An Introduction to Non-linear Mixed-effects Models

Marie Davidian
Department of Statistics
North Carolina State University

http://www.stat.ncsu.edu/∼davidian

Non-Clinical Statistics Conference | 2008

23 September 2008

Outline

1. Introduction
2. Applications
3. Model formulation
4. Model interpretation and inferential objectives

   Break

5. Inferential approaches
6. Implementation and examples
7. Extensions
8. Discussion

Some references

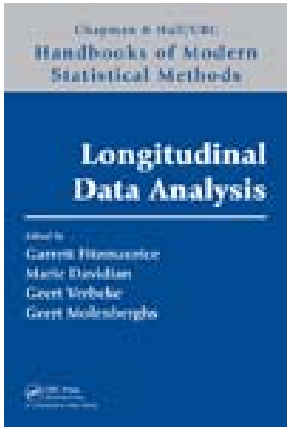
Material in this workshop is drawn from:

Davidian, M. and Giltinan, D.M. (1995). *Nonlinear Models for Repeated Measurement Data*. Chapman & Hall/CRC Press.

Davidian, M. and Giltinan, D.M. (2003). Nonlinear models for repeated measurement data: An overview and update. *Journal of Agricultural, Biological, and Environmental Statistics* **8**, 387–419.

Davidian, M. (2009). Non-linear mixed-effects models. In *Longitudinal Data Analysis*, G. Fitzmaurice, M. Davidian, G. Verbeke, and G. Molenberghs (eds). Chapman & Hall/CRC Press, ch. 5, 107–141.

Shameless promotion:



Introduction

Common situation in the biosciences:

- A *continuous response* evolves over *time* (or other condition) *within individuals* from a *population* of interest
- Scientific interest focuses on *features* or *mechanisms* that underlie *individual time trajectories* of the response and how these *vary* across the population
- A *theoretical* or *empirical model* for such individual profiles, typically *non-linear* in *parameters* that may be interpreted as representing such features or mechanisms, is available
- *Repeated measurements* over time are available on each individual in a *sample* drawn from the *population*
- *Inference* on the scientific questions of interest is to be made in the context of the *model* and its *parameters*

Introduction

Non-linear mixed-effects model:

- Also known as the *hierarchical non-linear model*
- A formal *statistical framework* for this situation
- Much *statistical methodological* research in the early 1990s
- Now *widely accepted* and *used*, with applications routinely reported and commercial and free *software* available
- *Extensions* and *methodological innovations* are still ongoing

Objectives of this workshop:

- Provide an *introduction* to the formulation, utility, and implementation of non-linear mixed models
- Focus on applications in *pharmaceutical* and *health sciences* research

Applications

Pharmacokinetics (PK): "*What the body does to the drug*"

- One of the most *important* application areas
- The area that *inspired* much of the methodological development for non-linear mixed-effects models
- *Broad goal*: Understand and characterize *intra-subject* processes of drug *absorption*, *distribution*, *metabolism* and *excretion* (*elimination*) governing achieved *drug concentrations*
- . . . and how these processes *vary* across subjects
- Critical for developing *dosing strategies*

An outstanding overview: "*Pharmacokinetics and pharmacodynamics*," by D.M. Giltinan, in *Encyclopedia of Biostatistics, 2nd edition*

Applications

PK studies in humans: Two types

- "*Intensive studies*"
  - *Small* number of subjects (often *healthy volunteers*)
  - *Frequent* samples over time, often following *single dose*
  - Usually *early* in drug development
  - Useful for gaining initial information on "*typical*" PK behavior in humans and for identifying an appropriate *PK model*...
- Preclinical PK studies in *animals* are generally *intensive* studies

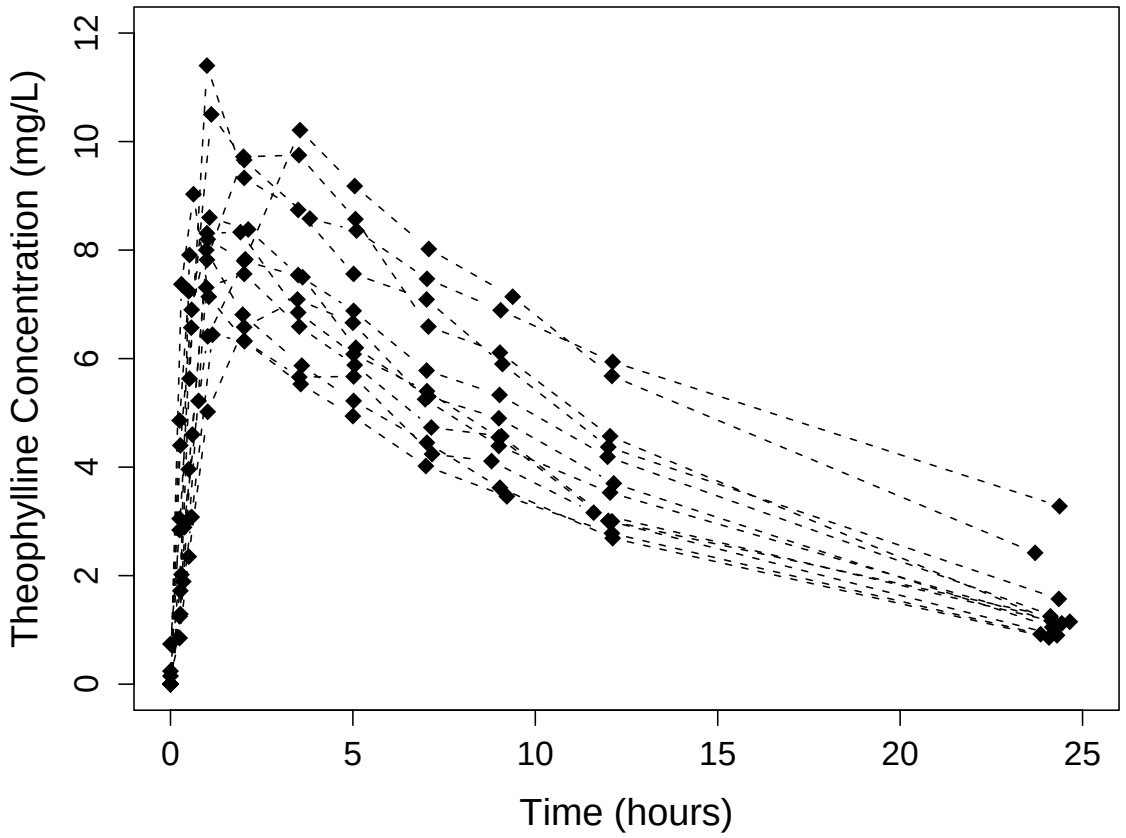
Applications

PK studies in humans: Two types

- "*Population studies*"
  - *Large* number of subjects (*heterogeneous* patients)
  - Often in *later stages* of drug development or after a drug is in *routine use*
  - *Haphazard* samples over time, *multiple dosing intervals*
  - Extensive *demographical* and *physiological* characteristics
  - Useful for understanding *associations* between *patient characteristics* and *PK behavior* $\Longrightarrow$ tailored *dosing recommendations*

Applications

Theophylline study: 12 subjects, same oral dose (mg/kg)



Applications

Features:

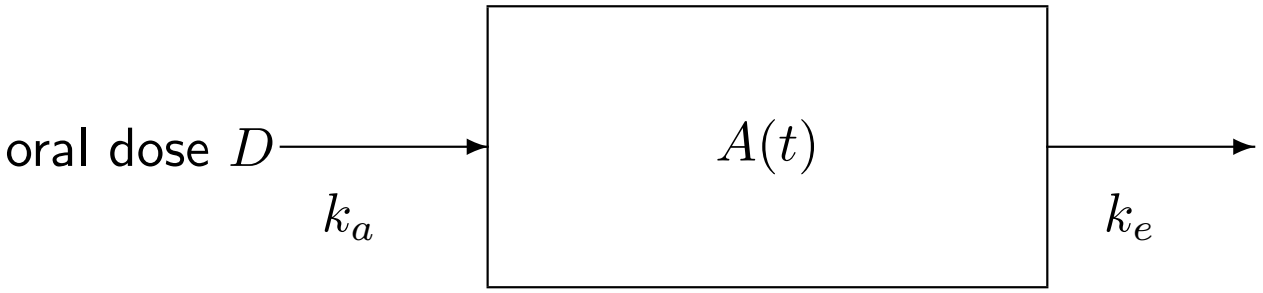
- *Intensive* study
- *Similarly shaped* concentration-time profiles across subjects
- . . . but peak, rise, decay *vary*
- Attributable to *inter-subject variation* in underlying PK behavior (absorption, distribution, elimination)

Standard: Represent the body by a simple system of *compartments*

- *Gross simplification* but extraordinarily useful. . .

Applications

One-compartment model with first-order absorption, elimination:



$$\begin{aligned} \frac{dA(t)}{dt} &= k_a A_a(t) - k_e A(t), \quad A(0) = 0 \\ \frac{dA_a(t)}{dt} &= -k_a A_a(t), \qquad\qquad A_a(0) = FA(0) \end{aligned}$$

$F =$ bioavailability, $A_a(t) =$ amount at absorption site

Concentration at t : $\quad m(t) = \dfrac{A(t)}{V} = \dfrac{k_a DF}{V(k_a - k_e)}\{\exp(-k_e t) - \exp(-k_a t)\},$

$k_e = Cl/V, \quad V =$ "*volume*" of compartment, $Cl =$ *clearance*

Applications

One-compartment model for theophylline:

- Single "*blood compartment*" with fractional rates of *absorption* k_a and *elimination* k_e
- *Deterministic mathematical model*
- *Individual PK behavior* characterized by *PK parameters* $\widehat{\boldsymbol{\theta}} = (k_a, V, Cl)'$

By-product:

- The PK model assumes PK processes are *dose-independent*
- $\Longrightarrow$ Knowledge of the values of $\widehat{\boldsymbol{\theta}} = (k_a, V, Cl)'$ allows *simulation* of concentrations achieved at any time t under *different doses*
- Can be used to develop *dosing regimens*

Applications

Objectives of analysis:

- Estimate "*typical*" values of $\boldsymbol{\theta} = (k_a, V, Cl)'$ and how they *vary* in the *population* of subjects based on the *longitudinal concentration* data from the *sample* of 12 subjects
- $\Longrightarrow$ Must *incorporate* the (*theoretical*) PK model in an appropriate *statistical model* (somehow...)

Applications

Argatroban study: Another *intensive* study

- Administered by *intravenous infusion* for 4 hours (240 min)
- $N = 37$ subjects assigned to different *constant infusion rates*
- *One-compartment model* with constant *intravenous infusion rate* D (μg/kg/min) for duration $t_{\text{inf}} = 240$ min

$$m(t) = \frac{D}{Cl}\left[\exp\left\{-\frac{Cl}{V}(t - t_{\text{inf}})_+\right\} - \exp\left(-\frac{Cl}{V}t\right)\right], \quad \boldsymbol{\theta} = (Cl, V)'$$

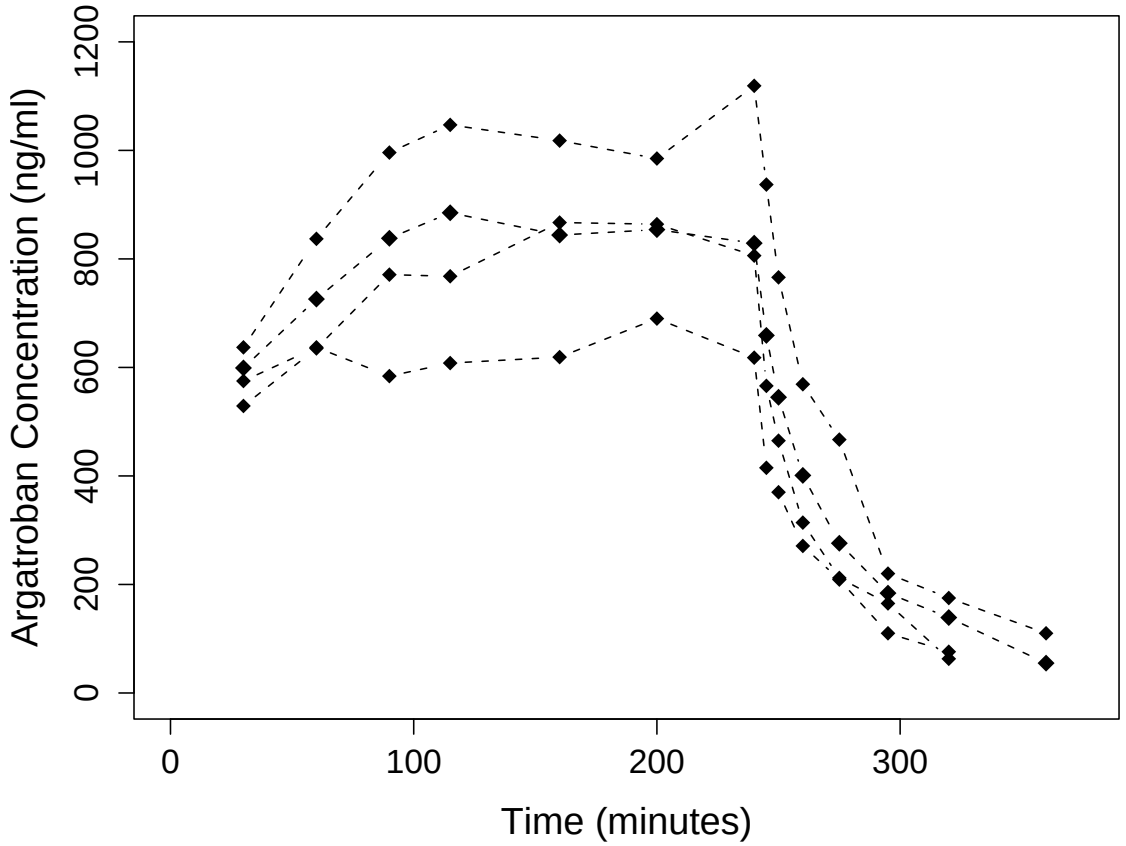
$$x_+ = 0 \text{ if } x \leq 0 \text{ and } x_+ = x \text{ if } x > 0$$

Objectives of analysis:

- Estimate "*typical*" values of $\boldsymbol{\theta} = (Cl, V)'$ and how they *vary* in the population of subjects
- Understand relationship between achieved concentrations and a clinical or other response (*pharmacodynamics*; more later...)

Applications

Profiles for 4 subjects receiving 4.5 μg/kg-min:



Applications

Quinidine population study: $N = 136$ patients undergoing treatment with oral quinidine for atrial fibrillation or arrhythmia

- *Demographical/physiological characteristics*: Age, weight, height, ethnicity/race, smoking status, ethanol abuse, congestive heart failure, creatinine clearance, α_1-acid glycoprotein concentration, ...
- Samples taken over *multiple dosing intervals* $\Longrightarrow$ (dose time, amount) $= (s_\ell, D_\ell)$ for the ℓth dose interval
- Standard assumption: "*Principle of superposition*" $\Longrightarrow$ multiple doses are "*additive*"
- *One compartment model* gives expression for concentration at time t...

Applications

For a subject not yet at a steady state:

$$\begin{aligned} A_a(s_\ell) &= A_a(s_{\ell-1})\exp\{-k_a(s_\ell - s_{\ell-1})\} + D_\ell, \\ m(s_\ell) &= m(s_{\ell-1})\exp\{-k_e(s_\ell - s_{\ell-1})\} + A_a(s_{\ell-1})\frac{k_a}{V(k_a-k_e)} \\ &\quad \times\Big[\exp\{-k_e(s_\ell - s_{\ell-1})\} - \exp\{-k_a(s_\ell - s_{\ell-1})\}\Big]. \end{aligned}$$

$$\begin{aligned} m(t) &= m(s_\ell)\exp\{-k_e(t-s_\ell)\} + A_a(s_\ell)\frac{k_a}{V(k_a-k_e)} \\ &\quad \times\Big[\exp\{-k_e(t-s_\ell)\} - \exp\{-k_a(t-s_\ell)\}\Big], \quad s_\ell < t < s_{\ell+1} \end{aligned}$$

$$k_e = Cl/V, \quad \boldsymbol{\theta} = (k_a, V, Cl)'$$

Objective of analysis: Characterize *typical values* of and *variation* in $\boldsymbol{\theta} = (k_a, V, Cl)'$ across the population and elucidate *systematic associations* between $\boldsymbol{\theta}$ and *patient characteristics*

Applications

Data for a representative subject:

time (hours)	conc. (mg/L)	dose (mg)	age (years)	weight (kg)	creat. (ml/min)	glyco. (mg/dl)
0.00	–	166	75	108	> 50	69
6.00	–	166	75	108	> 50	69
11.00	–	166	75	108	> 50	69
17.00	–	166	75	108	> 50	69
23.00	–	166	75	108	> 50	69
27.67	0.7	–	75	108	> 50	69
29.00	–	166	75	108	> 50	94
35.00	–	166	75	108	> 50	94
41.00	–	166	75	108	> 50	94
47.00	–	166	75	108	> 50	94
53.00	–	166	75	108	> 50	94
65.00	–	166	75	108	> 50	94
71.00	–	166	75	108	> 50	94
77.00	0.4	–	75	108	> 50	94
161.00	–	166	75	108	> 50	88
168.75	0.6	–	75	108	> 50	88

height=72 inches, Caucasian, smoker, no ethanol abuse, no CHF

Applications

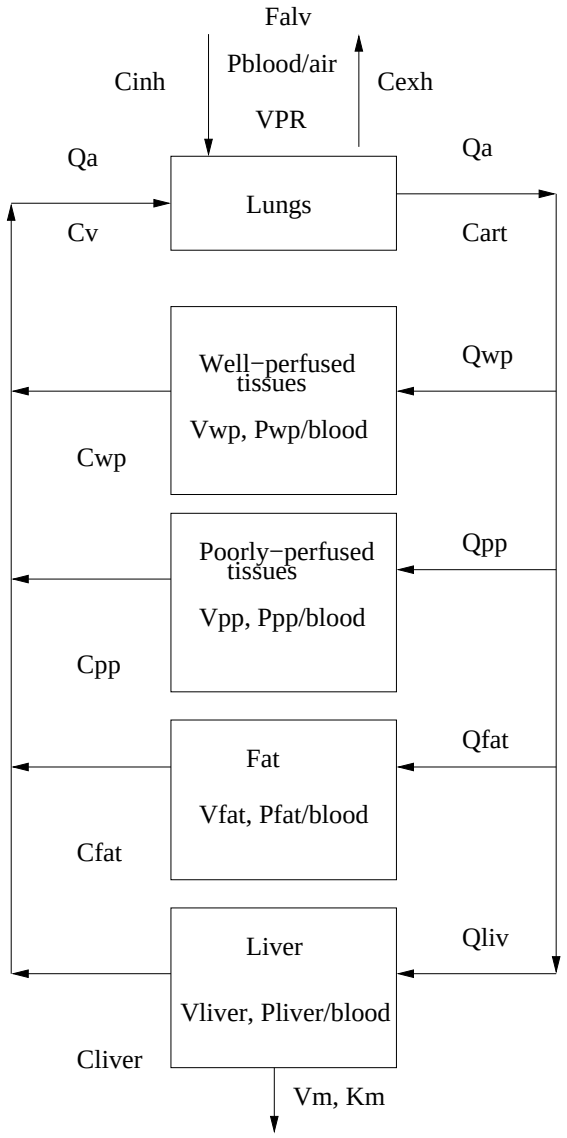
Toxicokinetics: *Physiologically-based pharmacokinetic* (*PBPK*) models

- PK of *environmental*, *chemical* agents; studies often in *animals*
- N animals exposed, *repeated concentrations* over time on each
- More "*realistic*" representation of the body (e.g., *organ, tissue compartments*)
- System of differential equations *cannot* be solved *analytically*
- *Lots* of PK parameters, some measurable, some *unknown*: Compartment volumes V, partition coefficients P, flow rates F, metabolic parameters V_{max}, K_m, etc

Objectives of analysis:

- Characterize in particular *metabolic mechanisms* (V_{max}, K_m) and how these *vary* in the population
- Understand relationship between *metabolic processes* and *toxicities*

Applications



$$C_{\text{art}} = \frac{F_{\text{card}} C_{\text{ven}} + F_{\text{alv}} C_{\text{inh}}}{F_{\text{card}} + F_{\text{alv}}/P_{\text{blood/air}}}, \quad C_{\text{ven}} = \sum_s \frac{F_s C_s}{F_{\text{card}}}$$

$$C_{\text{exh}} = (1-\delta)\frac{C_{\text{art}}}{P_{\text{blood/air}}} + \delta C_{\text{inh}}$$

$$\frac{dC_s}{dt} = \frac{F_s}{V_s}\left(C_{\text{art}} - \frac{C_s}{P_{s/\text{blood}}}\right), \quad s = \text{wp, pp, fat}$$

$$\frac{dC_{\text{liv}}}{dt} = \frac{F_{\text{liv}}}{V_{\text{liv}}}\left(C_{\text{art}} - \frac{C_{\text{liv}}}{P_{\text{liv/blood}}}\right) - R_{\text{liv}} \ (s = \text{liv}),$$

$$R_{\text{liv}} = \frac{V_{\text{max}} C_{\text{liv}}}{V_{\text{liv}}(K_m + C_{\text{liv}})},$$

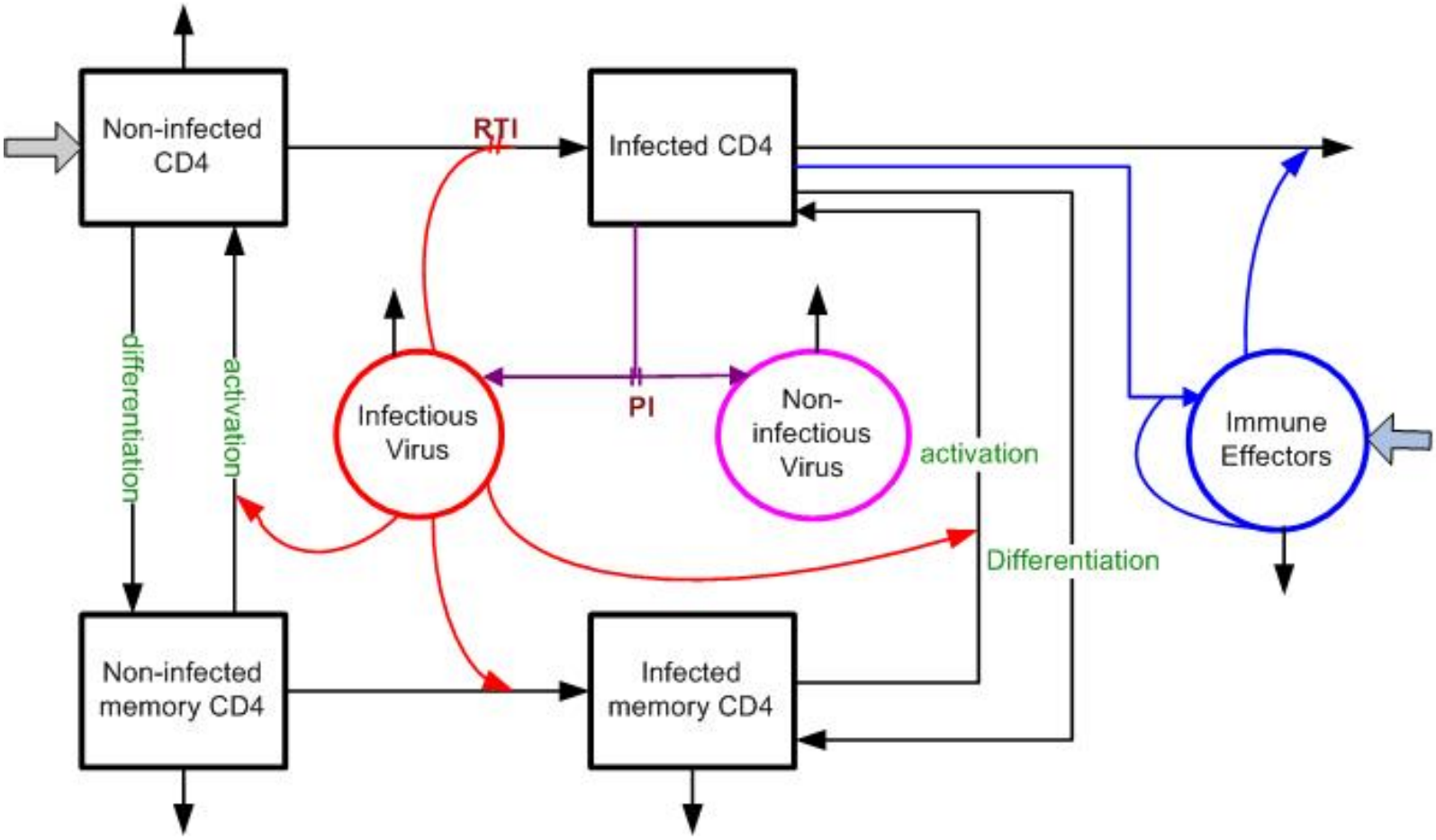
Applications

HIV dynamics: Human immunodeficiency virus (HIV), attacks the *immune system*

- *Broad goal*: Characterize *mechanisms* underlying the *interaction* between HIV and the immune system over time governing *disease progression* and the effects of *anti-retroviral treatments* (ART)
- *Typical study*: N subjects, *repeated measurements* on *viral load* (virologic status), *CD4+ T cell count* (immunologic status) over time (possibly on/off ART)
- *Compartmental* representation of *mechanisms* taking place within an *infected subject*
- System of (*deterministic*) *nonlinear ordinary differential equations*; $\Longrightarrow$ *viral load*, *CD4+ T cell count*, etc, at any time

Applications

Simple model for within-subject HIV dynamics:



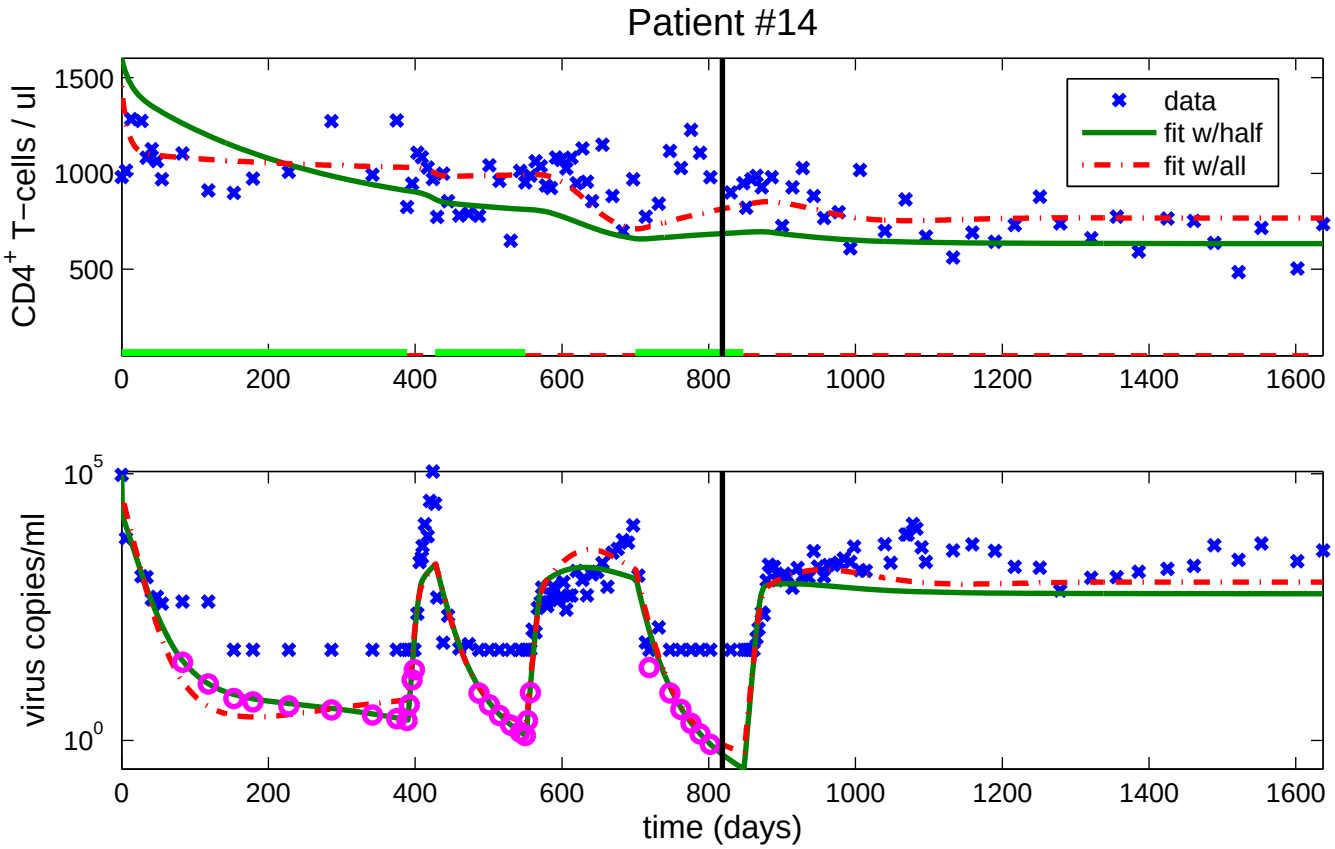
Applications

Differential equations:

$$
\begin{aligned}
\dot{T}_1 &= \lambda_1 - d_1 T_1 - \{1 - \epsilon_1 U(t)\} k_1 V_I T_1 \\
\dot{T}_2 &= \lambda_2 - d_2 T_2 - \{1 - f\epsilon_1 U(t)\} k_2 V_I T_2 \\
\dot{T}_1^* &= \{1 - \epsilon_1 U(t)\} k_1 V_I T_1 - \delta T_1^* - m_2 E T_1^* \\
\dot{T}_2^* &= \{1 - f\epsilon_1 U(t)\} k_2 V_I T_2 - \delta T_2^* - m_2 E T_2^* \\
\dot{V}_I &= \{1 - \epsilon_2 U(t)\} 10^3 N_T \delta (T_1^* + T_2^*) - c V_I \\
&\quad - \{1 - \epsilon_1 U(t)\} \rho_1 10^3 k_1 T_1 V_I - \{1 - f\epsilon_1 U(t)\} \rho_2 10^3 k_2 T_2 V_I \\
\dot{V}_{NI} &= \epsilon_2 U(t) 10^3 N_T \delta (T_1^* + T_2^*) - c V_{NI} \\
\dot{E} &= \lambda_E + \frac{b_E (T_1^* + T_2^*)}{(T_1^* + T_2^*) + K_b} E - \frac{d_E (T_1^* + T_2^*)}{(T_1^* + T_2^*) + K_d} E - \delta_E E
\end{aligned}
$$

- $\boldsymbol{\theta} = (\lambda_1, d_1, \epsilon_1, k_1, \ldots)'$ plus initial conditions
- Observable: *CD4 count* $= T_1 + T_1^*$, *viral load* $= V_I + V_{NI}$
- $U(t)$ = ART input at t ($0 \leq U(t) \leq 1$, 0 = off, 1 = on)

Applications



Objectives of analysis: Characterize *typical values* of and *variation* in $\boldsymbol{\theta}$ across the population, elucidate *systematic associations* between $\boldsymbol{\theta}$ and *patient characteristics*, *simulate* disease progression under different $U(t)$

Applications

Summary: Common *themes*

- A *response* (or responses) evolves over time (e.g., *concentration* in PK)
- Interest focuses on *underlying mechanisms/processes* taking place *within an individual* leading to response trajectories and how these *vary* across the population
- A (usually *deterministic*) *model* is available representing mechanisms explicitly by scientifically meaningful model *parameters*
- Mechanisms cannot be *observed directly*
- $\Longrightarrow$ Inference on mechanisms must be based on *repeated measurements* of the response over time on each of a sample of N individuals from the population

Applications

Other application areas:

- Stability testing
- Agriculture
- Forestry
- Dairy science
- Cancer dynamics
- Many more . . .

For definiteness: We will use *PK* as a running example

Model formulation

Non-linear mixed effects model: *Embed* the (*deterministic*) *model* describing *individual* trajectories in a *statistical model*

- Formalizes knowledge and *assumptions* about *variation* in responses and mechanisms *within* and *among* individuals
- Provides a framework for *inference* based on *repeated measurement* data from N individuals
- *For simplicity*: Focus on *univariate response* (= *drug concentration* in PK); some discussion of *multivariate response* at the end

Basic set-up: N individuals from a population of interest, $i = 1, \ldots, N$

- For individual i, observe n_i measurements of the response

$$Y_{i1}, Y_{i2}, \ldots, Y_{in_i} \quad \text{at times} \quad t_{i1}, t_{i2}, \ldots, t_{in_i}$$

- I.e., for individual i, Y_{ij} at time t_{ij}, $j = 1, \ldots, n_i$

Model formulation

Within-individual conditions of observation: For individual i, $\boldsymbol{U}_i$

- *Theophylline*: $\boldsymbol{U}_i = D_i =$ oral dose for i at time 0 (mg/kg)
- *Argatroban*: $\boldsymbol{U}_i = (D_i, t_{\text{inf}}) =$ infusion rate and duration for i
- *Quinidine*: For subject i observed over d_i dosing intervals, $\boldsymbol{U}_i$ has elements $(s_{i\ell}, D_{i\ell})'$, $\ell = 1, \ldots, d_i$
- *HIV dynamics*: $\boldsymbol{U}_i$ is continuous function $U_i(t)$ with subject i's known treatment status at any time t
- $\boldsymbol{U}_i$ are "*within-individual covariates*" – needed to describe response-time relationship at the *individual level*

Model formulation

Individual characteristics: For individual i, $\boldsymbol{A}_i$

- Age, weight, ethnicity, smoking status, etc. . .
- *For now*: Elements of $\boldsymbol{A}_i$ *do not change* over observation period (will discuss changing elements later)
- $\boldsymbol{A}_i$ are "*among-individual covariates*" – relevant only to how *individuals differ* but are *not needed* to describe response-time relationship at individual level

Observed data: $(\boldsymbol{Y}_i', \boldsymbol{X}_i')'$, $i = 1, \ldots, N$, *assumed independent* across i

- $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})'$
- $\boldsymbol{X}_i = (\boldsymbol{U}_i', \boldsymbol{A}_i')' =$ combined within- and among-individual covariates (for brevity later)

Basic model: A *two-stage hierarchy*

Model formulation

Stage 1 – Individual-level model:

$$Y_{ij} = m(t_{ij}, \boldsymbol{U}_i, \boldsymbol{\theta}_i) + e_{ij}, \;\; j = 1, \ldots, n_i, \quad \boldsymbol{\theta}_i \;\; (r \times 1)$$

- E.g., for *theophylline* $(F \equiv 1)$

$$m(t, \boldsymbol{U}_i, \boldsymbol{\theta}_i) = \frac{k_{ai} D_i}{V_i(k_{ai} - Cl_i/V_i)} \{\exp(-Cl_i t/V_i) - \exp(-k_{ai} t)\}$$

$$\boldsymbol{\theta}_i = (k_{ai}, V_i, Cl_i)' = (\theta_{i1}, \theta_{i2}, \theta_{i3})', \;\; r = \;\; 3, \quad \boldsymbol{U}_i = D_i$$

- *Assume* $e_{ij} = Y_{ij} - m(t_{ij}, \boldsymbol{U}_i, \boldsymbol{\theta}_i)$ satisfy $E(e_{ij} \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i) = 0$

$$\Longrightarrow \quad E(Y_{ij} \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i) = m(t_{ij}, \boldsymbol{U}_i, \boldsymbol{\theta}_i) \;\; \text{for each } j$$

- *Standard assumption*: e_{ij} and hence Y_{ij} are *conditionally normally distributed* (on $\boldsymbol{U}_i$, $\boldsymbol{\theta}_i$)

- *More shortly*. . .

Model formulation

Stage 2 – Population model:

$$\boldsymbol{\theta}_i = \boldsymbol{d}(\boldsymbol{A}_i, \boldsymbol{\beta}, \boldsymbol{b}_i), \;\; i = 1, \ldots, N, \quad (r \times 1)$$

- $\boldsymbol{d}$ is r-dimensional function describing *relationship* between $\boldsymbol{\theta}_i$ and $\boldsymbol{A}_i$ in terms of ...
- $\boldsymbol{\beta}$ $(p \times 1)$ *fixed parameter* ("*fixed effects*")
- $\boldsymbol{b}_i$ $(q \times 1)$ "*random effects*"
- Characterizes how elements of $\boldsymbol{\theta}_i$ *vary* across individual due to
  - *Systematic associations* with $\boldsymbol{A}_i$ (modeled via $\boldsymbol{\beta}$)
  - "*Unexplained variation*" in the population (represented by $\boldsymbol{b}_i$)
- *Usual assumptions*:

$$E(\boldsymbol{b}_i \,|\, \boldsymbol{A}_i) = E(\boldsymbol{b}_i) = \mathbf{0} \quad \text{and} \quad \text{Cov}(\boldsymbol{b}_i \,|\, \boldsymbol{A}_i) = \text{Cov}(\boldsymbol{b}_i) = G, \quad \boldsymbol{b}_i \sim N(\mathbf{0}, G)$$

Model formulation

Stage 2 – Population model:

$$\boldsymbol{\theta}_i = \boldsymbol{d}(\boldsymbol{A}_i, \boldsymbol{\beta}, \boldsymbol{b}_i), \;\; i = 1, \ldots, N$$

Example: *Quinidine*, $\boldsymbol{\theta}_i = (k_{ai}, V_i, Cl_i)'$ $(r = 3)$

- $\boldsymbol{A}_i = (w_i, \delta_i, a_i)'$, $w_i =$ weight, , $a_i =$ age,
  $\delta_i = I(\text{creatinine clearance} > 50 \text{ ml/min})$
- $\boldsymbol{b}_i = (b_{i1}, b_{i2}, b_{i3})'$ $(q = 3)$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_7)'$ $(p = 7)$

$$\begin{aligned}
k_{ai} &= \theta_{i1} = d_1(\boldsymbol{A}_i, \boldsymbol{\beta}, \boldsymbol{b}_i) = \exp(\beta_1 + b_{i1}), \\
V_i &= \theta_{i2} = d_2(\boldsymbol{A}_i, \boldsymbol{\beta}, \boldsymbol{b}_i) = \exp(\beta_2 + \beta_4 w_i + b_{i2}), \\
Cl_i &= \theta_{i3} = d_3(\boldsymbol{A}_i, \boldsymbol{\beta}, \boldsymbol{b}_i) = \exp(\beta_3 + \beta_5 w_i + \beta_6 \delta_i + \beta_7 a_i + b_{i3}),
\end{aligned}$$

- *Positivity* of k_{ai}, V_i, Cl_i enforced
- If $\boldsymbol{b}_i \sim N(\boldsymbol{0}, G)$, k_{ai}, V_i, Cl_i are each *lognormally distributed* in the population

Model formulation

Stage 2 – Population model:

$$\boldsymbol{\theta}_i = \boldsymbol{d}(\boldsymbol{A}_i, \boldsymbol{\beta}, \boldsymbol{b}_i), \;\; i = 1, \ldots, N$$

Example: *Quinidine*, continued, $\boldsymbol{\theta}_i = (k_{ai}, V_i, Cl_i)'$ $(r = 3)$

- "*Are elements of* $\boldsymbol{\theta}_i$ *fixed or random effects*?"
- "*Unexplained variation*" in one component of $\boldsymbol{\theta}_i$ "*small*" relative to others – *no* associated random effect, e.g., $r = 3$, $q = 2$

$$\begin{aligned} k_{ai} &= \exp(\beta_1 + b_{i1}) \\ V_i &= \exp(\beta_2 + \beta_4 w_i) \quad \text{(all population variation due to weight)} \\ Cl_i &= \exp(\beta_3 + \beta_5 w_i + \beta_6 \delta_i + \beta_7 a_i + b_{i3}) \end{aligned}$$

- *An approximation* – usually *biologically implausible*; used for *parsimony*, *numerical stability*

Model formulation

Stage 2 – Population model:

$$\boldsymbol{\theta}_i = \boldsymbol{d}(\boldsymbol{A}_i, \boldsymbol{\beta}, \boldsymbol{b}_i), \;\; i = 1, \ldots, N$$

- Allows *non-linear* (in $\boldsymbol{\beta}$ and $\boldsymbol{b}_i$) specifications for elements of $\boldsymbol{\theta}_i$
- May be more appropriate than *linear* specifications (*positivity* requirements, *skewed* distributions)

Some accounts: Restrict to *linear* specification

$$\boldsymbol{\theta}_i = A_i\boldsymbol{\beta} + B_i\boldsymbol{b}_i$$

- A_i $(r \times p)$ "*design matrix*" depending on elements of $\boldsymbol{A}_i$
- B_i $(r \times q)$ typically 0s and 1s (*identity* matrix if $r = q$)
- Mainly in the *statistical literature*

Model formulation

Stage 2 – Linear population model:

$$\boldsymbol{\theta}_i = A_i\boldsymbol{\beta} + B_i\boldsymbol{b}_i$$

Example: *Quinidine*, continued

- *Reparameterize* in terms of $\boldsymbol{\theta}_i = (k_{ai}^*, V_i^*, Cl_i^*)'$, $k_{ai}^* = \log(k_{ai})$, $V_i^* = \log(V_i)$, and $Cl_i^* = \log(Cl_i)$ $(r = 3)$

$$\begin{aligned} k_{ai}^* &= \beta_1 + b_{i1}, \\ V_i^* &= \beta_2 + \beta_4 w_i + b_{i2}, \\ Cl_i^* &= \beta_3 + \beta_5 w_i + \beta_6 \delta_i + \beta_7 a_i + b_{i3} \end{aligned}$$

$$A_i = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & w_i & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & w_i & \delta_i & a_i \end{pmatrix}, \qquad B_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Model formulation

Within-individual considerations: Complete the Stage 1 individual-level model

- *Assumptions* on the distribution of $\boldsymbol{Y}_i$ given $\boldsymbol{U}_i$ and $\boldsymbol{\theta}_i$
- Focus on a *single individual* i observed under conditions $\boldsymbol{U}_i$
- Y_{ij} at times t_{ij} viewed as *intermittent* observations on a *stochastic process*

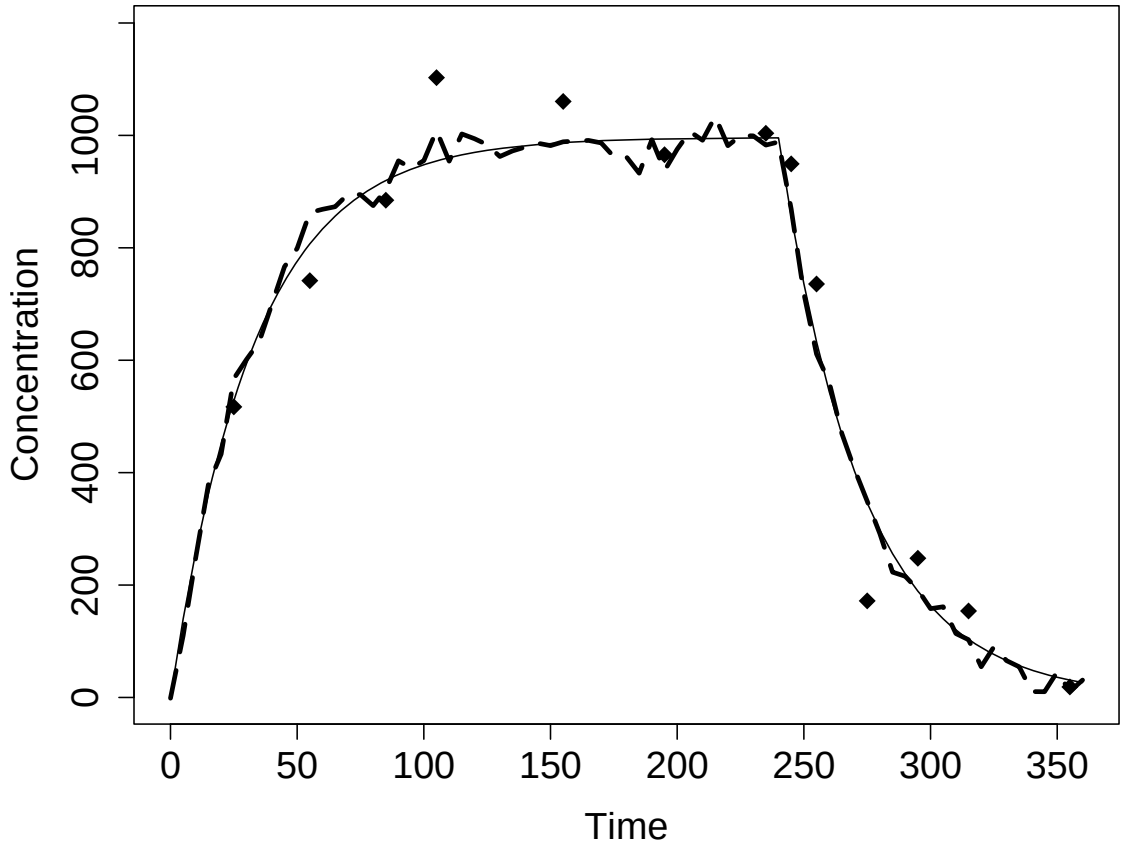
$$Y_i(t, \boldsymbol{U}_i) = m(t, \boldsymbol{U}_i, \boldsymbol{\theta}_i) + e_i(t, \boldsymbol{U}_i)$$

$E\{e_i(t, \boldsymbol{U}_i) \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i\} = 0, \quad E\{Y_i(t, \boldsymbol{U}_i) \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i\} = m(t, \boldsymbol{U}_i, \boldsymbol{\theta}_i)$ for all t

- $Y_{ij} = Y_i(t_{ij}, \boldsymbol{U}_i)$, $e_{ij} = e_i(t_{ij}, \boldsymbol{U}_i)$
- "*Deviation*" process $e_i(t, \boldsymbol{U}_i)$ represents *all sources of variation* acting *within an individual* causing a realization of $Y_i(t, \boldsymbol{U}_i)$ to deviate from the "*smooth*" trajectory $m(t, \boldsymbol{U}_i, \boldsymbol{\theta}_i)$

Model formulation

Conceptualization:



Model formulation

Conceptual interpretation:

- *Solid line*: $m(t, \boldsymbol{U}_i, \boldsymbol{\theta}_i)$ represents "*inherent tendency*" for i's response to evolve over time; depends on i's "*inherent characteristics*" $\boldsymbol{\theta}_i$
- *Dashed line*: Actual realization of the response – *fluctuates* about solid line because $m(t, \boldsymbol{U}_i, \boldsymbol{\theta}_i)$ is a *simplification* of complex truth
- *Symbols*: Actual, *intermittent measurements* of the dashed line – *deviate* from the dashed line due to *measurement error*

Result: *Two sources of intra-individual variation*

- "*Realization deviation*"
- *Measurement error variation*
- $m(t, \boldsymbol{U}_i, \boldsymbol{\theta}_i)$ is the *average* of *all possible realizations* of measured response trajectory that could be observed on i

Model formulation

To formalize: $e_i(t, \boldsymbol{U}_i) = e_{R,i}(t, \boldsymbol{U}_i) + e_{M,i}(t, \boldsymbol{U}_i)$

- Within-individual *stochastic process*

$$Y_i(t, \boldsymbol{U}_i) = m(t, \boldsymbol{U}_i, \boldsymbol{\theta}_i) + e_{R,i}(t, \boldsymbol{U}_i) + e_{M,i}(t, \boldsymbol{U}_i)$$

$$E\{e_{R,i}(t, \boldsymbol{U}_i) \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i\} = E\{e_{M,i}(t, \boldsymbol{U}_i) \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i\} = 0$$

- $\Longrightarrow Y_{ij} = Y_i(t_{ij}, \boldsymbol{U}_i)$, $e_{R,i}(t_{ij}, \boldsymbol{U}_i) = e_{R,ij}$, $e_{M,i}(t_{ij}, \boldsymbol{U}_i) = e_{M,ij}$

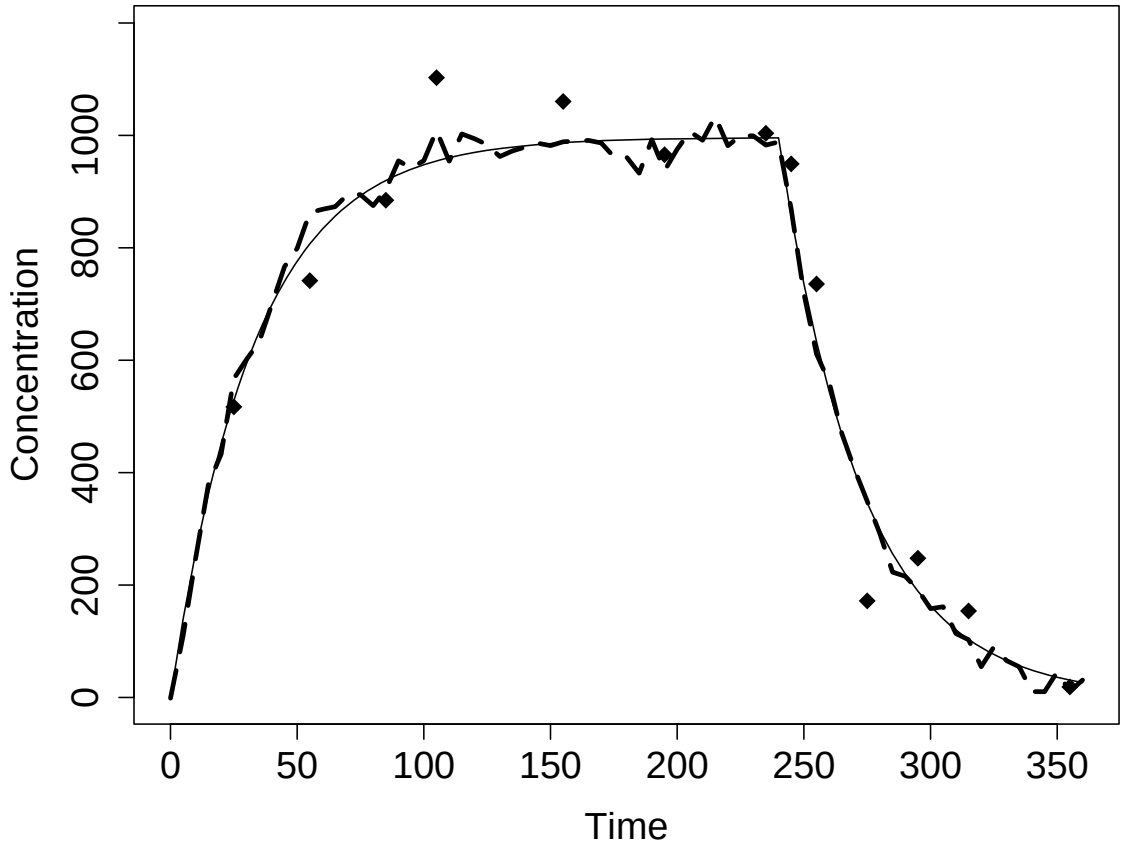
$$Y_{ij} = m(t_{ij}, \boldsymbol{U}_i, \boldsymbol{\theta}_i) + \underbrace{e_{R,ij} + e_{M,ij}}_{e_{ij}}$$

$$\boldsymbol{e}_{R,i} = (e_{R,i1}, \ldots, e_{R,in_i})', \quad \boldsymbol{e}_{M,i} = (e_{M,i1}, \ldots, e_{M,in_i})'$$

- $e_{R,i}(t, \boldsymbol{U}_i) =$ "*realization deviation process*"
- $e_{M,i}(t, \boldsymbol{U}_i) =$ "*measurement error deviation process*"
- *Assumptions* on $e_{R,i}(t, \boldsymbol{U}_i)$ and $e_{M,i}(t, \boldsymbol{U}_i)$ lead to a model for $\text{Cov}(\boldsymbol{e}_i \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i)$ and hence $\text{Cov}(\boldsymbol{Y}_i \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i)$

Model formulation

Conceptualization:



Model formulation

Realization deviation process:

- Natural to expect $e_{R,i}(t, \boldsymbol{U}_i)$ and $e_{R,i}(s, \boldsymbol{U}_i)$ at times t and s to be *positively correlated*, e.g.,

$$\text{corr}\{e_{R,i}(t, \boldsymbol{U}_i), e_{R,i}(s, \boldsymbol{U}_i) \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i\} = \exp(-\rho|t-s|), \quad \rho \geq 0$$

- Assume variation of realizations about $m(t, \boldsymbol{U}_i, \boldsymbol{\theta}_i)$ are of *similar magnitude* over time and individuals, e.g.,

$$\text{Var}\{e_{R,i}(t, \boldsymbol{U}_i) \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i\} = \sigma_R^2 \geq 0 \quad (\textit{constant} \text{ for all } t)$$

- *Or* assume variation depends on $m(t, \boldsymbol{U}_i, \boldsymbol{\theta}_i)$, e.g.,

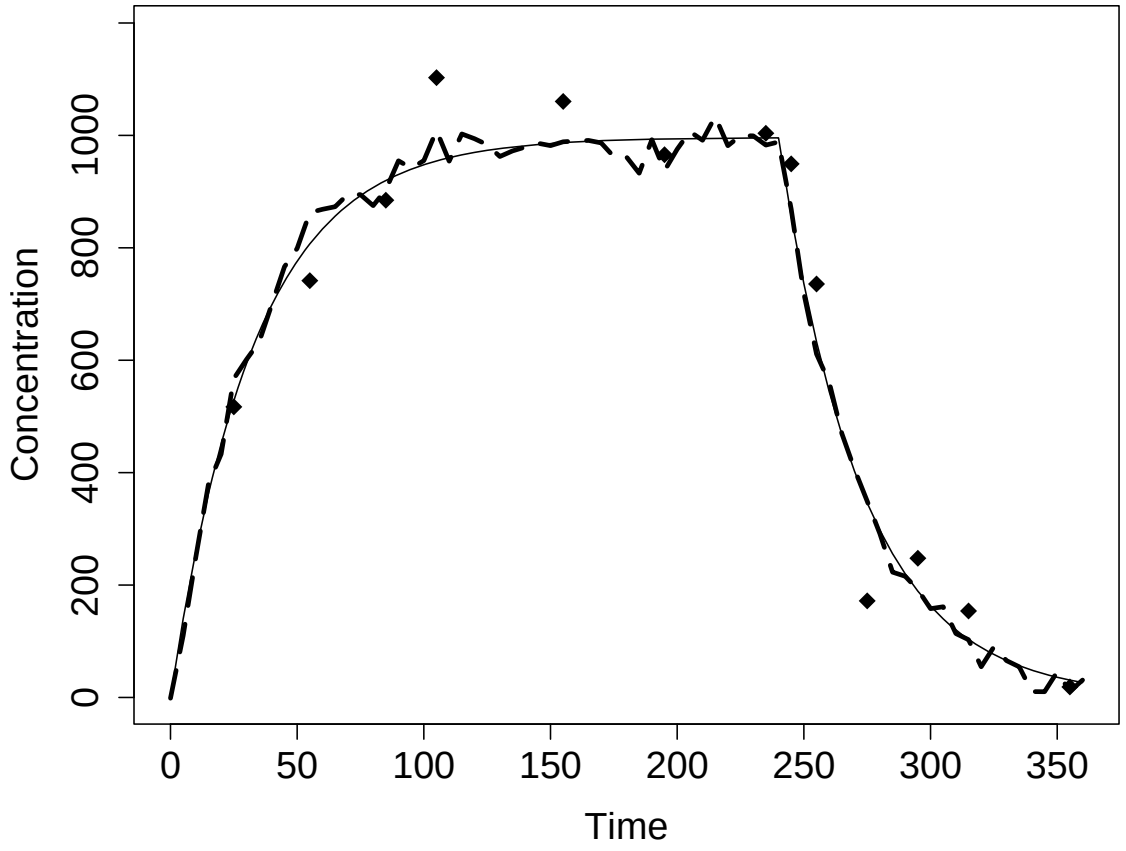
$$\text{Var}\{e_{R,i}(t, \boldsymbol{U}_i) \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i\} = \sigma_R^2 \{m(t, \boldsymbol{U}_i, \boldsymbol{\theta}_i)\}^{2\eta}, \quad \eta > 0$$

- *Result*: Assumptions imply a *covariance model* $(n_i \times n_i)$

$$\text{Cov}(\boldsymbol{e}_{R,i} \,|\, \boldsymbol{U})_i, \boldsymbol{\theta}_i) = V_{R,i}(\boldsymbol{U}_i, \boldsymbol{\theta}_i, \boldsymbol{\alpha}_R), \quad \boldsymbol{\alpha}_R = (\sigma_R^2, \rho)' \text{ or } \boldsymbol{\alpha}_R = (\sigma_R^2, \rho, \eta)'$$

Model formulation

Conceptualization:



Model formulation

Measurement error deviation process:

- Measuring devices commit *haphazard errors* $\Longrightarrow$

$$\text{corr}\{e_{M,i}(t, \boldsymbol{U}_i), e_{M,i}(s, \boldsymbol{U}_i) \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i\} = 0 \text{ for all } t > s$$

- Assume magnitude of errors is *similar* regardless of level, e.g.,

$$\text{Var}\{e_{M,i}(t, \boldsymbol{U}_i) \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i\} = \sigma_M^2 \geq 0 \quad (\textit{constant} \text{ for all } t)$$

- *Or* assume magnitude changes with level; often *approximated* under *assumption* $\text{Var}\{e_{R,i}(t, \boldsymbol{U}_i) \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i\} << \text{Var}\{e_{M,i}(t, \boldsymbol{U}_i) \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i\}$

$$\text{Var}\{e_{M,i}(t, \boldsymbol{U}_i) \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i\} = \sigma_M^2 \{m(t, \boldsymbol{U}_i, \boldsymbol{\theta}_i)\}^{2\zeta}, \quad \zeta > 0$$

- *Result*: Assumptions imply a *covariance model* $(n_i \times n_i)$ (*diagonal matrix*)

$$\text{Cov}(\boldsymbol{e}_{M,i} \,|\, \boldsymbol{U})_i, \boldsymbol{\theta}_i) = V_{M,i}(\boldsymbol{U}_i, \boldsymbol{\theta}_i, \boldsymbol{\alpha}_R), \quad \boldsymbol{\alpha}_M = \sigma_M^2 \text{ or } \boldsymbol{\alpha}_M = (\sigma_M^2, \zeta)'$$

Model formulation

Combining:

- *Standard assumption*: $e_{R,i}(t, \boldsymbol{U}_i)$ and $e_{M,i}(t, \boldsymbol{U}_i)$ are *independent*

$$\begin{aligned} \text{Cov}(\boldsymbol{e}_i \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i) &= \text{Cov}(\boldsymbol{e}_{R,i} \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i) + \text{Cov}(\boldsymbol{e}_{M,i} \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i) \\ &= V_{R,i}(\boldsymbol{U}_i, \boldsymbol{\theta}_i, \boldsymbol{\alpha}_R) + V_{M,i}(\boldsymbol{U}_i, \boldsymbol{\theta}_i, \boldsymbol{\alpha}_M) \\ &= V_i(\boldsymbol{U}_i, \boldsymbol{\theta}_i, \boldsymbol{\alpha}) \end{aligned}$$

$$\boldsymbol{\alpha} = (\boldsymbol{\alpha}'_R, \boldsymbol{\alpha}'_M)'$$

- This assumption *may or may not* be realistic

Practical considerations: Quite *complex intra-individual covariance models* can result from *faithful consideration* of the situation...

- ...But may be *difficult to implement*

Model formulation

Standard model simplifications: One or more might be adopted

- *Negligible* measurement error $\Longrightarrow$

$$V_i(\boldsymbol{U}_i, \boldsymbol{\theta}_i, \boldsymbol{\alpha}) = V_{R,i}(\boldsymbol{U}_i, \boldsymbol{\theta}_i, \boldsymbol{\alpha}_R)$$

- The t_{ij} may be at *widely spaced intervals* $\Longrightarrow$ autocorrelation among $e_{R,ij}$ *negligible* $\Longrightarrow$ $V_i(\boldsymbol{U}_i, \boldsymbol{\theta}_i, \boldsymbol{\alpha})$ is *diagonal*
- $\text{Var}\{e_{R,i}(t, \boldsymbol{U}_i) \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i\} << \text{Var}\{e_{M,i}(t, \boldsymbol{U}_i) \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i\} \Longrightarrow$ measurement error is *dominant source*
- Simplifications should be *justifiable* in the context at hand

Note: All of these considerations apply to *any* mixed-effects model formulation, not just *non-linear* ones!

Model formulation

Routine assumption: $V_i(\boldsymbol{U}_i, \boldsymbol{\theta}_i, \boldsymbol{\alpha}) = \sigma_e^2 I_{n_i}$ $\boldsymbol{\alpha} = \sigma_e^2$

- Often made by "*default*" with little consideration of the *assumptions it implies*!
- Assumes *autocorrelation* among $e_{R,ij}$ *negligible*
- Assumes *constant variances*, i.e., $\text{Var}\{e_{R,i}(t, \boldsymbol{U}_i) \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i\} = \sigma_R^2$ and $\text{Var}\{e_{M,i}(t, \boldsymbol{U}_i) \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i\} = \sigma_M^2 \Longrightarrow \sigma_e^2 = \sigma_R^2 + \sigma_M^2$
- If *measurement error is negligible* $\Longrightarrow \sigma_e^2 = \sigma_R^2$
- If $\text{Var}\{e_{R,i}(t, \boldsymbol{U}_i) \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i\} << \text{Var}\{e_{M,i}(t, \boldsymbol{U}_i) \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i\}$ $\Longrightarrow \sigma_e \approx \sigma_M^2$

Model formulation

Standard assumptions in PK:

- Sampling times are *sufficiently far apart* that *autocorrelation* among $e_{R,ij}$ *negligible* (*not* always justifiable!)
- Measurement error *dominates* realization error so that

$$\text{Var}(e_{R,ij} \mid \boldsymbol{U}_i, \boldsymbol{\theta}_i) << \text{Var}(e_{M,ij} \mid \boldsymbol{U}_i, \boldsymbol{\theta}_i)$$

  (often *reasonable*)

- Measurement error variance *depends on level*, approximated by

$$\text{Var}(e_{M,ij} \mid \boldsymbol{U}_i, \boldsymbol{\theta}_i) = \sigma_M^2 \{m(t_{ij}, \boldsymbol{U}_i, \boldsymbol{\theta}_i)\}^{2\zeta}$$

  so that $V_i(\boldsymbol{U}_i, \boldsymbol{\theta}_i, \boldsymbol{\alpha}) = V_{M,i}(\boldsymbol{U}_i, \boldsymbol{\theta}_i, \boldsymbol{\alpha}_M)$ is *diagonal* with these elements (*almost always* the case)

Model formulation

Distributional assumption:

- Specification for $E(\boldsymbol{Y}_i \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i) = \boldsymbol{m}_i(\boldsymbol{U}_i, \boldsymbol{\theta}_i)$,

$$\boldsymbol{m}_i(\boldsymbol{U}_i, \boldsymbol{\theta}_i) = \{m(t_{i1}, \boldsymbol{U}_i, \boldsymbol{\theta}_i), \ldots, m(t_{in_i}, \boldsymbol{U}_i, \boldsymbol{\theta}_i)\}' \quad (n_i \times 1)$$

- Specification for $\text{Cov}(\boldsymbol{Y}_i \,|\boldsymbol{U}_i, \boldsymbol{\theta}_i) = V_i(\boldsymbol{U}_i, \boldsymbol{\theta}_i, \boldsymbol{\alpha})$
- *Standard assumption*: Distribution of $\boldsymbol{Y}_i$ given $\boldsymbol{U}_i$ and $\boldsymbol{\theta}_i$ is *multivariate normal* with *these moments*
- Alternatively, model on the *log scale* $\Longrightarrow$ Y_{ij} are conditionally (on $\boldsymbol{U}_i$ and $\boldsymbol{\theta}_i$) *lognormal*
- *In what follows*: Y_{ij} denotes the response on the *original* or *transformed* scale as appropriate

Model formulation

Summary of the two-stage model: Recall $\boldsymbol{X}_i = (\boldsymbol{U}_i', \boldsymbol{A}_i')'$

- Substitute *population model* for $\boldsymbol{\theta}_i$ in *individual-level model*
- *Stage 1 – Individual-level model*:

$$\begin{aligned} E(\boldsymbol{Y}_i \,|\, \boldsymbol{X}_i, \boldsymbol{b}_i) &= E(\boldsymbol{Y}_i \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i) = \boldsymbol{m}_i(\boldsymbol{U}_i, \boldsymbol{\theta}_i) = \boldsymbol{m}_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i), \\ \text{Cov}(\boldsymbol{Y}_i \,|\, \boldsymbol{X}_i, \boldsymbol{b}_i) &= \text{Cov}(\boldsymbol{Y}_i \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i) = V_i(\boldsymbol{U}_i, \boldsymbol{\theta}_i, \boldsymbol{\alpha}) = V_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{\alpha}) \end{aligned}$$

- *Stage 2 – Population model*:

$$\boldsymbol{\theta}_i = \boldsymbol{d}(\boldsymbol{A}_i, \boldsymbol{\beta}, \boldsymbol{b}_i), \qquad \boldsymbol{b}_i \sim (\boldsymbol{0}, G)$$

- *Standard assumptions*:
  - $\boldsymbol{Y}_i$ given $\boldsymbol{X}_i$ and $\boldsymbol{b}_i$ *multivariate normal* (perhaps *transformed*)
  - $\boldsymbol{b}_i \sim N(\boldsymbol{0}, G)$
  - All of these can be *relaxed*

Model interpretation and inferential objectives

"Subject-specific" model:

- Individual behavior is *modeled explicitly* at Stage 1, depending on *individual-specific* parameters $\boldsymbol{\theta}_i$ that have *scientifically meaningful interpretation*
- Models for $E(\boldsymbol{Y}_i \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i)$ and $\boldsymbol{\theta}_i$, and *hence* $E(\boldsymbol{Y}_i \,|\, \boldsymbol{X}_i, \boldsymbol{b}_i)$, are specified . . .
- . . . in contrast to a "*population-averaged*" model, where a model for $E(\boldsymbol{Y}_i \,|\, \boldsymbol{X}_i)$ is specified *directly* (more on this momentarily. . .)
- This is consistent with the *inferential objectives*
- Interest is in "*typical*" values of $\boldsymbol{\theta}_i$ and how they *vary* in the population. . .

Model interpretation and inferential objectives

Main inferential objectives: May be *formalized* in terms of the model

- For a specific *population model* $\boldsymbol{d}$, the *fixed effect* $\boldsymbol{\beta}$ characterizes the *mean* or *median* ("*typical*") value of $\boldsymbol{\theta}_i$ in the population (perhaps for individuals with given value of $\boldsymbol{A}_i$)
- $\Longrightarrow$ Determining an *appropriate population model* $\boldsymbol{d}(\boldsymbol{A}_i, \boldsymbol{\beta}, \boldsymbol{b}_i)$ and inference on *elements* of $\boldsymbol{\beta}$ in it is of *central interest*
- *Variation* of $\boldsymbol{\theta}_i$ across individuals *beyond* that attributable to *systematic associations* with among-individual covariates $\boldsymbol{A}_i$ is described by G ("*unexplained variation*")
- $\Longrightarrow$ Inference on G is of interest (in particular, *diagonal elements*)

Model interpretation and inferential objectives

Additional inferential objectives: In some contexts

- Inference on $\boldsymbol{\theta}_i$ and/or $m(t_0, \boldsymbol{U}_i, \boldsymbol{\theta}_i)$ at some specific time t_0 for $i = 1, \ldots, N$ or for *future individuals* is of interest
- *Example*: "*Individualized*" dosing in PK
- The model is a natural framework for "*borrowing strength*" across similar individuals (more later)

Model interpretation and inferential objectives

"Subject-specific" vs. "Population-averaged":

- The *non-linear mixed model* is a "*subject-specific*" model $\Longrightarrow$ Interest is in "*typical*" *values of individual-specific parameters* (mechanisms), $\boldsymbol{\theta}_i$, and how they *vary* in the population
- A "*population-averaged*" model describes the "*typical*" *response pattern* (*averaged* over individuals in the population), $E(\boldsymbol{Y}_i \,|\, \boldsymbol{X}_i)$, and the *overall variation* in response patterns about it, $\text{Cov}(\boldsymbol{Y}_i \,|\, \boldsymbol{X}_i)$
- $\Longrightarrow$ In a "*population-averaged*" model, *individual-specific* behavior is *not acknowledged*; rather, it is "*averaged out*" in advance, i.e.,

$$E(\boldsymbol{Y}_i \,|\, \boldsymbol{X}_i) = \int E(\boldsymbol{Y}_i \,|\, \boldsymbol{X}_i, \boldsymbol{b}_i)\, dF_b(\boldsymbol{b}_i)$$

$\Longrightarrow$ $E(\boldsymbol{Y}_i \,|\, \boldsymbol{X}_i)$ is specified *directly*; a representation for $E(\boldsymbol{Y}_i \,|\, \boldsymbol{X}_i, \boldsymbol{b}_i)$ is *never* specified

Model interpretation and inferential objectives

"Subject-specific" vs. "Population-averaged":

- "Population-averaged" model cannot incorporate *theoretical assumptions* embedded in the model $m(t, \boldsymbol{U}_i, \boldsymbol{\theta}_i)$ for *individual behavior*
- *In fact*, using m as a model for $E(\boldsymbol{Y}_i \,|\, \boldsymbol{X}_i)$ makes *no scientific sense* (although it may provide a reasonable *empirical representation* of the "*typical*" *response pattern*) – *impossible* for

$$E(\boldsymbol{Y}_i \,|\, \boldsymbol{X}_i) = \int \boldsymbol{m}_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i)\, dF_b(\boldsymbol{b}_i) = m(\boldsymbol{X}_i, \boldsymbol{\beta})$$

- In the applications here, the response is of interest because it *carries information* on the $\boldsymbol{\theta}_i$, but *average response itself* is of *little or no importance* $\Longrightarrow$ "population-averaged" model is not *appropriate*

Model interpretation and inferential objectives

"Subject-specific" model $\Longrightarrow$ "population-averaged" model:

$$\begin{aligned} E(\boldsymbol{Y}_i|\boldsymbol{X}_i) &= \int \boldsymbol{m}_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i)\, dF_b(\boldsymbol{b}_i) \\ \text{Cov}(\boldsymbol{Y}_i|\boldsymbol{X}_i) &= E\{V_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{\alpha})|\boldsymbol{X}_i\} + \text{Cov}\{\boldsymbol{m}_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i)|\boldsymbol{X}_i\} \end{aligned}$$

- $E(\boldsymbol{Y}_i \,|\, \boldsymbol{X}_i)$ is *complicated* function of $\boldsymbol{\beta}$ and G $\Longrightarrow$ $\boldsymbol{\beta}$ alone *does not* describe the population average
- $E\{V_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{\alpha})|\boldsymbol{X}_i\}$ = average of realization/measurement variation over population $\Longrightarrow$ diagonal *only if* autocorrelation of within-individual realizations *negligible*
- $\text{Cov}\{\boldsymbol{m}_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i)|\boldsymbol{X}_i\}$ = population variation in "*inherent trajectories*" $\Longrightarrow$ *non-diagonal in general*
- $\Longrightarrow$ *Overall pattern* of variation/covariation in the response is the *aggregate* due to *both sources*
- I *prefer* "*aggregate*" covariance to "*within-individual*" covariance

Break

Inferential approaches

Reminder – summary of the two-stage model: $\boldsymbol{X}_i = (\boldsymbol{U}_i', \boldsymbol{A}_i')'$

- *Stage 1 – Individual-level model*:

$$\begin{aligned} E(\boldsymbol{Y}_i \,|\, \boldsymbol{X}_i, \boldsymbol{b}_i) &= E(\boldsymbol{Y}_i \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i) = \boldsymbol{m}_i(\boldsymbol{U}_i, \boldsymbol{\theta}_i) = \boldsymbol{m}_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i), \\ \mathrm{Cov}(\boldsymbol{Y}_i \,|\, \boldsymbol{X}_i, \boldsymbol{b}_i) &= \mathrm{Cov}(\boldsymbol{Y}_i \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i) = V_i(\boldsymbol{U}_i, \boldsymbol{\theta}_i, \boldsymbol{\alpha}) = V_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{\alpha}) \end{aligned}$$

- *Stage 2 – Population model*:

$$\boldsymbol{\theta}_i = \boldsymbol{d}(\boldsymbol{A}_i, \boldsymbol{\beta}, \boldsymbol{b}_i), \qquad \boldsymbol{b}_i \sim (\boldsymbol{0}, G)$$

- *Standard assumptions*:
  - $\boldsymbol{Y}_i$ given $\boldsymbol{X}_i$ and $\boldsymbol{b}_i$ *multivariate normal* (perhaps *transformed*)
    $\Longrightarrow$ probability density function $f_i(\boldsymbol{y}_i \,|\, \boldsymbol{x}_i, \boldsymbol{b}_i; \boldsymbol{\beta}, \boldsymbol{\alpha})$
  - $\boldsymbol{b}_i \sim N(\boldsymbol{0}, G) \Longrightarrow$ density $f(\boldsymbol{b}_i; G)$
- *Observed data*: $\{(\boldsymbol{Y}_i, \boldsymbol{X}_i),\ i = 1, \ldots, N\} = (\boldsymbol{Y}, \boldsymbol{X})$,
  $(\boldsymbol{Y}_i, \boldsymbol{X}_i)$ assumed *independent* across i

Inferential approaches

Natural basis for inference on β, G: *Maximum likelihood*

- *Joint density* of $\boldsymbol{Y}$ given $\boldsymbol{X}$ (by *independence*)

$$f(\boldsymbol{y} \,|\, \boldsymbol{x};\, \boldsymbol{\gamma}, G) = \prod_{i=1}^{N} f_i(\boldsymbol{y}_i \,|\, \boldsymbol{x}_i;\, \boldsymbol{\gamma}, G), \quad \boldsymbol{\gamma} = (\boldsymbol{\beta}', \boldsymbol{\alpha}')'$$

- $f_i(\boldsymbol{y}_i, \boldsymbol{b}_i \,|\, \boldsymbol{x}_i;\, \boldsymbol{\gamma}, G) = f_i(\boldsymbol{y}_i \,|\, \boldsymbol{x}_i, \boldsymbol{b}_i;\, \boldsymbol{\gamma}) f(\boldsymbol{b}_i;\, G)$
- *Log-likelihood* for $(\boldsymbol{\gamma}, G)$

$$\begin{aligned} \ell(\boldsymbol{\gamma}, G) &= \log\left\{\prod_{i=1}^{N} f_i(\boldsymbol{y}_i \,|\, \boldsymbol{x}_i;\, \boldsymbol{\gamma}, G)\right\} \\ &= \log\left\{\prod_{i=1}^{N} \int f_i(\boldsymbol{y}_i \,|\, \boldsymbol{x}_i, \boldsymbol{b}_i;\, \boldsymbol{\gamma})\, f(\boldsymbol{b}_i;\, G)\, d\boldsymbol{b}_i\right\} \end{aligned}$$

- Involves N $q-$*dimensional integrals*

Inferential approaches

$$\ell(\boldsymbol{\gamma}, G) = \log \left\{ \prod_{i=1}^{N} \int f_i(\boldsymbol{y}_i \,|\, \boldsymbol{x}_i, \boldsymbol{b}_i; \, \boldsymbol{\gamma}) \, f(\boldsymbol{b}_i; \, G) \, d\boldsymbol{b}_i \right\}$$

Major practical issue: These integrals are *analytically intractable* in general and may be *high-dimensional*

- Some means of *approximation* of the integrals required
- *Analytical approximation* (the approach used *historically*, first by PKists) – will discuss first
- *Numerical approximation* (more recent, as computational resources have *improved*)

Inferential approaches

Inference based on individual estimates: If $n_i \geq r$, can (in principle) obtain *individual regression estimates* $\widehat{\boldsymbol{\theta}}_i$

- E.g., if $V_i(\boldsymbol{U}_i, \boldsymbol{\theta}_i, \boldsymbol{\alpha}) = \sigma_e^2 I_{n_i}$ can use *ordinary least squares* for each i
- For *fancier* $V_i(\boldsymbol{U}_i, \boldsymbol{\theta}_i, \boldsymbol{\alpha})$ can use *generalized* (*weighted*) *least squares* for each i with an estimate of $\boldsymbol{\alpha}$ substituted
- $\boldsymbol{\alpha}$ can be estimated by "*pooling*" residuals across all N individuals
- *Realistically*: Require $n_i >> r$
- Described in Chapter 5 of Davidian and Giltinan (1995)

Idea: Use the $\widehat{\boldsymbol{\theta}}_i$, $i = 1, \ldots, N$, as "*data*" to estimate $\boldsymbol{\beta}$ and G...

Inferential approaches

Idea: Use the $\widehat{\boldsymbol{\theta}}_i$, $i = 1, \ldots, N$, as "*data*" to estimate $\boldsymbol{\beta}$ and G

- Consider *linear population model* $\boldsymbol{\theta}_i = A_i\boldsymbol{\beta} + B_i\boldsymbol{b}_i$
- Standard *large-n_i* asymptotic theory $\Longrightarrow$

$$\widehat{\boldsymbol{\theta}}_i \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i \dot{\sim} N(\boldsymbol{\theta}_i, C_i), \quad C_i \text{ depends on } \boldsymbol{\theta}_i, \boldsymbol{\alpha}$$

- Estimate C_i by substituting $\widehat{\boldsymbol{\theta}}_i$, $\widehat{\boldsymbol{\alpha}} \Longrightarrow \widehat{\boldsymbol{\theta}}_i \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i \dot{\sim} N(\boldsymbol{\theta}_i, \widehat{C}_i)$ and treat $\widehat{C}_i$ as *fixed*
- Write as $\widehat{\boldsymbol{\theta}}_i \approx \boldsymbol{\theta}_i + \boldsymbol{e}_i^*$, $\boldsymbol{e}_i^* \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i \dot{\sim} N(\boldsymbol{0}, \widehat{C}_i)$
- $\Longrightarrow$ *Approximate* "*linear mixed-effects model*" for "*response*" $\widehat{\boldsymbol{\theta}}_i$

$$\widehat{\boldsymbol{\theta}}_i \approx A_i\boldsymbol{\beta} + B_i\boldsymbol{b}_i + \boldsymbol{e}_i^*, \quad \boldsymbol{b}_i \sim N(\boldsymbol{0}, G), \quad \boldsymbol{e}_i^* \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i \dot{\sim} N(\boldsymbol{0}, \widehat{C}_i)$$

- Can be fitted (estimate $\boldsymbol{\beta}$, G) using *standard linear mixed model* methods (treating $\widehat{C}_i$ as *fixed*)

Inferential approaches

$$\widehat{\boldsymbol{\theta}}_i \approx A_i\boldsymbol{\beta} + B_i\boldsymbol{b}_i + \boldsymbol{e}_i^*, \quad \boldsymbol{b}_i \sim N(\boldsymbol{0}, G), \quad \boldsymbol{e}_i^* \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i \dot{\sim} N(\boldsymbol{0}, \widehat{C}_i)$$

Fitting the "linear mixed model":

- "*Global two-stage algorithm*" (*GTS*): Fit using the *EM algorithm*; see Davidian and Giltinan (1995, Chapter 5)
- Use *standard linear mixed model software* such as SAS `proc mixed`, R function `lme` – requires some *tweaking* to handle the fact that $\widehat{C}_i$ is regarded as *known*
- Appeal to usual *large-N asymptotic theory* for the "*linear mixed model*" to obtain *standard errors* for elements of $\widehat{\boldsymbol{\beta}}$, *confidence intervals* for elements of $\boldsymbol{\beta}$, etc (generally works *well*)

Common misconception: This method is often portrayed in the literature as having *no relationship* to the non-linear mixed-effects model

Inferential approaches

How does this approximate the integrals? Not *readily apparent*

- May view the $\widehat{\boldsymbol{\theta}}_i$ as approximate "*sufficient statistics*" for the $\boldsymbol{\theta}_i$
- *Change of variables* in the integrals and replace $f_i(\boldsymbol{y}_i \,|\, \boldsymbol{x}_i, \boldsymbol{b}_i; \boldsymbol{\gamma})$ by the (normal) density $f(\widehat{\boldsymbol{\theta}}_i \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i; \boldsymbol{\alpha})$ corresponding to the *asymptotic approximation*

Remarks:

- When *all* n_i are *sufficiently large* to justify the asymptotic approximation (e.g., *intensive* PK studies), I like this method!
- *Easy to explain* to collaborators
- Gives *similar answers* to other analytical approximation methods (coming up)
- *Drawback*: No standard *software* (although see my website for R/SAS code)

Inferential approaches

In many settings: "*Rich*" individual data *not available* for all i (e.g., *population* PK studies); i.e., n_i "*not large*" for some or all i

- Approximate the integrals *more directly* by approximating $f_i(\boldsymbol{y}_i \,|\, \boldsymbol{x}_i;\, \boldsymbol{\gamma}, G)$

Write model with normality assumptions at both stages:

$$\boldsymbol{Y}_i = \boldsymbol{m}_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i) + V_i^{1/2}(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{\alpha})\,\boldsymbol{\epsilon}_i, \quad \boldsymbol{b}_i \sim N(\boldsymbol{0}, G)$$

- $V_i^{1/2}$ $(n_i \times n_i)$ such that $V_i^{1/2}(V_i^{1/2})' = V_i$
- $\boldsymbol{\epsilon}_i \,|\, \boldsymbol{X}_i, \boldsymbol{b}_i \sim N(\boldsymbol{0}, I_{n_i})$ $(n_i \times 1)$
- First-order *Taylor series* about $\boldsymbol{b}_i = \boldsymbol{b}_i^*$ "close" to $\boldsymbol{b}_i$, *ignoring* cross-product $(\boldsymbol{b}_i - \boldsymbol{b}_i^*)\boldsymbol{\epsilon}_i$ as *negligible* $\Longrightarrow$

$$\boldsymbol{Y}_i \approx \boldsymbol{m}_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i^*) - Z_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i^*)\boldsymbol{b}_i^* + Z_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i^*)\boldsymbol{b}_i + V_i^{1/2}(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i^*, \boldsymbol{\alpha})\,\boldsymbol{\epsilon}_i$$

$$Z_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i^*) = \partial/\partial \boldsymbol{b}_i \{\boldsymbol{m}_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i)\}|_{\boldsymbol{b}_i = \boldsymbol{b}_i^*}$$

Inferential approaches

$$\boldsymbol{Y}_i \approx \boldsymbol{m}_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i^*) - Z_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i^*)\boldsymbol{b}_i^* + Z_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i^*)\boldsymbol{b}_i + V_i^{1/2}(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i^*, \boldsymbol{\alpha})\,\boldsymbol{\epsilon}_i$$

"First-order" method: Take $\boldsymbol{b}_i^* = \mathbf{0}$ (mean of $\boldsymbol{b}_i$)

- $\Longrightarrow$ Distribution of $\boldsymbol{Y}_i$ given $\boldsymbol{X}_i$ *approximately normal* with

$$\begin{aligned} E(\boldsymbol{Y}_i \,|\, \boldsymbol{X}_i) &\approx \boldsymbol{m}_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \mathbf{0}), \\ \mathrm{Cov}(\boldsymbol{Y}_i \,|\, \boldsymbol{X}_i) &\approx Z_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \mathbf{0})\, G\, Z_i'(\boldsymbol{X}_i, \boldsymbol{\beta}, \mathbf{0}) + V_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \mathbf{0}, \boldsymbol{\alpha}) \end{aligned}$$

- $\Longrightarrow$ *Approximate* $f_i(\boldsymbol{y}_i \,|\, \boldsymbol{x}_i; \boldsymbol{\gamma}, G)$ by a normal density with these moments, so that $\ell(\gamma, G)$ is in a *closed form*
- $\Longrightarrow$ Estimate $(\boldsymbol{\beta}, \boldsymbol{\alpha}, G)$ by *maximum likelihood* – because integrals are eliminated, is a *direct optimization* (but still *very messy*. . .)
- First proposed by Beal and Sheiner in early 1980s in the context of *population PK*

Inferential approaches

"First-order" method: *Software*

- `fo` method in the Fortran package `nonmem` (*widely used* by PKists)
- SAS `proc nlmixed` using the `method=firo` option (but *cannot handle* by default dependence of $V_i(\boldsymbol{U}_i, \boldsymbol{\theta}_i, \boldsymbol{\alpha}) = V_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{\alpha})$ on $\boldsymbol{\theta}_i$ and thus on $\boldsymbol{\beta}$, $\boldsymbol{b}_i$)

Alternative implementation: View as an *approximate* "*population-averaged*" model for mean and covariance

$$\begin{aligned} E(\boldsymbol{Y}_i \,|\, \boldsymbol{X}_i) &\approx \boldsymbol{m}_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{0}), \\ \mathrm{Cov}(\boldsymbol{Y}_i \,|\, \boldsymbol{X}_i) &\approx Z_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{0})\, G\, Z_i'(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{0}) + V_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{0}, \boldsymbol{\alpha}) \end{aligned}$$

- $\Longrightarrow$ Estimate $(\boldsymbol{\beta}, \boldsymbol{\alpha}, G)$ by solving a set of *generalized estimating equations* (GEEs; specifically, "*GEE-1*")
- Is a *different method* from maximum likelihood ("*GEE-2*")
- *Software*: SAS macro `nlinmix` with `expand=zero`

Inferential approaches

Problem: These *approximate moments* are clearly *poor approximations* to the true moments

- In particular, *poor approximation* to $E(\boldsymbol{Y}_i \,|\, \boldsymbol{X}_i) \Longrightarrow$ *biased estimators* for $\boldsymbol{\beta}$

"First-order conditional methods": Use a "*better*" approximation

- Take $\boldsymbol{b}_i^*$ "*closer*" to $\boldsymbol{b}_i$
- Natural choice: $\widehat{\boldsymbol{b}}_i =$ *mode* of the *posterior density*

$$f(\boldsymbol{b}_i \,|\, \boldsymbol{y}_i, \boldsymbol{x}_i; \boldsymbol{\gamma}, G) = \frac{f_i(\boldsymbol{y}_i \,|\, \boldsymbol{x}_i, \boldsymbol{b}_i; \boldsymbol{\gamma})\, f(\boldsymbol{b}_i; G)}{f_i(\boldsymbol{y}_i \,|\, \boldsymbol{x}_i; \boldsymbol{\gamma}, G)}$$

- $\Longrightarrow$ *Approximate moments*

$$\begin{aligned} E(\boldsymbol{Y}_i \,|\, \boldsymbol{X}_i) &\approx \boldsymbol{m}_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \widehat{\boldsymbol{b}}_i) - Z_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \widehat{\boldsymbol{b}}_i)\widehat{\boldsymbol{b}}_i \\ \text{Cov}(\boldsymbol{Y}_i \,|\, \boldsymbol{X}_i) &\approx Z_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \widehat{\boldsymbol{b}}_i)\, G\, Z_i'(\boldsymbol{X}_i, \boldsymbol{\beta}, \widehat{\boldsymbol{b}}_i) + V_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \widehat{\boldsymbol{b}}_i, \boldsymbol{\alpha}) \end{aligned}$$

Inferential approaches

Fitting algorithms: *Iterate* between

(i) *Update* $\widehat{\boldsymbol{b}}_i$, $i = 1, \ldots, N$, by maximizing the *posterior density* (or *approximation* to it) with $\widehat{\boldsymbol{\gamma}}$ and $\widehat{G}$ *substituted* and held fixed

(ii) Hold the $\widehat{\boldsymbol{b}}_i$ fixed and *update* estimation of $\boldsymbol{\gamma}$ and G by *either*

(a) *Maximizing* the approximate *normal log-likelihood* based on treating $\boldsymbol{Y}_i$ given $\boldsymbol{X}_i$ as normal with these moments, *OR*

(b) *Solving* a corresponding set of *GEEs*

- Usually "*converges*" (although no *guarantee*)

Software:

- `nonmem` with `foce` option implements (ii)(a)
- R function `nlme`, SAS macro `nlinmix` with `expand=blup` option implement (ii)(b)

Inferential approaches

Standard errors, etc: For *both* "*first-order*" approximations

- *Pretend* that the *approximate moments* are *exact* and use the *usual large-N asymptotic theory* for maximum likelihood or GEEs
- Provides *reliable inferences* in problems where N is *reasonably large* and the magnitude of *among-individual variation* is not huge

My experience:

- Even without the integration, these are *nasty computational problems*, and *good starting values* for the parameters are *required* (may have to try several sets of starting values).
- The "*first-order*" approximation is *too crude* and should be *avoided* in general (although can be a good way to get reasonable *starting values* for other methods)
- The "*first-order conditional*" methods often work well, are numerically *well-behaved*, and yield *reliable inferences*

Inferential approaches

$$\ell(\boldsymbol{\gamma}, G) = \log \left\{ \prod_{i=1}^{N} \int f_i(\boldsymbol{y}_i \,|\, \boldsymbol{x}_i, \boldsymbol{b}_i;\, \boldsymbol{\gamma})\, f(\boldsymbol{b}_i;\, G)\, d\boldsymbol{b}_i \right\}$$

Numerical approximation methods: *Approximate* the integrals using *deterministic* or *stochastic numerical integration* techniques (*$q-$dimensional numerical integration*) and *maximize the log-likelihood*

- *Issue*: For *each iteration* of the likelihood optimization algorithm, must approximate N *q-dimensional integrals*
- *Infeasible* until recently: Numerical integration embedded *repeatedly* in an optimization routine is *computationally intensive*
- Gets *worse* with *larger* q (the "*curse of dimensionality*")

Inferential approaches

Deterministic techniques:

- *Normality* of $\boldsymbol{b}_i \Longrightarrow$ *Gauss-Hermite quadrature*
- *Quadrature rule*: Approximate an integral by a *suitable weighted average* of the integrand evaluated at a $q-$*dimensional grid* of values $\Longrightarrow$ *accuracy increases* with more grid points, but so does *computational burden*
- *Adaptive Gaussian quadrature*: "Center" and "scale" the grid about $\widehat{\boldsymbol{b}}_i \Longrightarrow$ can *greatly reduce* the number of grid points needed

Software: SAS `proc nlmixed`

- *Adaptive Gaussian quadrature*: The *default*
- *Gaussian quadrature*: `method=gauss noad`
- As before, `proc nlmixed` *cannot handle* dependence of $V_i(\boldsymbol{U}_i, \boldsymbol{\theta}_i, \boldsymbol{\alpha}) = V_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{\alpha})$ on $\boldsymbol{\theta}_i$ and thus on $\boldsymbol{\beta}$, $\boldsymbol{b}_i$

Inferential approaches

$$\ell(\boldsymbol{\gamma}, G) = \log \left\{ \prod_{i=1}^{N} \int f_i(\boldsymbol{y}_i \,|\, \boldsymbol{x}_i, \boldsymbol{b}_i;\, \boldsymbol{\gamma})\, f(\boldsymbol{b}_i;\, G)\, d\boldsymbol{b}_i \right\}$$

Stochastic techniques:

- "*Brute force*" *Monte Carlo* integration: Represent integral for i by

$$B^{-1} \sum_{b=1}^{B} f_i(\boldsymbol{y}_i \,|\, \boldsymbol{x}_i, \boldsymbol{b}^{(b)};\, \boldsymbol{\gamma}),$$

  $\boldsymbol{b}^{(b)}$ are draws from $N(\mathbf{0}, G)$ (at the current estimates of $\boldsymbol{\gamma}, G$)

- Can require *very large* B for acceptable accuracy (*inefficient*)
- *Importance sampling*: Replace this by a *suitably weighted* version that is more efficient

Software: SAS `proc nlmixed` implements *importance sampling* (`method=isamp`)

Inferential approaches

My experience with SAS `proc nlmixed`:

- Good *starting values* are *essential* (may have to try many sets) – starting values are required for *all of* $\boldsymbol{\beta}, G, \boldsymbol{\alpha}$
- Could obtain starting values from an *analytical approximation method*
- Practically speaking, quadrature is *infeasible* for $q > 2$ almost always with the *mechanism-based* non-linear models in PK and other applications

Inferential approaches

Other methods: Maximize the log-likelihood via an *EM algorithm*

- For *non-linear mixed models*, the conditional expectation in the *E-step* is not available in a *closed form*
- *Monte Carlo EM algorithm*: Approximate the E-step by ordinary Monte Carlo integration
- *Stochastic approximation EM algorithm*: Approximate the E-step by Monte Carlo simulation and stochastic approximation
- *Software*?

Inferential approaches

Bayesian inference: Natural approach to *hierarchical models*

Big picture: In the *Bayesian paradigm*

- View $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, G, and $\boldsymbol{b}_i$, $i = 1, \ldots, N$, as *random parameters* (on equal footing) with *prior distributions* (priors for $\boldsymbol{b}_i$, $i = 1, \ldots, N$, are $N(\mathbf{0}, G)$)
- Bayesian inference on $\boldsymbol{\beta}$ and G is based on their *posterior distributions*
- The posterior distributions involve *high-dimensional integration* and *cannot* be derived analytically . . .
- . . . but *samples* from the posterior distributions can be obtained via *Markov chain Monte Carlo* (MCMC)

Inferential approaches

Bayesian hierarchy:

- *Stage 1 – Individual-level model*: Assume *normality*

$$\begin{aligned} E(\boldsymbol{Y}_i \,|\, \boldsymbol{X}_i, \boldsymbol{b}_i) &= E(\boldsymbol{Y}_i \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i) = \boldsymbol{m}_i(\boldsymbol{U}_i, \boldsymbol{\theta}_i) = \boldsymbol{m}_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i), \\ \mathrm{Cov}(\boldsymbol{Y}_i \,|\, \boldsymbol{X}_i, \boldsymbol{b}_i) &= \mathrm{Cov}(\boldsymbol{Y}_i \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i) = V_i(\boldsymbol{U}_i, \boldsymbol{\theta}_i, \boldsymbol{\alpha}) = V_i(\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{\alpha}) \end{aligned}$$

- *Stage 2 – Population model*: $\boldsymbol{\theta}_i = \boldsymbol{d}(\boldsymbol{A}_i, \boldsymbol{\beta}, \boldsymbol{b}_i), \quad \boldsymbol{b}_i \sim N(\boldsymbol{0}, G)$
- *Stage 3 – Hyperprior*: $(\boldsymbol{\beta}, \boldsymbol{\alpha}, G) \sim f(\boldsymbol{\beta}, \boldsymbol{\alpha}, G) = f(\boldsymbol{\beta}) f(\boldsymbol{\alpha}) g(G)$
- *Joint posterior density*

$$f(\gamma, G, \boldsymbol{b} \,|\, \boldsymbol{y}, \boldsymbol{x}) = \frac{\prod_{i=1}^{N} f_i(\boldsymbol{y}_i \,|\, \boldsymbol{x}_i, \boldsymbol{b}_i; \, \boldsymbol{\gamma}) \, f(\boldsymbol{b}_i; \, G) f(\boldsymbol{\beta}, \boldsymbol{\alpha}, G)}{f(\boldsymbol{y} \,|\, \boldsymbol{x})};$$

  denominator is *numerator integrated* wrt $(\gamma, G, \boldsymbol{b}_i, i = 1, \ldots, N)$

- E.g., *posterior* for $\boldsymbol{\beta}$, $f(\boldsymbol{\beta} \,|\, \boldsymbol{y}, \boldsymbol{x})$: Integrate out $\boldsymbol{\alpha}, G, \boldsymbol{b}_i, i = 1, \ldots, N$

Inferential approaches

Estimator for $\boldsymbol{\beta}$: *Mode* of posterior

- Uncertainty measured by *spread* of $f(\boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{x})$
- Similarly for $\boldsymbol{\alpha}$, G, and $\boldsymbol{b}_i$, $i = 1, \ldots, N$

Implementation: By *simulation* via MCMC

- Samples from the *full conditional distributions* (eventually) behave like samples from the posterior distributions
- The *mode* and measures of uncertainty may be calculated *empirically* from these samples
- *Issue*: Sampling from some of the full conditionals is *not entirely straightforward* because of *non-linearity* of m in $\boldsymbol{\theta}_i$ and hence $\boldsymbol{b}_i$
- $\Longrightarrow$ "*All-purpose*" software not available in general, but has been implemented for popular m in add-ons to `WinBUGS` (e.g., `PKBugs`)

Inferential approaches

Experience:

- With weak hyperpriors and "*good*" data, inferences are *very similar* to those based on *maximum likelihood* and *first-order conditional* methods
- *Convergence* of the chain must be monitored carefully; "*false convergence*" can happen
- *Advantage of Bayesian framework*: Natural mechanism to incorporate *known constraints* and prior *scientific knowledge*

Inferential approaches

Inference on individuals: Follows naturally from a Bayesian perspective

- *Goal*: "*Estimate*" $\boldsymbol{b}_i$ or $\boldsymbol{\theta}_i$ for a randomly chosen individual i from the population
- "*Borrowing strength*": Individuals sharing common characteristics can enhance inference
- $\Longrightarrow$ Natural "estimator" is the *mode* of the posterior $f(\boldsymbol{b}_i \,|\, \boldsymbol{y}, \boldsymbol{x})$ or $f(\boldsymbol{\theta}_i \,|\, \boldsymbol{y}, \boldsymbol{x})$
- *Frequentist perspective*: $(\boldsymbol{\gamma}, G)$ are *fixed* – relevant posterior is

$$f(\boldsymbol{b}_i \,|\, \boldsymbol{y}_i, \boldsymbol{x}_i; \, \boldsymbol{\gamma}, G) = \frac{f_i(\boldsymbol{y}_i \,|\, \boldsymbol{x}_i, \boldsymbol{b}_i; \, \boldsymbol{\gamma}) \, f(\boldsymbol{b}_i; \, G)}{f_i(\boldsymbol{y}_i \,|\, \boldsymbol{x}_i; \, \boldsymbol{\gamma}, G)}$$

  $\Longrightarrow$ *substitute* estimates for $(\boldsymbol{\gamma}, G)$
- $\widehat{\boldsymbol{\theta}}_i = \boldsymbol{d}(\boldsymbol{A}_i, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{b}}_i)$
- "*Empirical Bayes*"

Inferential approaches

Selecting the population model d: The foregoing is predicated on a *fixed* $\boldsymbol{d}(\boldsymbol{A}_i, \boldsymbol{\beta}, \boldsymbol{b}_i)$

- A key objective in many analyses (e.g., *population* PK) is to *identify* an appropriate $\boldsymbol{d}(\boldsymbol{A}_i, \boldsymbol{\beta}, \boldsymbol{b}_i)$
- Must identify *elements* of $\boldsymbol{A}_i$ to include in each component of $\boldsymbol{d}(\boldsymbol{A}_i, \boldsymbol{\beta}, \boldsymbol{b}_i)$ and the *functional form* of each component
- *Likelihood inference*: Use *nested hypothesis tests* or *information criteria* (AIC, BIC, etc)
- Challenging when $\boldsymbol{A}_i$ is *high-dimensional*. . .
- . . . Need a way of *selecting* among large number of variables and functional forms in each component (*still an open problem*. . .)

Inferential approaches

Selecting the population model d: Continued

- *Graphical methods*: Based on *Bayes* or *empirical Bayes* "estimates"
  - Fit an *initial* population model with *no covariates* (elements of $\boldsymbol{A}_i$ and obtain B/EB estimates $\widehat{\boldsymbol{b}}_i$, $i = 1 \ldots, N$
  - *Plot* components of $\widehat{\boldsymbol{b}}_i$ against elements of $\boldsymbol{A}_i$, look for *relationships*
  - Postulate and fit an *updated* population model $\boldsymbol{d}$ incorporating *relationships* and obtain updated B/EB estimates $\widehat{\boldsymbol{b}}_i$ and *re-plot*
  - If model is adequate, plots should show *haphazard scatter*; otherwise, *repeat*
  - *Issue 1*: "*Shrinkage*" of B/EB estimates could obscure relationships (especially if $\boldsymbol{b}_i$ really *aren't normally distributed*)
  - *Issue 2*: "*One-at-a-time*" assessment of relationships could *miss* important features

Inferential approaches

Normality of b_i: The assumption $\boldsymbol{b}_i \sim N(\mathbf{0}, G)$ is *standard* in mixed-effects model analysis; *however*

- Is it *always realistic*?
- *Unmeasured* binary *among-individual covariate* systematically associated with $\boldsymbol{\theta}_i \Longrightarrow \boldsymbol{b}_i$ has *bimodal distribution*
- Or a *normal* distribution may just *not* be the best model! *Heavy tails*, *skewness*...)
- *Consequences*?

Relaxing the normality assumption: Represent the *density* of $\boldsymbol{b}_i$ by a *flexible form*

- Estimate the *density* along with the *model parameters*
- $\Longrightarrow$ *Insight* into possible *omitted covariates*

Implementation and examples

Example 1: A basic analysis – *argatroban study*

- *Intensive PK study*, $N = 37$ subjects assigned to different *intravenous infusion* rates D_i for $t_{\text{inf}} = 240$ min
- $t_{ij} = 30,60,90,115,160,200,240,245,250,260,275,295,320,360$ min ($n_i = 14$)
- *One compartment model*

$$m(t, \boldsymbol{U}_i, \boldsymbol{\theta}_i) = \frac{D_i}{e^{Cl_i^*}} \left[\exp\left\{ -\frac{e^{Cl_i^*}}{e^{V_i^*}} (t - t_{\text{inf}})_+ \right\} - \exp\left(-\frac{e^{Cl_i^*}}{e^{V_i^*}} t \right) \right]$$

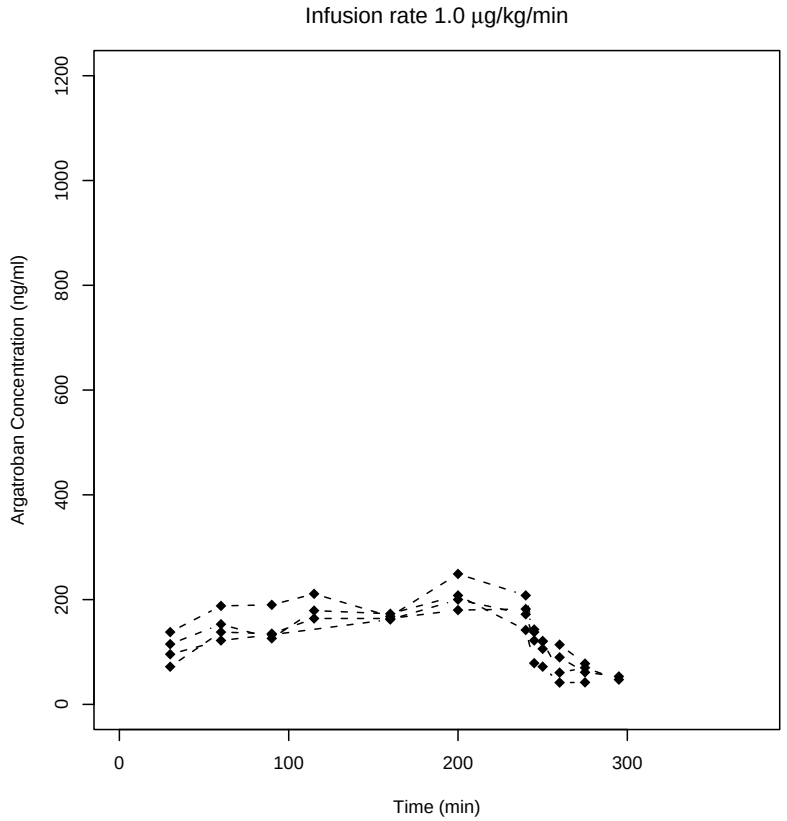
$$\boldsymbol{\theta}_i = (Cl_i^*, V_i^*)', \quad \boldsymbol{U}_i = (D_i, t_{\text{inf}})$$

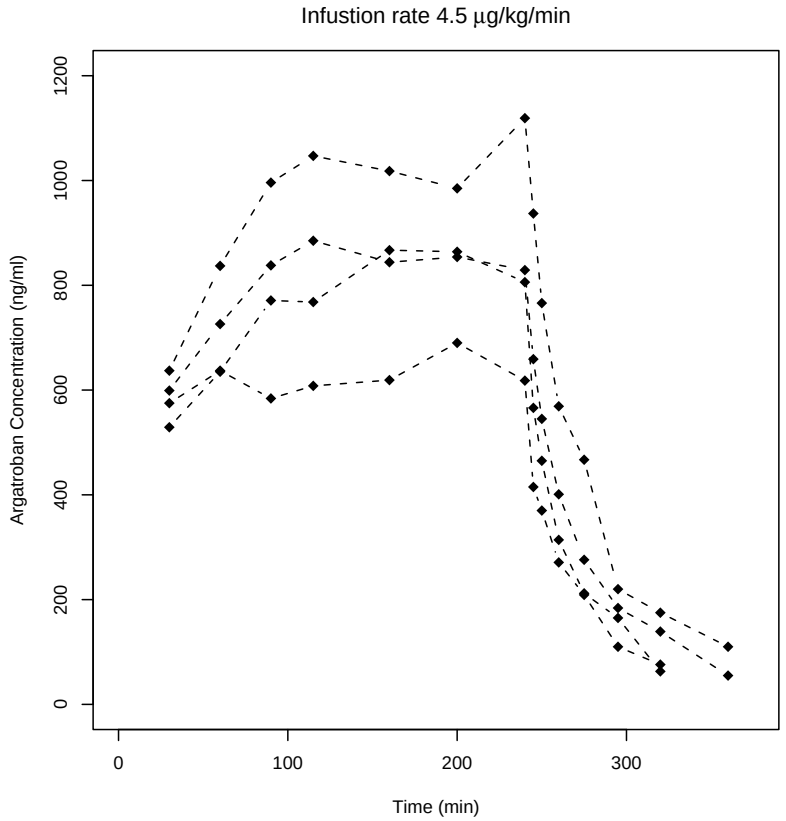
$$x_+ = 0 \text{ if } x \leq 0 \text{ and } x_+ = x \text{ if } x > 0$$

- *Parameterized* in terms of $Cl_i^* = \log(Cl_i)$, $V_i^* = \log(V_i)$ (population distributions of PK parameters likely *skewed*)
- No *among-individual covariates* $\boldsymbol{A}_i$

Applications

Profiles for subjects receiving 1.0 and 4.5 μg/kg-min:





Implementation and examples

Non-linear mixed model:

- *Stage 1 – Individual-level model*: Y_{ij} *normal* with

$$E(Y_{ij} \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i) = m(t_{ij}, \boldsymbol{U}_i, \boldsymbol{\theta}_i)$$

$$\text{Cov}(\boldsymbol{Y}_i \,|\, \boldsymbol{U}_i, \boldsymbol{A}_i) = V_i(\boldsymbol{U}_i, \boldsymbol{\theta}_i, \boldsymbol{\alpha}) = \sigma_e^2 \,\text{diag}\{m^{2\zeta}(t_{i1}, \boldsymbol{U}_i, \boldsymbol{\theta}_i), \ldots, m^{2\zeta}(t_{in_i}, \boldsymbol{U}_i, \boldsymbol{\theta}_i)\}$$

$\Longrightarrow$ *negligible autocorrelation*, measurement error *dominates*

- *Stage 2 – Population model*

$$\boldsymbol{\theta}_i = \boldsymbol{\beta} + \boldsymbol{b}_i, \quad \boldsymbol{\beta} = (\beta_1, \beta_2)', \quad \boldsymbol{b}_i \sim N(\boldsymbol{0}, G)$$

$\Longrightarrow$ β_1, β_2 represent *population means* of log clearance, volume; equivalently, $\exp(\beta_1), \exp(\beta_2)$ are *population medians*

$\Longrightarrow$ $\sqrt{G_{11}}, \sqrt{G_{22}} \approx$ *coefficients of variation* of clearance, volume

Implementation and examples

Implementation: Using

- *Individual estimates* $\widehat{\boldsymbol{\theta}}_i$ found using "*pooled*" *generalized least squares* including estimation of ζ (customized R code) followed by fitting the "*linear mixed model*" (SAS `proc mixed`)
- *First-order* method via version 8.01 of SAS macro `nlinmix` with `expand=zero` – fix $\zeta = 0.22$ (estimate from above)
- *First-order conditional* method via version 8.01 of SAS macro `nlinmix` with `expand=eblup` – fix $\zeta = 0.22$
- *First-order conditional* method via R function `nlme` (estimate ζ)
- *Maximum likelihood* via SAS `proc nlmixed` with *adaptive Gaussian quadrature* – does not support *non-constant intra-individual variance* $\Longrightarrow$ "*transform-both-sides*" with $\delta = 1 - \zeta \approx 0.75$

$$(Y_{ij}^{\delta} - 1)/\delta = [\{m(t_{ij}, \boldsymbol{U}_i, \boldsymbol{\theta}_i)\}^{\delta} - 1]/\delta + e_{ij}, \quad \boldsymbol{e}_i \,|\, \boldsymbol{U}_i, \boldsymbol{b}_i \sim N(\boldsymbol{0}, \sigma_e^2 I_{n_i})$$

Implementation and examples

Abridged code: Full code at website for *Longitudinal Data Analysis*

`http://www.biostat.harvard.edu/~fitzmaur/lda/`

First-order method: SAS `nlinmix` with `expand=zero`

First-order conditional method: SAS `nlinmix` with `expand=blup`

```
%inc 'nlmm801.sas' / nosource; *  nlinmix macro;

data arg; infile 'argconc.dat';
   input obsno indiv dose time conc;
   tinf=240;
   t1=1; if time>tinf then t1=0; t2=tinf*(1-t1)+t1*time;
run;
```

Implementation and examples

```
%nlinmix(data=arg,
   model=%str(
     logcl=beta1+b1; logv=beta2+b2; cl=exp(logcl); v=exp(logv);
     predv=(dose/cl)*(1-exp(-cl*t2/v))*exp(-cl*(1-t1)*(time-tinf)/v);
   ),
   derivs=%str( wt=1/predv**(2*0.22); ),
   parms=%str(beta1=-6.0 beta2=-2.0),
   stmts=%str(
      class indiv;
      model pseudo_conc = d_beta1 d_beta2 / noint notest solution;
      random d_b1 d_b2 / subject=indiv type=un solution;
      weight wt;
    ),
   expand=zero,   *  or expand=eblup,
   procopt=%str(maxiter=500 method=ml)
   )
run;
```

Implementation and examples

Abridged output: *First-order method*

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
UN(1,1)	indiv	0.1578
UN(2,1)	indiv	-0.00308
UN(2,2)	indiv	0.01676
Residual		699.80

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
d_beta1	-5.4889	0.06629	401	-82.80	<.0001
d_beta2	-1.8277	0.03429	401	-53.30	<.0001

Implementation and examples

Abridged output: *First-order conditional method*

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
UN(1,1)	indiv	0.1378
UN(2,1)	indiv	0.005669
UN(2,2)	indiv	0.004761
Residual		549.08

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
d_beta1	-5.4325	0.06212	401	-87.46	<.0001
d_beta2	-1.9256	0.02527	401	-76.19	<.0001

Implementation and examples

First-order conditional method: R function `nlme`

```
library(nlme)   #  access nlme()

thedat <- read.table("argconc.dat",col.names=c('obsno','indiv',
  'dose','time','conc'))

meanfunc <- function(x,b1,b2,dose){
   tinf <- 240; cl <- exp(logcl); v <- exp(logv)
   t1 <- x<=tinf; t2 <- tinf*(1-t1)+t1*x;
   f1 <- (dose/cl)*(1-exp(-cl*t2/v))*exp(-cl*(1-t1)*(x-tinf)/v)
   f1
}
```

Implementation and examples

```
arg.mlfit <- nlme(conc ~ meanfunc(time,logcl,logv,dose),
   fixed = list(logcl ~ 1,logv ~1),
   random = list(logcl ~ 1,logv ~ 1),
   groups = ~ indiv, data = thedat,
   start = list(fixed = c(-6.0,-2.0)),
   method="ML", verbose=T, weights=varPower(0.5))
```

Abridged output:

```
Nonlinear mixed-effects model fit by maximum likelihood
       AIC      BIC    logLik
  5738.429 5767.572 -2862.214

Random effects:   Formula: list(b1 ~ 1, b2 ~ 1)
 Level: indiv
 Structure: General positive-definite, Log-Cholesky parametrization
         StdDev      Corr
b1       0.37168333   b1
b2       0.06753254   0.268      Residual 20.42295300
```

Implementation and examples

```
Variance function:
 Structure: Power of variance covariate
 Formula: ~fitted(.)
 Parameter estimates:
    power
0.2432619
Fixed effects: list(b1 ~ 1, b2 ~ 1)
       Value  Std.Error  DF   t-value p-value
b1 -5.432546 0.06230325 437 -87.19522       0
b2 -1.917993 0.02513039 437 -76.32165       0
 Correlation:
   b1
b2 0.156

Number of Observations: 475
Number of Groups: 37
Estimate of sigma 20.42295
```

Implementation and examples

Maximum likelihood: SAS `proc nlmixed`

```
data arg; set arg; conctrans = conc**0.75; run;

proc nlmixed data=arg;
   parms beta1=-6.0 beta2=-2.0 s2b1=0.14 cb12=0.006
         s2b2=0.006 s2=23.0;
   logcl=beta1+b1;
   logv=beta2+b2;
   cl=exp(logcl);
   v=exp(logv);
   pred=((dose/cl)*(1-exp(-cl*t2/v))
               *exp(-cl*(1-t1)*(time-tinf)/v))**0.75;
   model conctrans ~ normal(pred,s2);
   random b1 b2 ~ normal([0,0],[s2b1,cb12,s2b2]) subject=indiv;
run;
```

Implementation and examples

Abridged output:

Fit Statistics

-2 Log Likelihood	4007.8
AIC (smaller is better)	4019.8

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|
beta1	-5.4237	0.06277	35	-86.40	<.0001
beta2	-1.9238	0.02972	35	-64.73	<.0001
s2b1	0.1411	0.03389	35	4.16	0.0002
cb12	0.006562	0.01020	35	0.64	0.5242
s2b2	0.006010	0.006141	35	0.98	0.3345
s2	192.72	13.6128	35	14.16	<.0001

Implementation and examples

Method	β_1	β_2	σ_e	ζ	G_{11}	G_{12}	G_{22}
Indiv. est.	-5.433	-1.927	23.47	0.22	0.137	6.06	6.17
	(0.062)	(0.026)					
First-order	-5.490	-1.828	26.45	–	0.158	-3.08	16.76
`nlinmix`	(0.066)	(0.034)					
First-order cond.	-5.432	-1.926	23.43	–	0.138	5.67	4.76
`nlinmix`	(0.062)	(0.026)					
First-order cond.	-5.433	-1.918	20.42	0.24	0.138	6.73	4.56
`nlme`	(0.063)	(0.025)					
ML	-5.424	-1.924	13.88	–	0.141	6.56	6.01
`nlmixed`	(0.063)	(0.030)					

Values for G_{12}, G_{22} are multiplied by 10^3

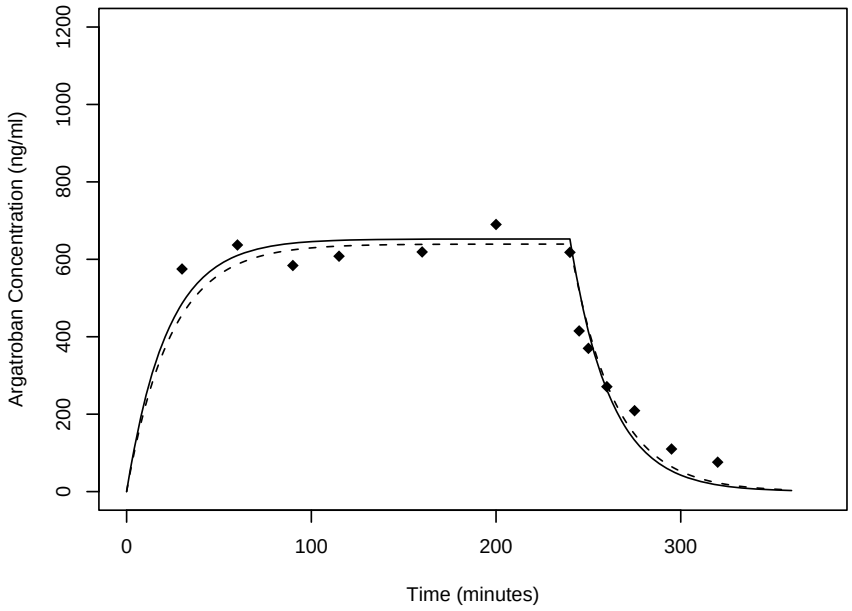
Implementation and examples

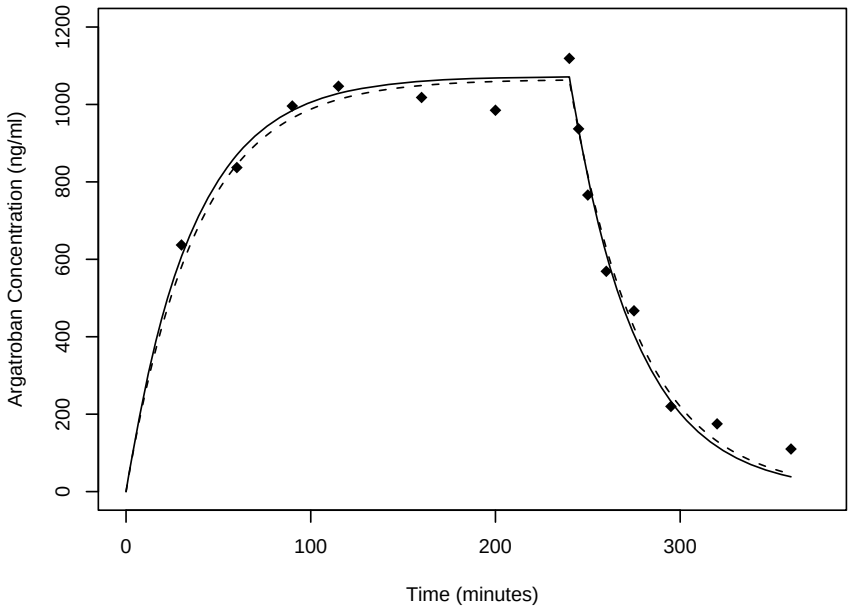
Interpretation: Concentrations measured in ng/ml $=$ 1000 μg/ml

- *Median* argatroban *clearance* $\approx$ 4.4 μg/ml/kg
  ($\approx \exp(-5.43) \times 1000$)
- *Median* argatroban *volume* $\approx$ 145.1 ml/kg $\Longrightarrow \approx$ 10 liters for a 70 kg subject
- Assuming Cl_i, V_i approximately *lognormal*
  - $G_{11} \approx \sqrt{0.14} \times 100 \approx 37\%$ *coefficient of variation* for *clearance*
  - $G_{22} \Longrightarrow$ 8% CV for *volume*

Implementation and examples

Individual inference: *Individual estimate* (dashed) and *empirical Bayes estimate* (solid)





Implementation and examples

Example 2: A simple *population PK study* analysis: *phenobarbital*

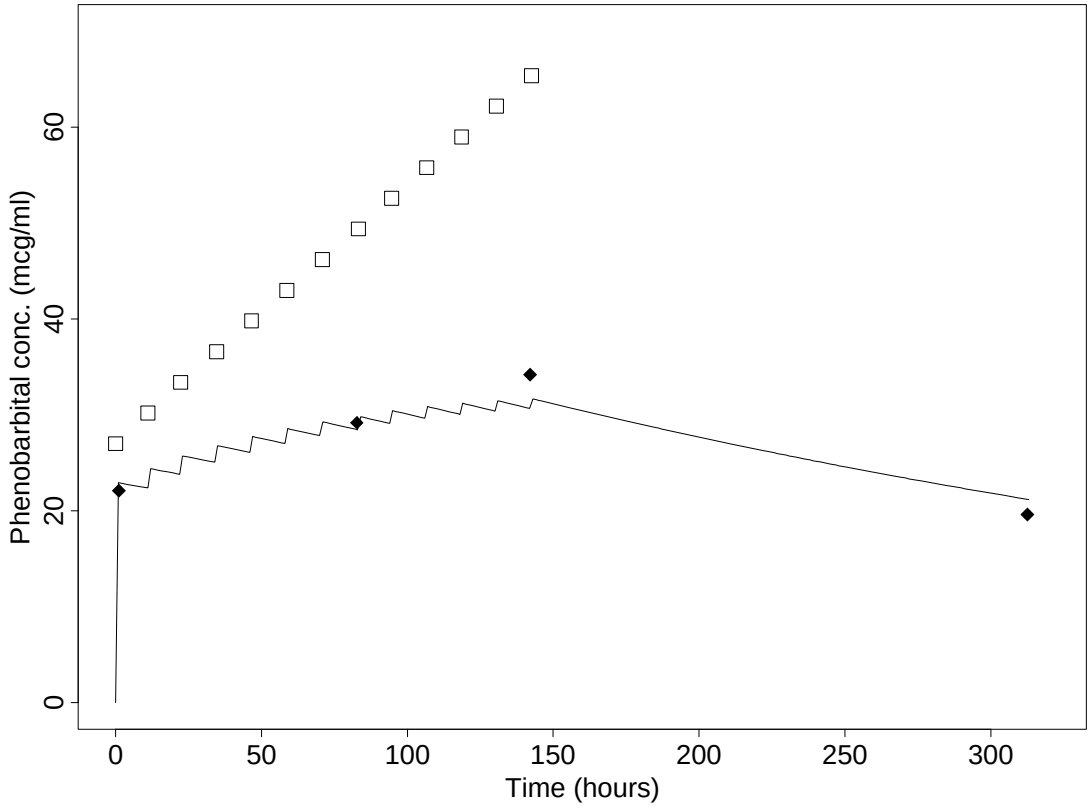
- *World-famous* example
- $N = 59$ preterm infants treated with *phenobarbital* for *seizures*
- $n_i = 1$ to 6 concentration measurements per infant, total of 155
- *Among-infant covariates* $(\boldsymbol{A}_i)$: *Birth weight* w_i (kg), 5-minute Apgar score $\delta_i = \mathrm{I}[\text{Apgar} < 5]$
- *Multiple intravenous doses*: $\boldsymbol{U}_i = (s_{i\ell}, D_{i\ell})$, $\ell = 1, \ldots, d_i$
- *One-compartment* model (*principle of superposition*)

$$m(t, \boldsymbol{U}_i, \boldsymbol{\theta}_i) = \sum_{\ell : s_{i\ell} < t} \frac{D_{i\ell}}{V_i} \exp\left\{ -\frac{Cl_i}{V_i}(t - s_{i\ell}) \right\}$$

- *Objectives*: Characterize PK and its variation – *Mean/median* Cl_i, V_i? *Systematic associations* with *among-infant covariates*? Extent of *unexplained variation*?

Implementation and examples

Dosing history and concentrations for one infant:



Implementation and examples

Non-linear mixed model:

- *Stage 1 – Individual-level model*

$$E(Y_{ij} \,|\, \boldsymbol{U}_i, \boldsymbol{\theta}_i) = m(t_{ij}, \boldsymbol{U}_i, \boldsymbol{\theta}_i), \quad \text{Cov}(\boldsymbol{Y}_i \,|\, \boldsymbol{U}_i, \boldsymbol{A}_i) = V_i(\boldsymbol{U}_i, \boldsymbol{\theta}_i, \boldsymbol{\alpha}) = \sigma_e^2 \, I_{n_i}$$

  $\Longrightarrow$ *negligible autocorrelation*, measurement error *dominates* and has *constant variance*

- *Stage 2 – Population model*
  - *Without* among-infant covariates $\boldsymbol{A}_i$

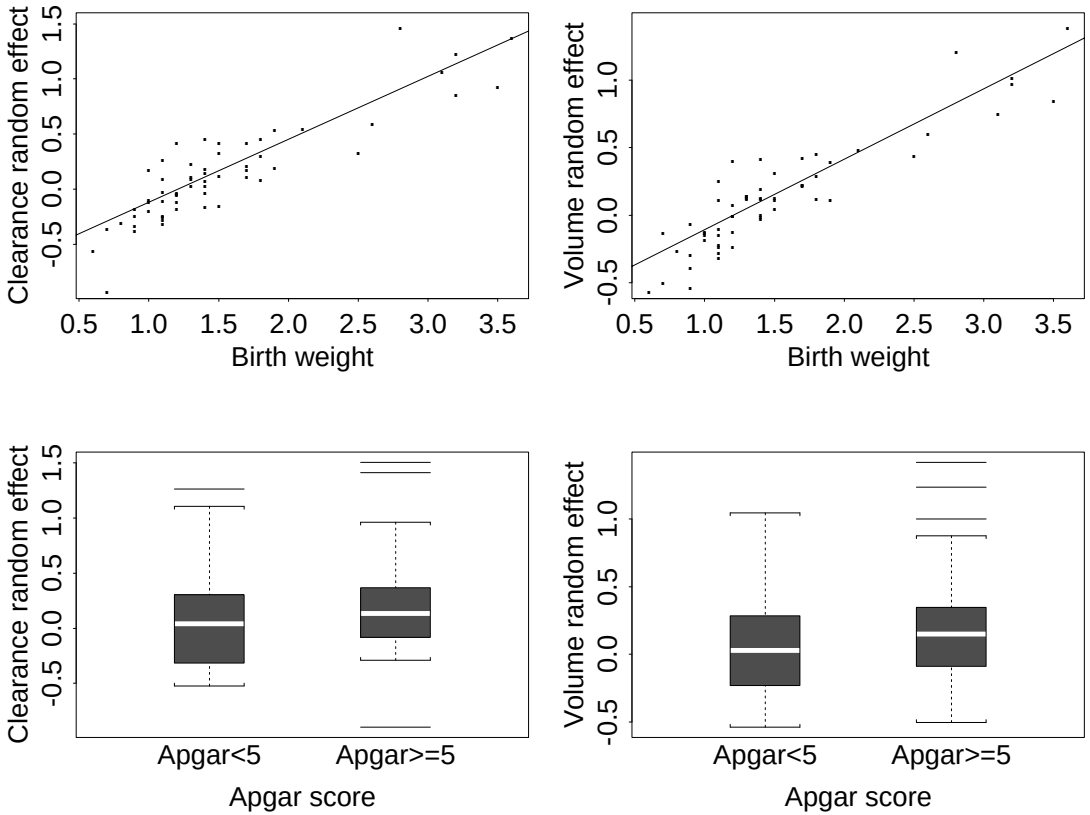
$$\log Cl_i = \beta_1 + b_{i1}, \quad \log V_i = \beta_2 + b_{i2}$$

  - *With* among-infant covariates $\boldsymbol{A}_i$

$$\log Cl_i = \beta_1 + +\beta_3 w_i + b_{i1}, \quad \log V_i = \beta_2 + +\beta_4 w_i + \beta_5 \delta_i + b_{i2}$$

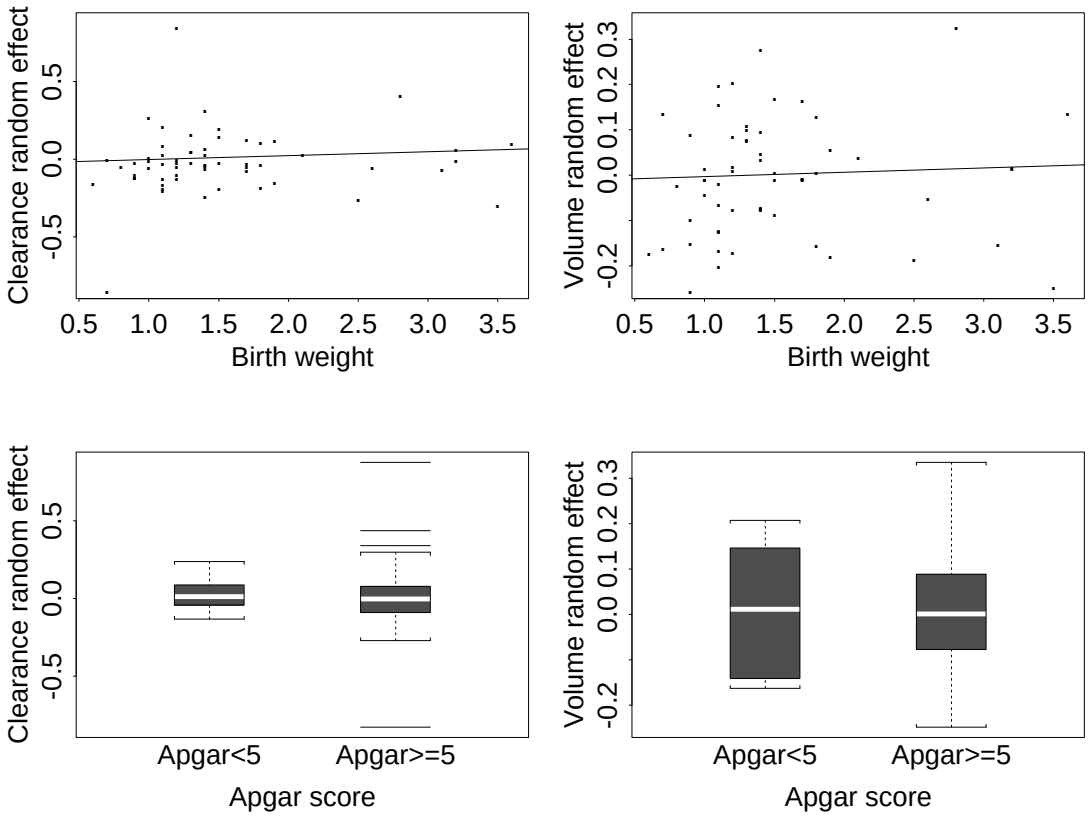
Implementation and examples

Empirical Bayes estimates vs. covariates: Fit *without*



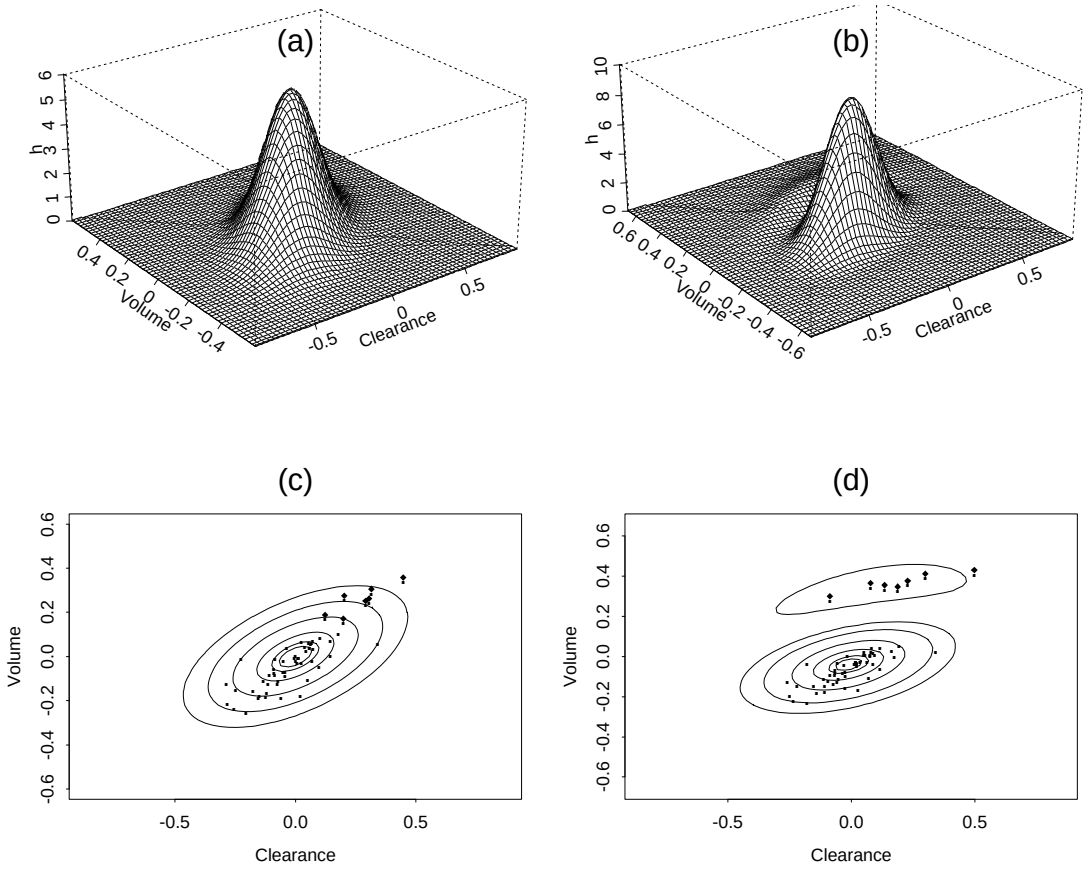
Implementation and examples

Empirical Bayes estimates vs. covariates: Fit *with*



Implementation and examples

Relaxing the normality assumption on b_i: Represent the density of $\boldsymbol{b}_i$ by a *flexible form*, fit by *maximum likelihood*



Extensions

Multivariate response: *More than one type of response* measured *longitudinally* on *each individual*

- *Objectives:* Understand the relationships between the response trajectories and the *processes underlying* them
- Key example: *pharmacokinetic/pharmacodynamic* (PK/PD) analysis
- PD – "*What the drug does to the body*"

Example: *Argatroban* study

- *In addition* to drug concentrations, samples at 5-9 time points from 0 to 540 min (not necessarily the *same* as for concentrations) $\Longrightarrow$ measure *activated partial thromboplastin time* (aPTT)
- aPTT is the *pharmacodynamic* response
- *Goal*: Elucidate the *relationships* between argatroban concentration and aPTT and among underlying PK and PD processes

Extensions

Required: A *joint model* for PK and PD

- *Data*:
  - Y_{ij}^{PK} at times t_{ij}^{PK} (*PK* concentrations)
  - Y_{ij}^{PD} at times t_{ij}^{PD} (*PD* aPTT responses)
- *One compartment model for PK*

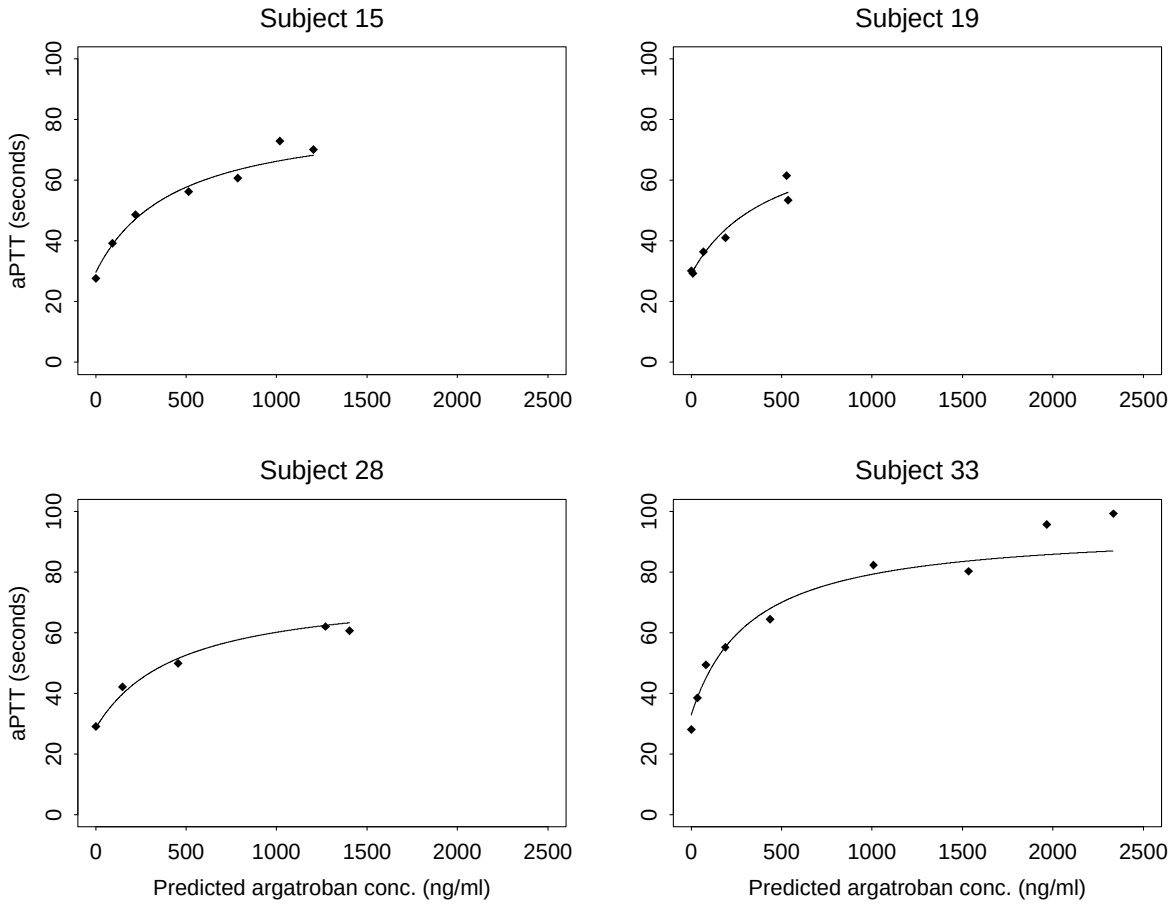
$$m^{PK}(t, \boldsymbol{U}_i, \boldsymbol{\theta}_i^{PK}) = \frac{D_i}{e^{Cl_i^*}} \left[\exp\left\{ \frac{e^{Cl_i^*}}{e^{V_i^*}} (t - t_{\mathsf{inf}})_+ \right\} - \exp\left(-\frac{e^{Cl_i^*}}{e^{V_i^*}} t \right) \right]$$

$$\boldsymbol{\theta}_i^{PK} = (Cl_i^*, V_i^*)', \quad \boldsymbol{U}_i = (D_i, t_{\mathsf{inf}})$$

- PK analysis $\Longrightarrow$ can obtain *individual estimates* $\widehat{\boldsymbol{\theta}}_i^{PK}$ and *predicted concentrations* $m(t_{ij}^{PD}, \widehat{\boldsymbol{\theta}}_i^{PK})$
- $\Longrightarrow$ *plot* Y_{ij}^{PD} vs. $m(t_{ij}^{PD}, \widehat{\boldsymbol{\theta}}_i^{PK})$

Extensions

Concentration-PD response relationship:



Extensions

Suggests: *Empirical model* for concentration-aPTT response relationship – sigmoidal "*E_{max} model*"

$$\text{aPTT} = m^{PD}(\text{conc}, \boldsymbol{\theta}^{PD}) = E_0 + \frac{E_{\max} - E_0}{1 + EC_{50}/\text{conc}}$$

$$\boldsymbol{\theta}^{PD} = (E_0, E_{\max}, EC_{50})'$$

Result: Assuming *measurement error dominates* realization variation, so "*true*" PK concentration for i at $t \approx m(t, \boldsymbol{U}_i, \boldsymbol{\theta}_i^{PK})$

- *Stage 1 – Individual-level model*

$$\begin{aligned} Y_{ij}^{PK} &= m^{PK}(t_{ij}^{PK}, \boldsymbol{U}_i, \boldsymbol{\theta}_i^{PK}) + e_{ij}^{PK} \\ Y_{ij}^{PD} &= m^{PD}\{m^{PK}(t_{ij}^{PD}, \boldsymbol{U}_i, \boldsymbol{\theta}_i^{PK}), \boldsymbol{\theta}_i^{PD}\} + e_{ij}^{PD} \end{aligned}$$

- e_{ij}^{PK}, e_{ij}^{PD} mutually *independent* (primarily *measurement error*)

Extensions

Full model: *Combined* responses $\boldsymbol{Y}_i = (\boldsymbol{Y}_i^{PK\prime}, \boldsymbol{Y}_i^{PD\prime})'$

$$\boldsymbol{\theta}_i = (\boldsymbol{\theta}_i^{PK\prime}, \boldsymbol{\theta}_i^{PD\prime})' = (Cl_i^*, V_i^*, E_{0i}, E_{\mathsf{max},i}, EC_{50i})'$$

- *Stage 1 – Individual-level model*

$$E(Y_{ij}^{PK}|\boldsymbol{U}_i, \boldsymbol{\theta}_i) = m^{PK}(t_{ij}^{PK}, \boldsymbol{U}_i, \boldsymbol{\theta}_i^{PK})$$

$$E(Y_{ij}^{PD}|\boldsymbol{U}_i, \boldsymbol{\theta}_i) = m^{PD}\{m^{PK}(t_{ij}^{PD}, \boldsymbol{U}_i, \boldsymbol{\theta}_i^{PK}), \boldsymbol{\theta}_i^{PD}\}$$

$$\mathsf{Cov}(\boldsymbol{Y}_i|\boldsymbol{U}_i, \boldsymbol{\theta}_i) = \text{block diag}\{V_i^{PK}(\boldsymbol{U}_i, \boldsymbol{\theta}_i, \boldsymbol{\alpha}^{PK}), V_i^{PD}(\boldsymbol{U}_i, \boldsymbol{\theta}_i, \boldsymbol{\alpha}^{PD})\}$$

$$V_i^{PK}(\boldsymbol{U}_i, \boldsymbol{\theta}_i, \boldsymbol{\alpha}^{PK}) = \sigma_{e,PK}^2 \mathsf{diag}[\ldots, \{m^{PK}(t_{ij}^{PK}, \boldsymbol{U}_i, \boldsymbol{\theta}_i^{PK})\}^{2\zeta^{PK}}, \ldots]$$

$$V_i^{PD}(\boldsymbol{U}_i, \boldsymbol{\theta}_i, \boldsymbol{\alpha}^{PD}) = \sigma_{e,PD}^2 \mathsf{diag}\Big(\ldots, [m^{PD}\{m^{PK}(t_{ij}^{PD}, \boldsymbol{U}_i, \boldsymbol{\theta}_i^{PK}), \boldsymbol{\theta}_i^{PD}\}]^{2\zeta^{PD}}, \ldots\Big)$$

- *Stage 2 – Population model*

$$\boldsymbol{\theta}_i = \boldsymbol{\beta} + \boldsymbol{b}_i, \quad \boldsymbol{\beta} = (\beta_1, \ldots, \beta_5)', \quad \boldsymbol{b}_i \sim N(\boldsymbol{0}, G)$$

Extensions

Time-dependent among-individual covariates: Among-individual covariates *change* over time *within an individual*

- *In principle*, one could write $\boldsymbol{\theta}_{ij}$ for each t_{ij}; *however*...
- *Key issue*: Does this make *scientific sense*?
- *PK*: Do *pharmacokinetic processes* vary *within* an individual?

Example: *Quinidine* study

- Creatinine clearance, α_1-acid glycoprotein concentration, etc, *change* over *dosing intervals*
- How to incorporate *dependence of* Cl_i, V_i on α_1-acid glycoprotein concentration?

Extensions

Data for a representative subject:

time (hours)	conc. (mg/L)	dose (mg)	age (years)	weight (kg)	creat. (ml/min)	glyco. (mg/dl)
0.00	–	166	75	108	> 50	69
6.00	–	166	75	108	> 50	69
11.00	–	166	75	108	> 50	69
17.00	–	166	75	108	> 50	69
23.00	–	166	75	108	> 50	69
27.67	0.7	–	75	108	> 50	69
29.00	–	166	75	108	> 50	94
35.00	–	166	75	108	> 50	94
41.00	–	166	75	108	> 50	94
47.00	–	166	75	108	> 50	94
53.00	–	166	75	108	> 50	94
65.00	–	166	75	108	> 50	94
71.00	–	166	75	108	> 50	94
77.00	0.4	–	75	108	> 50	94
161.00	–	166	75	108	> 50	88
168.75	0.6	–	75	108	> 50	88

height=72 inches, Caucasian, smoker, no ethanol abuse, no CHF

Extensions

Population model: *Standard approach in PK*

- For subject i: α_1-acid glycoprotein concentration likely measured *intermittently* at times 0, 29, 161 hours and *assumed constant* over the intervals (0,29), (29,77), (161,·) hours
- For intervals I_k, $k = 1, \ldots, a$ ($a = 3$ here), $\boldsymbol{A}_{ik}$ = among-individual covariates for $t_{ij} \in I_k \Longrightarrow$ e.g., linear model

$$\boldsymbol{\theta}_{ij} = \boldsymbol{A}_{ik}\boldsymbol{\beta} + \boldsymbol{b}_i$$

- This *population model* assumes "*within subject inter-interval variation*" entirely "*explained*" by changes in covariate values
- *Alternatively*: *Nested* random effects

$$\boldsymbol{\theta}_{ij} = \boldsymbol{A}_{ik}\boldsymbol{\beta} + \boldsymbol{b}_i + \boldsymbol{b}_{ik}, \quad \boldsymbol{b}_i, \boldsymbol{b}_{ik} \text{ independent}$$

Extensions

Multi-level models: More *generally*

- *Nesting*: E.g., responses Y_{ikj}, $j = 1, \ldots, n_{ik}$, on several trees $(k = 1, \ldots, v_i)$ within each of several plots $(i = 1, \ldots, N)$

$$\boldsymbol{\theta}_{ik} = \boldsymbol{A}_{ik}\boldsymbol{\beta} + \boldsymbol{b}_i + \boldsymbol{b}_{ik}, \quad \boldsymbol{b}_i, \boldsymbol{b}_{ik} \text{ independent}$$

Missing/mismeasured covariates: $\boldsymbol{A}_i$, $\boldsymbol{U}_i$, t_{ij}

Censored response: E.g., due to an *assay quantification limit*

Semiparametric models: Allow $m(t, \boldsymbol{U}_i, \boldsymbol{\theta}_i)$ to depend on an *unspecified function* $g(t, \boldsymbol{\theta}_i)$

- *Flexibility*, *model misspecification*

Clinical trial simulation: "*Virtual*" subjects simulated from a *non-linear mixed-effects model* for *PK/PD/disease progression* linked to a *clinical end-point*

Discussion

Summary:

- The *non-linear mixed-effects model* is now a *standard statistical framework* in many areas of application
- Is appropriate when scientific interest focuses on *within-individual mechanisms/processes* that can be represented by *parameters* in a *non-linear* (often *theoretical*) *model* for *individual time course*
- Free and commercial *software* is available, but implementation is still *complicated*
- *Specification* of models and assumptions, particularly the *population model*, is somewhat an *art-form*
- *Current challenge*: *High-dimensional* $\boldsymbol{A}_i$ (e.g., *genomic* information)
- Still plenty of *methodological research* to do

Discussion

See the references on slide 3 for an extensive bibliography