Mixed models for the analysis of categorical repeated measures

Geert Verbeke

geert.verbeke@med.kuleuven.be

Biostatistical Centre, K.U.Leuven, Belgium

Joint work with Geert Molenberghs and many others

Overview

- Example: Rat data
- The linear mixed model
- Example: Toenail data
- The generalized linear mixed model
- Estimation methods
- Parameter interpretation
- Example: Theophylline data
- The (generalized) non-linear mixed model

Example: Rat data

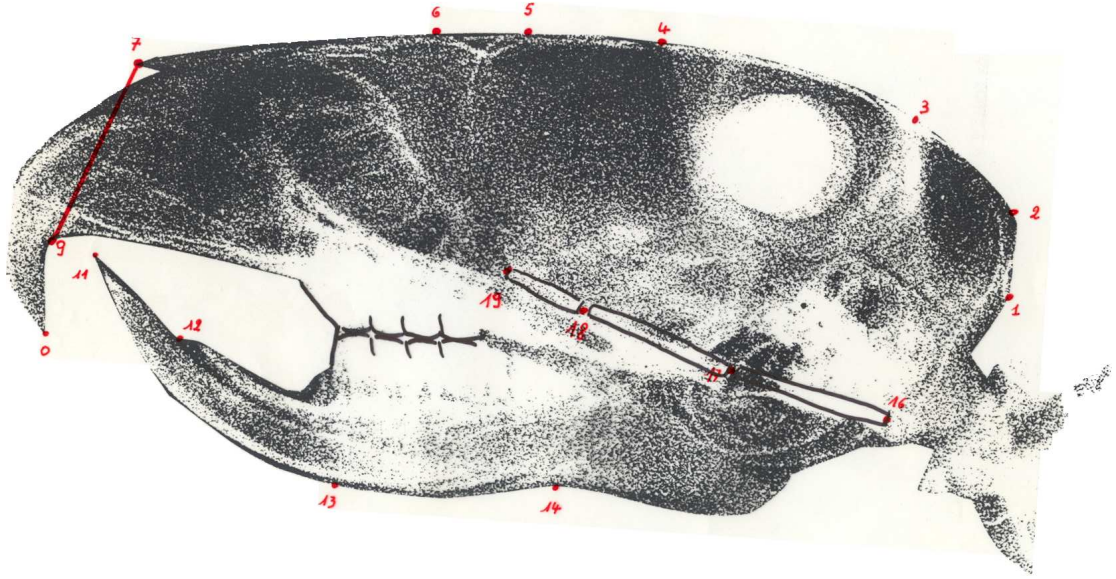
- Research question (Dentistry, K.U.Leuven):

> How does craniofacial growth
> depend on testosteron production ?

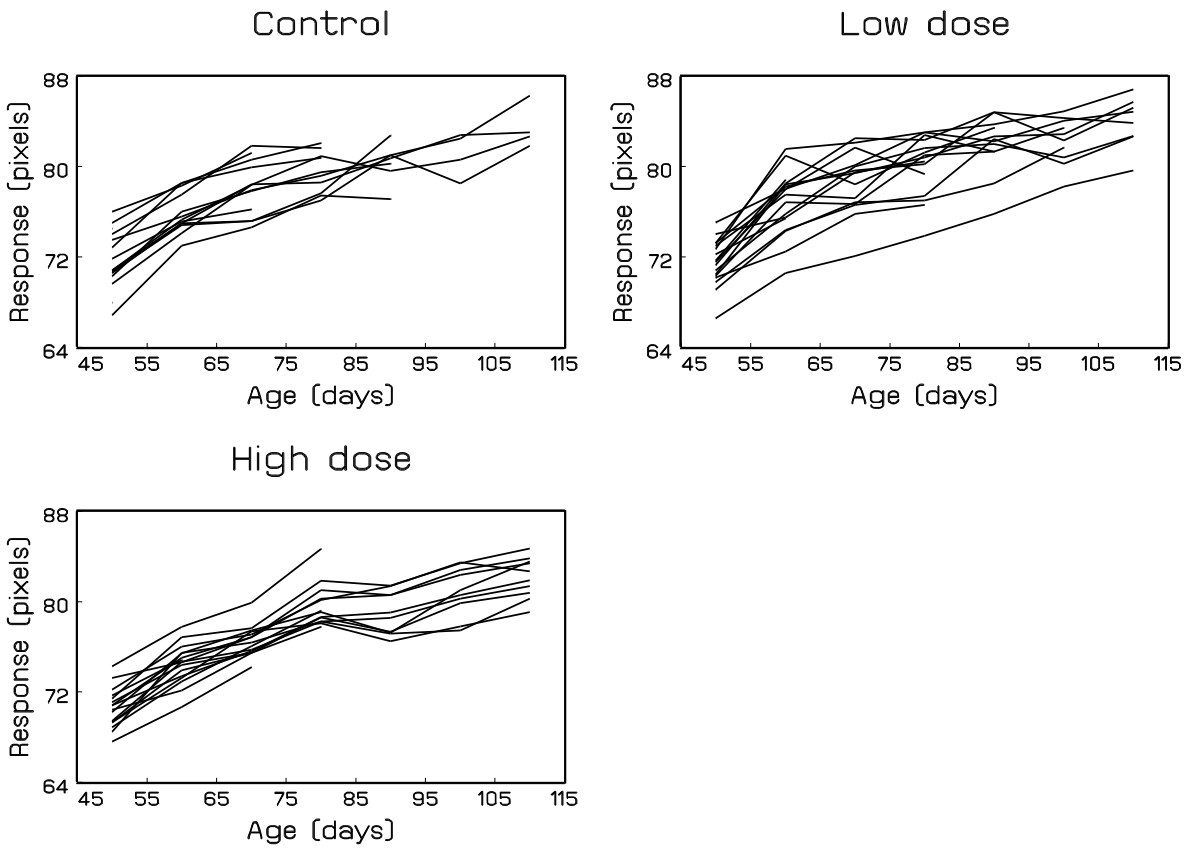
- Randomized experiment in which 50 male Wistar rats are randomized to:
  - Control (15 rats)
  - Low dose of Decapeptyl (18 rats)
  - High dose of Decapeptyl (17 rats)

Measured outcome(s)

- Treatment starts at the age of 45 days; measurements taken every 10 days, from day 50 on.
- The responses are distances (pixels) between well defined points on x-ray pictures of the skull of each rat:



Individual profiles



Complication: Dropout due to anaesthesia (56%)

A statistical model

- Transformation of the time scale to linearize the profiles:

$$\text{Age} \longrightarrow t = \ln[1 + (\text{Age} - 45)/10)]$$

- A linear mixed model:

$$Y_i(t) = \begin{cases} (\beta_0 + b_{1i}) + (\beta_1 + b_{2i})t + \varepsilon_{ij}, & \text{if low dose,} \\ (\beta_0 + b_{1i}) + (\beta_2 + b_{2i})t + \varepsilon_{ij}, & \text{if high dose,} \\ (\beta_0 + b_{1i}) + (\beta_3 + b_{2i})t + \varepsilon_{ij}, & \text{if control} \end{cases}$$

- β_0: average response at the start of the treatment
- β_1, β_2, and β_3: average time effect for each treatment group

The linear mixed model

$$\boldsymbol{Y_i} = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b_i} + \boldsymbol{\varepsilon_i}$$

$$\boldsymbol{b_i} \sim N(\mathbf{0}, D),$$

$$\boldsymbol{\varepsilon_i} \sim N(\mathbf{0}, \sigma^2 I),$$

$$\boldsymbol{b_1}, \ldots, \boldsymbol{b_N}, \boldsymbol{\varepsilon_1}, \ldots, \boldsymbol{\varepsilon_N}$$

independent

Terminology:

- Fixed effects: $\boldsymbol{\beta}$
- Random effects: $\boldsymbol{b_i}$
- Variance components:

  elements in D and σ^2

The implied marginal model

$$\boldsymbol{Y_i} = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b_i} + \boldsymbol{\varepsilon}_i$$

$$\begin{cases} \boldsymbol{b_i} \sim N(\mathbf{0}, D) \\ \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \sigma^2 I) \end{cases} \Longrightarrow \boxed{\boldsymbol{Y_i} \sim N\left[X_i\boldsymbol{\beta}, V_i = Z_i D Z_i' + \sigma^2 I\right]}$$

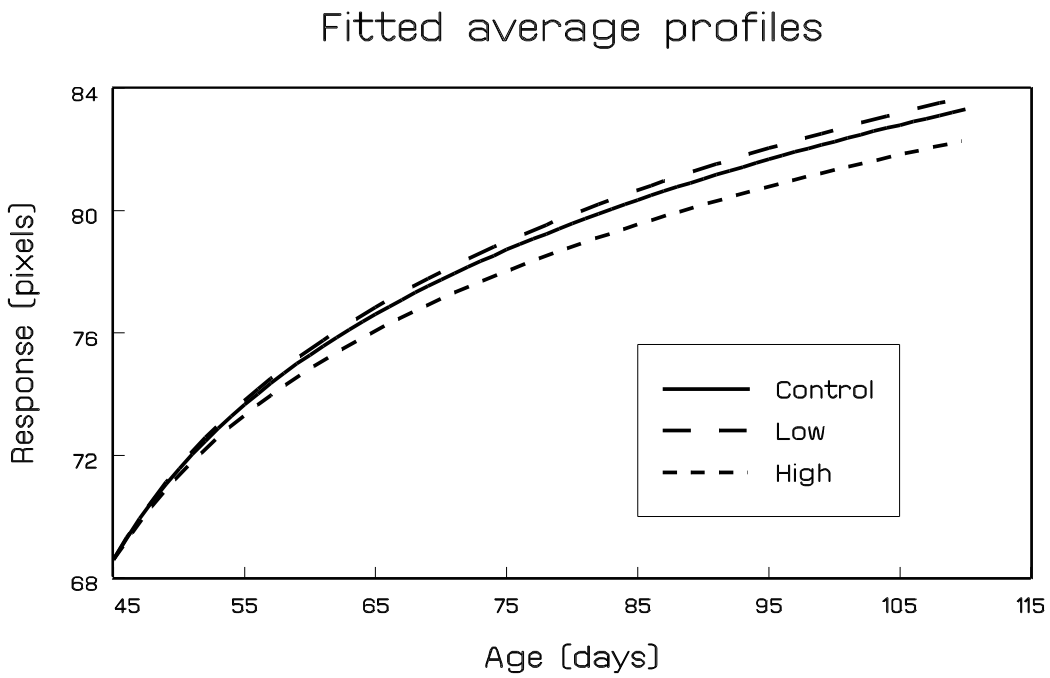
$$\begin{cases} f(\boldsymbol{y_i}|b_i) \\ f(\boldsymbol{b_i}) \end{cases} \Longrightarrow \boxed{f(\boldsymbol{y_i})}$$

Mixed model and marginal model are **NOT** equivalent !

Estimation and inference

- Based on marginal model: $\boldsymbol{Y_i} \sim N(X_i\boldsymbol{\beta}, V_i = Z_i D Z_i' + \sigma^2 I)$
- Independence across subjects
- Estimation based on likelihood principles
- Inference:
  - Wald tests, t-tests, F-tests
  - LR tests

Results for rat data



H_0 : equal slopes

(p =0.1013)

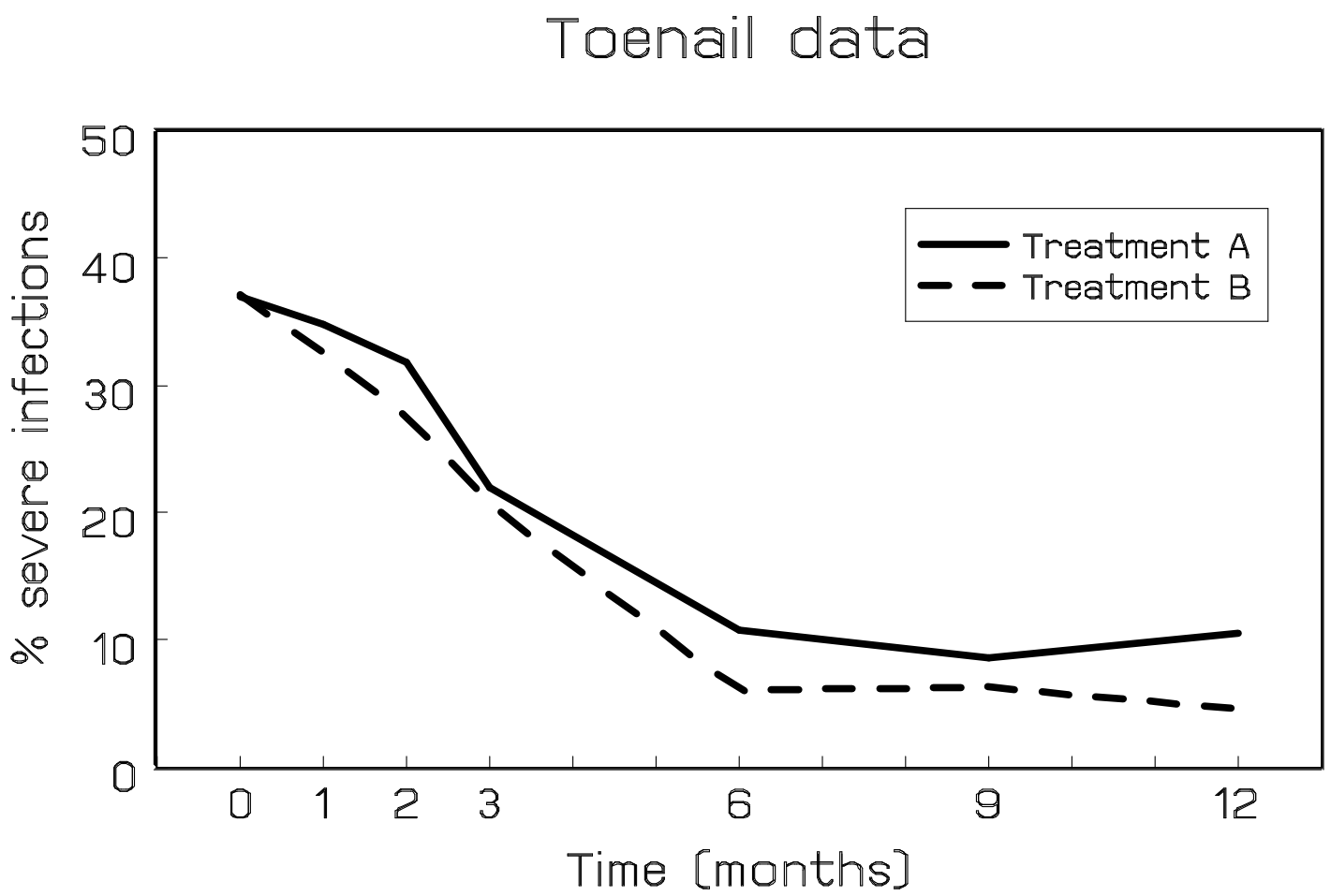
Example: Toenail data

- **T**oenail **D**ermatophyte **O**nychomycosis
- Randomized, double-blind, parallel group, comparing 2 oral compounds (A and B), 2×189 patients
- Research question:

  Severity relative to treatment of TDO ?

- 12 months of follow up, 3 months of treatment
- Measurements at months 0, 1, 2, 3, 6, 9, 12.

Frequencies at each visit



A statistical model

- Y_{ij} is binary severity indicator for subject i at visit j.
- Model:

$$Y_{ij}|b_i \sim \text{Bernoulli}(\pi_{ij}),$$

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + b_i + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij}$$

- Notation:
  - T_i: treatment indicator for subject i
  - t_{ij}: time point at which jth measurement is taken for ith subject

Distributional assumptions

- As for the linear model:
  - Measurements are assumed independent, conditional on the random effects:

$$f_i(\boldsymbol{y_i}|\boldsymbol{b_i},\boldsymbol{\beta}) \quad = \quad \prod_{j=1}^{n_i} f_{ij}(y_{ij}|\boldsymbol{b_i},\boldsymbol{\beta})$$

  - Random effects $\boldsymbol{b_i}$ are assumed $N(\mathbf{0}, D)$
  - The random effects generate an association structure for the repeated measurements
- Estimation and inference will again be based on the marginal likelihood

The marginal likelihood

- Assuming independent subjects,

$$\begin{aligned} L(\boldsymbol{\beta}, D) &= \prod_{i=1}^{N} f_i(\boldsymbol{y_i}|\boldsymbol{\beta}, D) \\ &= \prod_{i=1}^{N} \int f_i(\boldsymbol{y_i}|\boldsymbol{b_i}, \boldsymbol{\beta}) f(\boldsymbol{b_i}|D) d\boldsymbol{b_i} \end{aligned}$$

- Unlike in the normal linear model, the integrals can no longer be worked out analytically, and approximations are required:
  - Approximation of integrand
  - Approximation of data
  - Approximation of integral

Laplace approximation of integrand

- Integrals in $L(\boldsymbol{\beta}, D)$ can be written in the form $I = \int e^{Q(\boldsymbol{b})} d\boldsymbol{b}$
- Second-order Taylor expansion of $Q(\boldsymbol{b})$ around the mode yields

$$Q(\boldsymbol{b}) \quad \approx \quad Q(\widehat{\boldsymbol{b}}) + \frac{1}{2}(\boldsymbol{b} - \widehat{\boldsymbol{b}})' Q''(\widehat{\boldsymbol{b}})(\boldsymbol{b} - \widehat{\boldsymbol{b}}),$$

- Quadratic term leads to re-scaled normal density. Hence,

$$I \approx (2\pi)^{q/2} \left| -Q''(\widehat{\boldsymbol{b}}) \right|^{-1/2} e^{Q(\widehat{\boldsymbol{b}})}.$$

- Exact approximation in case of normal kernels
- Good approximation in case of many repeated measures per subject

Approximation of data

- Re-write GLMM as:

$$Y_{ij} \quad = \quad \mu_{ij} + \varepsilon_{ij} \quad = \quad h(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b_i}) + \varepsilon_{ij}$$

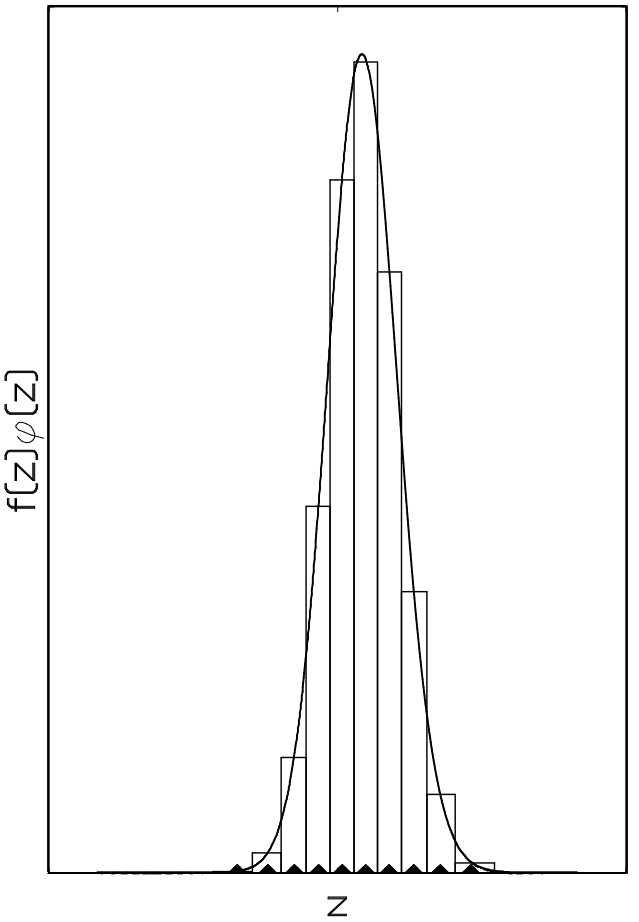
- Linear Taylor expansion for μ_{ij}:
  - Penalized quasi-likelihood (PQL): Around current $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{b_i}}$
  - Marginal quasi-likelihood (MQL): Around current $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{b_i} = \boldsymbol{0}$
- An approximate linear mixed model is obtained which yields updates for $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{b_i}}$

PQL versus MQL

- MQL only performs reasonably well if random-effects variance is (very) small
- Both perform bad for binary outcomes with few repeated measurements per cluster
- With increasing number of measurements per subject:
  - MQL remains biased
  - PQL consistent
- Improvements possible with higher-order Taylor expansions

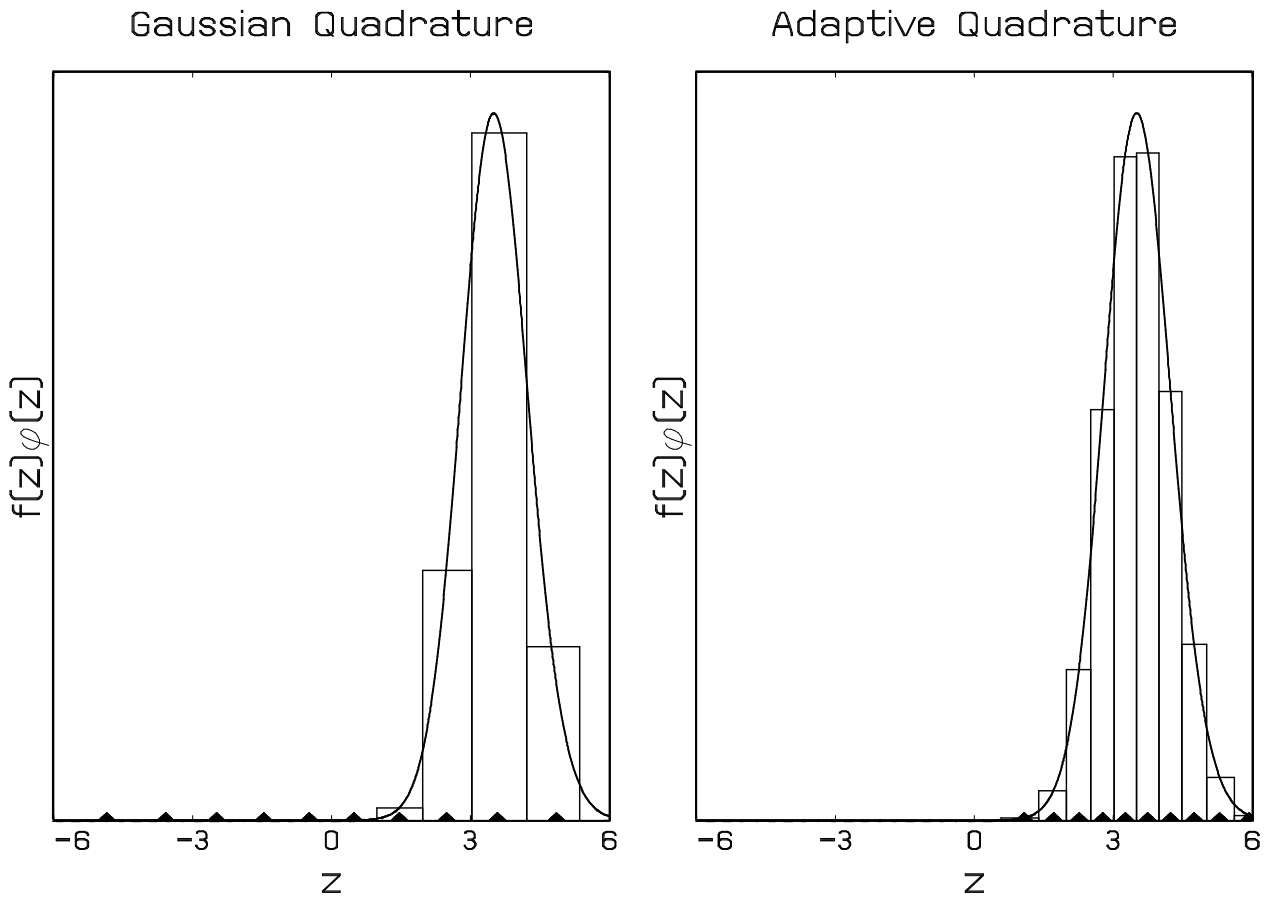
Approximation of integral

- Approximate each integral by the surface of rectangles
- The higher the number Q of intervals, the more accurate the approximation will be
- 'Gaussian quadrature' is optimal in our situation



Adaptive Gaussian quadrature

Adapt nodes and weights to the 'support' of the function to be integrated:



Adaptive verus non-adaptive Gaussian quadrature

- Typically, adaptive Gaussian quadrature needs (much) less quadrature points than classical Gaussian quadrature.
- On the other hand, adaptive Gaussian quadrature is much more time consuming.
- Adaptive Gaussian quadrature of order one is equivalent to Laplace transformation.

Example: Quadrature for toenail Data

	Gaussian quadrature				
	$Q = 3$	$Q = 5$	$Q = 10$	$Q = 20$	$Q = 50$
β_0	-1.52 (0.31)	-2.49 (0.39)	-0.99 (0.32)	-1.54 (0.69)	-1.65 (0.43)
β_1	-0.39 (0.38)	0.19 (0.36)	0.47 (0.36)	-0.43 (0.80)	-0.09 (0.57)
β_2	-0.32 (0.03)	-0.38 (0.04)	-0.38 (0.05)	-0.40 (0.05)	-0.40 (0.05)
β_3	-0.09 (0.05)	-0.12 (0.07)	-0.15 (0.07)	-0.14 (0.07)	-0.16 (0.07)
τ	2.26 (0.12)	3.09 (0.21)	4.53 (0.39)	3.86 (0.33)	4.04 (0.39)
	Adaptive Gaussian quadrature				
	$Q = 3$	$Q = 5$	$Q = 10$	$Q = 20$	$Q = 50$
β_0	-2.05 (0.59)	-1.47 (0.40)	-1.65 (0.45)	-1.63 (0.43)	-1.63 (0.44)
β_1	-0.16 (0.64)	-0.09 (0.54)	-0.12 (0.59)	-0.11 (0.59)	-0.11 (0.59)
β_2	-0.42 (0.05)	-0.40 (0.04)	-0.41 (0.05)	-0.40 (0.05)	-0.40 (0.05)
β_3	-0.17 (0.07)	-0.16 (0.07)	-0.16 (0.07)	-0.16 (0.07)	-0.16 (0.07)
τ	4.51 (0.62)	3.70 (0.34)	4.07 (0.43)	4.01 (0.38)	4.02 (0.38)

Conclusions

- Different Q can lead to considerable differences in estimates and standard errors:
  - For example, using non-adaptive quadrature, with $Q = 3$, we found no difference in time effect between both treatment groups ($t = -0.09/0.05, p = 0.0833$).
  - Using adaptive quadrature, with $Q = 50$, we find a significant interaction between the time effect and the treatment ($t = -0.16/0.07, p = 0.0255$).
- Assuming that $Q = 50$ is sufficient, the 'final' results are well approximated with smaller Q under adaptive quadrature, but not under non-adaptive quadrature.

Comparison of approximations: Toenail data

- Adaptive Gaussian Quadrature, $Q = 50$
- MQL and PQL

Parameter	QUAD	PQL	MQL
Intercept group A	-1.63 (0.44)	-0.72 (0.24)	-0.56 (0.17)
Intercept group B	-1.75 (0.45)	-0.72 (0.24)	-0.53 (0.17)
Slope group A	-0.40 (0.05)	-0.29 (0.03)	-0.17 (0.02)
Slope group B	-0.57 (0.06)	-0.40 (0.04)	-0.26 (0.03)
Var. random intercepts (τ^2)	15.99 (3.02)	4.71 (0.60)	2.49 (0.29)

Fitting generalized linear mixed models in SAS

- **MQL/PQL:**

```
proc glimmix data=test method=RSPL ;
class idnum;
model onyresp (event='1') = treatn time treatn*time
                          / dist=binary solution;
random intercept / subject=idnum;
run;
```

- **(Adaptive) quadrature / Laplace:**

```
proc nlmixed data=test noad qpoints=3;
parms beta0=-1.6 beta1=0 beta2=-0.4 beta3=-0.5 sigma=3.9;
teta = beta0 + b + beta1*treatn + beta2*time + beta3*timetr;
expteta = exp(teta);
p = expteta/(1+expteta);
model onyresp ~ binary(p);
random b ~ normal(0,sigma**2) subject=idnum;
run;
```

Toenail data: Fitted model

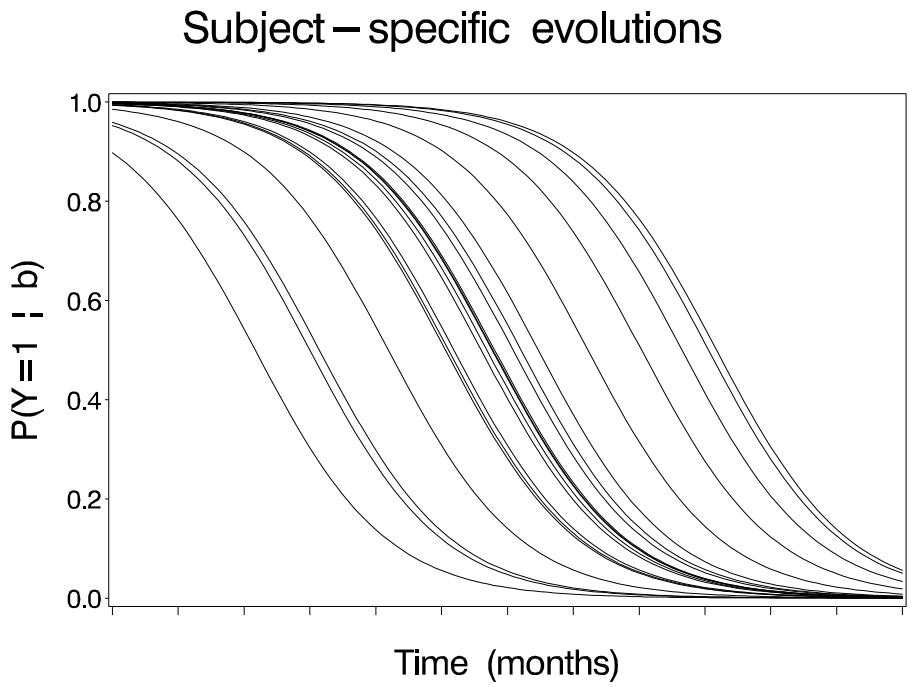
- The fitted model for the toenail data is given by

$$P(Y_{ij}=1|b_i) = \begin{cases} \dfrac{\exp(-1.6308+b_i-0.4043t_{ij})}{1+\exp(-1.6308+b_i-0.4043t_{ij})} \\ \dfrac{\exp(-1.7454+b_i-0.5657t_{ij})}{1+\exp(-1.7454+b_i-0.5657t_{ij})} \end{cases}$$

- Parameters need to be interpreted with care !
- This will be explained in the context of the logistic mixed model with random intercepts.

The logistic mixed model with random intercepts

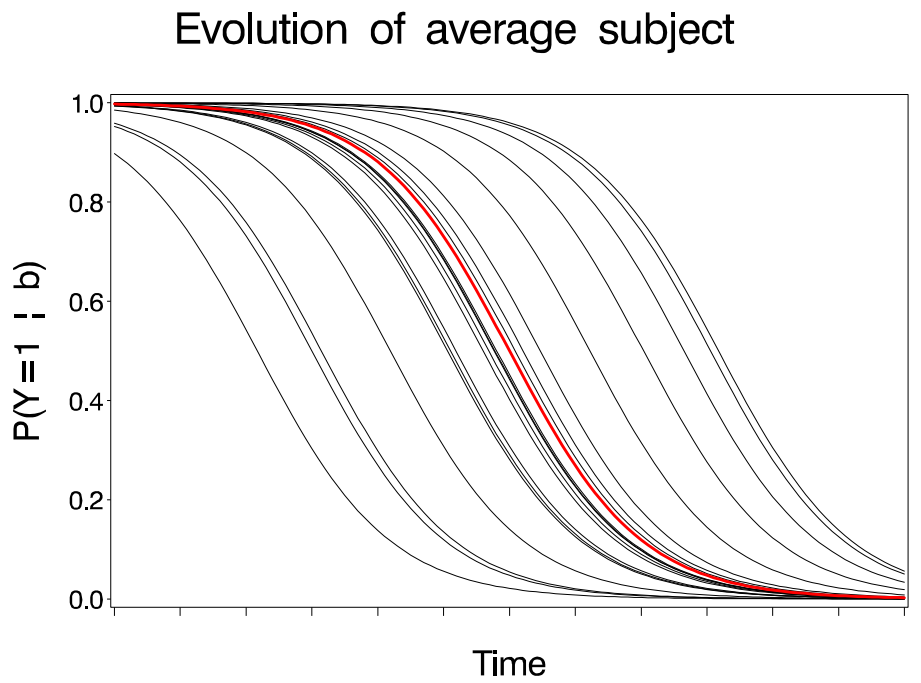
$$P(Y_i(t) = 1|b_i) = \frac{\exp(\beta_0 + b_i + \beta_1 t)}{1 + \exp(\beta_0 + b_i + \beta_1 t)}$$



Average subject

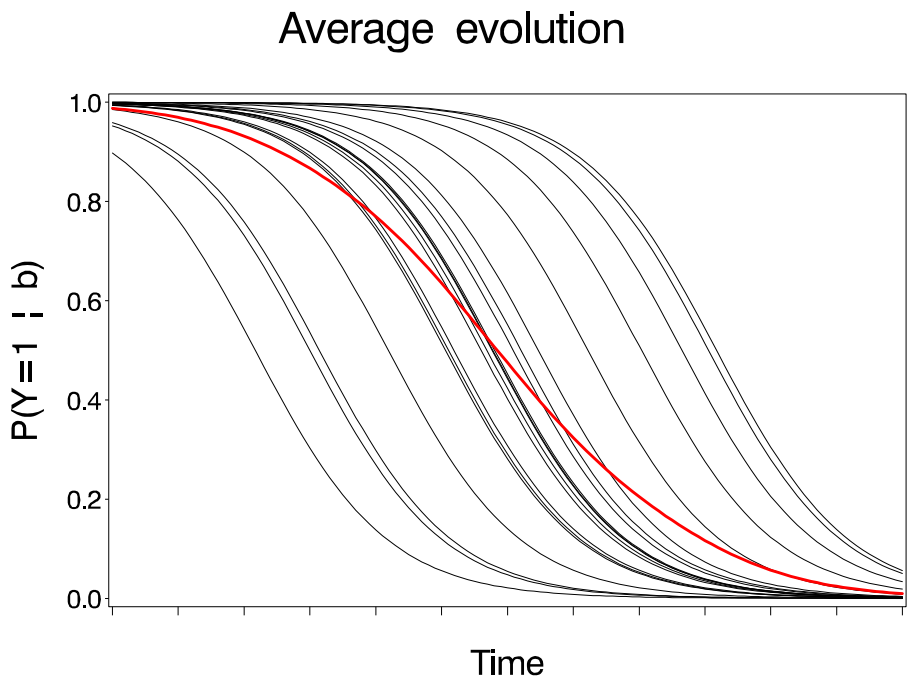
Subject with average regression coefficients, i.e., $b_i = 0$

$$P(Y_i(t) = 1 | b_i = 0) = \frac{\exp(\beta_0 + \beta_1 t)}{1 + \exp(\beta_0 + \beta_1 t)}$$



Average evolution

$$P(Y_i(t)=1) \quad = \quad E[P(Y_{ij}=1|b_i)] \; = \; E\left[\frac{\exp(\beta_0+b_i+\beta_1 t)}{1+\exp(\beta_0+b_i+\beta_1 t)}\right]$$



Conclusion

Average evolution $\neq$ Evolution average subject

- Parameters in the mixed model have a subject-specific interpretation, not a population-averaged one.
- The problem arises from the fact that, $E[g(Y)] \neq g[E(Y)]$, unless for linear functions, such as in the case of linear mixed models:
  - Conditional mean: $E(\boldsymbol{Y_i}|\boldsymbol{b_i}) = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b_i}$
  - Average subject: $E(\boldsymbol{Y_i}|\boldsymbol{b_i} = \mathbf{0}) = X_i\boldsymbol{\beta}$
  - Marginal mean: $E(\boldsymbol{Y_i}) = X_i\boldsymbol{\beta}$

How to derive the marginal evolution ?

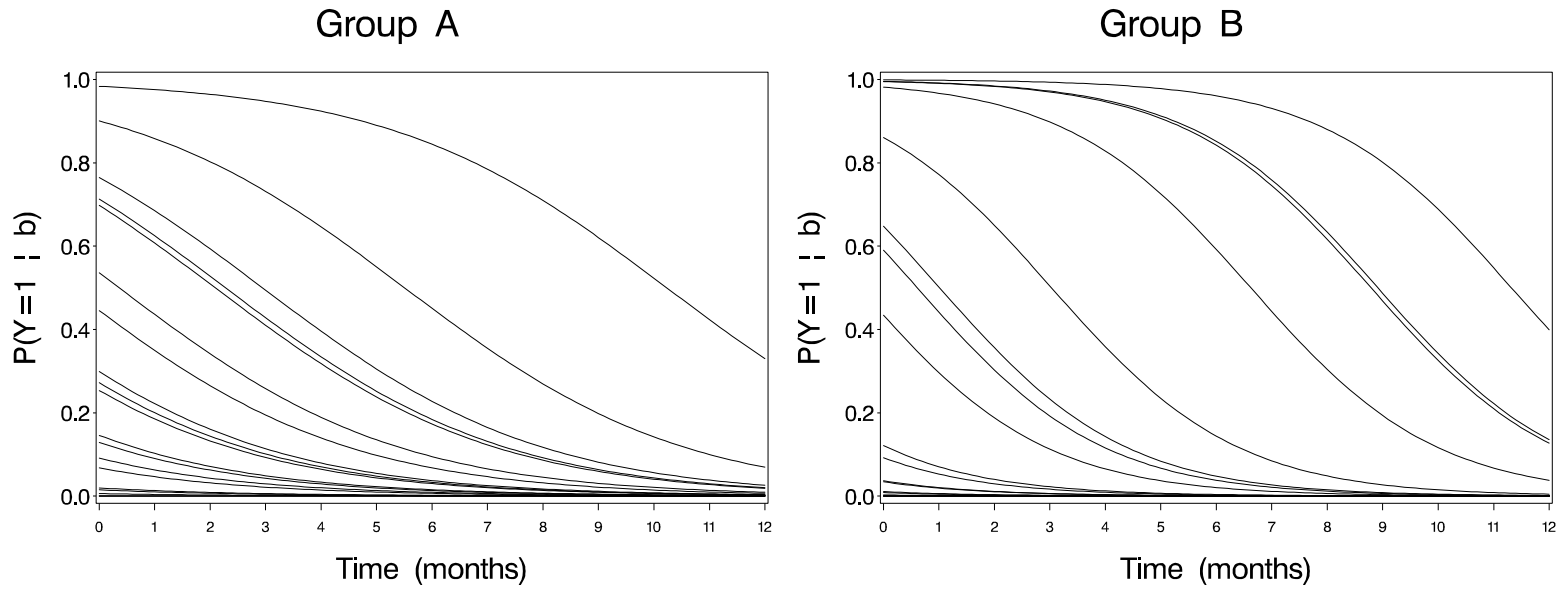
- Directly fit a marginal model (e.g., GEE)
- Based on a mixed model, calculation of average evolution requires evaluation of

$$P(Y_i(t)=1) \quad = \quad E[P(Y_{ij}=1|b_i)] \; = \; E\left[\frac{\exp(\beta_0+b_i+\beta_1 t)}{1+\exp(\beta_0+b_i+\beta_1 t)}\right]$$

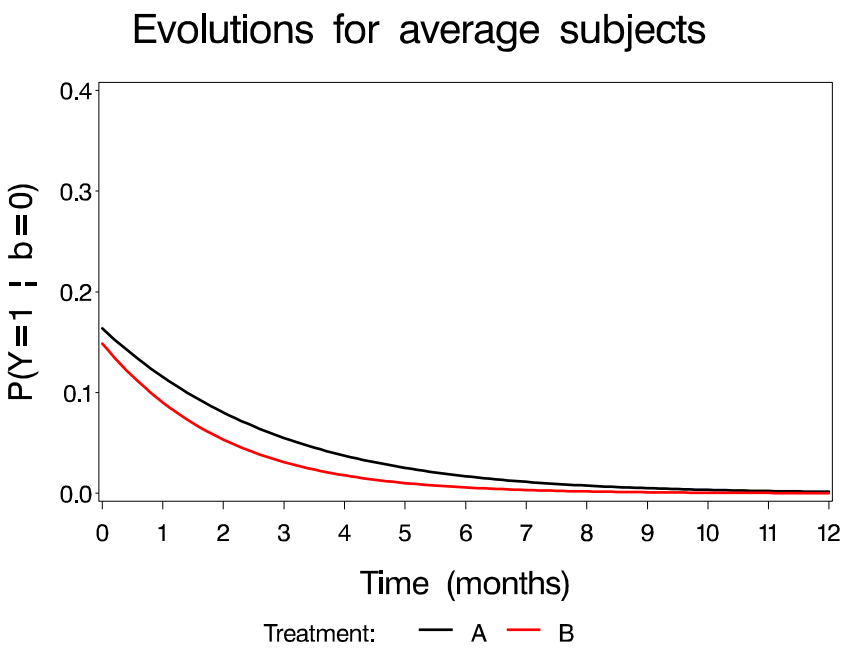
- This cannot be done analytically. Hence, approximations are needed:
  - Numerical quadrature
  - Sampling techniques

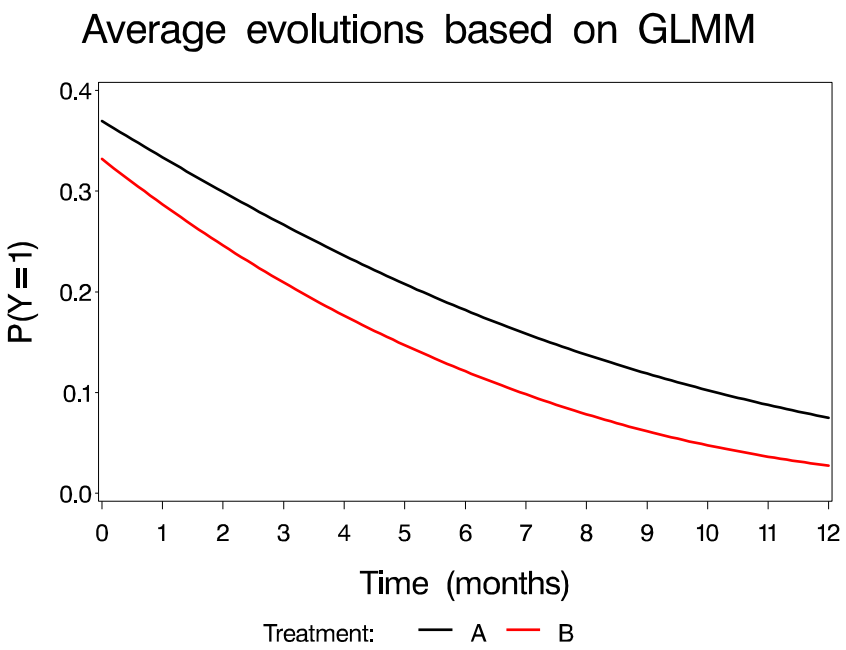
Toenail data: Fitted model

$$P(Y_{ij}=1|b_i) = \begin{cases} \dfrac{\exp(-1.6308+b_i-0.4043t_{ij})}{1+\exp(-1.6308+b_i-0.4043t_{ij})} \\ \dfrac{\exp(-1.7454+b_i-0.5657t_{ij})}{1+\exp(-1.7454+b_i-0.5657t_{ij})} \end{cases}$$



Toenail data: Average subject / average evolution

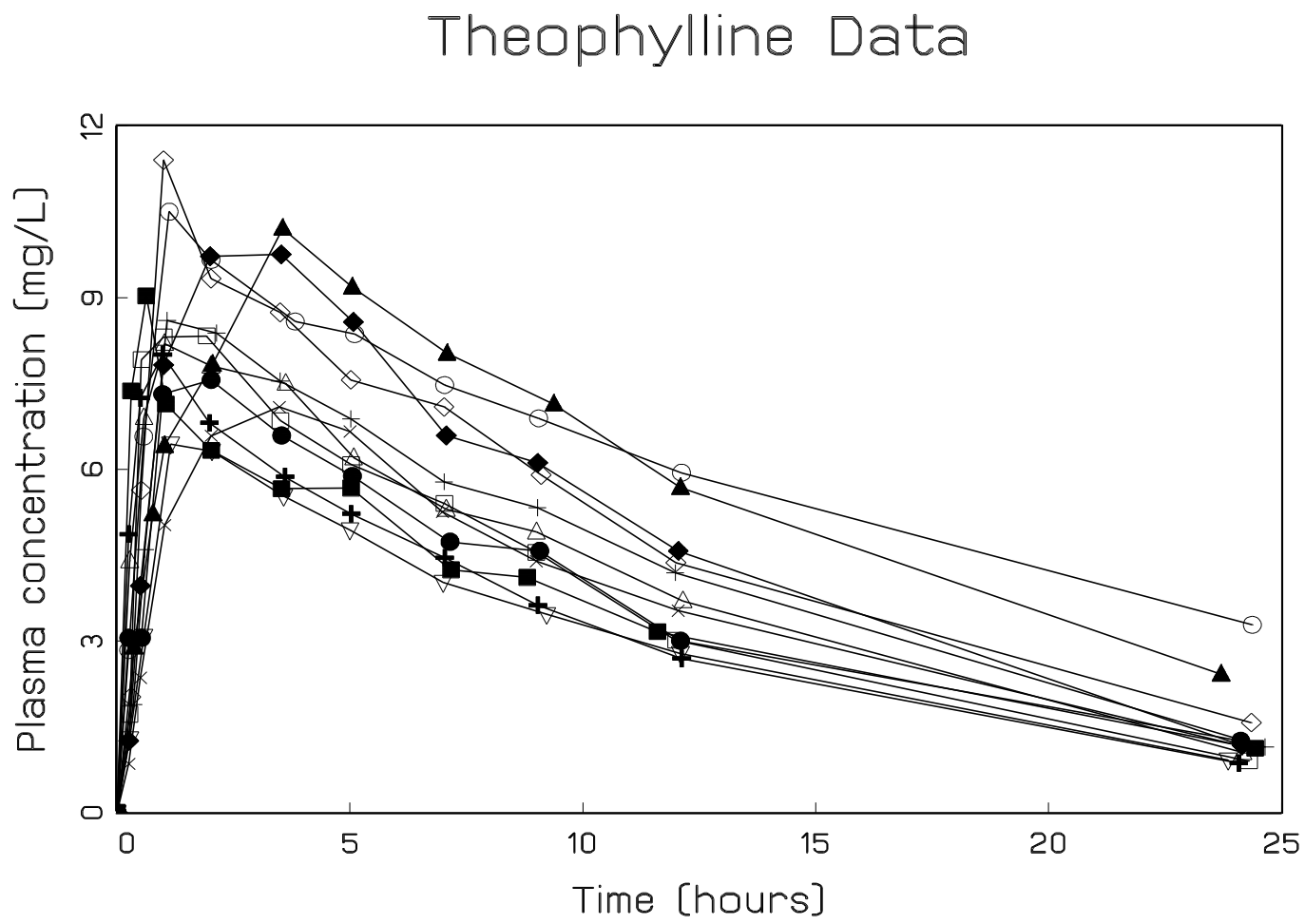




Theophylline data

- Theophylline: anti-asthmatic agent, administered orally
- 12 subjects, dose at $t = 0$
- Blood samples at 10 time points over the following 25 hours
- Outcome of interest: Theophylline concentration

Individual profiles



A statistical model

- A one-compartment open model with first-order absorption and elimination

$$Y_{ij} = C_i(t_{ij}) \quad = \quad \frac{k_{ai}k_{ei}d_i}{C\ell_i(k_{ai} - k_{ei})} \times [\exp(-k_{ei}t_{ij}) - \exp(-k_{ai}t_{ij})] + \varepsilon_{ij}$$

- Parameter interpretation:
  - k_{ai}: fractional absorption rate for subject i
  - k_{ei}: fractional elimination rate for subject i
  - $C\ell_i$: clearance for subject i

Reparameterization

- In order to restrict k_{ai}, k_{ei}, and $C\ell_i$ to be positive:

$$\begin{aligned} C\ell_i &= \exp(\beta_1 + b_{i1}), \\ k_{a,i} &= \exp(\beta_2 + b_{i2}), \\ k_{e,i} &= \exp(\beta_3 + b_{i3}). \end{aligned}$$

- b_{i1}, b_{i2}, and b_{i3} are assumed multivariate normal with mean $\mathbf{0}$

Model fitting

- As for the generalized linear model:
  - Measurements are assumed independent, conditional on the random effects:

$$f_i(\boldsymbol{y_i}|\boldsymbol{b_i}, \boldsymbol{\beta}) = \prod_{j=1}^{n_i} f_{ij}(y_{ij}|\boldsymbol{b_i}, \boldsymbol{\beta})$$

  - Assuming independent subjects,

$$L(\boldsymbol{\beta}, D) = \prod_{i=1}^{N} f_i(\boldsymbol{y_i}|\boldsymbol{\beta}, D) = \prod_{i=1}^{N} \int f_i(\boldsymbol{y_i}|\boldsymbol{b_i}, \boldsymbol{\beta}) f(\boldsymbol{b_i}|D) d\boldsymbol{b_i}$$

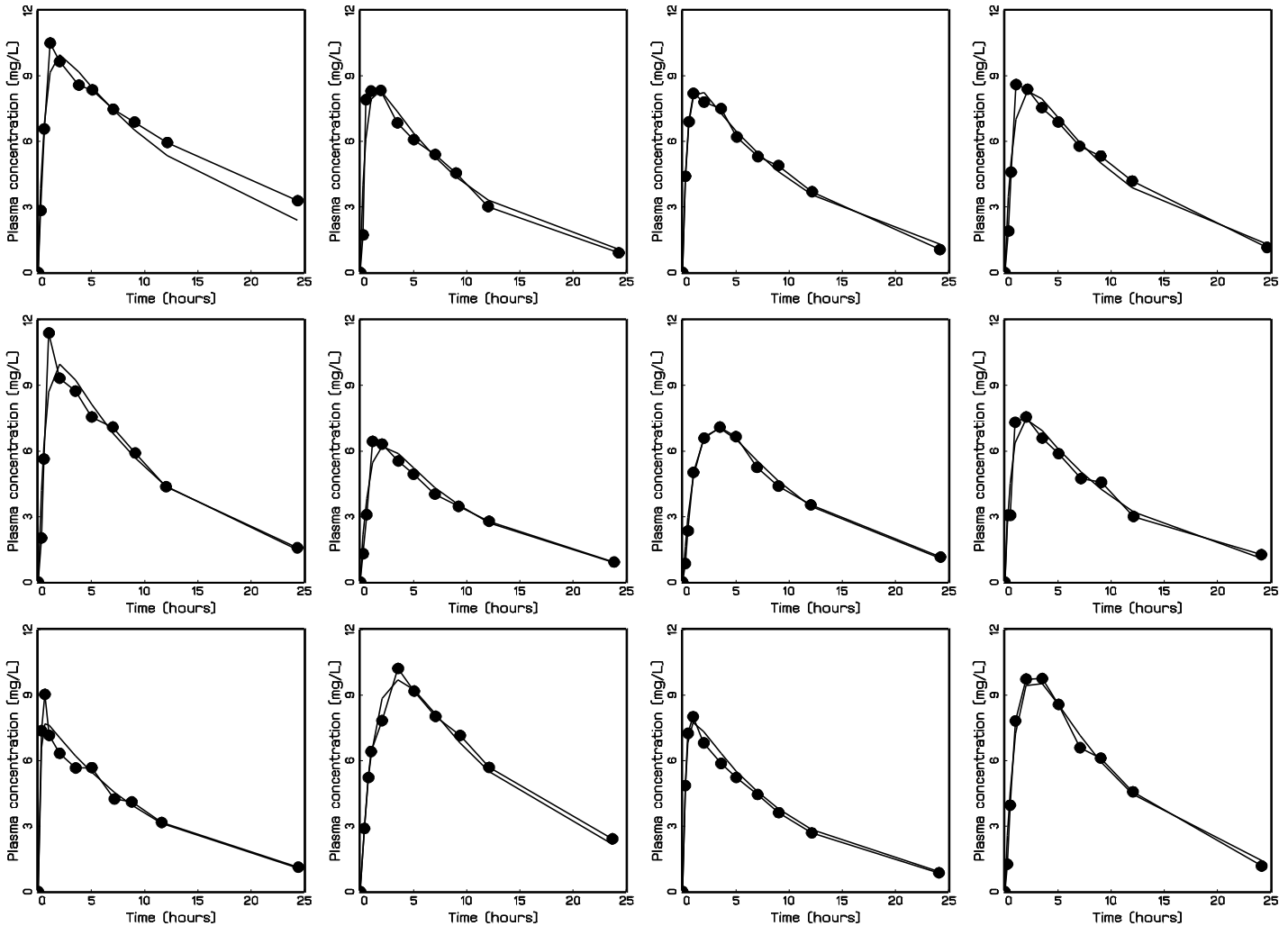
- ML estimation using Gaussian quadrature methods

Results

Parameter	Estimate (s.e.)
Fixed effects:	
β_1 (Cl)	-3.277 (0.046)
β_2 (k_a)	0.537 (0.063)
β_3 (k_e)	-2.454 (0.064)

Parameter	Estimate (s.e.)
Residual variance:	
σ^2	0.623 (0.083)
Random-effect variances:	
d_{11}	0.057 (0.022)
d_{12}	-0.012 (0.018)
d_{22}	0.264 (0.054)
d_{13}	0.030 (0.020)
d_{23}	-0.025 (0.017)
d_{33}	0.035 (0.017)

Observed and fitted profiles



Remarks

- The non-linear nature of the model implies that the parameters have subject-specific interpretations
- Calculation of marginal averages again requires numerical integration or sampling methods
- Generalized linear mixed models can also be extended to accommodate non-linear predictors.

Conclusions

- Mixed models provide a general framework for the analysis of continuous and discrete repeated measurements, based on linear and non-linear models
- In general, parameters in mixed models do not immediately yield population-based inferences
- Mixed models specify the full distribution of $\boldsymbol{Y}_i$:
  - Calculation of joint probabilities
  - Missing data issues
- Mixed models are more sensitive to model miss-specification than most models for cross-sectional data