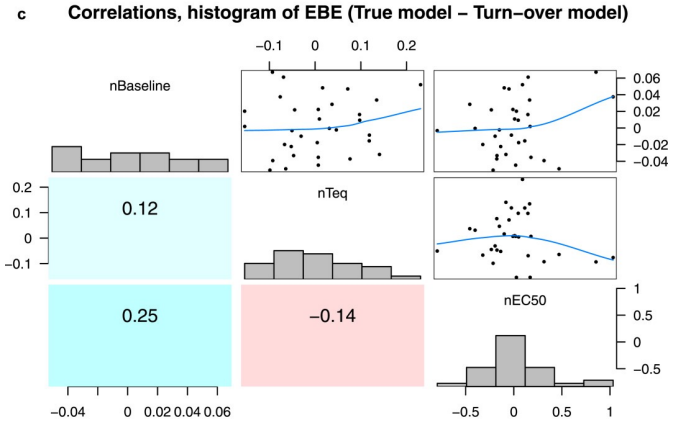
THT Nguyen[1], M-S Mouksassi[2], N Holford[3], N Al-Huniti[4], I Freedman[5], AC Hooker[6], J John[7], MO Karlsson[6], DR Mould[8], JJ Pérez Ruixo[9], EL Plan[10], R Savic[11], JGC van Hasselt[12], B Weber[13], C Zhou[14], E Comets[1,15] and F Mentré[1]*
for the Model Evaluation Group of the International Society of Pharmacometrics (ISoP) Best Practice Committee

This article represents the first in a series of tutorials on model evaluation in nonlinear mixed effect models (NLMEMs), from the International Society of Pharmacometrics (ISoP) Model Evaluation Group. Numerous tools are available for evaluation of NLMEM, with a particular emphasis on visual assessment. This first basic tutorial focuses on presenting graphical evaluation tools of NLMEM for continuous data. It illustrates graphs for correct or misspecified models, discusses their pros and cons, and recalls the definition of metrics used.

CPT Pharmacometrics Syst. Pharmacol. (2017) **6**, 87–109; doi:10.1002/psp4.12161; published online 24 November 2016.







- DV vs PRED/IPRED
- ETAs vs covariates
- VPCs
- NPDE vs time/covariates
- etc.

METRUM RESEARCH GROUP

NONMEM output needs some tweaking

- NONMEM will output means of parameter estimates
  - Probably OK for THETAs, can introduce bias for variance terms
- $TABLE outputs
  - ETA values not derived across posterior distribution, but post hoc estimates using mean of THETAs/OMEGAs
    - May result in spurious correlations

Better to derive estimates/diagnostics using the full posterior: PREDs and IPREDs

- Simulate S replicates:
  - $y_{ijs}^{\text{sim,PRED}}$
    - include all variability, sample from posterior at each replicate
  - $y_{ijs}^{\text{sim,IPRED}}$
    - include within-subject variability, posterior samples of population parameters and ETAs
- Calculate:
  - $\text{PRED}_{ij} = \frac{1}{S}\sum_{s=1}^{S} y_{ijs}^{\text{sim,PRED}}$
  - $\text{IPRED}_{ij} = \frac{1}{S}\sum_{s=1}^{S} y_{ijs}^{\text{sim,IPRED}}$

Better to derive estimates/diagnostics using the full posterior: Shrinkage

- $\text{Shrinkage} = 1 - \frac{SD_{k=1}^{K}(\overline{\eta_k})}{\sqrt{\overline{\Omega}}}$
  - $\overline{\eta_k}$ is mean of ETA posterior samples for subject k
  - $SD_{k=1}^{K}$ is standard deviation across K subjects
  - $\overline{\Omega}$ is mean of OMEGA estimates across posterior samples

Better to derive estimates/diagnostics using the full posterior: NPDE

- Can be calculated with npde R package
- Reuse output from PRED simulations:

$$y_{ijs}^{\text{sim,PRED}} \Rightarrow \mathbf{Y}_i^{\text{sim}(k)}$$

$$\mathbf{Y}_i^{\text{sim}(k)*} = \text{var}(\mathbf{Y}_i)^{-1/2}(\mathbf{Y}_i^{\text{sim}(k)} - E(\mathbf{Y}_i))$$

$$\mathbf{Y}_i^{*} = \text{var}(\mathbf{Y}_i)^{-1/2}(\mathbf{Y}_i - E(\mathbf{Y}_i))$$

$$\text{pde}_{ij} = F_{ij}^*(y_{ij}^*) \approx \frac{1}{K}\sum_{k=1}^{K}\delta_{ijk}^*$$

where $\delta_{ijk}^* = 1$ if $y_{ij}^{\text{sim}(k)*} < y_{ij}^*$ and 0 otherwise.

Emmanuelle Comets, Karl Brendel, France Mentré, “Computing normalised prediction distribution errors to evaluate nonlinear mixed-effect models: The npde add-on package for R”, *Computer Methods and Programs in Biomedicine*, Volume 90, Issue 2, 2008, Pages 154-166, ISSN 0169-2607, https://doi.org/10.1016/j.cmpb.2007.12.002

Key messages

- Bayesian estimation can and should be used for pharmacometric models
- NONMEM speaks Bayes
- Some thought is required
  - Selecting priors
  - Processing output

the end

Backup

METRUM
RESEARCH GROUP

Model selection criteria: what not to use

- Traditional objective function comparison not appropriate
- Alternatives: AIC, DIC, WAIC, cross-validation
  - AIC: not suitable for strong informative priors
  - DIC: unreliable for non-Gaussian posteriors
  - WAIC: not robust with weak priors or influential observations
  - Cross-validation: too computationally demanding

Model selection: Use PSIS-LOO

- PSIS: Pareto smoothed importance sampling
- LOO: leave-one-out cross-validation

$$\widehat{\text{elpd}}_{\text{psis-loo}} = \sum_{i=1}^{n} \log \left(\frac{\sum_{s=1}^{S} w_i^s p(y_i|\theta^s)}{\sum_{s=1}^{S} w_i^s} \right)$$

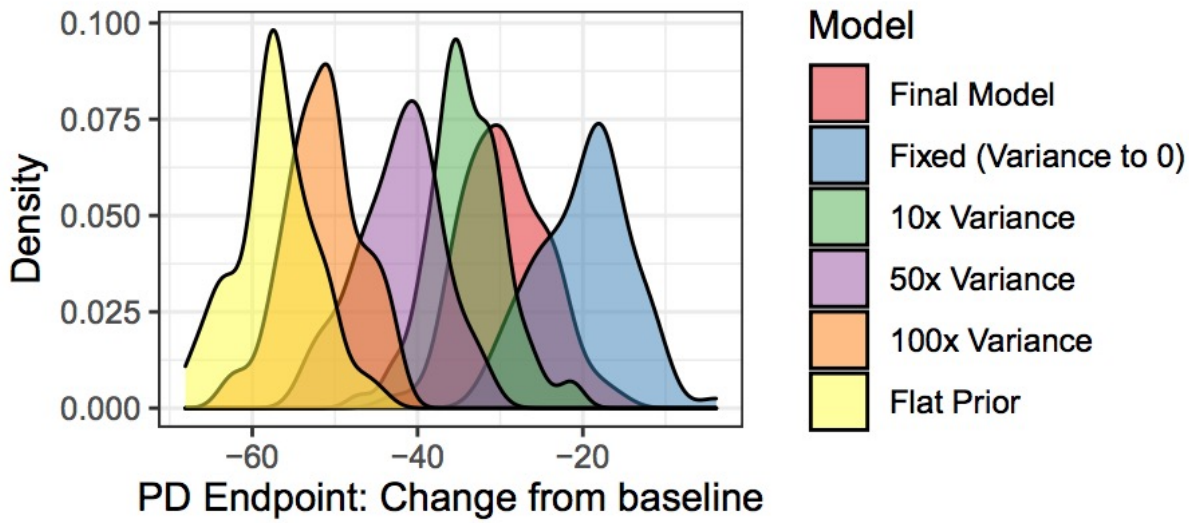
- Available using `loo` R package
- $p(y_i|\theta^s)$ is likelihood for a ***subject*** or ***observation*** at a given posterior sample
  - Requires post hoc calculation with posterior ETAs

Vehtari, A., Gelman, A. & Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput* **27**, 1413–1432 (2017). https://doi.org/10.1007/s11222-016-9696-4

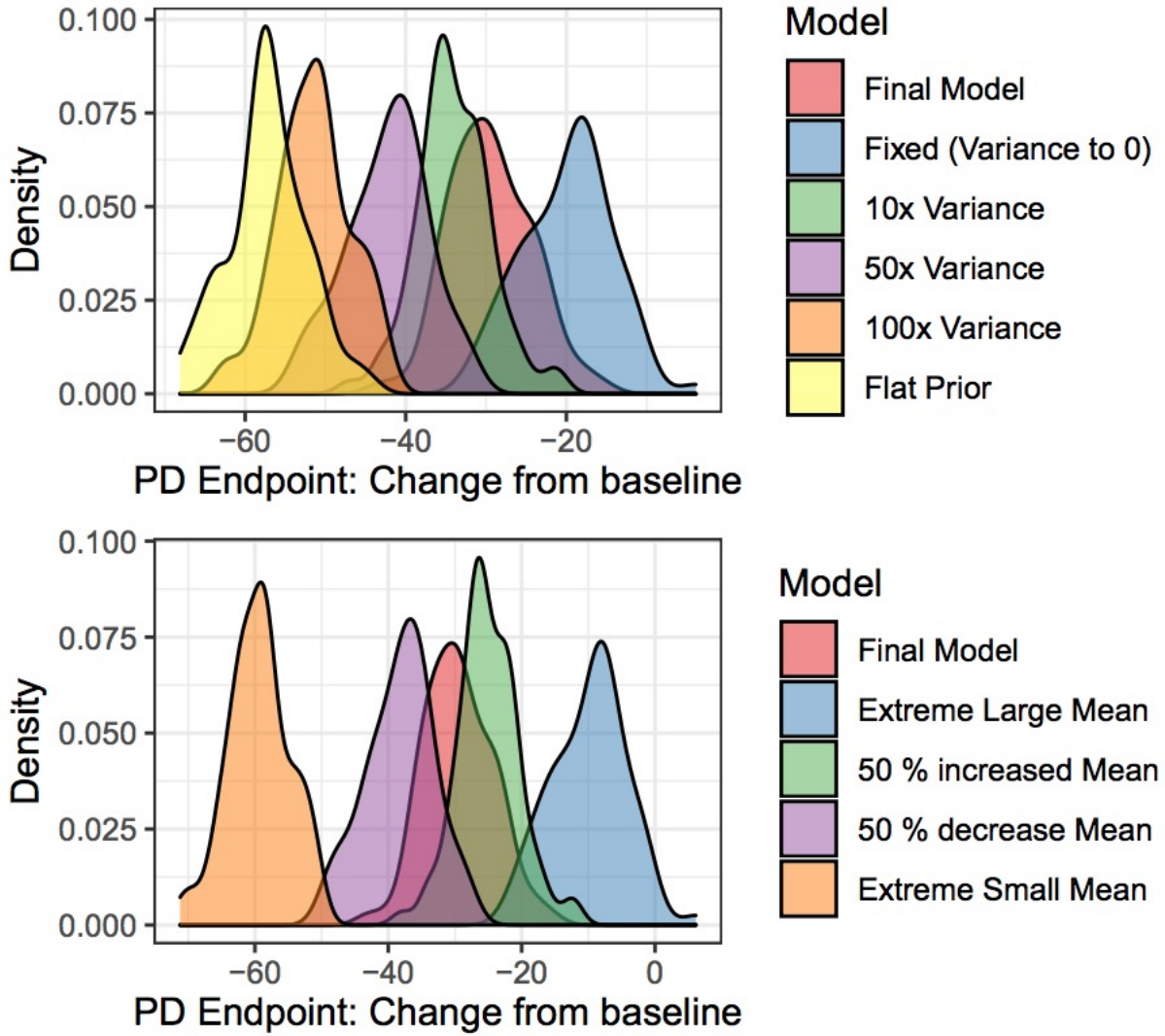
Prior impact assessment

- Sensitivity analysis: re-estimate with alternative priors
- Based on intended use of posterior inferences
- Alternatives should include changes to both variance and location of priors
- Goal is **not** to show limited impact, but to
  - Provide insight into any impact
  - Support use of prior to characterize external data

Example: Informative prior for AUC_{50}



Example: Informative prior for AUC_{50}



METRUM
RESEARCH GROUP