NONLINEAR MIXED EFFECTS MODELS

An Overview and Update

Marie Davidian
Department of Statistics
North Carolina State University

http://www.stat.ncsu.edu/~davidian

Based on: Davidian, M. and Giltinan, D.M. (2003), "Nonlinear Models for Repeated Measurement Data: An Overview and Update," *JABES* 8, 387–419

Outline

- Introduction
- The Setting
- The Model
- Inferential Objectives and Model Interpretation
- Implementation
- Extensions and Recent Developments
- Discussion

Introduction

Common situation in agricultural, environmental, and biomedical applications:

- A *continuous response* evolves over *time* (or other condition) *within* individuals from a population of interest
- Inference focuses on *features* or *mechanisms* that underlie individual profiles of *repeated measurements* of the response and how these *vary* in the population
- A *theoretical or empirical model* for individual profiles with parameters that may be interpreted as representing such features or mechanisms is available

Introduction

Nonlinear mixed effects model: aka *hierarchical nonlinear model*

- A formal *statistical framework* for this situation
- A "*hot*" methodological research area in the early 1990s
- Now *widely accepted* as a suitable approach to inference, with applications routinely reported and commercial *software* available
- Many recent *extensions*, *innovations*

Introduction

Nonlinear mixed effects model: aka *hierarchical nonlinear model*

- A formal *statistical framework* for this situation
- A "*hot*" methodological research area in the early 1990s
- Now *widely accepted* as a suitable approach to inference, with applications routinely reported and commercial *software* available
- Many recent *extensions*, *innovations*

Objective of this talk: An updated review of the model and survey of recent advances

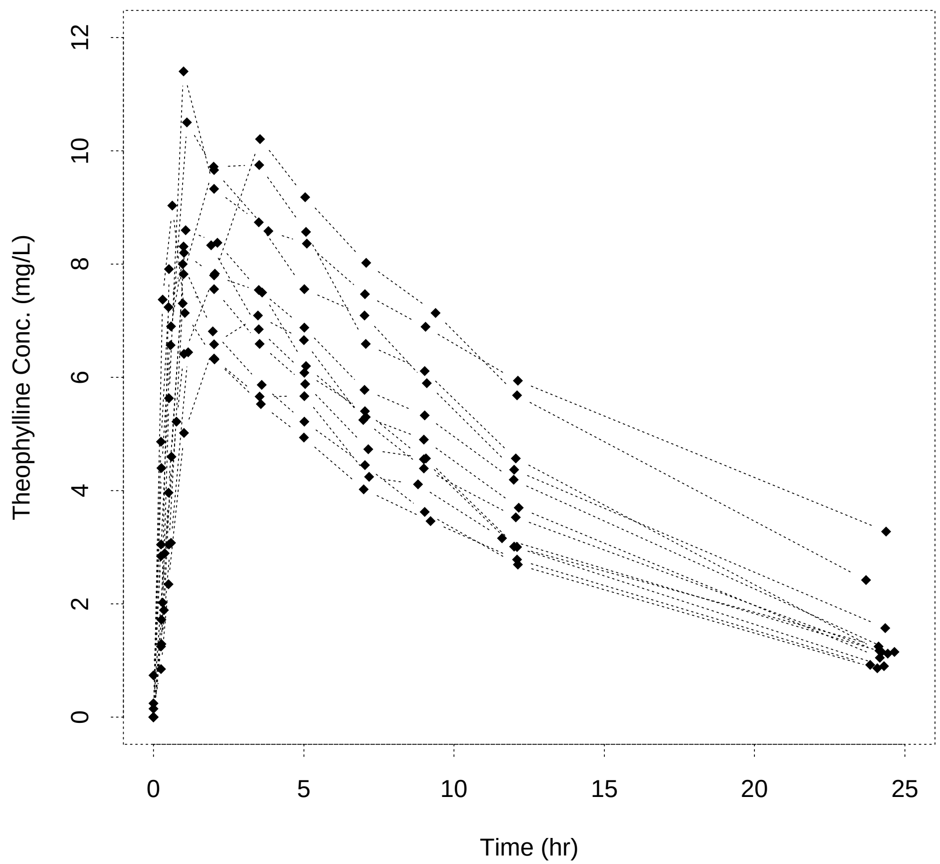
The Setting

Example 1: *Pharmacokinetics*

- *Broad goal*: Understand *intra-subject* processes of drug *absorption*, *distribution*, and *elimination* governing achieved concentrations
- . . . and how these *vary* across subjects
- Critical for developing *dosing* strategies and guidelines

The Setting

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



The Setting

Example 1: *Pharmacokinetics* (PK)

- *Similarly-shaped* concentration-time profiles across subjects
- . . . but peak, rise, decay *vary* considerably
- Attributable to *inter-subject variation* in underlying PK processes (absorption, etc)

The Setting

Example 1: *Pharmacokinetics* (PK)

- Standard: approximate representation of the body by simple *compartment models* (differential equations)
- *One-compartment model* for theophylline following oral dose D at time $t = 0$ leads to description of concentration $C(t)$ at time $t \geq 0$

$$C(t) = \frac{Dk_a}{V(k_a - Cl/V)} \left\{ \exp(-k_a t) - \exp\left(-\frac{Cl}{V} t\right) \right\}$$

k_a	fractional rate of absorption (1/time)
Cl	clearance rate (volume/time)
V	volume of distribution

- (k_a, Cl, V) summarize *PK processes* underlying observed concentration profiles for a given subject

The Setting

Example 1: *Pharmacokinetics* (PK)

- *Goal, more precisely stated*: Determine mean/median values of (k_a, Cl, V) and how they vary in the *population of subjects*
- Elucidate whether some of this variation is *associated* with *subject characteristics* (e.g. weight, age, renal function)
- Develop dosing strategies for *subpopulations* with certain characteristics (e.g. the elderly)

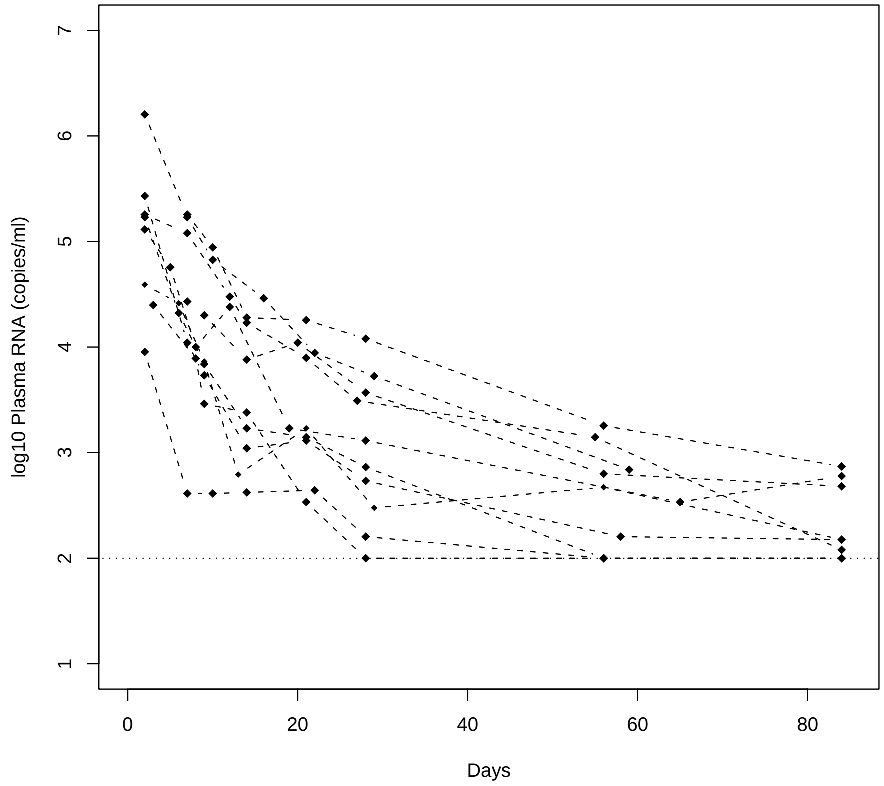
The Setting

Example 2: *HIV Dynamics*

- Monitoring of "*viral load*" (concentration of virus) is now routine for HIV-infected patients
- *Broad goal*: Characterize mechanisms underlying the *interaction* between *HIV virus* and the *immune system* governing decay (and rebound) of virus levels following treatment with Highly Active AntiRetroviral Therapy (HAART)

The Setting

ACTG 315: (log) Viral load profiles for 10 subjects following HAART



The Setting

Example 2: *HIV Dynamics*

- *Similarly-shaped* profiles with different decay patterns
- *Complication* – Viral load assay has *lower limit of quantification*

The Setting

Example 2: *HIV Dynamics*

- Represent body by *system of ordinary differential equations*, e.g.

$$\frac{dX}{dt} = (1-\epsilon)kVT - \delta X$$
$$\frac{dV}{dt} = pX - cV$$

X, T	size of infected, uninfected immune cell populations		
V	size of viral population,	c	viral clearance
δ	infected cell death rate,	p	viral production rate
k	probability of infection,	ϵ	treatment efficacy

- *Parameters* characterize intra-subject mechanisms related to interaction between virus and immune system

The Setting

Example 2: *HIV Dynamics*

- *Complication* – Expression for V (viral load) may not be available in a *closed form*
- *Further complication* – All *states* of the system of ODEs may not have been *measured*
- *Goal, more precisely stated*: Elucidate "*typical*" parameter values (mean/median), *variation* across subjects, *associations* with measures of pre-treatment disease status

The Setting

Example 3: *Forestry*

- Interest in impact of *silvicultural treatments* and *soil types* on features of profiles of forest *growth yield*
- Individual-tree *growth model*, e.g. *Richards model* for dominant height $H(t)$ at stand age t

$$H(t) = A\{1 - \exp(-bt)\}^c$$

A	asymptotic value of dominant height
b	rate parameter
c	shape parameter

- Goal: Determine "*typical*" values, whether *variation* in parameters is associated with factors such as treatments and soil types

The Setting

Further applications:

- Dairy science
- Wildlife science
- Fisheries science
- Biomedical science

The Model

Basic model: The *data* are *repeated measurements* on each of m subjects

y_{ij}	response at jth "time" t_{ij} for subject i
$\boldsymbol{u}_i$	vector of additional conditions under which i is observed
$\boldsymbol{a}_i$	vector of characteristics for subject i

$$i = 1, \ldots, m, \quad j = 1, \ldots, n_i, \quad \boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})^T$$

$$(\boldsymbol{y}_i, \boldsymbol{u}_i, \boldsymbol{a}_i) \text{ are independent across } i$$

Example: *Theophylline pharmacokinetics*

- y_{ij} is drug concentration for subject i at time t_{ij} post-dose
- $\boldsymbol{u}_i = D_i$ is dose given to subject i at time zero
- $\boldsymbol{a}_i$ contains subject characteristics such as weight, age, renal function, smoking status, etc.

The Model

Basic model: *Stage 1* – *Individual-level* model

$$y_{ij} = f(t_{ij}, \boldsymbol{u}_i, \boldsymbol{\beta}_i) + e_{ij}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n_i$$

f	function governing *within-individual* behavior
$\boldsymbol{\beta}_i$	parameters of f specific to individual i $(p \times 1)$
e_{ij}	satisfy $E(e_{ij} \vert \boldsymbol{u}_i, \boldsymbol{\beta}_i) = 0$

Example: *Theophylline pharmacokinetics*

- f is the one-compartment model with dose $\boldsymbol{u}_i = D_i$
- $\boldsymbol{\beta}_i = (k_{ai}, V_i, Cl_i)^T = (\beta_{1i}, \beta_{2i}, \beta_{3i})^T$, where k_{ai}, V_i, and Cl_i are absorption rate, volume, and clearance for subject i

$$f(t, \boldsymbol{u}_i, \boldsymbol{\beta}_i) = \frac{D_i k_{ai}}{V_i(k_{ai} - Cl_i/V_i)} \left\{ \exp(-k_{ai}t) - \exp\left(-\frac{Cl_i}{V_i} t \right) \right\}$$

The Model

Basic model: *Stage 2* – *Population* model

$$\boldsymbol{\beta}_i = \boldsymbol{d}(\boldsymbol{a}_i, \boldsymbol{\beta}, \boldsymbol{b}_i), \quad i = 1, \ldots, m$$

$\boldsymbol{d}$ p-dimensional function

$\boldsymbol{\beta}$ *fixed effects* $(r \times 1)$

$\boldsymbol{b}_i$ *random effects* $(k \times 1)$

Characterizes how elements of $\boldsymbol{\beta}_i$ *vary* across individuals due to

- *Systematic association* with $\boldsymbol{a}_i$ (modeled via $\boldsymbol{\beta}$)
- *Unexplained variation* in the population (represented by $\boldsymbol{b}_i$)
- *Usual assumption* $E(\boldsymbol{b}_i|\boldsymbol{a}_i) = E(\boldsymbol{b}_i) = \mathbf{0}$, var$(\boldsymbol{b}_i|\boldsymbol{a}_i) =$ var$(\boldsymbol{b}_i) = \boldsymbol{D}$ (can be *relaxed*)

The Model

Basic model: *Stage 2* – *Population* model

$$\boldsymbol{\beta}_i = \boldsymbol{d}(\boldsymbol{a}_i, \boldsymbol{\beta}, \boldsymbol{b}_i), \quad i = 1, \ldots, m$$

Example: *Theophylline pharmacokinetics*

- E.g. $\boldsymbol{a}_i = (c_i, w_i)^T$, $c_i = I($ creatinine clearance > 50 ml/min $)$, $w_i =$ weight (kg)
- $\boldsymbol{b}_i = (b_{1i}, b_{2i}, b_{3i})^T$ $(p = k = 3)$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_7)^T$ $(r = 7)$

$$\begin{aligned} k_{ai} &= \exp(\beta_1 + b_{1i}) \\ V_i &= \exp(\beta_2 + \beta_3 w_i + b_{2i}) \\ Cl_i &= \exp(\beta_4 + \beta_5 w_i + \beta_6 c_i + \beta_7 w_i c_i + b_{3i}) \end{aligned}$$

- If $\boldsymbol{b}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{D})$, k_{ai}, V_i, Cl_i are *lognormal*

The Model

Basic model: *Stage 2* – *Population* model

$$\boldsymbol{\beta}_i = \boldsymbol{d}(\boldsymbol{a}_i, \boldsymbol{\beta}, \boldsymbol{b}_i), \quad i = 1, \ldots, m$$

Example: *Theophylline pharmacokinetics*, continued

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

$$\begin{aligned} k_{ai} &= \exp(\beta_1 + b_{1i}) \\ V_i &= \exp(\beta_2 + \beta_3 w_i) \quad \text{(all population variation due to weight)} \\ Cl_i &= \exp(\beta_4 + \beta_5 w_i + \beta_6 c_i + \beta_7 w_i c_i + b_{3i}) \end{aligned}$$

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

The Model

Basic model: *Stage 2* – *Population* model

$$\boldsymbol{\beta}_i = \boldsymbol{d}(\boldsymbol{a}_i, \boldsymbol{\beta}, \boldsymbol{b}_i), \quad i = 1, \ldots, m$$

Example: *Theophylline pharmacokinetics*, continued

- *Alternative* parameterization – *reparameterize* f in terms of $(k_a^*, V^*, Cl^*)^T = (\log k_a, \log V, \log Cl)^T$, $\boldsymbol{\beta}_i = (k_{ai}^*, V_i^*, Cl_i^*)^T$,

$$\begin{aligned} k_{ai}^* &= \beta_1 + b_{1i} \\ V_i^* &= \beta_2 + \beta_3 w_i + b_{2i} \\ Cl_i^* &= \beta_4 + \beta_5 w_i + \beta_6 c_i + \beta_7 w_i c_i + b_{3i} \end{aligned}$$

- Common *special case* – *linear* population model

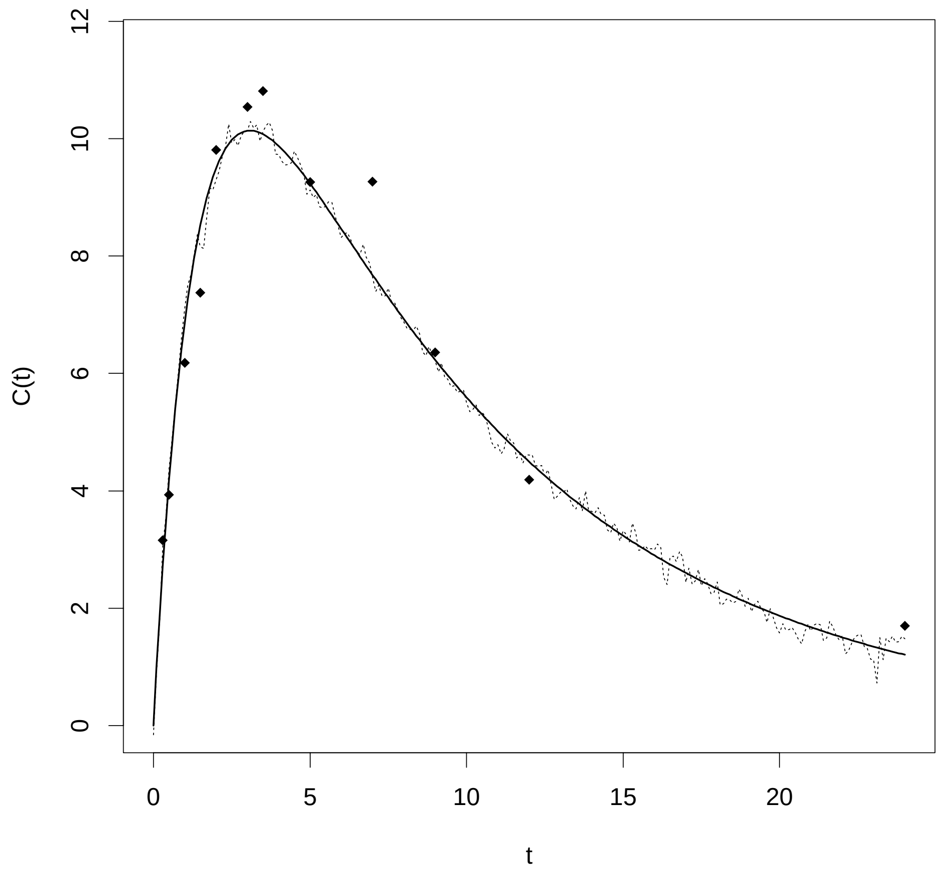
$$\boldsymbol{\beta}_i = \boldsymbol{A}_i \boldsymbol{\beta} + \boldsymbol{B}_i \boldsymbol{b}_i$$

The Model

Within-individual variation: Often *misunderstood*

The Model

Within-individual variation: Often *misunderstood*



The Model

Within-individual variation: *Conceptual* perspective

- $E(y_{ij}|\boldsymbol{u}_i, \boldsymbol{\beta}_i) = f(t_{ij}, \boldsymbol{u}_i, \boldsymbol{\beta}_i) \Longrightarrow f$ represents i's "*on-average*" profile (smooth curve)
- f may not capture all within-individual processes *perfectly*, "*local fluctuations*" $\Longrightarrow$ *actual realized* profile (jittery line)
- $f(t, \boldsymbol{u}_i, \boldsymbol{\beta}_i)$ is average over *all possible realizations* $\Longrightarrow$ "*inherent tendency*" for i's profile evolution
- $\Longrightarrow \boldsymbol{\beta}_i$ is "*inherent characteristic*" of i
- $\Longrightarrow$ Interest focuses on *inherent properties* of individuals rather than actual response realizations

The Model

Within-individual variation: *Conceptual* perspective

- Within-individual *stochastic process*

$$y_i(t, \boldsymbol{u}_i) = f(t, \boldsymbol{u}_i, \boldsymbol{\beta}_i) + e_{R,i}(t, \boldsymbol{u}_i) + e_{M,i}(t, \boldsymbol{u}_i)$$

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

- Thus $y_{ij} = y_i(t_{ij}, \boldsymbol{u}_i)$, $e_{R,i}(t_{ij}, \boldsymbol{u}_i) = e_{R,ij}$, $e_{M,i}(t_{ij}, \boldsymbol{u}_i) = e_{M,ij}$

$$y_{ij} = f(t_{ij}, \boldsymbol{u}_i, \boldsymbol{\beta}_i) + \underbrace{e_{R,ij} + e_{M,ij}}_{e_{ij}}$$

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

- $e_{R,i}(t, \boldsymbol{u}_i)$ = "*realization deviation process*"
- $e_{M,i}(t, \boldsymbol{u}_i)$ = "*measurement error process*"

The Model

Within-individual variation: *Conceptual* perspective

- Model for $e_{R,i}(t, \boldsymbol{u}_i)$ and hence $\boldsymbol{e}_{R,i}$ based on assumptions about actual realization *variance*, *correlation*

$$\text{var}(\boldsymbol{e}_{R,i}|\boldsymbol{u}_i, \boldsymbol{\beta}_i) = \boldsymbol{T}_i^{1/2}(\boldsymbol{u}_i, \boldsymbol{\beta}_i, \boldsymbol{\delta})\boldsymbol{\Gamma}_i(\boldsymbol{\rho})\boldsymbol{T}_i^{1/2}(\boldsymbol{u}_i, \boldsymbol{\beta}_i, \boldsymbol{\delta}), \quad (n_i \times n_i)$$

- Model for $e_{M,i}(t, \boldsymbol{u}_i)$ and hence $e_{M,i}$ based on assumptions about measurement error *variance*

$$\text{var}(\boldsymbol{e}_{M,i}|\boldsymbol{u}_i, \boldsymbol{\beta}_i) = \boldsymbol{\Lambda}_i(\boldsymbol{u}_i, \boldsymbol{\beta}_i, \boldsymbol{\theta}), \quad (n_i \times n_i) \text{ diagonal matrix}$$

- Common *assumption* – realization, measurement error processes *independent* $\Longrightarrow$

$$\text{var}(\boldsymbol{y}_i|\boldsymbol{u}_i, \boldsymbol{\beta}_i) = \text{var}(\boldsymbol{e}_{R,i}|\boldsymbol{u}_i, \boldsymbol{\beta}_i) + \text{var}(\boldsymbol{e}_{M,i}|\boldsymbol{u}_i, \boldsymbol{\beta}_i) = \boldsymbol{R}_i(\boldsymbol{u}_i, \boldsymbol{\beta}_i, \boldsymbol{\xi})$$

$$\boldsymbol{\xi} = (\boldsymbol{\delta}^T, \boldsymbol{\rho}^T, \boldsymbol{\theta}^T)^T$$

The Model

$$\begin{aligned} \text{var}(\boldsymbol{y}_i|\boldsymbol{u}_i,\boldsymbol{\beta}_i) &= \text{var}(\boldsymbol{e}_{R,i}|\boldsymbol{u}_i,\boldsymbol{\beta}_i) + \text{var}(\boldsymbol{e}_{M,i}|\boldsymbol{u}_i,\boldsymbol{\beta}_i) \\ &= \boldsymbol{T}_i^{1/2}(\boldsymbol{u}_i,\boldsymbol{\beta}_i,\boldsymbol{\delta})\boldsymbol{\Gamma}_i(\boldsymbol{\rho})\boldsymbol{T}_i^{1/2}(\boldsymbol{u}_i,\boldsymbol{\beta}_i,\boldsymbol{\delta}) + \boldsymbol{\Lambda}_i(\boldsymbol{u}_i,\boldsymbol{\beta}_i,\boldsymbol{\theta}) \\ &= \boldsymbol{R}_i(\boldsymbol{u}_i,\boldsymbol{\beta}_i,\boldsymbol{\xi}) \end{aligned}$$

Example: *Theophylline pharmacokinetics*

- *Usual assumption* – t_{ij} are sufficiently far apart that *correlation* among $e_{R,ij}$ is *negligible* $(\boldsymbol{\Gamma}_i(\boldsymbol{\rho}) = \boldsymbol{I})$
- *Usual assumption* – *Local fluctuations* are *negligible*, measurement error *dominates* realization error
- $\boldsymbol{R}_i(\boldsymbol{u}_i,\boldsymbol{\beta}_i,\boldsymbol{\xi}) = \boldsymbol{\Lambda}_i(\boldsymbol{u}_i,\boldsymbol{\beta}_i,\boldsymbol{\theta})$ *diagonal* with diagonal elements

$$\text{var}(e_{ij}|\boldsymbol{u}_i,\boldsymbol{\beta}_i) = \text{var}(e_{M,ij}|\boldsymbol{u}_i,\boldsymbol{\beta}_i) = \sigma_M^2 f^{2\theta}(t_{ij},\boldsymbol{u}_i,\boldsymbol{\beta}_i)$$

The Model

Summary: $\boldsymbol{f}_i(\boldsymbol{u}_i, \boldsymbol{\beta}_i) = \{f(\boldsymbol{x}_{i1}, \boldsymbol{\beta}_i), \ldots, f(\boldsymbol{x}_{in_i}, \boldsymbol{\beta}_i)\}^T$, $\boldsymbol{z}_i = (\boldsymbol{u}_i^T, \boldsymbol{a}_i^T)^T$

- *Stage 1* – *Individual-level* model

$$
\begin{aligned}
E(\boldsymbol{y}_i | \boldsymbol{z}_i, \boldsymbol{b}_i) &= \boldsymbol{f}_i(\boldsymbol{u}_i, \boldsymbol{\beta}_i) = \boldsymbol{f}_i(\boldsymbol{z}_i, \boldsymbol{\beta}, \boldsymbol{b}_i) \\
\text{var}(\boldsymbol{y}_i | \boldsymbol{z}_i, \boldsymbol{b}_i) &= \boldsymbol{R}_i(\boldsymbol{u}_i, \boldsymbol{\beta}_i, \boldsymbol{\xi}) = \boldsymbol{R}_i(\boldsymbol{z}_i, \boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{\xi})
\end{aligned}
$$

- *Stage 2* – *Population* model

$$
\boldsymbol{\beta}_i = \boldsymbol{d}(\boldsymbol{a}_i, \boldsymbol{\beta}, \boldsymbol{b}_i), \quad \boldsymbol{b}_i \sim (\boldsymbol{0}, \boldsymbol{D})
$$

The Model

“Within-individual correlation”

- Implies *marginal moments*

$$
\begin{aligned}
E(\boldsymbol{y}_i|\boldsymbol{z}_i) &= \int \boldsymbol{f}_i(\boldsymbol{z}_i, \boldsymbol{\beta}, \boldsymbol{b}_i)\, dF_b(\boldsymbol{b}_i) \\
\text{var}(\boldsymbol{y}_i|\boldsymbol{z}_i) &= E\{\boldsymbol{R}_i(\boldsymbol{z}_i, \boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{\xi})|\boldsymbol{z}_i\} + \text{var}\{\boldsymbol{f}_i(\boldsymbol{z}_i, \boldsymbol{\beta}, \boldsymbol{b}_i)|\boldsymbol{z}_i\}
\end{aligned}
$$

- $E\{\boldsymbol{R}_i(\boldsymbol{z}_i, \boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{\xi})|\boldsymbol{z}_i\}$ = average of realization/measurement variation over population $\Longrightarrow$ diagonal *only if* correlation of within-individual realizations *negligible*
- $\text{var}\{\boldsymbol{f}_i(\boldsymbol{z}_i, \boldsymbol{\beta}, \boldsymbol{b}_i)|\boldsymbol{z}_i\}$ = population variation in “*inherent trajectories*” $\Longrightarrow$ *non-diagonal in general*
- *Result* – Overall pattern of *marginal correlation* is the *aggregate* of correlation due to *both* sources
- Prefer “*aggregate correlation*” to “*within-individual correlation*”

Inferential Objectives and Model Interpretation

Main goal:

- Elements of $\boldsymbol{\beta}_i$ represent *underlying features*
- "*Typical*" values of underlying features, *variation* in these, and *association* with individual characteristics $\Longrightarrow$ inference on $\boldsymbol{\beta}$ and $\boldsymbol{D}$
- $\Longrightarrow$ Deduce an appropriate $\boldsymbol{d}$

Additional goal: "*Individual-level prediction*

- Inference on $\boldsymbol{\beta}_i$, $f(t_0, \boldsymbol{u}_i, \boldsymbol{\beta}_i)$
- "*Borrow strength*" across similar subjects

Inferential Objectives and Model Interpretation

Subject-specific model:

- *Not the same as* the *population averaged* approach of modeling $E(\boldsymbol{y}_i|\boldsymbol{z}_i)$, $\text{var}(\boldsymbol{y}_i|\boldsymbol{z}_i)$ *directly*
- Explicitly *acknowledges* individual behavior
- Interest in the "typical value," variation of underlying features $\boldsymbol{\beta}_i$, *not* in the "typical response profile" and overall variation about it
- Incorporates *scientific assumptions* embedded in the model f for individual behavior

Implementation

Likelihood: With distributional assumptions on $(\boldsymbol{y}_i|\boldsymbol{z}_i, \boldsymbol{b}_i)$ and $\boldsymbol{b}_i$ (almost always *normal*)

$$L(\boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{D}) = \prod_{i=1}^{m} \int p(\boldsymbol{y}_i, \boldsymbol{b}_i|\boldsymbol{z}_i, ; \boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{D})\, d\boldsymbol{b}_i = \prod_{i=1}^{m} \int p(\boldsymbol{y}_i|\boldsymbol{z}_i, \boldsymbol{b}_i; \boldsymbol{\beta}, \boldsymbol{\xi}) p(\boldsymbol{b}_i; \boldsymbol{D})\, d\boldsymbol{b}_i$$

- *Maximize* jointly in $(\boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{D})$
- *Intractable integrations* in general
- Potentially *high-dimensional*, computationally *expensive*
- $\Longrightarrow$ *Approximate* $L(\boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{D})$ by *analytical approximation* to

$$p(\boldsymbol{y}_i|\boldsymbol{z}_i; \boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{D}) = \int p(\boldsymbol{y}_i|\boldsymbol{z}_i, \boldsymbol{b}_i; \boldsymbol{\beta}, \boldsymbol{\xi}) p(\boldsymbol{b}_i; \boldsymbol{D})\, d\boldsymbol{b}_i$$

Implementation

First-order methods: *Combine* both stages as

$$\boldsymbol{y}_i = \boldsymbol{f}_i(\boldsymbol{z}_i, \boldsymbol{\beta}, \boldsymbol{b}_i) + \boldsymbol{R}_i^{1/2}(\boldsymbol{z}_i, \boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{\xi})\boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i | \boldsymbol{z}_i, \boldsymbol{b}_i \sim (\boldsymbol{0}, \boldsymbol{I}_{n_i})$$

- *Taylor series* about $\boldsymbol{b}_i = \boldsymbol{0}$ to linear terms

$$\boldsymbol{y}_i \approx \boldsymbol{f}_i(\boldsymbol{z}_i, \boldsymbol{\beta}, \boldsymbol{0}) + \boldsymbol{Z}_i(\boldsymbol{z}_i, \boldsymbol{\beta}, \boldsymbol{0})\boldsymbol{b}_i + \boldsymbol{R}_i^{1/2}(\boldsymbol{z}_i, \boldsymbol{\beta}, \boldsymbol{0}, \boldsymbol{\xi})\boldsymbol{\epsilon}_i$$

$$\boldsymbol{Z}_i(\boldsymbol{z}_i, \boldsymbol{\beta}, \boldsymbol{b}^*) = \partial/\partial \boldsymbol{b}_i \{\boldsymbol{f}_i(\boldsymbol{z}_i, \boldsymbol{\beta}, \boldsymbol{b}_i)\}|_{b_i = b^*}$$

- Implies
$$\begin{aligned} E(\boldsymbol{y}_i|\boldsymbol{z}_i) &\approx \boldsymbol{f}_i(\boldsymbol{z}_i, \boldsymbol{\beta}, \boldsymbol{0}) \\ \text{var}(\boldsymbol{y}_i|\boldsymbol{z}_i) &\approx \boldsymbol{Z}_i(\boldsymbol{z}_i, \boldsymbol{\beta}, \boldsymbol{0})\boldsymbol{D}\boldsymbol{Z}_i^T(\boldsymbol{z}_i, \boldsymbol{\beta}, \boldsymbol{0}) + \boldsymbol{R}_i(\boldsymbol{z}_i, \boldsymbol{\beta}, \boldsymbol{0}, \boldsymbol{\xi}) \end{aligned}$$

- Estimate $(\boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{D})$ by fitting this approximate *marginal model*

Implementation

First-order methods: *Software*

- SAS macro `nlinmix` with `expand=zero` – Solve a set of *generalized estimating equations* ("*GEE-1*") based on these marginal moments
- `nonmem fo` method, SAS `proc nlmixed` with `method=firo` – Maximize *normal likelihood* with these marginal moments ("*GEE-2*")
- `proc nlmixed` cannot handle dependence of $\boldsymbol{R}_i$ on $\boldsymbol{\beta}_i$, $\boldsymbol{\beta}$
- Obvious potential for *bias*

Implementation

First-order conditional methods: More "*refined*" approximation for "n_i *large*" (several variations)

$$\begin{aligned} E(\boldsymbol{y}_i|\boldsymbol{z}_i) &\approx \boldsymbol{f}_i(\boldsymbol{z}_i,\boldsymbol{\beta},\widehat{\boldsymbol{b}}_i) - \boldsymbol{Z}_i(\boldsymbol{z}_i,\boldsymbol{\beta},\widehat{\boldsymbol{b}}_i)\widehat{\boldsymbol{b}}_i \\ \text{var}(\boldsymbol{y}_i|\boldsymbol{z}_i) &\approx \boldsymbol{Z}_i(\boldsymbol{z}_i,\boldsymbol{\beta},\widehat{\boldsymbol{b}}_i)\boldsymbol{D}\boldsymbol{Z}_i^T(\boldsymbol{z}_i,\boldsymbol{\beta},\widehat{\boldsymbol{b}}_i) + \boldsymbol{R}_i(\boldsymbol{z}_i,\boldsymbol{\beta},\widehat{\boldsymbol{b}}_i,\boldsymbol{\xi}) \end{aligned}$$

$$\widehat{\boldsymbol{b}}_i = \boldsymbol{D}\boldsymbol{Z}_i^T(\boldsymbol{z}_i,\boldsymbol{\beta},\widehat{\boldsymbol{b}}_i)\boldsymbol{R}_i(\boldsymbol{z}_i,\boldsymbol{\beta},\widehat{\boldsymbol{b}}_i,\boldsymbol{\xi})\{\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{z}_i,\boldsymbol{\beta},\widehat{\boldsymbol{b}}_i)\}$$

- May be derived by *Taylor series* argument or invoking *Laplace's approximation*
- Suggests *iterative scheme* – alternate between update of $\widehat{\boldsymbol{b}}_i$ and fitting the approximate *marginal model*

Implementation

First-order conditional methods: *Software*

- `nonmem foce` – Based on normal likelihood ("*GEE-2*")
- SAS macro `nlinmix` with `expand=eblup` and R/Splus function `nlme()` – Solve a set of *generalized estimating equations* ("*GEE-1*") based on these marginal moments

Performance: Work well even for n_i not large *as long as* within-individual variation is not large

Implementation

"Exact likelihood" methods: Maximize likelihood "*directly*" using *deterministic* or *stochastic* approximation to the integrals

- *Deterministic* approximation – *Quadrature*, *Adaptive Gaussian quadrature*
- *Stochastic* approximation – Importance sampling, brute-force *Monte Carlo* integration

"Exact likelihood" methods: *Software*

- `proc nlmixed` – quadrature methods, importance sampling when $\boldsymbol{b}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{D})$
- Other *non-commercial* software

Implementation

Bayesian formulation: *Stage 3* – *Hyperprior*

$$(\boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{D}) \sim p(\boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{D})$$

- *Markov chain Monte Carlo* (MCMC) techniques to simulate samples from *posterior distributions* for $\boldsymbol{\beta}$, $\boldsymbol{\xi}$, $\boldsymbol{D}$
- Not possible in general in `WinBUGS` because *nonlinearity* of f may require tailored approach
- `PKBugs` has tailored implementation for *compartment models* for f used in PK
- *Attractive feature* – natural way to incorporate *constraints* and *subject-matter information*

Extensions and Recent Developments

Multi-level models: In many applications

- *Nesting* – response profiles $(y_{ihj},\ j = 1, \ldots, n_{ih})$ on several trees $(h = 1, ..., p_i)$ within each of several plots $(i = 1, \ldots, m)$, e.g.,

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

Multivariate response: More than one type of response profile $(\ell = 1, \ldots, q)$ on each individual

- $y_{ij\ell} = f_\ell(t_{ij\ell}, \boldsymbol{u}_i, \boldsymbol{\beta}_{i\ell}) + e_{ij\ell}$
- *Pharmacokinetics* (concentration-time) and pharmacodynamics (response-concentration)

$$\begin{aligned} y_{ij,PK} &= f_{PK}(t_{ij,PK}, \boldsymbol{u}_i, \boldsymbol{\beta}_{i,PK}) + e_{ij,PK} \\ y_{ij,PD} &= f_{PD}\{ f_{PK}(t_{ij,PK}, \boldsymbol{u}_i, \boldsymbol{\beta}_{i,PK}), \boldsymbol{\beta}_{i,PD} \} + e_{ij,PD} \end{aligned}$$

Extensions and Recent Developments

Missing/mismeasured covariates: $\boldsymbol{a}_i$, $\boldsymbol{u}_i$, and t_{ij}

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

Semiparametric models: Model misspecification, flexibility

- f depends on unspecified function $g(t, \boldsymbol{\beta}_i)$

Clinical trial simulation: Hypothetical subjects simulated from *nonlinear mixed models* for *population PK/PD*, linked to clinical endpoint

Discussion

- The *nonlinear mixed model* is now a standard inferential tool used routinely in many applications
- For extensive *references* and more *details* see

  Davidian, M. and Giltinan, D.M. (2003), "Nonlinear Models for Repeated Measurement Data: An Overview and Update," *JABES* 8, 387–419