

Advanced Use of Stan, RStan and Torsten for Pharmacometric Applications

Bill Gillespie
Kyle Baron
Metrum Research Group Staff

PAGE Meeting
Montreux, Switzerland
29 May 2018



Course Outline

Advanced Use of Stan, RStan and Torsten for Pharmacometric Applications

Objective

Provide a guided hands-on experience in the advanced use of **Stan**, **RStan** and **Torsten** for Bayesian PKPD modeling.

Primary intended audience: Pharmacometrists

Background assumed

- Population PKPD modeling
- Use of R
- Basic understanding of Bayesian principles
- Knowledge of Stan comparable to the content of our previous introductory workshops.

Advanced Use of Stan, RStan and Torsten for Pharmacometric Applications

- 0800: Start
- Brief review of the use of Stan and RStan
 - Modeling workflow using RStan
 - User-defined functions
 - Implementing popPKPD models
- Models with systems of ODEs
 - Linear case
 - General case
- Torsten: Prototype library of PKPD functions for Stan
 - Built-in models: 1 and 2 compartment models with 1st order absorption
 - Numerical solution of user-specified ODEs
- 1000–1015: Break
- Hands-on session 1: PopPK using Torsten

Advanced Use of Stan, RStan and Torsten for Pharmacometric Applications

- Review of HMC/NUTS
- Optimizing Stan code
 - Improving computation efficiency
 - Diagnosing and remedying sampling problems encountered with Stan
 - Reparameterization, e.g., centered vs non-centered parameterizations for hierarchical models
 - Prior distributions
- 1200–1300: Lunch
- Hands-on session 2: PKPD using a model based on a linear system of ODEs

Advanced Use of Stan, RStan and Torsten for Pharmacometric Applications

- Hands-on session 3: Semi-mechanistic popPKPD model
- 1500–1530: Break
- Dealing with censored data in Stan
- User-defined probability functions and likelihoods
- Hands-on session 4: Parametric time-to-event model
- What didn't we cover?
- Appendices
 - Model evaluation and comparison
 - Use of informative prior distributions in pharmacometrics applications
 - Additional examples
- 1700: Adjourn

This is an accelerated 1 day workshop

- This is an abridged and accelerated version of a 2–3 day workshop.
- Course assistant
 - Kyle Baron
- Hands-on sessions will be demos that you run as I give step-by-step instructions rather than independent programming and problem-solving exercises.



Computer resources

Cloud-based computer access via  metworx

- Each participant may access an 8 core cloud instance using Metworx.
- All software needed for the workshop is pre-installed.
- We will use Rstudio and the R package RStan for the course exercises.
- System requirements: Laptop with wifi and a browser than can access Metworx at
<https://metworx-eu-central-stage.metworx.com/>

Accessing metworx

url: <https://metworx-eu-central-stage.metworx.com/>
username: To be assigned
password: To be assigned

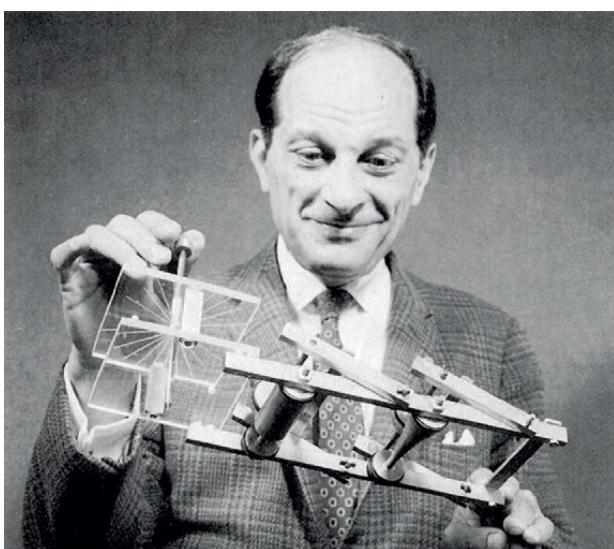
- ➊ Sign in
- ➋ Click on Rstudio

username: To be assigned
password: Same as Metworx password

Bayes & Stan at PAGE

- D-01: David Ternant. Population and Bayesian kinetic modelling of necrosis biomarkers to assess the effect of conditioning therapies on infarct size
- IV-40: Eric Novik. Stan: an open source probabilistic programming language for high-performance statistical computing
- I-16: Eunjung Song. Bayesian estimation of parameters in the pharmacokinetic model
- I-56: Sebastian Weber. Supporting drug development as a Bayesian in due time?!
- II-03: Elvira Erhardt. Bayesian knowledge integration for an in vitroin vivo correlation (IVIVC) model
- II-06: Felix Held. Bayesian hierarchical model of oscillatory cortisol response during drug intervention
- IV-20: Paolo Magni. Evaluation of software tools for Bayesian estimation on population models: an update based on current software versions

Brief review of the use of Stan and RStan



Stanislaw Ulam, co-inventor of Monte Carlo methods, holding an analog computer known as the FERMIAC that performed a mechanical simulation of random diffusion of neutrons (<http://fas.org/sgp/othergov/doe/lanl/pubs/00326866.pdf>).

Modeling workflow using RStan

We will use R to implement data analysis workflows:

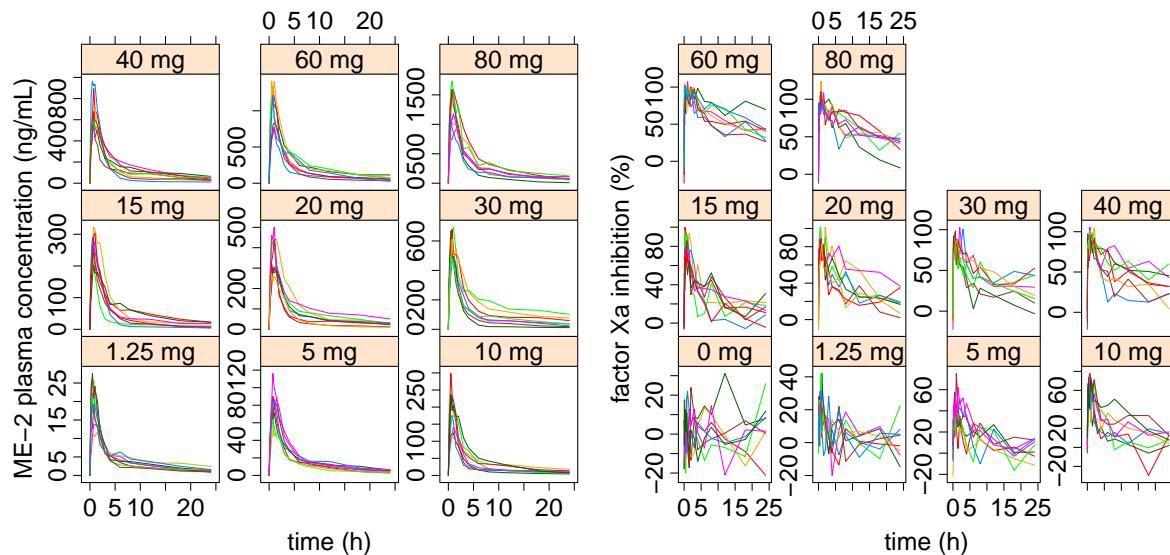
- Data management
- Launching Stan via the RStan package
- Summarizing and analyzing the MCMC samples generated by Stan
 - Tabulation of parameter estimates using an RStan function
 - Graphical diagnostics and summaries using functions from the RStan and bayesplot packages
 - Posterior predictive checking via R code and ggplot2 plots

Stan workflow example

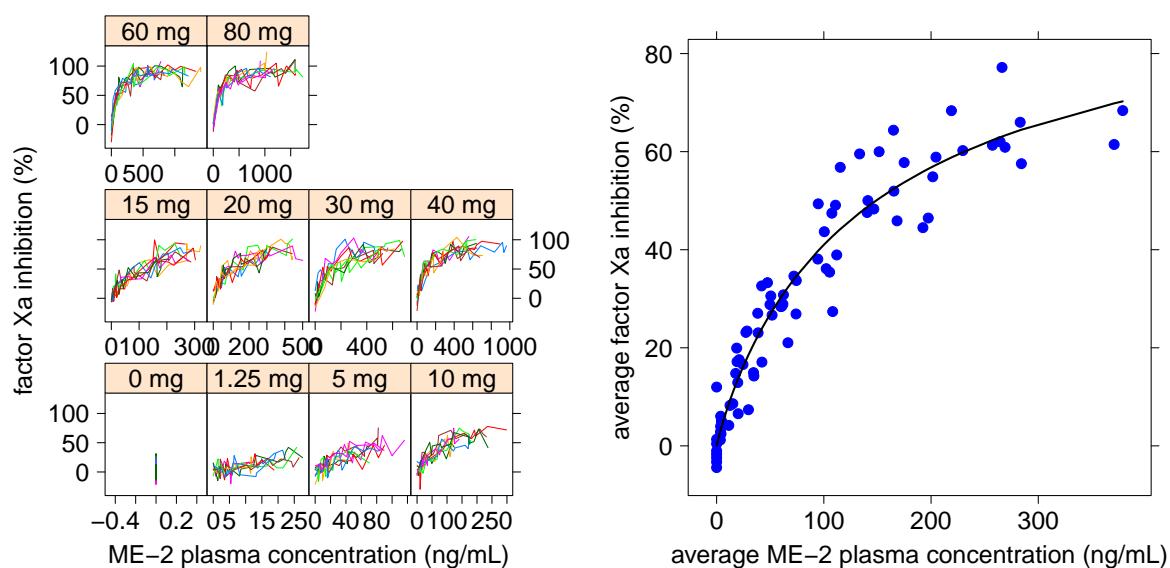
PK-PD modeling of time-averaged biomarker and PK data

- Phase 1 single dose study in healthy volunteers
 - Parallel dose-escalation design
 - 8 subjects per dose arm
 - Single doses of ME-2
 - Placebo, 1.25, 5, 10, 15, 20, 30, 40, 60 and 80 mg
 - PK: plasma concentrations of parent drug
 - Biomarker: ex vivo inhibition of factor Xa activity in plasma
 - PK and biomarker measured at 0, 0.083, 0.167, 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4, 6, 8, 12, 18 and 24 hours after dose.
- Hands-on exercise:
 - Model relationship between time-averaged factor Xa inhibition and time-averaged ME-2 plasma concentrations

EDA: PK and biomarker data



EDA: Relationship between biomarker and PK data



Model

- Sigmoid Emax model relating time-averaged % inhibition of factor Xa activity to time-averaged ME-2 plasma concentration in the i^{th} subject:

$$\bar{E}_{24,i} \sim N(\hat{E}_{24,i}, \sigma)$$

$$\hat{E}_{24,i} = \frac{E_{max} \bar{C}_{24,i}^\gamma}{EC_{50}^\gamma + \bar{C}_{24,i}^\gamma}$$

- Weakly informative prior distributions:

$$E_{max} \sim U(0, 100)$$

$$EC_{50} \sim \text{half-}N(0, 250)$$

$$\gamma \sim \text{half-}N(0, 5)$$

$$\sigma \sim \text{half-Cauchy}(0, 10)$$

See [19] for half-Cauchy rationale.

Files

- Data: data/derived/fxa.data.avg.csv
- Stan model: model/fxalInhibitAvg.stan
- R script: script/fxalInhibitAvg.R

User-defined functions

New Stan language functions may be defined in a program block called functions.

```
functions{
  real oneCpt(real time, real dose, real CL, real V){
    real k = CL / V;
    return dose * exp(-k * time) / V;
  }
}
```

- Such functions may be used just like built-in functions.
- It is also possible to define new probability distributions and random number generating functions.

User-defined function syntax

- The functions block must precede all other blocks.
- Function arguments and return types may be any Stan data type, e.g., int, int[], real, real[], vector, row_vector, matrix.
- Function arguments and return types are not specified with sizes or constraints, e.g.,
 - Two dimensional real array: y[,];
 - Vector: vector y;
- Void functions are permitted, i.e., functions with no return value.

Implementing popPKPD models

Let's focus on 2 programming tasks:

- Pharmacokinetic compartmental models
- Dosing and observation event schedules

Stan demo: ME-2 single dose PK

- Two compartment model with first order absorption describing ME-2 plasma concentration on the i^{th} occasion in the j^{th} subject as a function of time, dose and body weight:

$$\begin{aligned} \log(c_{ij}) &\sim N(\log(\hat{c}_{ij}), \sigma) \\ \hat{c}_{ij} &= f_{2cpt}(t_{ij}, D_j, CL_j, Q_j, V_{1j}, V_{2j}, k_{aj}) \\ \log(CL_j, Q_j, V_{1j}, V_{2j}, k_{aj}) &\sim N\left(\log\left(\widehat{CL}\left(\frac{bw_j}{70}\right)^{0.75}, \widehat{Q}\left(\frac{bw_j}{70}\right)^{0.75}, \widehat{V}_1\left(\frac{bw_j}{70}\right), \widehat{V}_2\left(\frac{bw_j}{70}\right), \widehat{k}_a\right), \Omega\right) \end{aligned}$$

- Weakly informative prior distributions:

$$\begin{aligned} \widehat{CL} &\sim \text{half-}N(0, 25) \quad \widehat{Q} \sim \text{half-}N(0, 50) \quad \widehat{V}_1 \sim \text{half-}N(0, 100) \\ \widehat{V}_2 &\sim \text{half-}N(0, 200) \quad \widehat{k}_a \sim \text{half-}N(0, 5) \quad \sigma \sim \text{half-Cauchy}(0, 1) \\ \Omega &= \text{diag}(\omega) P \text{ diag}(\omega) \\ \omega_i &\sim \text{half-Cauchy}(0, 1), i \in \{1, 2, 3, 4, 5\} \quad P \sim \text{LKJCorr}(1) \end{aligned}$$

Stan demo: ME-2 single dose PK Files

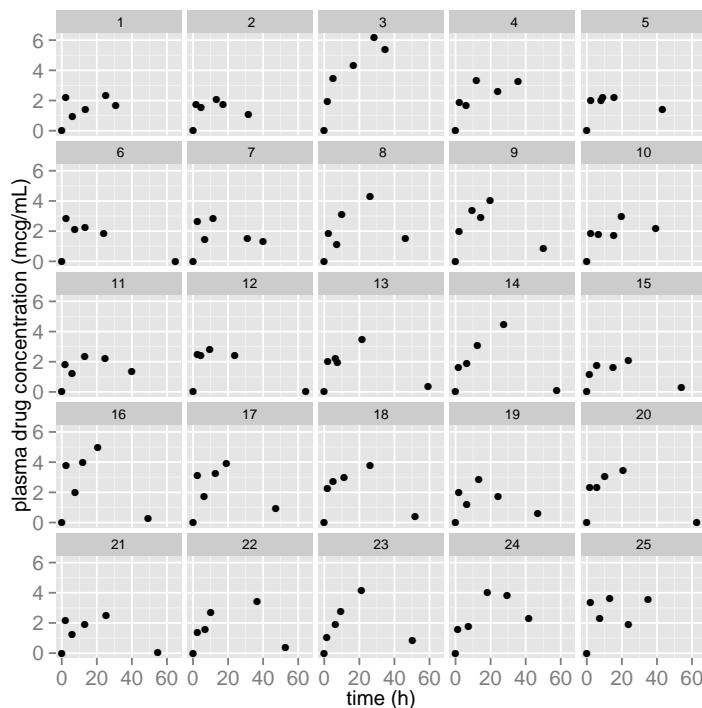
- Data: data/derived/fxa.data.csv
- Stan model: model/singleDosePK1.stan
- R script: script/singleDosePK1.R
- Version with vectorized function
 - Stan model: model/singleDosePK2.stan
 - R script: script/singleDosePK2.R

Recursive approach

Dealing with dosing and observation event schedules

- Data format: Time-ordered event records for each individual ala NONMEM
- For each event time calculate the amount in each compartment given the compartment amounts plus doses at the previous event time.
- This also allows for time-varying (piece-wise constant) parameter values.
- Same approach works for compartmental models described in terms of analytic or numerical solutions.

Stan demo: Multiple dose PK with sparse sampling



Sparsely sampled plasma drug concentrations resulting from administration of 200 mg of drug every 8 hours for 5 doses.

Stan demo: Multiple dose PK with sparse sampling

- One compartment model with first order absorption describing plasma drug concentration on the i^{th} occasion in the j^{th} subject as a function of time and dose:

$$\begin{aligned}\log(c_{ij}) &\sim N(\log(\hat{c}_{ij}), \sigma^2) \\ \hat{c}_{ij} &= f_{1cpt}(t_{ij}, D_j, \tau_j, CL_j, V_j, k_{aj}) \\ \log(CL_j, V_j, k_{aj}) &\sim N(\log(\widehat{CL}, \widehat{V}, \widehat{k}_a), \Omega)\end{aligned}$$

- Prior distributions: strongly informative for k_a ; weakly informative for the remaining parameters

$$\begin{aligned}\widehat{CL} &\sim \text{half-}N(0, 20^2) \quad \widehat{V} \sim \text{half-}N(0, 100^2) \\ \widehat{k}_a &\sim \text{lognormal}(\log(0.45), 0.2^2) \quad \sigma \sim \text{half-Cauchy}(0, 2) \\ \Omega &= \text{diag}(\omega) P \text{ diag}(\omega) \\ \omega_i &\sim \text{half-Cauchy}(0, 2), i \in \{1, 2\} \quad \omega_3 \sim \text{lognormal}(\log(0.25), 0.3^2) \\ P &\sim \text{LKJCorr}(1)\end{aligned}$$

Stan demo: Multiple dose PK with sparse sampling Files

- Data: data/derived/prob_2fixed2.csv
- Stan model: model/multiDosePK1.stan
- R script: script/multiDosePK1.R

Models with systems of ODEs

Linear ODEs: matrix exponential

- matrix_exp(A)

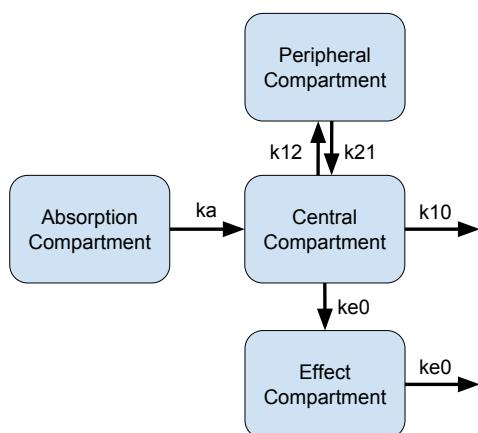
- Calculates a matrix exponential e^A where A is a square matrix.
- Uses a Padé approximation with scaling and squaring [20].
- Typical PMX use case is the solution of a linear system of ODEs with constant coefficients, e.g., an initial value problem that may be expressed:

$$\begin{aligned}x' &= Kx \\x(t_0) &= x_0\end{aligned}$$

where x is a vector and K is a square matrix. The solution may be expressed in terms of the matrix exponential:

$$x(t) = e^{(t-t_0)K} x_0$$

PKPD modeling example using the matrix exponential



- PK: Two compartment model with 1st order absorption
- PD: Response described by an Emax function of "concentration" in an effect compartment.

$$x' = Kx$$

$$x(t_0) = x_0$$

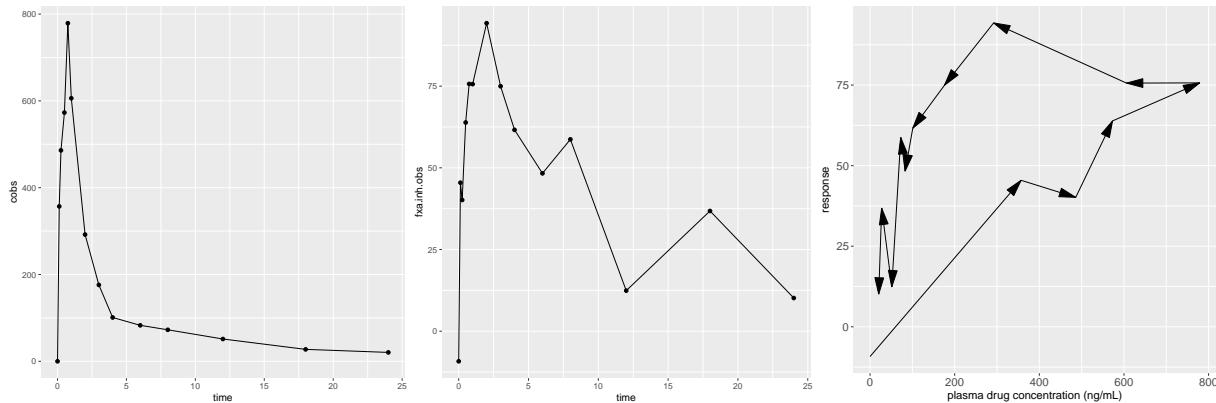
$$K = \begin{bmatrix} -k_a & 0 & 0 & 0 \\ ka & -\frac{CL+Q}{V_1} & \frac{Q}{V_2} & 0 \\ 0 & \frac{Q}{V_1} & -\frac{Q}{V_2} & 0 \\ 0 & k_{e0} & 0 & -k_{e0} \end{bmatrix}$$

$$C = \frac{x_2}{V_1}$$

$$C_e = \frac{x_4}{V_1}$$

$$E = \frac{100C_e}{EC_{50} + C_e}$$

PKPD modeling example using the matrix exponential



- Data: data/derived/phase1effcpt.csv
- Stan model for fitting: model/effCptSinglePatient.stan
- R script: script/effCptSinglePatient.R

effCptSinglePatient.stan excerpt

```

transformed parameters{
  vector<lower = 0>[nObs] cHat;
  vector<lower = 0>[nObs] respHat;
  vector<lower = 0>[nObs] ceHat;
  matrix[nObs, 4] x;
  matrix[4, 4] K;

  K = rep_matrix(0, 4, 4);

  K[1, 1] = -ka;
  K[2, 1] = ka;
  K[2, 2] = -(CL + Q) / V1;
  K[2, 3] = Q / V2;
  K[3, 2] = Q / V1;
  K[3, 3] = -Q / V2;
  K[4, 2] = ke0;
  K[4, 4] = -ke0;

  :
}

}

```

effCptSinglePatient.stan excerpt

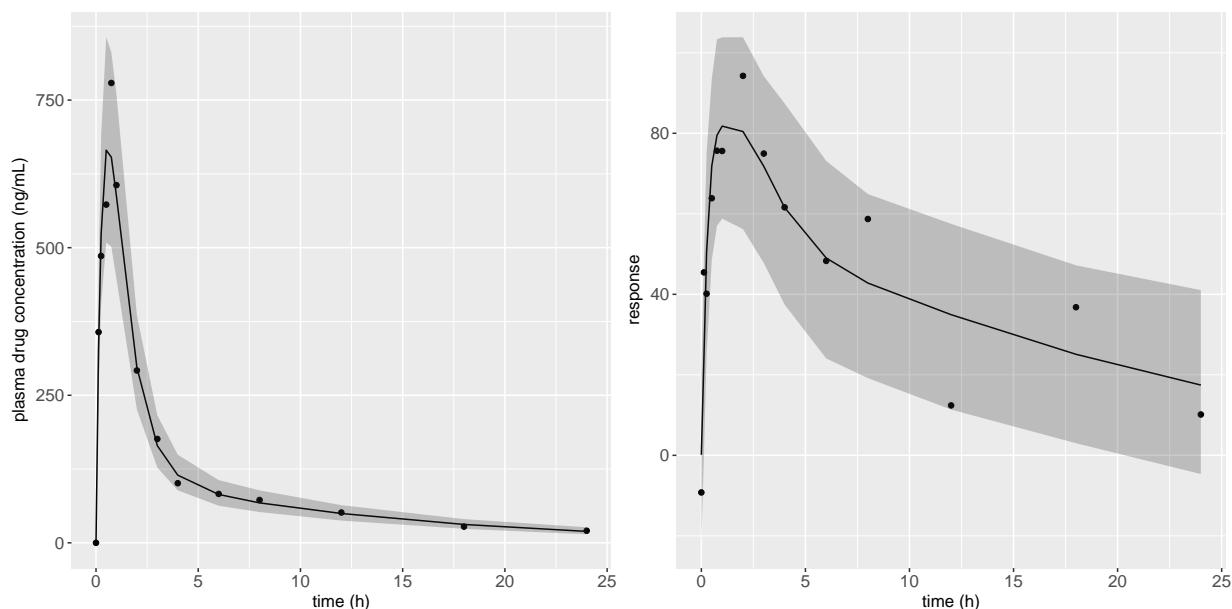
```

    :
for(i in 1:nObs)
  x[i, ] = (matrix_exp(time[i] * K) * to_vector(dose, 0, 0, 0))';

cHat = 1000 * x[,2] ./ V1;
ceHat = 1000 * x[,4] ./ V1;
respHat = 100 * ceHat ./ (EC50 + ceHat);
}

```

PKPD modeling example using the matrix exponential



General ODEs

Stan currently provides 2 functions for numerical solution of ODEs:

```
real[,] integrate_ode_rk45(function ode,
    real[] initial_state, real initial_time, real[] times,
    real[] theta, real[] x_r, int[] x_i,
    real rel_tol, real abs_tol, int max_num_steps)
```

- Runge Kutta Doprri 4th/5th order algorithm with the implementation from Boost
- Suitable for non-stiff ODEs

```
real[,] integrate_ode_bdf(function ode,
    real[] initial_state, real initial_time, real[] times,
    real[] theta, real[] x_r, int[] x_i,
    real rel_tol, real abs_tol, int max_num_steps)
```

- Backward differentiation formula (BDF) method with the implementation from SUNDIALS (CVODES)
- Designed for stiff ODEs

User-specified system of ODEs

- User specifies the system of ODEs in the function block with a function of the form:

```
real[] ode(real time,
    real[] state,
    real[] theta,
    real[] x_r,
    int[] x_i)
```

- time, the time to evaluate the ODE system
- state, the state of the ODE system at the time specified
- theta, parameter values used to evaluate the ODE system
- x_r, data values used to evaluate the ODE system
- x_i, integer data values used to evaluate the ODE system.

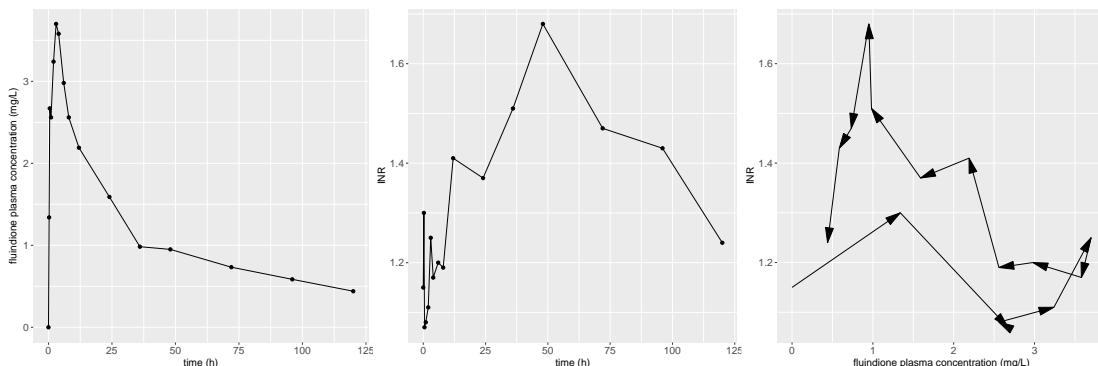
Other ODE solver arguments

```
real[,] integrate_ode_bdf(function ode,
                           real[] initial_state, real initial_time, real[] times,
                           real[] theta, real[] x_r, int[] x_i,
                           real rel_tol, real abs_tol, int max_num_steps)
```

- `initial_state`: initial state, type `real[]`,
- `initial_time`: initial time, type `int` or `real`, data only,
- `times`: solution times, type `real[]`, data only,
- `theta`: parameters, type `real[]`,
- `x_r`: real data, type `real[]`, data only,
- `x_i`: integer data, type `int[]`, data only
- `rel_tol`: relative tolerance for the ODE solver, type `real`, data only (default = 10^{-6}),
- `abs_tol`: absolute tolerance for the ODE solver, type `real`, data only (default = 10^{-6}),
- `max_num_steps`: maximum number of steps to take in the ODE solver, type `int`, data only (default = 10^6).

Example using an indirect action PKPD model

Let's analyze (simulated) plasma fluindione concentrations and INR measurements (international normalized ratio), a measure of blood coagulation following a single 20 mg dose of fluindione.



Indirect action model for fluindione effect on INR

Let's use the model reported by Verstuyft *et al* [21] based on one reported earlier by Mentré *et al* [22].

$$\begin{aligned}x_1' &= -k_a x_1 \\x_2' &= k_a x_1 - (k_{10} + k_{12}) x_2 + k_{21} x_3 \\x_3' &= k_{12} x_2 - k_{21} x_3 \\x_4' &= k_{in} (1 - E_{\text{drug}}) - k_{out} x_4\end{aligned}$$

where

$$x_4 = \frac{1}{\widehat{\text{INR}}} \quad E_{\text{drug}} = \frac{E_{\max} \widehat{c}^\gamma}{EC_{50}^\gamma + \widehat{c}^\gamma} \quad \widehat{c} = \frac{x_2}{V_1}$$

Indirect action model for fluindione effect on INR

To model the non-zero initial value of INR, use the change of variable $\delta_4 = x_4 - x_4(0)$, so the fourth differential equation becomes:

$$\delta_4' = k_{in} (1 - E_{\text{drug}}) - k_{out} \left(\delta_4 + \frac{1}{\widehat{\text{INR}}(0)} \right)$$

Model residual variability as log-normal

$$\log(c_i) \sim N(\widehat{c}_i, \sigma_{PK}) \quad \log(\text{INR}_i) \sim N(\widehat{\text{INR}}_i, \sigma_{PK}) \quad \widehat{\text{INR}}_i = \frac{1}{\delta_{4i} + \frac{1}{\widehat{\text{INR}}(0)}}$$

Weakly-moderately informative priors

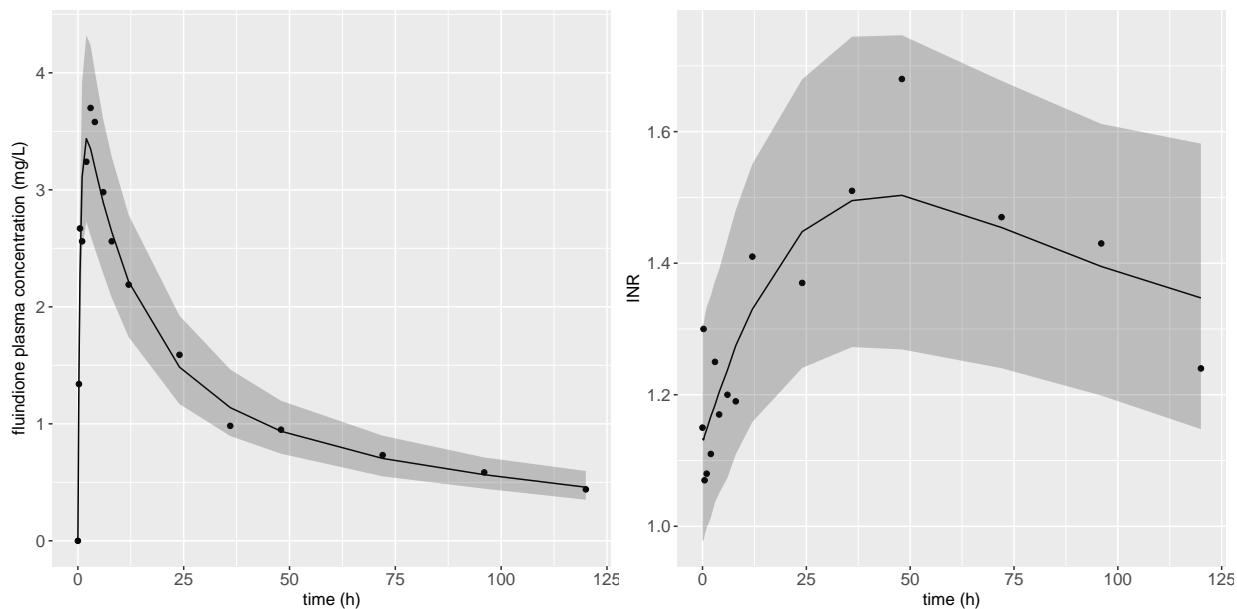
$$\begin{array}{lll} CL \sim \text{half-}N(0, 0.5) & Q \sim \text{half-}N(0, 0.5) & V_1 \sim \text{half-}N(0, 10) \\ V_2 \sim \text{half-}N(0, 10) & k_a \sim \text{half-}N(0, 5) & E_{\max} \sim \text{beta}(3, 1) \\ EC_{50} \sim \text{half-}N(0, 10) & \text{INR}(0) \sim \text{half-}N(0, 2) & k_{out} \sim \text{half-}N(0, 1) \\ \gamma \sim \text{half-}N(0, 5) & \sigma_{PK} \sim \text{half-Cauchy}(0, 1) & \sigma_{PD} \sim \text{half-Cauchy}(0, 1) \end{array}$$

Files

- Data

- Fluindione plasma concentration and INR data:
data/derived/fluindione.csv
- Stan model: model/fluindione1.stan
- R script: script/fluindione1.R

Indirect action model for fluindione effect on INR



Torsten: Library of PKPD functions for Stan

Torsten: Library of PKPD functions for Stan

A set of Stan functions that provides functionality similar to NONMEM's PREDPP library

Core functions in the current version:

- One & two compartment PK models with 1st order absorption (analytic solutions)
 - PKModelOneCpt(time, amt, rate, ii, evid, cmt, addl, ss, theta, F, tlag)
 - PKModelTwoCpt(time, amt, rate, ii, evid, cmt, addl, ss, theta, F, tlag)
- Linear compartment model specified as a rate constant matrix
 - linOdeModel(time, amt, rate, ii, evid, cmt, addl, ss, Kmatrix, F, tlag)
- General compartmental model specified as a system of 1st order ODEs
 - generalOdeModel_rk45(odeFunction, nCmt, time, amt, rate, ii, evid, cmt, addl, ss, theta, F, tlag, rel_tol, abs_tol, max_num_steps)
 - generalOdeModel_bdf(odeFunction, nCmt, time, amt, rate, ii, evid, cmt, addl, ss, theta, F, tlag, rel_tol, abs_tol, max_num_steps)

Torsten Teorell

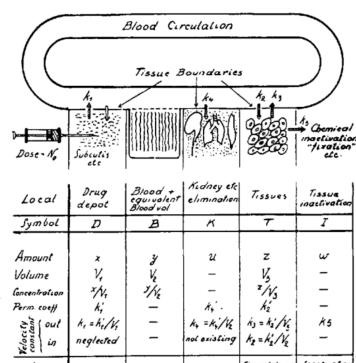


FIG. 1
Schems of the Change of Drug Distribution used in this paper.
Instead the injection pictured in the figure, the administration of the drug depot can be made per os, per rectum, by inhalation, etc.

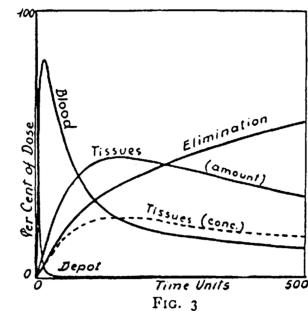


FIG. 3

Typical Case of Extravascular Administration in the absence of tissue inactivation.

($k_1 = 0.2$; $k_2 = 0.01$; $k_3 = 0.005$; i.e. "blood" volume/tissue volume is 1:2; $k_4 = 0.005$; $k_5 = 0$).

T. Teorell. Kinetics of distribution of substances administered to the body. I. The extravascular modes of administration. Arch Int Pharmacodyn et Ther 57: 205-225, 1937.

Torsten PMX functions

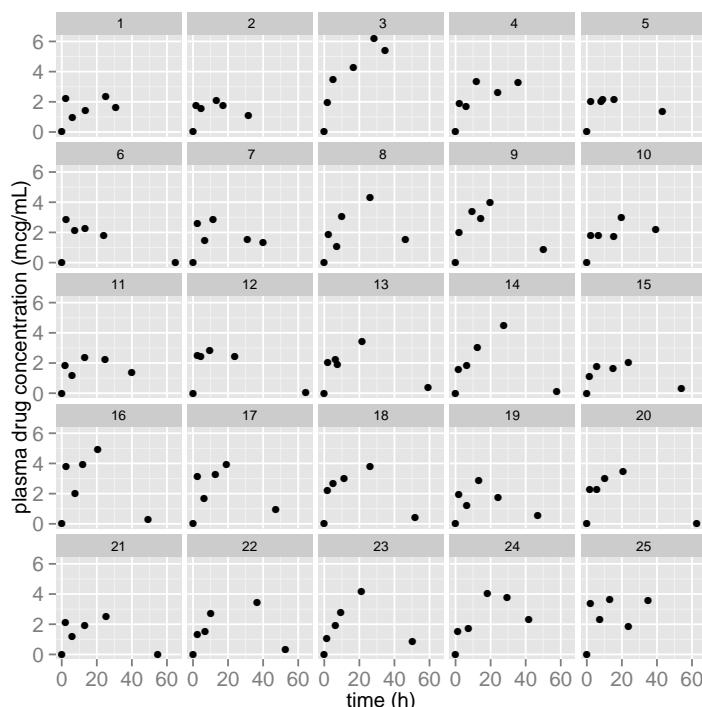
- Uses NONMEM/PREDPP conventions for data specification and event handling
- Data format: Time-ordered event records for each individual à la NONMEM
- Implemented NONMEM data types: TIME, CMT, AMT, RATE, EVID, II, ADDL, SS
- Recursive calculation: For each event time calculate the amount in each compartment given the compartment amounts plus doses at the previous event time.
- Allows for time-varying (piece-wise constant) parameter values.

Torsten: Prototype library of PKPD functions for Stan

- Current version of Torsten is available at:
<https://github.com/metrumresearchgroup/Torsten>
- Includes installation instructions for use with CmdStan and RStan.
- Documentation: https://github.com/metrumresearchgroup/Torsten/blob/master/docs/torsten_manual.pdf

Hands-on session 1

Hands-on session 1: PopPK using a 1 compartment model with 1st order absorption



Let's revisit the previous popPK example except we will now use a Torsten function.

Sparsely sampled plasma drug concentrations resulting from administration of 200 mg of drug every 8 hours for 5 doses.

Hands-on session 1: Multiple dose PK with sparse sampling

- One compartment model with first order absorption describing plasma drug concentration on the i^{th} occasion in the j^{th} subject as a function of time and dose:

$$\begin{aligned}\log(c_{ij}) &\sim N(\log(\hat{c}_{ij}), \sigma^2) \\ \hat{c}_{ij} &= f_{1cpt}(t_{ij}, D_j, \tau_j, CL_j, V_j, k_{aj}) \\ \log(CL_j, V_j, k_{aj}) &\sim N(\log(\widehat{CL}, \widehat{V}, \widehat{k}_a), \Omega)\end{aligned}$$

- Prior distributions: strongly informative for k_a ; weakly informative for the remaining parameters

$$\begin{aligned}\widehat{CL} &\sim \text{half-}N(0, 20^2) \quad \widehat{V} \sim \text{half-}N(0, 100^2) \\ \widehat{k}_a &\sim \text{lognormal}(\log(0.45), 0.2^2) \quad \sigma \sim \text{half-Cauchy}(0, 2) \\ \Omega &= \text{diag}(\omega) P \text{ diag}(\omega) \\ \omega_i &\sim \text{half-Cauchy}(0, 2), i \in \{1, 2\} \quad \omega_3 \sim \text{lognormal}(\log(0.25), 0.3^2) \\ P &\sim \text{LKJCorr}(1)\end{aligned}$$

Hands-on session 1: Multiple dose PK with sparse sampling Files

- Data: data/derived/prob_2fixed2.csv
- Stan model: model/multiDosePK1Torsten.stan
- R script: script/multiDosePK1Torsten.R
- Version illustrating simulation of additional concentration values
 - Stan model: model/multiDosePK1Torsten2.stan
 - R script: script/multiDosePK1Torsten2.R

Review of HMC/NUTS

Hamiltonian Monte Carlo (HMC) simulation

Physical analogy to motivate HMC

- In classical mechanics the Hamiltonian equations describe the evolution of a system over time.
- The state of the system is described in terms of kinetic energy as a function of momentum ($\text{mass} \times \text{velocity}$) and potential energy as a function of position.
- For the analogy equate the model parameters θ to position and equate a set of auxiliary parameters ρ to momentum.
- Now define a Hamiltonian in terms of the joint posterior distribution of θ and ρ :

$$\begin{aligned} H(\theta, \rho) &= -\log(p(\theta, \rho|y)) = -\log(p(\theta|y)p(\rho|\theta, y)) \\ &= -\log(p(\theta|y)) - \log(p(\rho|\theta, y)) \\ &= V(\theta) + T(\rho|\theta) \end{aligned}$$

$$V(\theta) = -\log(p(\theta|y)) = \text{potential energy}$$

$$T(\rho|\theta) = -\log(p(\rho|\theta, y)) = \text{kinetic energy}$$

Hamiltonian Monte Carlo (HMC) simulation

- θ is what we really care about.
- ρ allows the use of Hamiltonian mechanics to more efficiently move through the relevant parts of the parameter space.
- Usually the distribution of ρ is chosen to be independent of θ , e.g., $p(\rho|\theta) = p(\rho) = N(0, \Sigma)$.
- Suppose we place a frictionless particle on the potential energy surface ($-\log(p(\theta|y))$) at some position θ^{t-1} .
 - We give it a shove that imparts a momentum ρ^{t-1} to that particle at time $t - 1$.
 - The particle moves over that surface according to Hamiltonian dynamics.
 - Now stop the particle at time t and measure its position θ^t .
 - Now randomly sample a new momentum from $p(\rho)$ and give the particle another shove, and so on...

Hamiltonian Monte Carlo (HMC) simulation

- Though the initial momentum at each step is random, the subsequent path will favor regions of lower potential energy (higher probability density).
- The set of sampled positions are distributed according to the target posterior density.
- In practice the Hamiltonian equations are solved numerically. As a result some error is introduced in the estimated path.
- A Metropolis step is used to assure that the position samples converge in distribution to the target distribution.

The HMC algorithm

Repeat the following steps:

- ① Sample $\rho^{t-1} \sim N(0, \Sigma)$
- ② Simultaneously update θ and ρ by numerically solving the Hamiltonian equations using the leapfrog method to generate a proposal θ^* for θ^t .
- ③ Apply a Metropolis step to decide whether to accept or reject the proposal θ^* as θ^t .

The leapfrog method

Using the starting values θ^{t-1} and ρ^{t-1} the leapfrog algorithm alternates half-step updates of ρ with full step updates of θ :

$$\begin{aligned}\rho &\leftarrow \rho - \frac{\epsilon}{2} \frac{\partial V}{\partial \theta} = \rho + \frac{\epsilon}{2} \frac{d \log(p(\theta|Y))}{d\theta} \\ \theta &\leftarrow \theta + \epsilon \Sigma \rho \\ \rho &\leftarrow \rho - \frac{\epsilon}{2} \frac{\partial V}{\partial \theta} = \rho + \frac{\epsilon}{2} \frac{d \log(p(\theta|Y))}{d\theta}\end{aligned}$$

For each HMC iteration repeat this L times to yield the proposal values θ^* and ρ^* .

The Metropolis step

- Compute the ratio:

$$\begin{aligned} r &= \exp \left(H \left(\theta^{t-1}, \rho^{t-1} \right) - H \left(\theta^*, \rho^* \right) \right) \\ &= \frac{p(\theta^*|y) p(\rho^*)}{p(\theta^{t-1}|y) p(\rho^{t-1})} \end{aligned}$$

- Accept/reject step:

$$\theta^t = \begin{cases} \theta^*, & \text{with probability } \min(r, 1) \\ \theta^{t-1}, & \text{otherwise} \end{cases}$$

- Since ρ is sampled independently of θ and previous values of ρ , we just discard ρ^* and sample a new value for the next HMC iteration.

HMC algorithm parameters

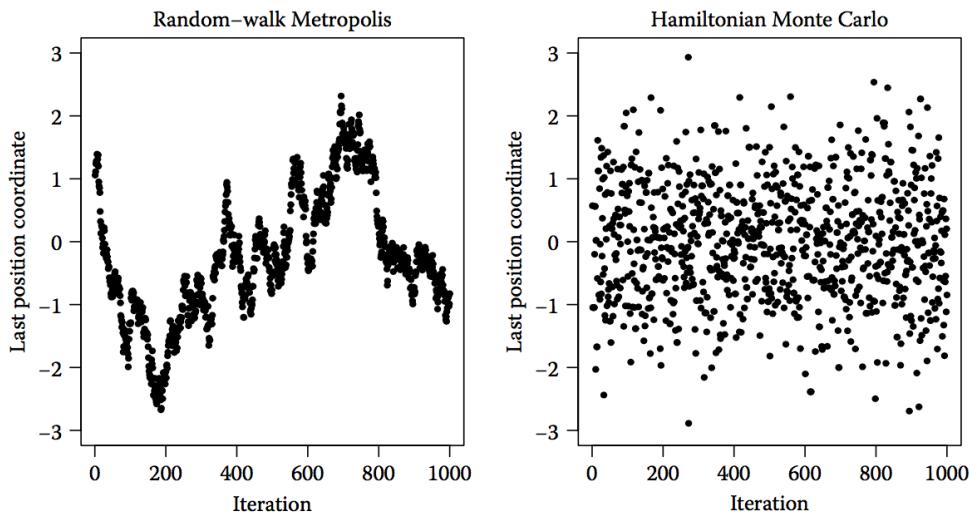
Parameters that must be set: discretization time ϵ , number of leapfrog steps L and mass matrix Σ^{-1} .

Sampling efficiency is very sensitive to those parameters:

- ϵ too large \rightarrow too many proposals rejected
- ϵ too small \rightarrow long simulation times
- L too large \rightarrow too much work for each iteration
- L too small \rightarrow devolves to a random walk
- If Σ^{-1} is poorly tuned to the problem, ϵ needs to be decreased and L increased to maintain precision and efficiency.

Stan automatically optimizes those parameters using the NUTS (no U-turn sampling) algorithm [23].

HMC performance

**FIGURE 5.6**

Values for the variable with largest standard deviation for the 100-dimensional example, from a random-walk Metropolis run and an HMC run with $L = 150$. To match computation time, 150 updates were counted as one iteration for random-walk Metropolis.

from RM Neal. MCMC Using Hamiltonian Dynamics (2011) [24]

HMC performance

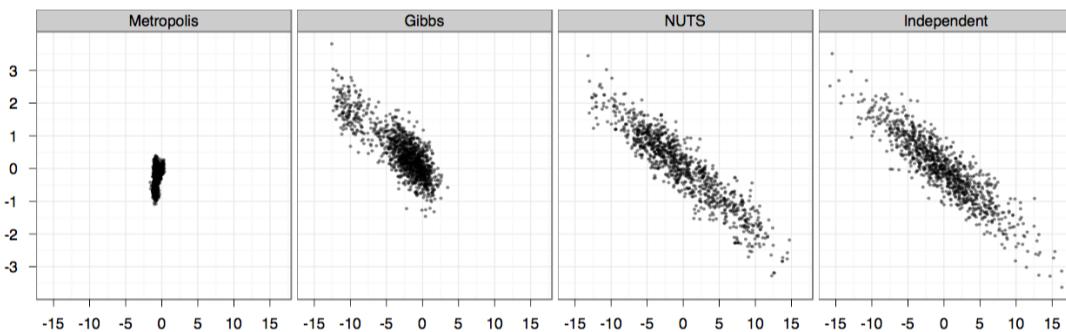


Figure 7: Samples generated by random-walk Metropolis, Gibbs sampling, and NUTS. The plots compare 1,000 independent draws from a highly correlated 250-dimensional distribution (right) with 1,000,000 samples (thinned to 1,000 samples for display) generated by random-walk Metropolis (left), 1,000,000 samples (thinned to 1,000 samples for display) generated by Gibbs sampling (second from left), and 1,000 samples generated by NUTS (second from right). Only the first two dimensions are shown here.

from MD Hoffman and A Gelman. The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo (2014) [23]

HMC issues/limitations

- Requires calculation of the gradient $\frac{d \log(p(\theta|Y))}{d\theta}$
- Suitable for sampling of continuous parameters only
 - Cannot sample discrete parameters
 - Discrete data is OK as long as the likelihood depends only on continuous parameters.
 - Models with discrete parameters, e.g., finite mixture models, can often be implemented by marginalizing out the discrete parameters.

Stan implementation of HMC

Automatic parameter tuning:

- Optimizes discretization time (ϵ) to match a target acceptance rate (default = 0.8) during warmup.
- Estimates the mass matrix (Σ^{-1}) based on warmup samples.
- Dynamically adapts the number of leapfrog steps throughout sampling using the NUTS algorithm.
 - Seeks to maximize the distance traversed before the path doubles back on itself (U-turns).
- Users can override any of these automatic settings by explicitly setting the parameters.

See reference [25] for details.

Optimizing Stan code

Optimizing Stan code

- Non-optimal Stan code can result in computational inefficiency or sampling problems such as:
 - Poor convergence and mixing
 - Divergent transitions
- We briefly discuss some remedies in this section.
- See Chapter 28 of the *Stan Modeling Language User's Guide and Reference Manual* [26] for more details and approaches.

Optimizing Stan code

“A well-behaved Stan program will have low variance between runs with different random initializations and differently seeded random number generators. But sometimes an algorithm can get stuck in one part of the posterior, typically due to high curvature. Such sticking almost always indicates the need to reparameterize the model. . . . This problem with getting stuck can often be overcome by lowering the initial step size to avoid getting stuck during adaptation and increasing the target acceptance rate in order to target a lower step size.”

Stan Modeling Language User’s Guide and Reference Manual [26, pp 337–338]

Improving computational efficiency and sampling performance

The main strategies are:

- Vectorization
- Reparameterization
- Adjusting HMC tuning parameters
- Weakly informative priors to regularize fitting of hierarchical models

Vectorization

- Vectorized calculations can markedly speed up calculations primarily by increasing the efficiency of the gradient calculations in Stan.
- Particularly true for probability functions.

Reparameterization

Dealing with those pesky divergent transitions

- A divergent transition occurs when the leapfrog integrator fails to follow the true Hamiltonian trajectory with sufficient accuracy—sometimes catastrophically.
 - Increase the target acceptance rate (`adapt_delta`) and/or decrease initial step size (`stepsize`).
 - Switch from centered to non-centered parameterization (or vice versa) of random effects in hierarchical models.
 - Other reparameterizations that reduce posterior curvature and correlation.
- Let's explore the first two approaches with a couple examples.

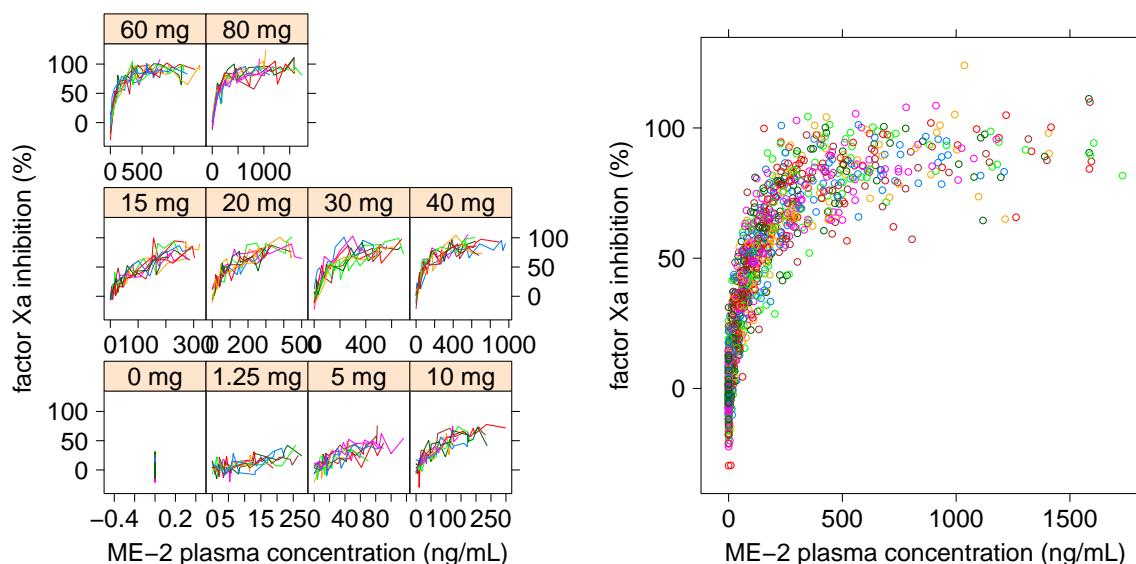
Non-centered parameterization: Univariate case

Population PK-PD modeling of time-matched biomarker and PK data

Now we analyze the time course data from the same Phase I study we analyzed in the Stan workflow example:

- Phase 1 single dose study in healthy volunteers
 - Parallel dose-escalation design
 - 8 subjects per dose arm
 - Single doses of ME-2
 - Placebo, 1.25, 5, 10, 15, 20, 30, 40, 60 and 80 mg
 - PK: plasma concentrations of parent drug
 - Biomarker: ex vivo inhibition of factor Xa activity in plasma
 - PK and biomarker measured at 0, 0.083, 0.167, 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4, 6, 8, 12, 18 and 24 hours after dose.
- Hands-on exercise:
 - Apply a direct action PK/PD model to the time-matched factor Xa inhibition and ME-2 plasma concentrations.

EDA: Relationship between biomarker and PK data



Model

- Sigmoid Emax model relating % inhibition of factor Xa activity to ME-2 plasma concentration on the i^{th} occasion in the j^{th} subject:

$$\begin{aligned} E_{ij} &\sim N(\hat{E}_{ij}, \sigma) \\ \hat{E}_{ij} &= \frac{E_{max} c_{ij}^\gamma}{EC_{50,j}^\gamma + c_{ij}^\gamma} \\ \log(EC_{50,j}) &\sim N(\log(\widehat{EC}_{50}), \omega_{EC_{50}}) \end{aligned}$$

- Some possible weakly informative prior distributions:

$$\begin{aligned} E_{max} &\sim U(0, 100) \quad \widehat{EC}_{50} \sim \text{half-}N(0, 250) \\ \gamma &\sim \text{half-}N(0, 5) \\ \omega_{EC_{50}} &\sim \text{half-Cauchy}(0, 1) \quad \sigma \sim \text{half-Cauchy}(0, 10) \end{aligned}$$

Files

- Data: data/derived/fxa.data.csv
- Stan model: model/fxalInhibit1.stan
- R script: script/fxalInhibit1.R
- Non-centered parametrization to improve sampling efficiency and prevent divergent transitions
 - Stan model: model/fxalInhibit1Ncp.stan
 - R script: script/fxalInhibit1Ncp.R

Non-centered parameterization: Multivariate case

- Use the Cholesky decomposition of the covariance matrix to generate the multivariate normal random effects from standard normal random variables.

Files

- Data: data/derived/prob_2fixed2.csv
- Centered parametrization
 - Stan model: model/multiDosePK1Torsten2.stan
 - R script: script/multiDosePK1Torsten2.R
- Non-centered parametrization
 - Stan model: model/multiDosePK1Ncp.stan
 - R script: script/multiDosePK1Ncp.R

Brief discussion on prior distributions

- Think of prior distributions as part of the model.
- Priors should be chosen and subjected to scrutiny much like other model components.
- Model checking should ideally include sensitivity analysis of the priors.
- Choice of priors is most critical with sparse or limited data.

See <https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations>

What is the function of a prior distribution

- Represent prior knowledge
- Regularization to facilitate computation
 - Typically weakly-moderately informative
 - E.g., Cauchy with most of its mass in a plausible range, but heavy tails allow for diagnosis of prior-posterior discrepancies.

What does it mean to be informative, uninformative or weakly informative?

Not well defined, but here's an attempt at some loose definitions:

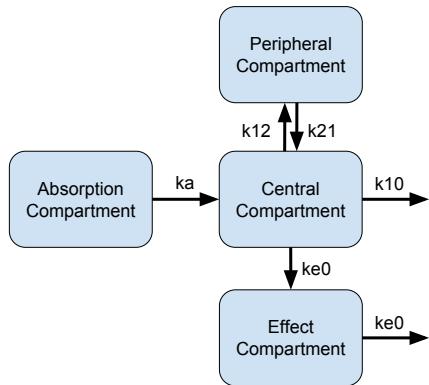
- Weakly informative prior: A prior that rules out unreasonable parameter values but is not so strong as to rule out values that might make sense
- Informative prior: A prior that purposely represents information intended to influence the posterior distribution
 - To capture prior knowledge
 - To challenge the analysis with competing points of view, e.g., use of pessimistic or optimistic priors.
- Uninformative prior: Ostensibly a prior that represents no information and therefore “let’s the data tell the story.”
 - E.g., a constant over the entire real line—an improper prior

Beware: That “uninformative” prior might not be!

- Suppose you use an improper prior for a standard deviation—a constant over the positive real line.
- That means all positive values are equally likely. Sounds like a reasonable definition of uninformative doesn't it?
- But that means that the prior assigns infinitely more probability to the set of values greater than any fixed value you care to choose.
- This will tend to bias the posterior to high values.
- Bottom line: A uniform distribution does not automatically confer uninformativeness.

Hands-on session 2: PKPD using a model based on a linear system of ODEs

Let's apply Torsten's linOdeModel function to the previous matrix exponential example.



- PK: Two compartment model with 1st order absorption
- PD: Response described by an Emax function of “concentration” in an effect compartment.

$$\dot{x} = Kx$$

$$x(t_0) = x_0$$

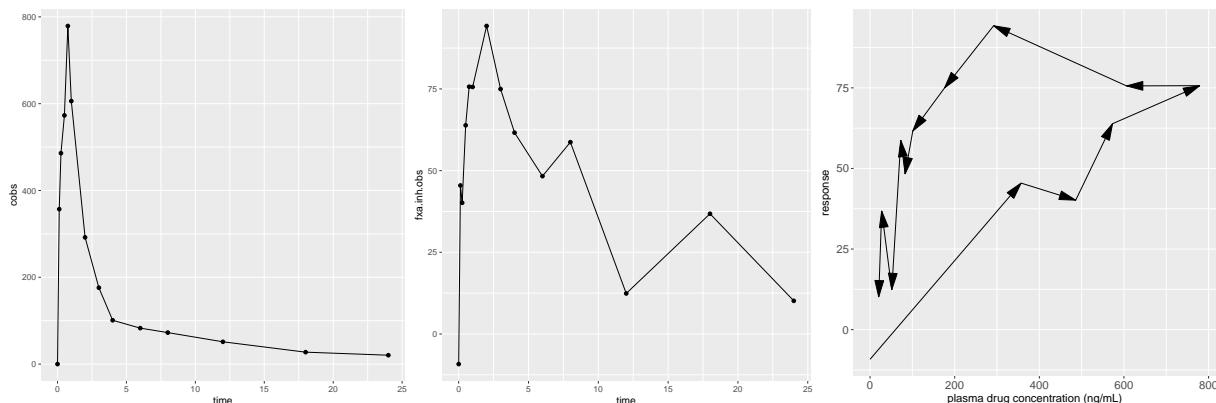
$$K = \begin{bmatrix} -k_a & 0 & 0 & 0 \\ ka & -\frac{CL+Q}{V_1} & \frac{Q}{V_2} & 0 \\ 0 & \frac{Q}{V_1} & -\frac{Q}{V_2} & 0 \\ 0 & k_{e0} & 0 & -k_{e0} \end{bmatrix}$$

$$c = \frac{x_2}{V_1}$$

$$c_e = \frac{x_4}{V_1}$$

$$E = \frac{100c_e}{EC_{50} + c_e}$$

Hands-on session 2: PKPD using a model based on a linear system of ODEs

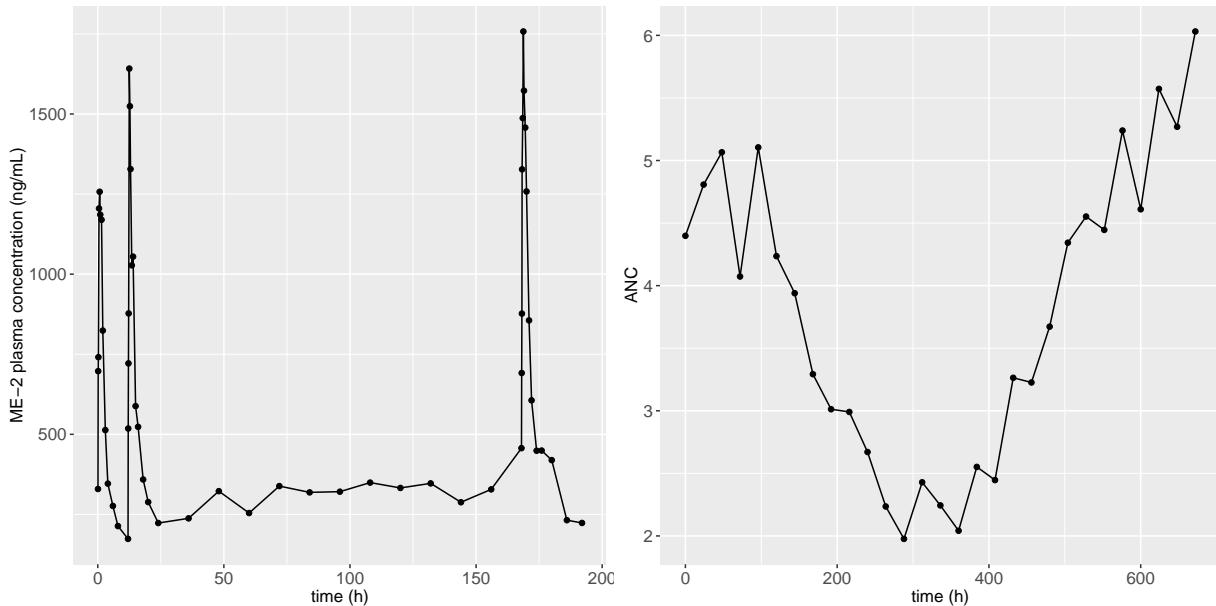


- Data: data/derived/phase1effcpt.csv
- Stan model for fitting: model/effCptSinglePatientTorsten.stan
- R script: script/effCptSinglePatientTorsten.R

Hands-on session 3

PKPD model based on nonlinear ODEs

PKPD modeling of ME-2 induced neutropenia in a single patient



Hands-on session 3

Friberg-Karlsson semi-mechanistic model for drug-induced myelosuppression

- Friberg-Karlsson semi-mechanistic model for drug-induced myelosuppression [27, 28, 29, 30, 31, 14]

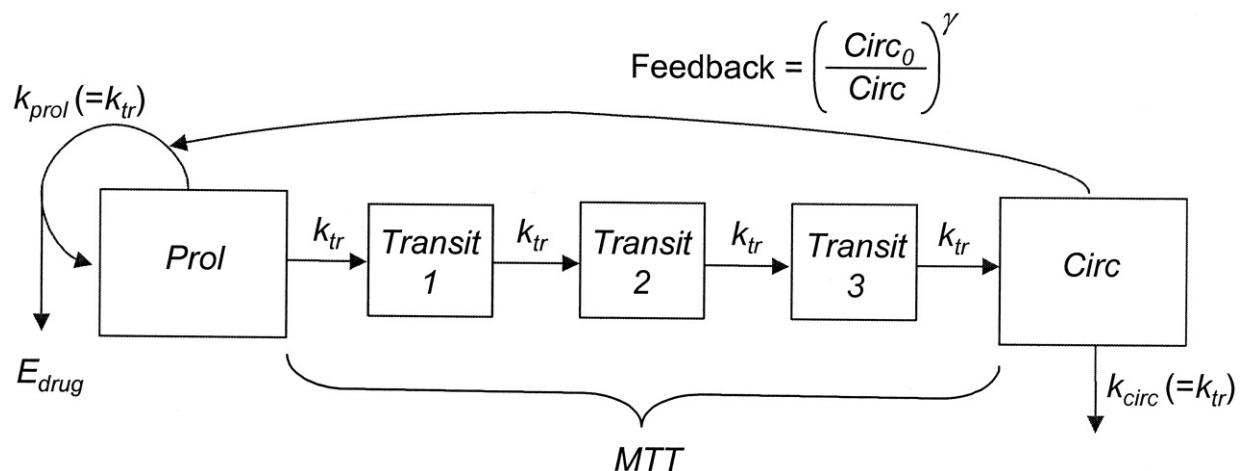


Figure 2 of reference [27]

Hands-on session 3

Friberg-Karlsson semi-mechanistic model for drug-induced myelosuppression

$$\begin{aligned}
 \frac{dProl}{dt} &= k_{prol} Prol (1 - E_{drug}) \left(\frac{Circ_0}{Circ} \right)^\gamma - k_{tr} Prol \\
 \frac{dTransit1}{dt} &= k_{tr} Prol - k_{tr} Transit1 \\
 \frac{dTransit2}{dt} &= k_{tr} Transit1 - k_{tr} Transit2 \\
 \frac{dTransit3}{dt} &= k_{tr} Transit2 - k_{tr} Transit3 \\
 \frac{dCirc}{dt} &= k_{tr} Transit3 - k_{circ} Circ
 \end{aligned}$$

\hat{c} \equiv plasma drug concentration
 $Circ$ \equiv absolute neutrophil count (ANC)

$$\begin{aligned}
 E_{drug} &= \alpha \hat{c} \\
 k_{prol} &= k_{circ} = k_{tr} \\
 MTT &= \frac{n + 1}{k_{tr}}
 \end{aligned}$$

Parameters in red are *system* parameters, i.e., drug-independent.

Hands-on session 3

Model

$$\log(c_i) \sim N(\log(\hat{c}_i), \sigma_{PK}) \quad \log(ANC_i) \sim N(\log(Circ_i), \sigma_{PD})$$

$$\hat{c}_i = \frac{x_{2i}}{V_1} \quad Circ_i = x_{8i} + Circ_0$$

$$x'_1 = -k_a x_1$$

$$x'_2 = k_a x_1 - \frac{CL + Q}{V_1} x_2 + \frac{Q}{V_2} x_3$$

$$x'_3 = \frac{CL + Q}{V_1} x_2 - \frac{Q}{V_2} x_3$$

$$x'_4 = k_{tr} Prol \left((1 - E_{drug}) \left(\frac{Circ_0}{Circ} \right)^\gamma - 1 \right)$$

$$x'_5 = k_{tr} (Prol - Transit_1)$$

$$x'_6 = k_{tr} (Transit_1 - Transit_2)$$

$$x'_7 = k_{tr} (Transit_2 - Transit_3)$$

$$x'_8 = k_{tr} (Transit_3 - Circ)$$

where

$$x_4 = Prol - Circ_0 \quad x_5 = Transit_1 - Circ_0 \quad x_6 = Transit_2 - Circ_0$$

$$x_7 = Transit_3 - Circ_0 \quad x_8 = Circ - Circ_0 \quad k_{tr} = \frac{4}{MTT}$$

Hands-on session 3

Model

- Mix of moderately to strongly informative prior distributions:

$$\begin{aligned} CL &\sim LN(\log(10), 0.5) & Q &\sim LN(\log(15), 0.5) & V_1 &\sim LN(\log(35), 0.5) \\ V_2 &\sim LN(\log(105), 0.5) & k_a &\sim LN(\log(2), 0.5) \\ Circ_0 &\sim LN(\log(5), 0.2) & MTT &\sim LN(\log(125), 0.2) \\ \gamma &\sim LN(\log(0.7), 0.2) & \alpha &\sim LN\left(\log(3 \times 10^{-4}), 1\right) \\ \sigma_{PK} &\sim \text{half-Cauchy}(0, 1) & \sigma_{PD} &\sim \text{half-Cauchy}(0, 1) \end{aligned}$$

Hands-on session 3

Friberg-Karlsson semi-mechanistic model for drug-induced myelosuppression

- Stan model for simulation:
model/neutropeniaSinglePatient1Sim.stan
- Stan model for fitting: model/neutropeniaSinglePatient1.stan
- R script: script/neutropeniaSinglePatient1.R
- The example also illustrates how to simulate data using Stan.
- Model implemented using our new mixed model functions
 - Stan model for fitting: model/neutropeniaSinglePatientMix1.stan
 - R script: script/neutropeniaSinglePatientMix1.R

Dealing with censored data in Stan

- The Stan language strongly separates the roles of data and parameters (random variables).
- This has consequences for handling of censored data.
 - Observed data are known values and thus true data.
 - Censored data are unknown values and thus random variables.
- The user must explicitly separate and declare observed and censored data:
 - Observed data in the data block.
 - Censored data in the parameters block.
 - Only need to declare censored data if you want to simulate predicted values.
 - Information such as censoring bounds and whether the data is censored is data that should be declared in the data block.
- And also specify different likelihoods for observed and censored data.

A little aside on Stan machinery

A Stan language model is essentially a set of instructions for Stan to calculate the log probability given the data and a set of parameter values.

- For (conditionally) independent data the total log probability is the sum of the individual log probabilities.
- In Stan

```
y ~ normal(mu, sigma);
```

is equivalent to

```
target += normal_lpdf(y | mu, sigma);
```

- More generally you can use “target +=” to specify a user-defined log probability when none of the built-in distributions are appropriate.

Likelihoods for censored normally distributed data

Let's use “target +=” to specify the log probability function for censored data.

- Right censored data

- Observation whose value exceeds some upper bound U .
- The appropriate likelihood (probability) function is $\Pr(y > U)$ or 1 minus the normal CDF evaluated at U in the normal case.

```
target += normal_lccdf(U | mu, sigma);
```

- Left censored data

- Observation whose value is less than some lower bound L , e.g., BQL data.
- The appropriate likelihood (probability) function is $\Pr(y \leq L)$ or the normal CDF evaluated at L in the normal case.

```
target += normal_lcdf(L | mu, sigma);
```

Stan demo: PK modeling with BQL data

Let's introduce some BQL data into the hands-on session 1 example.
Assume LOQ = 0.5 $\mu\text{g/mL}$.

- Data: data/derived/prob_2fixed2.csv
- Stan model: model/multiDosePK2Ncp.stan
- R script: script/multiDosePK2Ncp.R

User-defined probability functions and likelihoods

- User-defined functions are not limited to deterministic calculations.
- You can write your own functions for probability densities (or masses) and random number generation.
- You can also specify any likelihood via target += statements.
 - Demonstrated in hands-on session 4

User-defined log probability density function

- Recall that Stan probability functions actually calculate log probability and use them to increment the overall log probability for a model.
- A user-defined probability function must also calculate log probability.
- The function must end with the suffix _lpdf.
- It may be used in sampling or target += statements, e.g., for a function defined as real myfun_lpdf(real y, real theta1, real theta2) you can write:

```
y ~ myfun(theta1, theta2);
      or
target += myfun_lpdf(y | theta1, theta2);
```

Example: User-defined truncated normal probability function

```
functions{
  real normalTrunc_lpdf(real y, real mu, real sigma, real lower,
    real upper){
    real result;

    if(y >= lower && y <= upper){
      result = normal_lpdf(y | mu, sigma) -
        log(normal_cdf(upper, mu, sigma) -
          normal_cdf(lower, mu, sigma));
    }else{
      result = negative_infinity();
    }
    return result;
  }
}
```

Example: Model using user-defined truncated normal probability function

- Data: data/derived/fxa.data.avg.csv
- Stan model: model/fxaInhibitAvgTrunc.stan
- R script: script/fxaInhibitAvgTrunc.R

Hands-on session 4

Parametric time-to-event model

- Let's analyze the time to the first grade 2+ peripheral neuropathy (PN) event in patients treated with an antibody-drug conjugate (ADC) delivering monomethyl auristatin E (MMAE).
- We will simulate and analyze data using a simplified version of the model reported in reference [32]

Citation: CPT Pharmacometrics Syst Pharmacol. (2017) 00, 00; doi:10.1002/psp4.12192
 © 2017 ASCPT All rights reserved

ORIGINAL ARTICLE

Time-to-Event Analysis of Polatuzumab Vedotin-Induced Peripheral Neuropathy to Assist in the Comparison of Clinical Dosing Regimens

D Lu^{1*}, WR Gillespie², S Girish¹, P Agarwal¹, C Li¹, J Hirata¹, Y-W Chu¹, M Kagedal¹, L Leon¹, V Maiya¹ and JY Jin¹

©2018

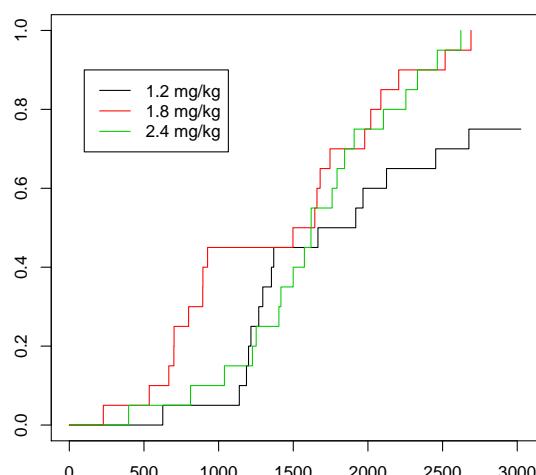
Stan, RStan and Torsten for PMX

29 May 2018

95 / 175

Hands-on session 4

Time to first PN grade 2+ during fauxlatuzumab vedotin treatment



- 3 treatment arms:
fauxlatuzumab vedotin 1.2, 1.8 and 2.4 mg/kg iv boluses q3w x 6 doses
- 20 patients per treatment arm

Hands-on session 4

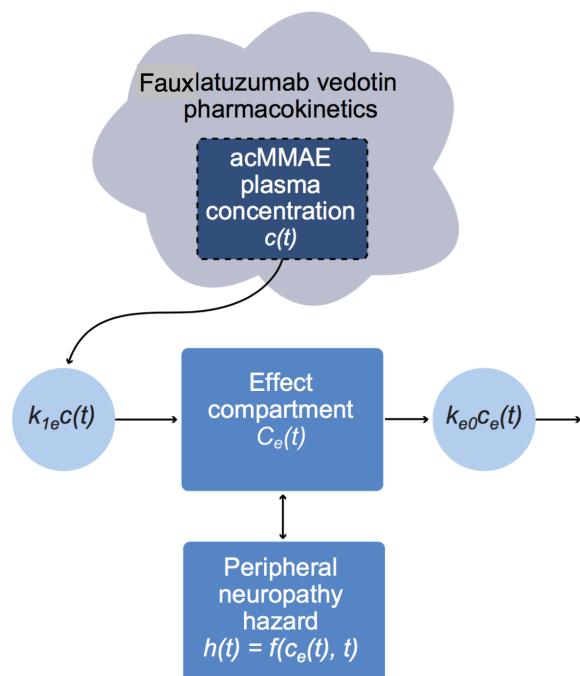
Time to first PN grade 2+ during fauxlatuzumab vedotin treatment

- PK

- One compartment model
- Ideally we would simultaneously fit both PK and PD, but we won't. To keep things simpler, just use the simulated individual CL and V values.

- PD

- PN hazard is substantially delayed relative to PK exposure.
- Hazard increases over time to an extent not completely described by PK.



Hands-on session 4

Model

Likelihood for time to first PN ≥ 2 event in the i^{th} patient:

$$L(\theta | t_{\text{PN},i}, \text{censor}_i, X_i) = \begin{cases} h_i(t_{\text{PN},i} | \theta, X_i) e^{- \int_0^{t_{\text{PN},i}} h_i(u | \theta, X_i) du}, & \text{censor}_i = 0 \\ e^{- \int_0^{t_{\text{PN},i}} h_i(u | \theta, X_i) du}, & \text{censor}_i = 1 \end{cases}$$

where

t_{PN} \equiv time to first PN ≥ 2 or right censoring event

θ \equiv model parameters

X \equiv independent variables / covariates

$$\text{censor} \equiv \begin{cases} 1, & \text{PN} \geq 2 \text{ event is right censored} \\ 0, & \text{PN} \geq 2 \text{ event is observed} \end{cases}$$

Hands-on session 4

Model

- Hazard of PN grade 2+ based on the Weibull distribution
- Drug effect proportional to effect site concentration of MMAE

$$h_j(t) = \beta E_{\text{drug}j}(t)^\beta t^{(\beta-1)}$$

$$E_{\text{drug}j}(t) = \alpha c_{ej}(t)$$

$$c'_{ej}(t) = k_{e0} (c_j(t) - c_{ej}(t))$$

- Overall ODE system including integration of the hazard function:

$$x'_1 = -\frac{CL}{V} x_1$$

$$x'_2 = k_{e0} \left(\frac{x_1}{V} - x_2 \right)$$

$$x'_3 = h(t)$$

where $x_2(t) = c_e(t)$ and $x_3(t) = \int_0^t h(u)du$ aka cumulative hazard.

Hands-on session 4

Files

- Stan model for simulation: model/tppn2Sim.stan
- Stan model for fitting: model/tppn2.stan
- R script: script/tppn2.R

Other Stan pharmacometrics resources

- Stan wiki page on “Complex ODE Based Models”
 - <https://github.com/stan-dev/stan/wiki/Complex-ODE-Based-Models>
- PMXStan
 - By Yuan Xiong & Wenping Wang, Novartis
 - Similar objectives to Torsten
 - R package + Stan functions
 - Uses a modified version of LSODA for numerical solution of ODEs
 - <http://discuss.go-isop.org/t/introduction-to-pmxstan-an-r-library-to-facilitate-pkpd-modeling-with-stan/554>
 - <https://github.com/yxiong1/pmxstan>
- Examples of models written in Stan language
 - by Sebastian Weber, Novartis
 - <https://github.com/stan-dev/example-models/tree/feature/issue-72-stan-pkpdlib/misc/pkpd>

Current role of Stan/Torsten for PMX applications

- Very flexible platform for fully Bayesian analyses that cannot be implemented in standard PMX platforms, e.g., NONMEM or Monolix, without substantial compromises.
- You can do more routine popPK and popPKPD analyses with Stan, particularly with the Torsten extensions, but
 - Computation times make it non-optimal for such applications
- Bottom line: For the moment save it for problems where
 - Fully Bayesian methods are particularly useful, e.g., use of informative priors.
 - A more flexible model specification language is needed.

Barriers to routine use of Stan for PMX applications

- Computation time required for adequate MCMC sampling
- Programming time required to implement typical popPKPD models

Stan & Torsten development plans

- Fast approximate Bayesian methods
 - Gradient-based marginal optimization (GMO) for marginal maximum penalized likelihood estimation
 - Data-streaming variational Bayes via stochastic automatic differentiation variational inference (ADVI)
 - Data-parallel variational Bayes via expectation propagation (EP)
- Within chain parallel computation
- DAEs
- PDEs
- SDEs
- R package to simplify implementation of pharmacometrics models
- R package for specialized visualization and reporting of PKPD model analyses

Addition of those features will make Stan/Torsten a superior open source alternative to (your favorite PMX platform here).

What didn't we cover?

- Categorical and count data
- Population and trial simulations based on MCMC results
- Mixture models
- Working with CmdStan and other Stan interfaces
- Interfacing with external C++ code (<https://cran.r-project.org/web/packages/rstan/vignettes/external.html>)
- Change of variable
- Using ADVI
- Using Stan for point estimation
 - Penalized maximized likelihood
 - Posterior modes
- ...

Acknowledgments

Credit where it is due...

Metrum Research Group

- Charles Margossian
- Yi Zhang
- Marc Gastonguay

Stan developers

- Andrew Gelman, Columbia Univ.
- Bob Carpenter, Columbia Univ.
- Mike Betancourt, Columbia Univ.
- Daniel Lee, Columbia Univ.
- Sebastian Weber, Novartis
- + contributions from many more

Funding sources:

- Bill & Melinda Gates Foundation
- Office of Naval Research STTR Program
 - contract no. N00014-16-P-2039
 - contract no. N6833518C0110

References I

- [1] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin.
Bayesian Data Analysis.
 CRC Press, Boca Raton, FL, third edition, 2014.
- [2] Christian P. Robert.
The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation.
 Springer, second edition, 2007.
- [3] B. P. Carlin and T. A. Louis.
Bayesian Methods for Data Analysis.
 Chapman & Hall/CRC, third edition, 2008.
- [4] J.O. Berger.
Statistical Decision Theory and Bayesian Analysis.
 Springer, second edition, 1993.
- [5] D.J. Lunn, A. Thomas, N. Best, and D. Spiegelhalter.
 Winbugs – a bayesian modelling framework: concepts, structure, and extensibility.
Statistics and Computing, 10:325–337, 2000.

References II

- [6] D. J. Spiegelhalter, K. R. Abrams, and J. P. Myles.
Bayesian Approaches to Clinical Trials and Health-Care Evaluation.
 Wiley, 2004.
- [7] Christian P. Robert and George Casella.
Introducing Monte Carlo Methods with R.
 Springer, 2010.
- [8] Christian P. Robert and George Casella.
Monte Carlo Statistical Methods.
 Springer, second edition, 2004.
- [9] C. Dansirikul, R. G. Morris, S. E. Tett, and S. B. Duffull.
 A bayesian approach for population pharmacokinetic modelling of sirolimus.
Br.J Clin.Pharmacol., 62(4):420–434, 2006.
- [10] M. G. Dodds and P. Vicini.
 Assessing convergence of markov chain monte carlo simulations in hierarchical bayesian models for population pharmacokinetics.
Ann.Biomed.Eng, 32(9):1300–1313, 2004.

References III

- [11] A. Dokoumetzidis and L. Aarons.
Propagation of population pharmacokinetic information using a bayesian approach: comparison with meta-analysis.
J Pharmacokinet Pharmacodyn., 32(3-4):401–418, 2005.
- [12] L. E. Friberg, G. K. Isbister, L. P. Hackett, and S. B. Duffull.
The population pharmacokinetics of citalopram after deliberate self-poisoning: a bayesian approach.
J Pharmacokinet Pharmacodyn, 32(3-4):571–605, 2005.
- [13] I. Gueorguieva, L. Aarons, and M. Rowland.
Diazepam pharmacokinetics from preclinical to phase i using a bayesian population physiologically based pharmacokinetic model with informative prior distributions in winbugs.
J Pharmacokinet Pharmacodyn., 33(5):571–594, 2006.
- [14] Steven J Kathman, Daphne H Williams, Jeffrey P Hodge, and Mohammed Dar.
A bayesian population pk-pd model for ipilimumab/docetaxel combination-induced myelosuppression.
Cancer Chemother Pharmacol, 63(3):469–476, 2009 Feb.
- [15] D. J. Lunn, N. Best, A. Thomas, J. Wakefield, and D. Spiegelhalter.
Bayesian analysis of population pk/pd models: general concepts and software.
Journal of Pharmacokinetics and Pharmacodynamics, 29(3):271–307, 2002.

References IV

- [16] S. Mu and T. M. Ludden.
Estimation of population pharmacokinetic parameters in the presence of non-compliance.
Journal of Pharmacokinetics and Pharmacodynamics, 30(1):53–81, 2003.
- [17] M. K. Smith, I. Jones, M. F. Morris, A. P. Grieve, and K. Tan.
Implementation of a bayesian adaptive design in a proof of concept study.
Pharm Stat, 5(1):39–50, 2006.
- [18] M. K. Smith and S. Marshall.
A bayesian design and analysis for dose-response using informative prior information.
J Biopharm Stat, 16(5):695–709, 2006.
- [19] Andrew Gelman.
Prior distributions for variance parameters in hierarchical models.
Bayesian Analysis, 1:515–533, 2006.
- [20] Cleve Moler and Charles Van Loan.
Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later.
SIAM review, 45(1):3–49, 2003.

References V

- [21] Céline Verstuyft, Xavier Delavenne, Alexandra Rousseau, Annie Robert, Michel Tod, Bertrand Diquet, Martine Lebot, Patrice Jaillon, and Laurent Becquemont. A pharmacokinetic-pharmacodynamic model for predicting the impact of cyp2c9 and vkorc1 polymorphisms on fluindione and acenocoumarol during induction therapy. *Clinical pharmacokinetics*, 51(1):41–53, 2012.
- [22] F. Mentre, F. Pousset, E. Comets, B. Plaud, B. Diquet, G. Montalescot, A. Ankri, A. Mallet, and P. Lechat. Population pharmacokinetic-pharmacodynamic analysis of fluindione in patients. *Clinical Pharmacology and Therapeutics*, 63(1):64–78, 1998.
- [23] Matthew D Hoffman and Andrew Gelman. The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *The Journal of Machine Learning Research*, 15(1):1593–1623, 2014.
- [24] Radford M. Neal. MCMC using hamiltonian dynamics. In Steve Brooks, Andrew Gelman, Galin L. Jones, and Xiao-Li Meng, editors, *Handbook of Markov Chain Monte Carlo*, chapter 5, pages 113–162. Chapman & Hall/CRC, Boca Raton, FL, 2011.

References VI

- [25] Matthew D Hoffman and Andrew Gelman. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(1):1593–1623, 2014.
- [26] Stan Modeling Language Users Guide and Reference Manual, Version 2.17.0, 2017. <http://mc-stan.org/>.
- [27] L. E. Friberg, A. Henningsson, H. Maas, L. Nguyen, and M. O. Karlsson. Model of chemotherapy-induced myelosuppression with parameter consistency across drugs. *J Clin Oncol*, 20(24):4713–21, 2002.
- [28] L. E. Friberg and M. O. Karlsson. Mechanistic models for myelosuppression. *Invest New Drugs*, 21(2):183–194, 2003.
- [29] J. E. Latz, M. O. Karlsson, J. J. Rusthoven, A. Ghosh, and R. D. Johnson. A semimechanistic-physiologic population pharmacokinetic/pharmacodynamic model for neutropenia following pemetrexed therapy. *Cancer Chemotherapy and Pharmacology*, 57(4):412–426, 2006.

References VII

- [30] I. F. Troconiz, M. J. Garrido, C. Segura, J. M. Cendros, P. Principe, C. Peraire, and R. Obach.
Phase i dose-finding study and a pharmacokinetic/pharmacodynamic analysis of the neutropenic response of intravenous diflomotecan in patients with advanced malignant tumours.
Cancer Chemother Pharmacol, 57(6):727–35, 2006.
- [31] S. J. Kathman, D. H. Williams, J. P. Hodge, and M. Dar.
A bayesian population pk-pd model of ispinesib-induced myelosuppression.
Clin Pharmacol Ther, 81(1):88–94, 2007.
- [32] D Lu, WR Gillespie, S Girish, P Agarwal, C Li, J Hirata, Y-W Chu, M Kagedal, L Leon, V Maiya, et al.
Time-to-event analysis of polatuzumab vedotin-induced peripheral neuropathy to assist in the comparison of clinical dosing regimens.
CPT: Pharmacometrics & Systems Pharmacology, 2017.
- [33] Aki Vehtari, Andrew Gelman, and Jonah Gabry.
Efficient implementation of leave-one-out cross-validation and waic for evaluating fitted bayesian models.
arXiv preprint arXiv:1507.04544, 2015.

Appendix

Bayesian model evaluation & comparison: The ideal

- Mixture of all possible “true” models including all known substantive information
- Extends the notion of Bayesian modeling of uncertainty to include uncertainty in model structure
- Under such a grand mixture model the tasks of model evaluation, model comparison and sensitivity analysis are all handled within a single analysis
 - Each model in the mixture is associated with a posterior probability that it is the “true” (or “best”) model. This provides a metric for comparing models.
 - Sensitivity to model choice is obtained by comparing model-specific posterior predictions
- Like most ideals it cannot be achieved in full, only approached via practical compromises

Bayesian model evaluation & comparison

Philosophical consequences of the ideal

- Model selection is often avoided in favor of using a “fuller” model that contains potential competing models as special cases
 - Particularly avoided is the classic approach of performing a sequence of hypothesis tests
 - The “fuller” model approach acknowledges and quantifies the uncertainty in the relative “correctness” of competing models.
- Types of “fuller” models
 - Discrete mixture models, e.g.,

$$p(y|\theta) = \Pr(M_1)p(y|\theta, M_1) + \Pr(M_2)p(y|\theta, M_2) + \dots + \Pr(M_n)p(y|\theta, M_n)$$

- Continuous mixture models, e.g., models containing all covariates of potential interest

Bayesian model evaluation & comparison

Typical practical Bayesian model development

- Propose initial model structure based on available prior information.
 - Realistically exploratory analysis of the new data also influences this process,
- Assess whether model is consistent with the data and prior information
 - Posterior predictive checking
 - Are model inferences, predictions and values of parameters or other derived quantities consistent with other knowledge?
- Assess sensitivity to potentially influential assumptions, e.g., choice of prior distributions
- If deficiencies are discovered that could adversely affect important inferences then explore model revisions and reassess as before.

Bayesian model evaluation & comparison

Typical practical Bayesian model development (cont.)

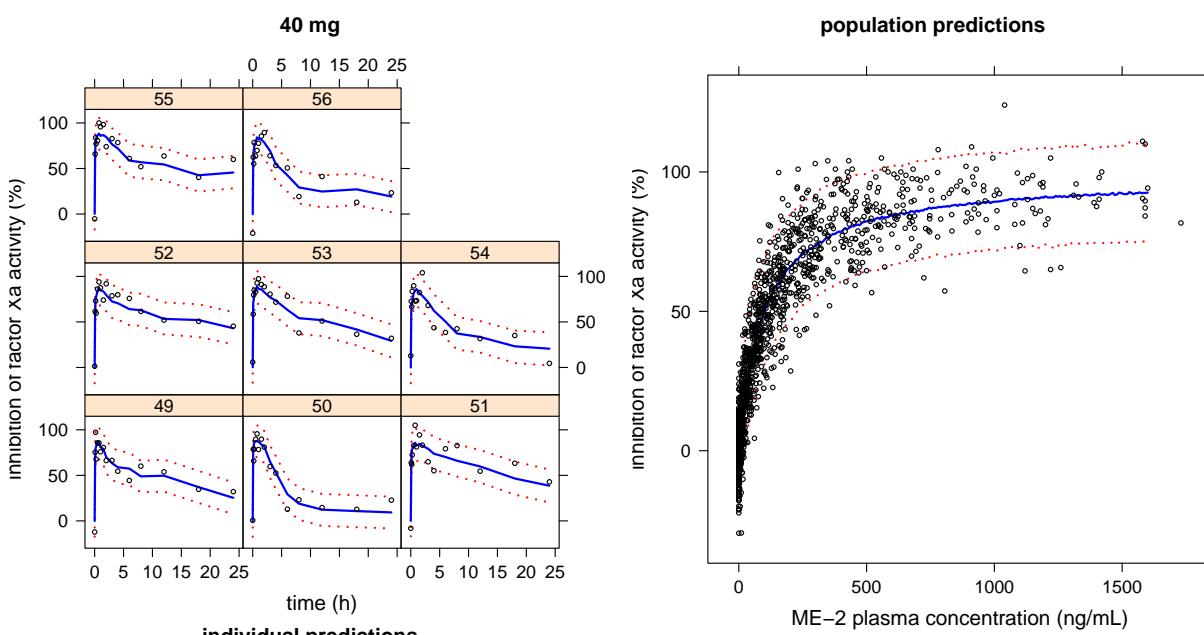
- Use the model resulting from this process for pre-planned inferences.
- Optionally there may be additional hypothesis-generating activities, e.g.,
 - Data mining efforts to explore the influence of covariates not considered in the original model
 - Further exploration of alternative model structures not considered in the previous sensitivity analyses

Posterior predictive checking (PPC)

- Graphical checks
 - Comparing data & predictions
 - Comparing data summaries and model predictions/inferences
 - Residuals
- Formal numerical checks
 - Posterior predictive p-values based on a test statistic

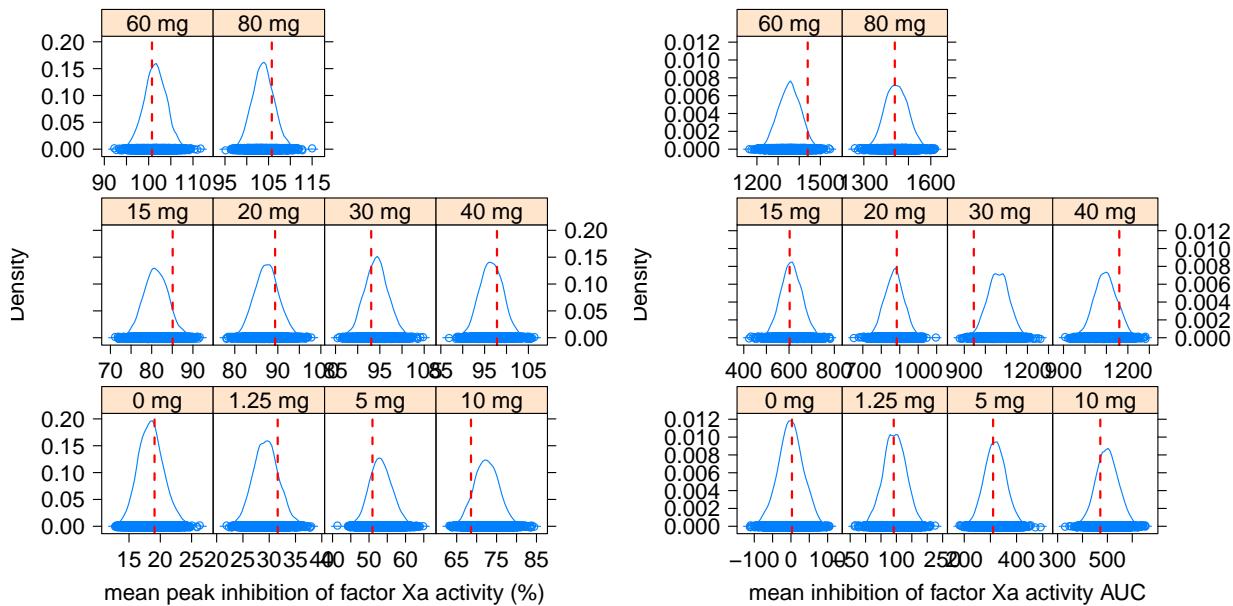
Graphical checks: Comparing data & predictions

Posterior medians & 90% prediction intervals compared to observed data



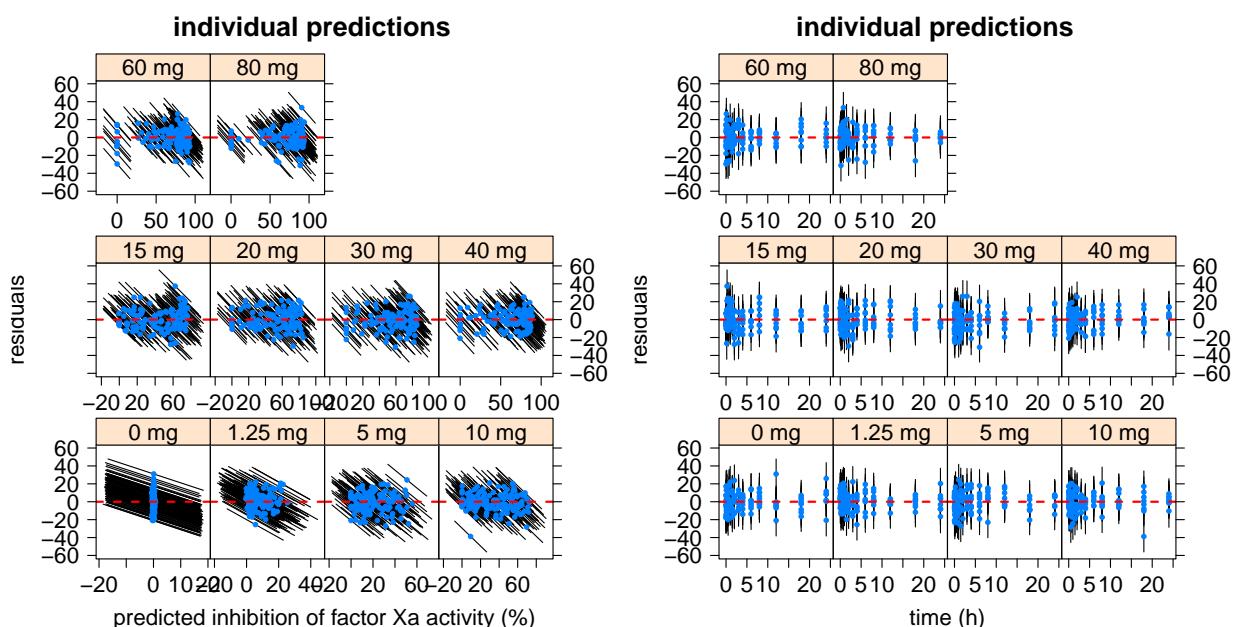
Graphical checks: Comparing data summaries and model predictions/inferences

Posterior densities of predicted mean peak effect and area under the effect curve compared to observed means.



Graphical checks: Residuals

Uncertainty in residuals and predictions can be depicted with lines representing prediction intervals



Sensitivity analysis

- Generally desirable to assess sensitivity of key inferences to model assumptions
 - True for any modeling strategy, not just Bayesian approaches
- Model assumptions to be evaluated
 - Model structure, i.e., functional forms of the likelihood and prior distribution
 - Quantitative assumptions, i.e., values of the parameters of the prior distribution
 - Let's focus on the prior distribution since that is unique to Bayesian modeling

Sensitivity analysis: Prior distribution

- When your intention is for the prior distribution to be non-informative
 - Compare effect of different prior distributions on important inferences or predictions
 - Different degrees of weak informativeness, e.g., normal distributions with various large variances
 - Different distribution types, e.g., t-distributions with various degrees of freedom instead of a normal distribution

Sensitivity analysis: Prior distribution

When your intention is for the prior distribution to be informative

- Desirable to assess the relative influence of prior knowledge and new data on important inferences.
- Sensitivity to the prior distribution is not necessarily a bad thing, but it is desirable to know the extent to which the prior affects your inferences.
- Should explore a range of priors consistent with the range of knowledge and beliefs held by other stakeholders in the analysis.
- For example:
 - When evaluating the probable benefit of some treatment, consider priors that reflect more or less skepticism than the one used for the primary analysis.
 - If the primary analysis concludes the treatment is more beneficial than a comparator:
 - Consider identifying the least skeptical prior that would make the conclusion unconvincing.
 - Assess whether that skeptical prior is plausible.

Bayesian model comparison

- Competing models may be compared with respect to the various modeling evaluation methods just described, i.e.,
 - Posterior predictive checks
 - Consistency with other substantive knowledge
- Specific metrics for comparing models include:
 - Measures of out of sample predictive accuracy [1, Chapter 7]
 - Watanabe-Akaike Information Criterion (WAIC)
 - Leave one out cross validation (LOO-CV)
 - Both may be estimated from Stan output using the loo R package [33].

Bayesian model comparison

- Demo: Using the hands-on session example, let's compare the previously specified model with a model that assumes Emax = 100%.
 - Data: data/derived/fxa.data.csv
 - Stan model 1: model/fxalInhibit1Ncp.stan
 - R script for model 1: script/fxalInhibit1Ncp.R
 - Stan model 2: model/fxalInhibit2Ncp.stan
 - R script for model 2: script/fxalInhibit2Ncp.R
 - R script for model comparison: fxalInhibitModelCompare.R

Other approaches not covered

- Other ICs, e.g., Akaike IC, Bayesian IC, Schwartz IC,
- Bayes factors
 - For model comparison
 - For model averaging
- Other model averaging or mixture model approaches

Use of informative priors in clinical pharmacology applications

- Integration of prior modeling results
 - Biomarker-to-clinical outcome models
 - Informative priors for parameters not identifiable with new data set, e.g., k_a for a rapidly absorbed drug in popPK modeling of sparse data
- Physiologically-based modeling
 - PB-PK/TK
 - Mechanistically-based models of xenobiotic response
 - Toxicology applications
 - Drug development/treatment applications

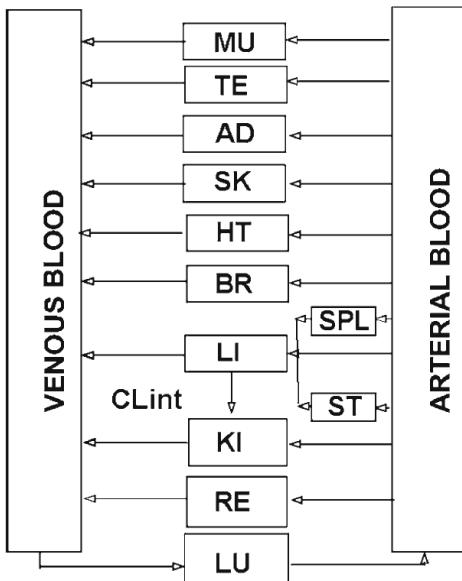
Use of informative priors in clinical pharmacology applications

- Design and analysis of clinical trials
 - Informative priors for historical controls
 - Sequentially combining knowledge from multiple trials
 - Combining prior knowledge about exposure-response for existing drugs with data for a new drug candidate
 - Combining prior knowledge about biomarker-clinical outcome relationships for existing drugs with biomarker data for a new drug candidate to make inferences about clinical outcomes
- Individual patient treatment
 - Adaptive adjustment of treatment regimen via Bayesian analysis of individual data using informative priors from a population model.

Diazepam Pharmacokinetics from Preclinical to Phase I Using a Bayesian Population Physiologically Based Pharmacokinetic Model with Informative Prior Distributions in Winbugs

Ivelina Gueorguieva,^{1,2} Leon Aarons,¹ and Malcolm Rowland¹

J Pharmacokinet Pharmacodyn. 2007 Jun;34(3):313-31.



Illustrates potential for applying Bayesian methods + mechanistic modeling to early clinical development

- A complex physiologically-based PK (PBPK) model is used for:
 - Analyzing preclinical (rat) PK data
 - Predicting human PK
 - Analyzing human PK data
- But how can you apply such a large, apparently over-parameterized model?

Bayesian modeling + informative priors make it feasible

- Model structure largely determined by physiologic knowledge
- Informative prior distributions for mechanistically interpretable parameters obtained from publications or preclinical experiments:
 - Physiologic parameters, e.g., blood flows and tissue volumes
 - Drug-specific parameters, e.g., tissue-to-plasma partition coefficients, fraction unbound in plasma, blood-to-plasma ratio and intrinsic clearances
- Potential for more reliable prediction of human PK than empirical allometric scaling methods
- Posterior predictive distributions provide a clear statement of the uncertainty in such predictions
- Model may be continuously updated as human data becomes available

PB-TK example: Assessing cancer risk of dichloromethane exposure

A Bayesian Analysis of the Influence of GSTT1 Polymorphism on the Cancer Risk Estimate for Dichloromethane

Fredrik Jonsson* and Gunnar Johanson*†

Toxicology and Applied Pharmacology **174**, 99–112 (2001)

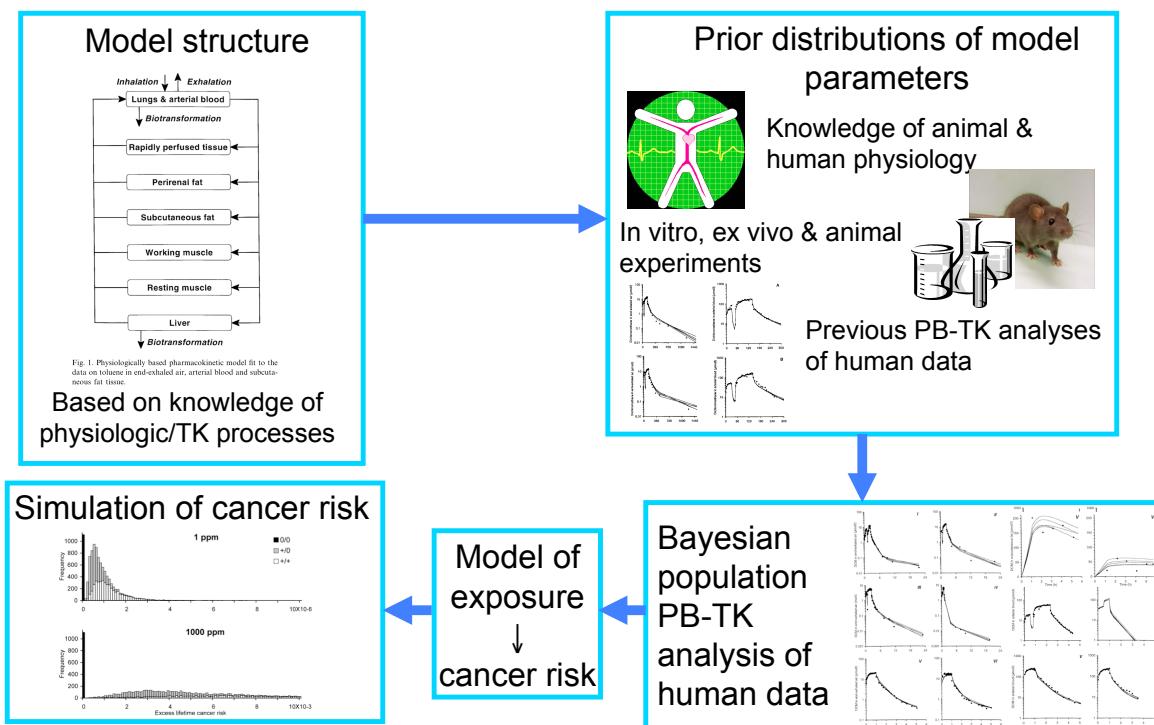
- Objective

- Estimate excess cancer risk as a function of exposure to inhaled dichloromethane (DCM).

- Considerations

- Very limited data regarding blood or organ DCM concentrations or regarding cancer risk in humans as a function of DCM exposure.
- DCM carcinogenicity related to metabolic activation mediated by glutathione transferase theta 1 (GSTT1).
- Genetic polymorphism in GSTT1 mediated metabolism.

Bayesian modeling of DCM TK + Monte Carlo simulation of cancer risk



For this case prior knowledge makes inferences possible that could not otherwise be done

- Illustrates the value of using
 - Mechanistic knowledge, prior quantitative knowledge and limited human data to predict xenobiotic exposure in humans.
 - Population models that consider multiple sources of variability.
 - Bayesian methods to quantify uncertainty in prior knowledge and the resulting uncertainty in predicted exposure.
- The resulting predictions and inferences about human cancer risk appropriately consider that uncertainty and variability.

PB-PD example: Mechanistically-based model for drug-induced myelosuppression

A Bayesian Population PK-PD Model of Ispinesib-induced Myelosuppression

SJ Kathman¹, DH Williams¹, JP Hodge¹ and M Dar¹

Clinical Pharmacology & Therapeutics 81: 88–94 (2007)

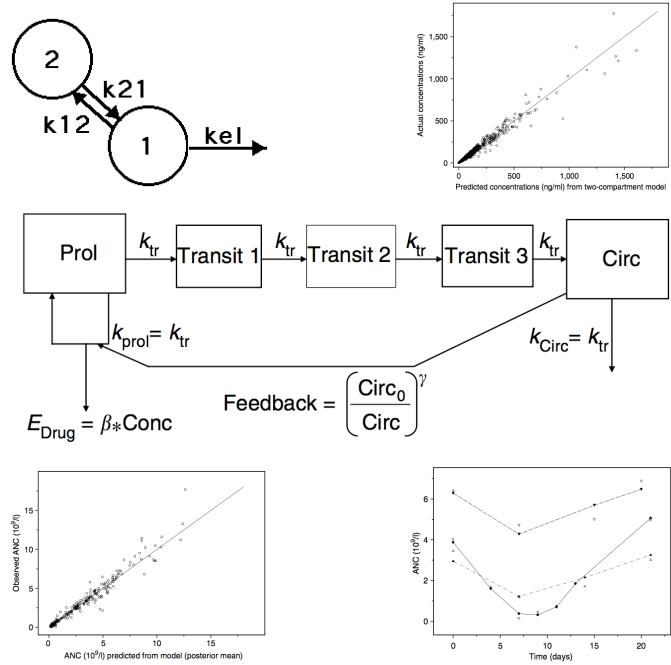
- Objective
 - Model relationship between ispinesib exposure and neutrophil counts—to support future efforts to optimize treatment regimens.

PB-PD example: Mechanistically-based model for drug-induced myelosuppression

- Data
 - Phase I dose-escalation study in 45 patients
 - Repeated Ispinesib concentrations for 48 h following first dose
 - Absolute neutrophil counts (ANC) weekly for 3 weeks
- Prior knowledge
 - Semi-mechanistic model of drug-induced neutropenia (L Friberg, M Karlsson, Invest. New Drugs 21, 183–194 (2003).)
 - Subset of parameters are thought to be drug-independent.
- Can we construct a model for ispiniseb-induced myelosuppression conditioned on both new data and prior knowledge? How?

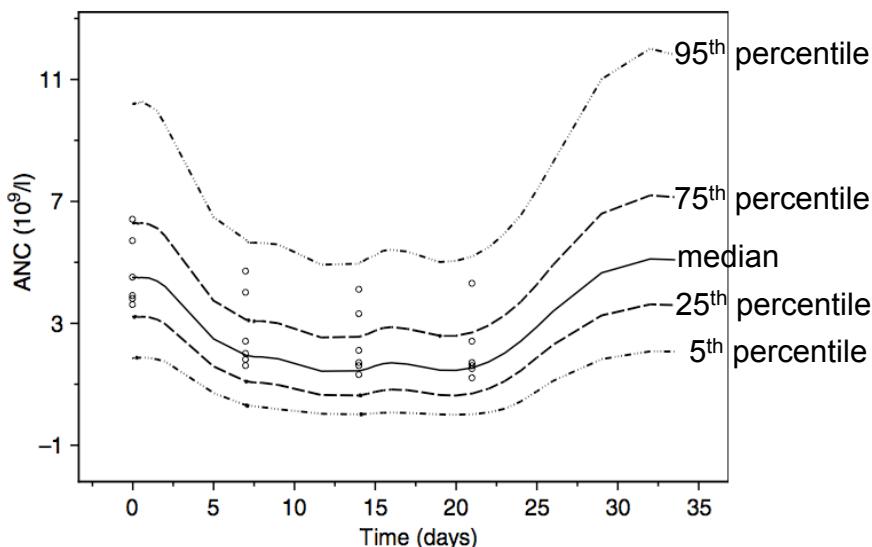
Bayesian modeling using a mix of informative and uninformative prior distributions

- Population PK model
 - Two compartment model
 - Uninformative priors
- Population PKPD model
 - Mechanistic model of the neutrophil life cycle.
 - Drug-independent parameters: Informative priors constructed from published estimates of the model parameters
 - Drug-dependent parameters: Uninformative priors



Resulting model-based predictions reflect both new data and prior knowledge

- May be used to explore probable range of myelosuppression that may result from different dosing regimens.
 - E.g., predicted and observed ANC for 7 mg/m² weekly x 3:



Combine prior knowledge about exposure-response for existing drugs with data for a new drug candidate

Scenario

- New drug candidate
 - Shares a common mechanism with one or more existing drugs
 - Currently in Phase I
 - No therapeutically-relevant exposure-response information
- Exposure-response information available for existing drugs
 - Models available or can be constructed from available data
- Objective: Design an efficient Phase II strategy for PoC and dose-finding
 - PoC: Is there a well-tolerated dose of the new agent with efficacy \geq active comparator?
 - Dose-finding:
 - What new drug dose is equally efficacious to active comparator?
 - What new drug dose(s) should be used for patient treatment and Phase III trials?

How can knowledge about exposure-response of existing drugs be used to enhance the efficiency and informativeness of the next Phase II trial?

- Such knowledge can reduce the amount of information that needs to be obtained about the active comparator:
 - Fewer patients required in active comparator arm(s).
- What qualitative and quantitative elements of the exposure-response models for existing drugs are applicable to the new drug?
- For example, suppose that the existing drugs:
 - Produce efficacy responses described by sigmoid Emax models,
 - Share a common mechanism, and
 - Are full agonists.
- Then it can be argued that they may share common values for Emax and γ .
- Informative priors for Emax and γ as well as some variance parameters may be used for analyzing the trial and addressing the key PoC and dose-finding inferences.

Such a strategy is illustrated by ...

A BAYESIAN DESIGN AND ANALYSIS FOR DOSE-RESPONSE USING INFORMATIVE PRIOR INFORMATION

Michael K. Smith

*Pharmacometrics, Pfizer Global Research and Development,
Sandwich, Kent, UK*

Scott Marshall

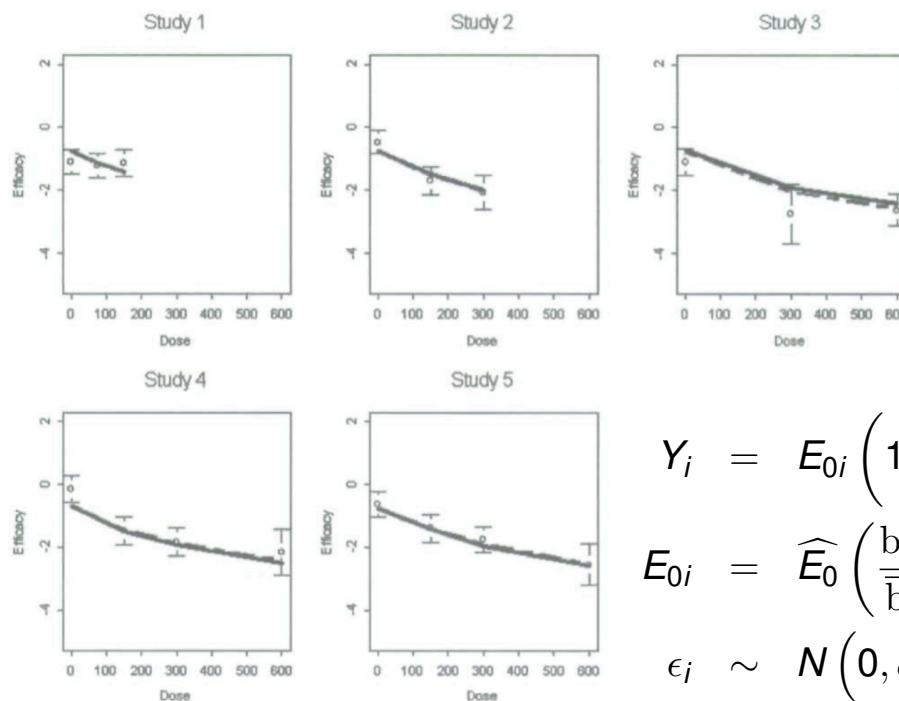
*Clinical Pharmacology, Pfizer Global Research and Development,
Sandwich, Kent, UK*

Journal of Biopharmaceutical Statistics, 16: 695–709, 2006

Trial simulations to optimize Phase II PoC & dose-finding trial

- Focus on:
 - Optimizing sample size and treatment allocation.
 - Assessing impact of informative priors based on data for existing drug with same MOA.
- Compared operating characteristics of 4 treatment arm study designs:
 - Treatment 1: Placebo
 - Treatment 2: New drug dose A
 - Treatment 3: New drug dose B
 - Treatment 4: Existing drug

Dose-response modeling of efficacy data for existing drug



$$\begin{aligned}
 Y_i &= E_{0i} \left(1 + \frac{E_{\max} D_i}{ED_{50} + D_i} + \epsilon_i \right) \\
 E_{0i} &= \widehat{E}_0 \left(\frac{\text{baseline}_i}{\overline{\text{baseline}}} \right)^{\gamma} \\
 \epsilon_i &\sim N(0, \sigma^2)
 \end{aligned}$$

Figure 1 Dose-response estimates for the existing compound across the five available studies.

Bayesian analysis of simulated trials

- Prior distributions for E_0 (placebo response), E_{\max} and ED_{50} for the existing drug were obtained by modeling data for the existing drug.
- A relatively vague prior was assigned to relative potency (RP) of the new drug based on preclinical information: lognormal(median = 1, CV = 158%).

Informative prior reduces sample size requirements by ~ half

- Results for case where true relative potency (RP) = 4
- Success defined as $\Pr(RP > 1) > 0.9$ and $E(RP) > 3$

Design (sample size per arm)				total sample size	Pr(success)	
placebo	new drug		existing drug		informative prior	uninformative prior
	dose A	dose B				
25	25	25	25	100	0.73	0.55
10	40	40	10	100	0.87	0.33
15	60	60	15	150	0.89	0.63
25	50	50	25	150	0.87	0.71
10	10	10	10	40	0.5	0.14
80	80	80	80	320	0.91	0.90
40	40	40	40	160	0.84	0.72

Adapted from Table 8 of MK Smith, S Marshall. J Biopharm Stat 16: 695–709 (2006).

Constructing informative prior distributions

- Elicitation of subjective opinion
 - Limited utility for clinical pharmacology applications
- Summarizing external evidence
 - Relevant approach for most clinical pharmacology applications
- Examples
 - Dansirikul, Morris, Tett & Duffull (2005): Priors constructed using summary stats from published reports.
 - Jonsson, Jonsson, Bois & Marshall (2007): Excellent description of prior construction from mechanistic and empirical knowledge.

Summary of external evidence

Different methods are appropriate depending on the relationship between the prior evidence and new data:

- Irrelevant
 - Prior evidence should not be used
- Equal
 - Prior evidence has the same form and is equally informative about the parameter(s) of interest
- Exchangeable
 - Hierarchical model in which study-specific parameters are samples from a common distribution
- Equal but discounted
 - Prior variance is inflated to reflect uncertainty in the relevance of prior evidence
- Potentially biased
 - Hierarchical model used to accommodate potential unexplained differences among prior studies and new data.
- Functionally dependent
 - Systematic differences between prior evidence and new data are explicitly modeled

See Chapter 5 of Spiegelhalter, Abrams, Myles (2004)

Considerations/cautions

Relevance/exchangeability of prior knowledge/data

- Are past studies really exchangeable with new ones or should the past info be discounted in some way to account for differences (known or unknown)?
 - Are the patients comparable?
 - Has the standard of care or measurement changed?
- If using published info, is there a risk of significant publication bias?

Considerations/cautions

Critical evaluation / Sensitivity analysis

- Assessing relative influence of new data and priors
- Robustness of key inferences to priors
- Consideration of beliefs/knowledge/values of key stakeholders
 - Is the prior acceptable to the key stakeholders in the inferences and decisions resulting from the analysis?
 - Construct a prior by consensus, or
 - Consider a community of priors¹, e.g., do a sensitivity analysis to explore how the inference changes for different priors preferred by particular stakeholders.

¹Kass & Greenhouse. Statistical Science 4: 310–317 (1989)

PopPK using PKModelTwoCpt

PopPK using a 2 compartment model with 1st order absorption

We will analyze the combined PK data from:

- A phase 1 single dose study of ME-2 safety and PK in healthy volunteers,
- A phase 1 multiple dose study of ME-2 safety and PK in healthy volunteers,
- A phase IIa PoC trial of ME-2 in patients for prevention of post-op VTEs.

PopPK using PKModelTwoCpt

Population PK of ME-2

Phase 1 single dose study in healthy volunteers

- Parallel dose-escalation design
- 8 subjects per dose arm
- Single doses of ME-2
 - Placebo, 1.25, 5, 10, 15, 20, 30, 40, 60 and 80 mg
- PK: plasma concentrations of parent drug
 - PK measured at 0, 0.083, 0.167, 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4, 6, 8, 12, 18 and 24 hours after dose.

PopPK using PKModelTwoCpt

Population PK of ME-2

Phase 1 multiple dose study in healthy volunteers

- Parallel dose-escalation design
- 8 subjects per dose arm
- Placebo or ME-2 5, 10, 20, 40 or 80 mg bid (q12h) x 7 days
- PK: plasma concentrations of parent drug
 - PK measured at 0, 0.083, 0.167, 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4, 6, 8, 12, 12.1, 12.2, 12.5, 12.8, 13, 13.5, 14, 15, 16, 18, 20, 24, 36, 48, 60, 72, 84, 96, 108, 120, 132, 144, 156, 168, 168, 168, 168, 169, 169, 170, 170, 171, 172, 174, 176, 180, 186 and 192 hours after the first dose.

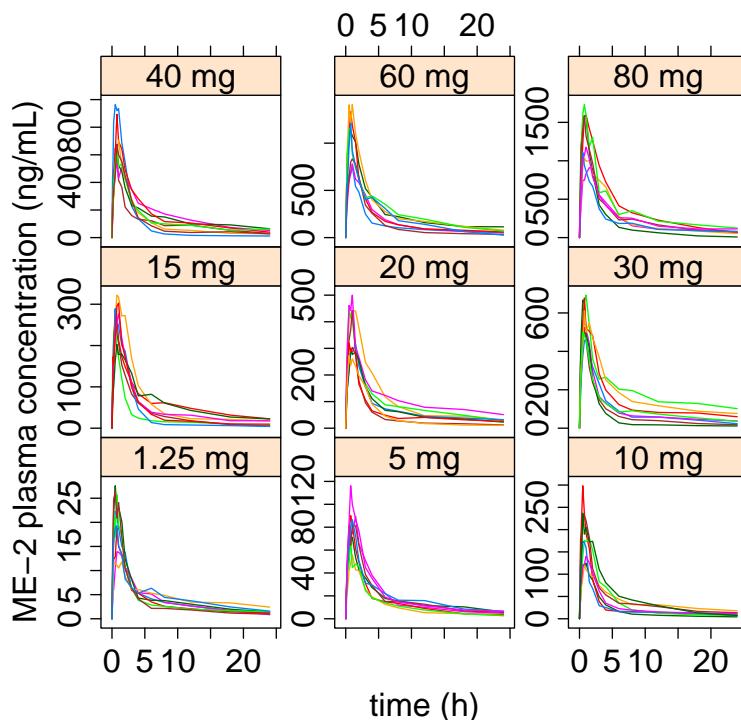
PopPK using PKModelTwoCpt

Population PK of ME-2

Phase IIa PoC trial of ME-2 for prevention of post-op VTEs

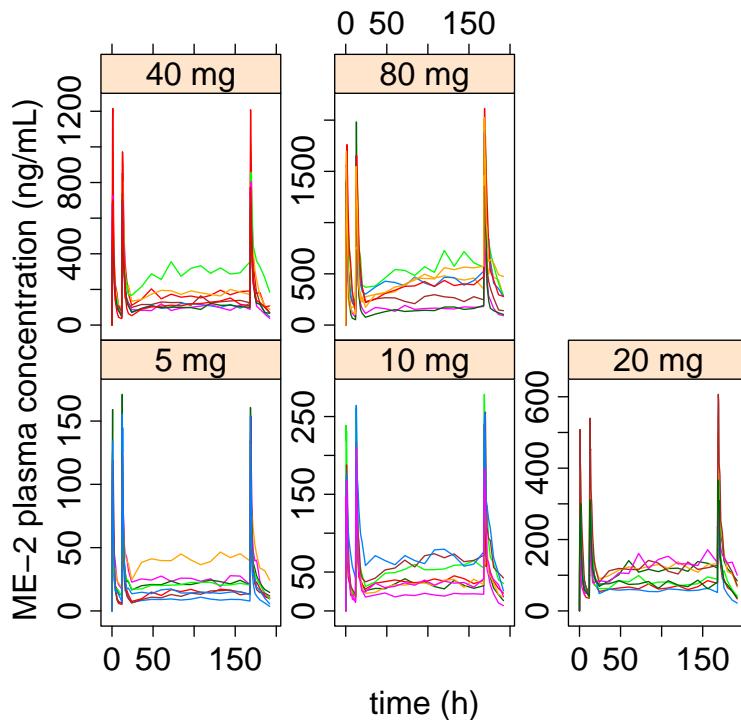
- Treatments
 - ME-2 20 mg bid (q12h) x 7 days
 - Enoxaparin 30 mg bid (q12h) x 7 days
- 100 patients per treatment arm
- Sparse ME-2 PK data (3-6 samples/patient)
 - LOQ = 10 ng/mL

PopPK using PKModelTwoCpt



ME-2 PK data from Phase I SD trial

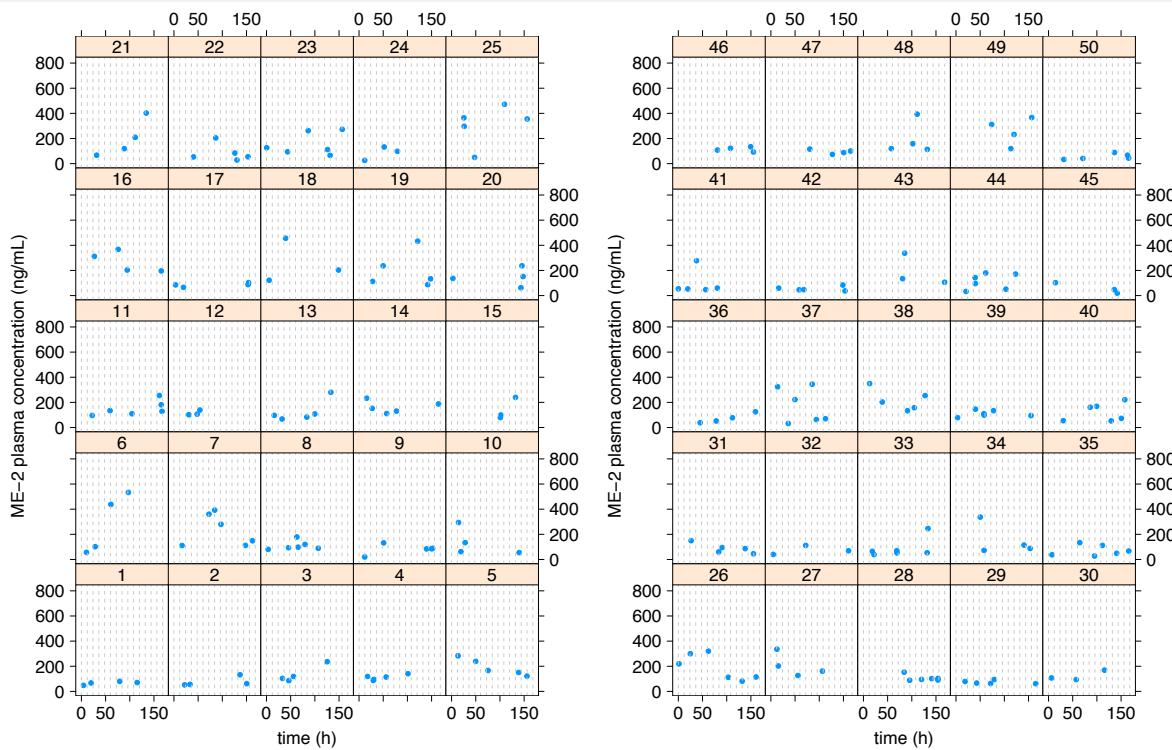
PopPK using PKModelTwoCpt



ME-2 PK data from Phase I MD trial

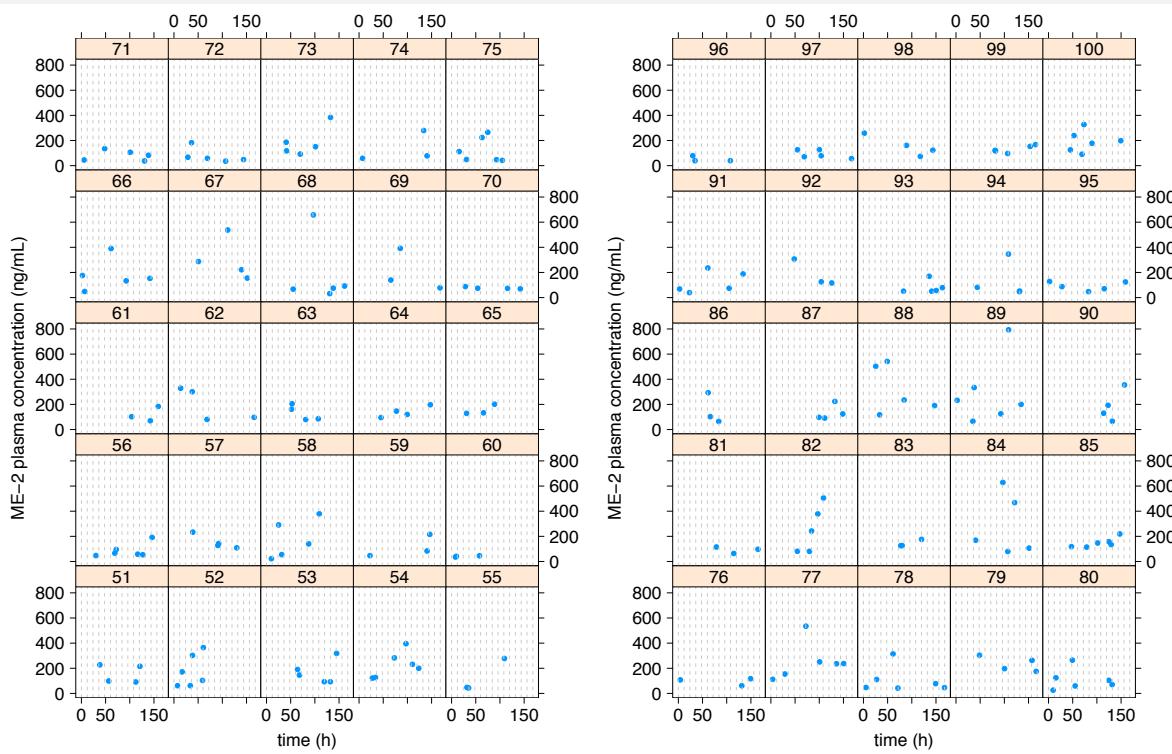
PopPK using PKModelTwoCpt

ME-2 PK data from Phase IIa trial



PopPK using PKModelTwoCpt

ME-2 PK data from Phase IIa trial



PopPK using PKModelTwoCpt

Proposed base model

- Two compartment model with first order absorption describing ME-2 plasma concentration on the i^{th} occasion in the j^{th} subject as a function of time, dose and body weight:

$$\begin{aligned}\log(c_{ij}) &\sim N(\log(\hat{c}_{ij}), \sigma) \\ \hat{c}_{ij} &= f_{2cpt}(t_{ij}, D_j, \tau_j, CL_j, Q_j, V_{1j}, V_{2j}, k_{aj}) \\ \log(CL_j, Q_j, V_{1j}, V_{2j}, k_{aj}) &\sim N\left(\log\left(\widehat{CL}\left(\frac{bw_j}{70}\right)^{0.75}, \widehat{Q}\left(\frac{bw_j}{70}\right)^{0.75}, \widehat{V}_1\left(\frac{bw_j}{70}\right), \widehat{V}_2\left(\frac{bw_j}{70}\right), \widehat{k}_a\right), \Omega\right)\end{aligned}$$

- Weakly informative prior distributions:

$$\begin{aligned}\widehat{CL} &\sim \text{half-}N(0, 25) \quad \widehat{Q} \sim \text{half-}N(0, 50) \quad \widehat{V}_1 \sim \text{half-}N(0, 100) \\ \widehat{V}_2 &\sim \text{half-}N(0, 200) \quad \widehat{k}_a \sim \text{half-}N(0, 5) \quad \sigma \sim \text{half-Cauchy}(0, 1) \\ \Omega &= \text{diag}(\omega) P \text{diag}(\omega) \\ \omega_i &\sim \text{half-Cauchy}(0, 1), i \in \{1, 2, 3, 4, 5\} \quad P \sim \text{LKJCorr}(1)\end{aligned}$$

PopPK using PKModelTwoCpt

Files

- Data
 - ME-2 plasma concentration data (including all covariates) in NONMEM format: data/derived/fxaNONMEMData.csv
- Stan model: model/multiDoseME2PK1.stan
- R script: script/multiDoseME2PK1.R

Non-centered parameterization: Multivariate case

- Use the Cholesky decomposition of the covariance matrix to generate the multivariate normal random effects from standard normal random variables.

Files

- Data
 - ME-2 plasma concentration data (including all covariates) in NONMEM format: data/derived/fxaNONMEMData.csv
- Stan model: model/multiDoseME2PK1.stan
- R script: script/multiDoseME2PK1.R
- Non-centered parametrization
 - Stan model: model/multiDoseME2PK1Ncp.stan
 - R script: script/multiDoseME2PK1Ncp.R

PK modeling with BQL data

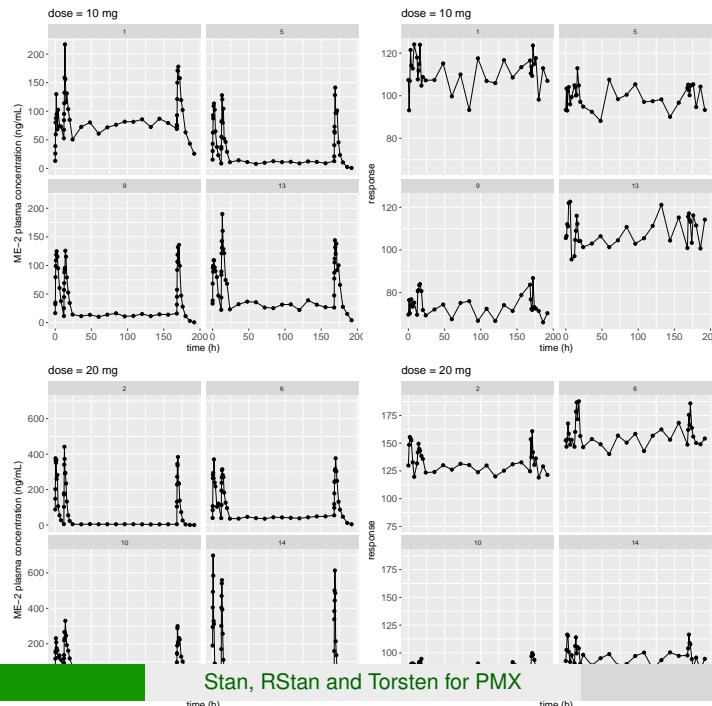
For the previous examples we excluded the BQL data.

Let's return to that example and appropriately incorporate the BQL data in the analysis.

- Data: data/derived/fxaNONMEMData.csv
- Stan model: model/multiDoseME2PK2Ncp.stan
- R script: script/multiDoseME2PK2Ncp.R

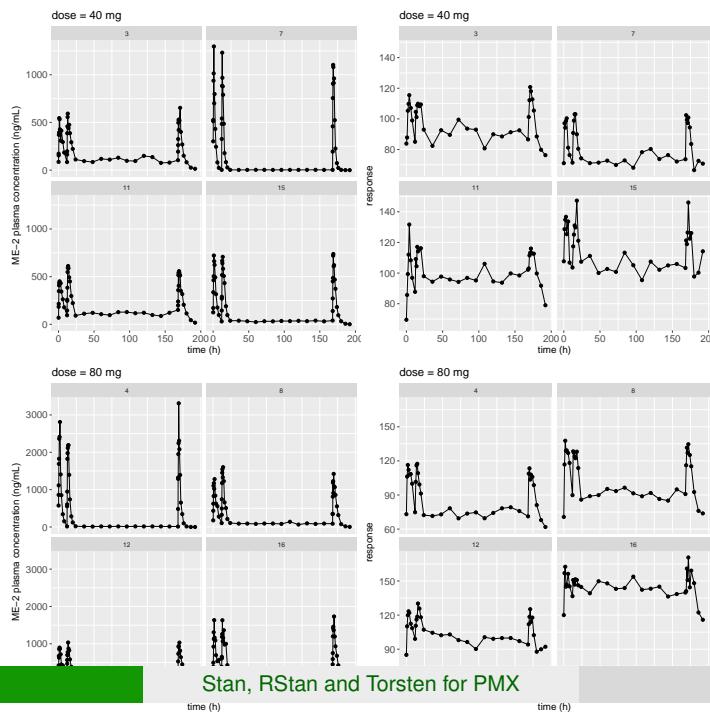
PopPKPD using a model based on a linear system of ODEs

PopPKPD using a model based on a linear system of ODEs



PopPKPD using a model based on a linear system of ODEs

PopPKPD using a model based on a linear system of ODEs



©2018

Stan, RStan and Torsten for PMX

29 May 2018

167 / 175

PopPKPD using a model based on a linear system of ODEs

Model

- PK: One compartment model with first order absorption
- PD: Sigmoid Emax function of an effect compartment concentration

$$\log(c_{ij}) \sim N(\log(\hat{c}_{ij}), \sigma_{PK}) \quad \log(E_{ij}) \sim N(\log(\hat{E}_{ij}), \sigma_{PD})$$

$$\hat{c}_{ij} = \frac{x_{2ij}}{V_{1j}} \quad \hat{E}_{ij} = E_{0j} + \frac{E_{\max j} c_{eij}^\gamma}{EC_{50}^\gamma + c_{eij}^\gamma} \quad c_{eij} = x_{4ij}$$

$$x'_j = K_j x_j \quad K = \begin{bmatrix} -k_a & 0 & 0 \\ ka & -\frac{CL}{V} & 0 \\ 0 & k_{e0} & -k_{e0} \end{bmatrix}$$

$$(CL_j, V_j, k_{aj}, k_{e0j}, E_{\max j}, E_{0j})$$

$$\sim N \left(\log \left(\widehat{CL} \left(\frac{bw_j}{70} \right)^{0.75} \right), \widehat{V} \left(\frac{bw_j}{70} \right), \widehat{k}_a, \widehat{k}_{e0}, \widehat{E}_{\max}, \widehat{E}_0 \right), \Omega \right)$$

PopPKPD using a model based on a linear system of ODEs

Model

- Weakly-moderately informative prior distributions:

$$\widehat{CL} \sim LN(\log(10), 1) \quad \widehat{V} \sim LN(\log(35), 1)$$

$$\widehat{k}_a \sim LN(\log(2), 1), \quad \widehat{k}_a > \frac{\widehat{CL}}{\widehat{V}}$$

$$\widehat{k_{e0}} \sim LN(\log(0.5), 1) \quad \widehat{E_0} \sim LN(\log(80), 1) \quad \widehat{E_{\max}} \sim LN(\log(40), 1)$$

$$EC_{50} \sim LN(\log(250), 1) \quad \gamma \sim LN(\log(2), 1)$$

$$\sigma_{PK} \sim \text{half-Cauchy}(0, 1) \quad \sigma_{PD} \sim \text{half-Cauchy}(0, 1) \quad \Omega = \text{diag}(\omega) P \text{ diag}(\omega)$$

$$\omega_i \sim \text{half-Cauchy}(0, 1), i \in \{1, 2, 3, 4, 5, 6\} \quad P \sim \text{LKJCorr}(1)$$

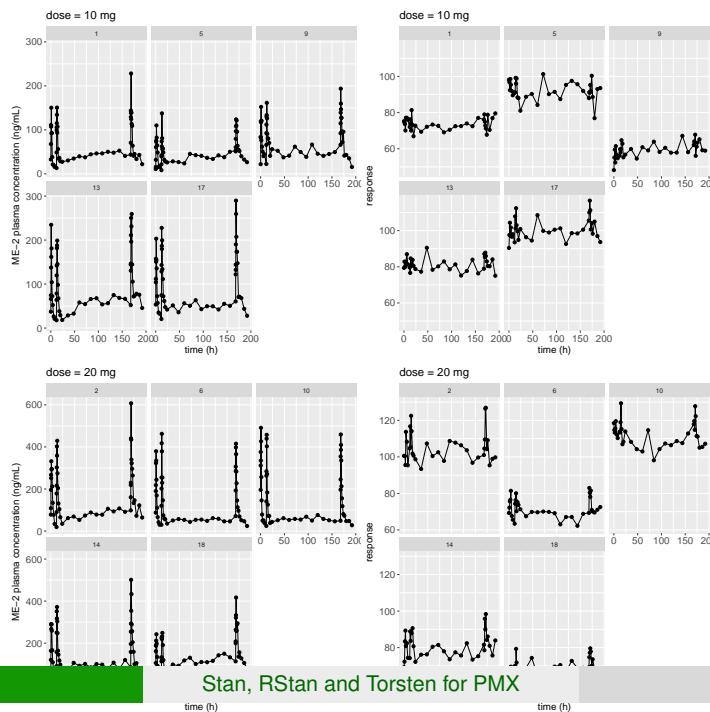
PopPKPD using a model based on a linear system of ODEs

Files

- Stan model for simulation: model/effectCpt2NcpSim.stan
- Stan model for fitting: model/effectCpt2Ncp.stan
- R script: script/effectCpt2Ncp.R
- The example also illustrates how to simulate data using Stan.

PopPKPD using a model based on a linear system of ODEs - alternate example

PopPKPD using a model based on a linear system of ODEs



©2018

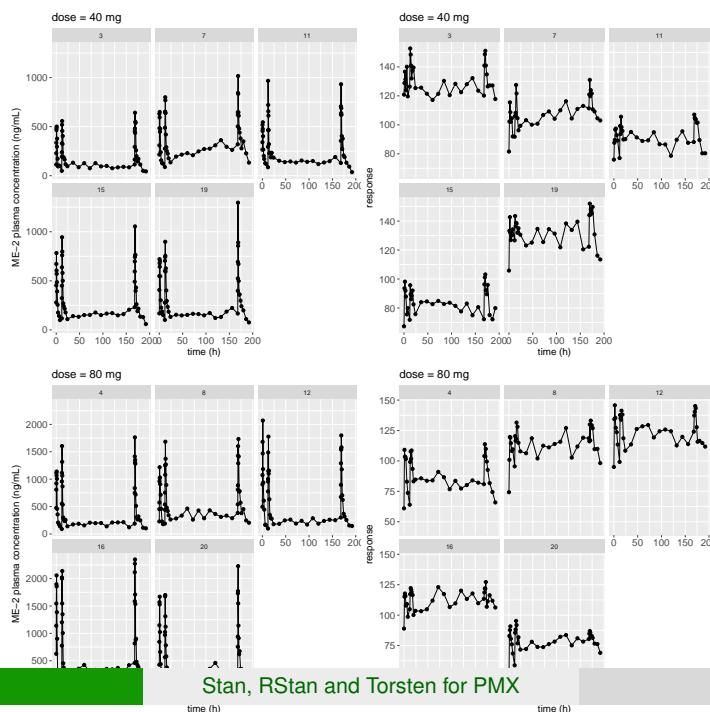
Stan, RStan and Torsten for PMX

29 May 2018

171 / 175

PopPKPD using a model based on a linear system of ODEs

PopPKPD using a model based on a linear system of ODEs



©2018

Stan, RStan and Torsten for PMX

29 May 2018

172 / 175

PopPKPD using a model based on a linear system of ODEs

Model

- PK: Two compartment model with first order absorption
- PD: Sigmoid Emax function of an effect compartment concentration

$$\log(c_{ij}) \sim N(\log(\hat{c}_{ij}), \sigma_{PK}) \quad \log(E_{ij}) \sim N(\log(\hat{E}_{ij}), \sigma_{PD})$$

$$\hat{c}_{ij} = \frac{x_{2ij}}{V_{1j}} \quad \hat{E}_{ij} = E_{0j} + \frac{E_{\max j} c_{eij}^\gamma}{EC_{50}^\gamma + c_{eij}^\gamma} \quad c_{eij} = x_{4ij}$$

$$x'_j = K_j x_j \quad K = \begin{bmatrix} -k_a & 0 & 0 & 0 \\ ka & -\frac{CL+Q}{V_1} & \frac{Q}{V_2} & 0 \\ 0 & \frac{Q}{V_1} & -\frac{Q}{V_2} & 0 \\ 0 & k_{e0} & 0 & -k_{e0} \end{bmatrix}$$

$$(CL_j, Q_j, V_{1j}, V_{2j}, k_{aj}, k_{e0j}, E_{\max j}, E_{0j}) \\ \sim N\left(\log\left(\widehat{CL}\left(\frac{bw_j}{70}\right)^{0.75}\right), \widehat{Q}\left(\frac{bw_j}{70}\right)^{0.75}, \widehat{V}_1\left(\frac{bw_j}{70}\right), \right. \\ \left. \widehat{V}_2\left(\frac{bw_j}{70}\right), \widehat{k}_a, \widehat{k}_{e0}, \widehat{E_{\max}}, \widehat{E_0}\right), \Omega$$

PopPKPD using a model based on a linear system of ODEs

Model

- Weakly-moderately informative prior distributions:

$$\widehat{CL} \sim LN(\log(10), 1) \quad \widehat{Q} \sim LN(\log(15), 1) \quad \widehat{V}_1 \sim LN(\log(35), 1)$$

$$\widehat{V}_2 \sim LN(\log(105), 1) \quad \widehat{k}_a \sim LN(\log(2), 1)$$

$$\widehat{k_{e0}} \sim LN(\log(0.5), 1) \quad \widehat{E_0} \sim LN(\log(80), 1) \quad \widehat{E_{\max}} \sim LN(\log(40), 1)$$

$$EC_{50} \sim LN(\log(250), 1) \quad \gamma \sim LN(\log(2), 1)$$

$$\sigma_{PK} \sim \text{half-Cauchy}(0, 1) \quad \sigma_{PD} \sim \text{half-Cauchy}(0, 1) \quad \Omega = \text{diag}(\omega) P \text{diag}(\omega)$$

$$\omega_i \sim \text{half-Cauchy}(0, 1), i \in \{1, 2, 3, 4, 5, 6, 7, 8\} \quad P \sim \text{LKJCorr}(1)$$

PopPKPD using a model based on a linear system of ODEs

Files

- Stan model for simulation: model/effectCptNcpSim.stan
- Stan model for fitting: model/effectCptNcp.stan
- R script: script/effectCptNcp.R
- The example also illustrates how to simulate data using Stan.