Longitudinal and Incomplete Data

Geert Verbeke

geert.verbeke@kuleuven.be

Geert Molenberghs

geert.molenberghs@uhasselt.be

geert.molenberghs@kuleuven.be

Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat)

Katholieke Universiteit Leuven & Universiteit Hasselt, Belgium

http://www.ibiostat.be

Parexel, 2017

Contents

Chapter 0
Related References

- Aerts, M., Geys, H., Molenberghs, G., and Ryan, L.M. (2002). Topics in Modelling of Clustered Data. London: Chapman & Hall.

- Brown, H. and Prescott, R. (1999). *Applied Mixed Models in Medicine*. New York: John Wiley & Sons.

- Carpenter, J.R. and Kenward, M.G. (2013). *Multiple Imputation and its Application*. New York: John Wiley & Sons.

- Crowder, M.J. and Hand, D.J. (1990). *Analysis of Repeated Measures.* London: Chapman & Hall.

- Davidian, M. and Giltinan, D.M. (1995). *Nonlinear Models For Repeated Measurement Data.* London: Chapman & Hall.

- Davis, C.S. (2002). *Statistical Methods for the Analysis of Repeated Measurements.* New York: Springer.

- Demidenko, E. (2004). *Mixed Models: Theory and Applications.* New York: John Wiley & Sons.

- Diggle, P.J., Heagerty, P.J., Liang, K.Y. and Zeger, S.L. (2002). *Analysis of Longitudinal Data.* (2nd edition). Oxford: Oxford University Press.

- Fahrmeir, L. and Tutz, G. (2002). *Multivariate Statistical Modelling Based on Generalized Linear Models (2nd ed)*. New York: Springer.

- Fitzmaurice, G.M., Davidian, M., Verbeke, G., and Molenberghs, G.(2009). *Longitudinal Data Analysis. Handbook.* Hoboken, NJ: John Wiley & Sons.

- Fitzmaurice, G.M., Laird, N.M., and Ware, J.H. (2004). *Applied Longitudinal Analysis.* New York: John Wiley & Sons.

- Gałecki, A. and Burzykowski, T. (2013). *Linear Mixed-Effects Models Using R.* New York: Springer.

- Goldstein, H. (1979). *The Design and Analysis of Longitudinal Studies.* London: Academic Press.

- Goldstein, H. (1995). *Multilevel Statistical Models.* London: Edward Arnold.

- Hand, D.J. and Crowder, M.J. (1995). *Practical Longitudinal Data Analysis.* London: Chapman & Hall.

- Hedeker, D. and Gibbons, R.D. (2006). *Longitudinal Data Analysis*. New York: John Wiley & Sons.

- Hogan, J. and Daniels, M. (2008) *Missing Data in Longitudinal Studies:*

Strategies for Bayesian modelling and sensitivity analysis. Boca Raton: CRC/Chapman & Hall.

- Jones, B. and Kenward, M.G. (1989). *Design and Analysis of Crossover Trials.* London: Chapman & Hall.

- Kshirsagar, A.M. and Smith, W.B. (1995). *Growth Curves.* New York: Marcel Dekker.

- Leyland, A.H. and Goldstein, H. (2001) *Multilevel Modelling of Health Statistics.* Chichester: John Wiley & Sons.

- Lindsey, J.K. (1993). *Models for Repeated Measurements.* Oxford: Oxford University Press.

- Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., and Schabenberger, O. (2005). *SAS for Mixed Models (2nd ed.).* Cary: SAS Press.

- Little, R.J.A., D'Agostino, R., Dickerson, K., Emerson, S.S., Farrar, J.T., Frangakis, C., Hogan, J.W., Molenberghs, G., Murphy, S.A., Neaton, J.D., Rotnitzky, A., Scharfstein, D., Shih, W., Siegel, J.P., and Stern, H. National Research Council (2010). *The Prevention and Treatment of Missing Data in Clinical Trials. Panel on Handling Missing Data in Clinical Trials.* Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, D.C.: The National Academies Press.

- Little, R.J.A. and Rubin, D.B. (1987, 2002, 2014). *Statistical Analysis with Missing Data (2nd ed.)*. New York: John Wiley & Sons.

- Longford, N.T. (1993). *Random Coefficient Models.* Oxford: Oxford University Press.

- Mallinckrodt, C.H. (2013). *Preventing and Treating Missing Data in Longitudinal Clinical Trials: A Practical Guide.* New York: Cambridge University Press.

- McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models* (second

edition). London: Chapman & Hall.

- Mallinckrodt, C.H. (2013). *Preventing and Treating Missing Data in Longitudinal Clinical Trials: A Practical Guide.* New York: Cambridge University Press.

- Molenberghs, G., Fitzmaurice, G., Kenward, M.G., Tsiatis, A.A., and Verbeke, G. (2015). *Handbook of Missing Data*. Boca Raton: Chapman & Hall/CRC.

- Molenberghs, G. and Kenward, M.G. (2007). *Missing Data in Clinical Studies.* Chichester: John Wiley & Sons.

- Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data.* New York: Springer.

- O'Kelly, M. and Ratitch, B. (2014). *A Guide to Planning for Missing Data.* New York: John Wiley & Sons.

- Pinheiro, J.C. and Bates D.M. (2000). *Mixed effects models in S and S-Plus*.

New York: Springer.

- Raghunathan, T. (2017). *Missing Data Analysis in Practice*. London: CRC/Chapman & Hall.

- Rizopoulos, D. (2012). *Joint Models for Longitudinal and Time-to-Event Data. With Applications in R.* Boca Raton: Chapman & Hall/CRC.

- Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.

- Schafer J.L. (1997). *Analysis of Incomplete Multivariate Data*. London: CRC/Chapman & Hall.

- Searle, S.R., Casella, G., and McCulloch, C.E. (1992). *Variance Components*. New-York: Wiley.

- Senn, S.J. (1993). *Cross-over Trials in Clinical Research.* Chichester: Wiley.

- Tan, M.T., Tian, G.-L., and Ng, K.W. (2010). *Bayesian Missing Data Problems.* Boca Raton: Chapman & Hall/CRC.

- van Buuren, S. (2012). *Flexible Imputation of Missing Data.* Boca Raton: Chapman & Hall/CRC.

- Verbeke, G. and Molenberghs, G. (1997). *Linear Mixed Models In Practice: A SAS Oriented Approach*, Lecture Notes in Statistics 126. New-York: Springer.

- Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer Series in Statistics. New-York: Springer.

- Vonesh, E.F. and Chinchilli, V.M. (1997). *Linear and Non-linear Models for the Analysis of Repeated Measurements*. Basel: Marcel Dekker.

- Weiss, R.E. (2005). *Modeling Longitudinal Data.* New York: Springer.

- West, B.T., Welch, K.B., and Gałecki, A.T. (2007). *Linear Mixed Models: A*

Practical Guide Using Statistical Software. Boca Raton: Chapman & Hall/CRC.

- Wu, H. and Zhang, J.-T. (2006). *Nonparametric Regression Methods for Longitudinal Data Analysis.* New York: John Wiley & Sons.

- Wu, L. (2010). *Mixed Effects Models for Complex Data.* Boca Raton: Chapman & Hall/CRC.

Part I

Continuous Longitudinal Data

Chapter 1
Introduction

- Repeated Measures / Longitudinal data
- Examples

1.1 Repeated Measures / Longitudinal Data

Repeated measures are obtained when a response is measured repeatedly on a set of units

- Units:
  - ▷ Subjects, patients, participants, ...
  - ▷ Animals, plants, ...
  - ▷ Clusters: families, towns, branches of a company,...
  - ▷ ...
- Special case: Longitudinal data

1.2 Rat Data

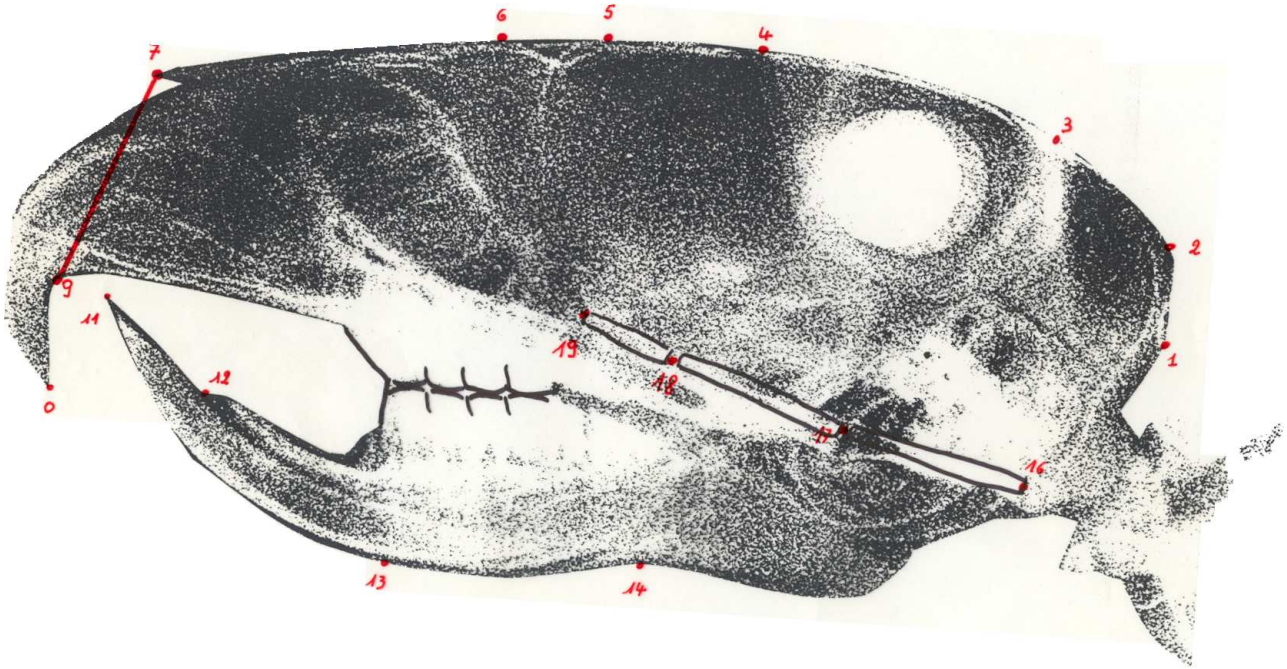
- Research question (Dentistry, K.U.Leuven):

> 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)

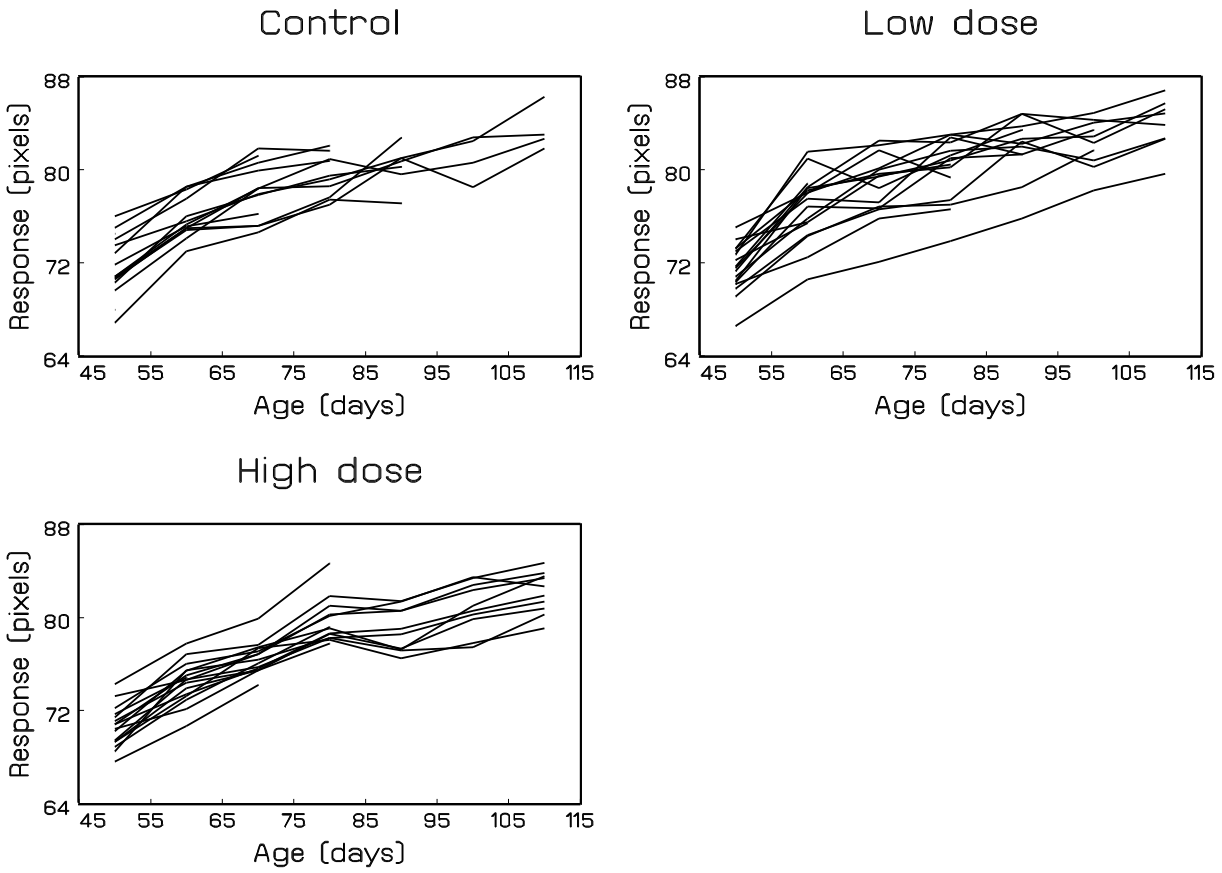
- Treatment starts at the age of 45 days; measurements taken every 10 days, from day 50 on.

- The responses are distances (pixels) between well defined points on x-ray pictures of the skull of each rat:



- Measurements with respect to the roof, base and height of the skull. Here, we consider only one response, reflecting the height of the skull.

- Individual profiles:



- Complication: Dropout due to anaesthesia (56%):

	# Observations			
Age (days)	Control	Low	High	Total
50	15	18	17	50
60	13	17	16	46
70	13	15	15	43
80	10	15	13	38
90	7	12	10	29
100	4	10	10	24
110	4	8	10	22

- Remarks:
  - ▷ Much variability between rats, much less variability within rats
  - ▷ Fixed number of measurements scheduled per subject, but not all measurements available due to dropout, for known reason.
  - ▷ Measurements taken at fixed time points

1.3 Prostate Data

- References:
  - ▷ Carter *et al* (1992, Cancer Research).
  - ▷ Carter *et al* (1992, Journal of the American Medical Association).
  - ▷ Morrell *et al* (1995, Journal of the American Statistical Association).
  - ▷ Pearson *et al* (1994, Statistics in Medicine).

- Prostate disease is one of the most common and most costly medical problems in the United States

- Important to look for markers which can detect the disease at an early stage

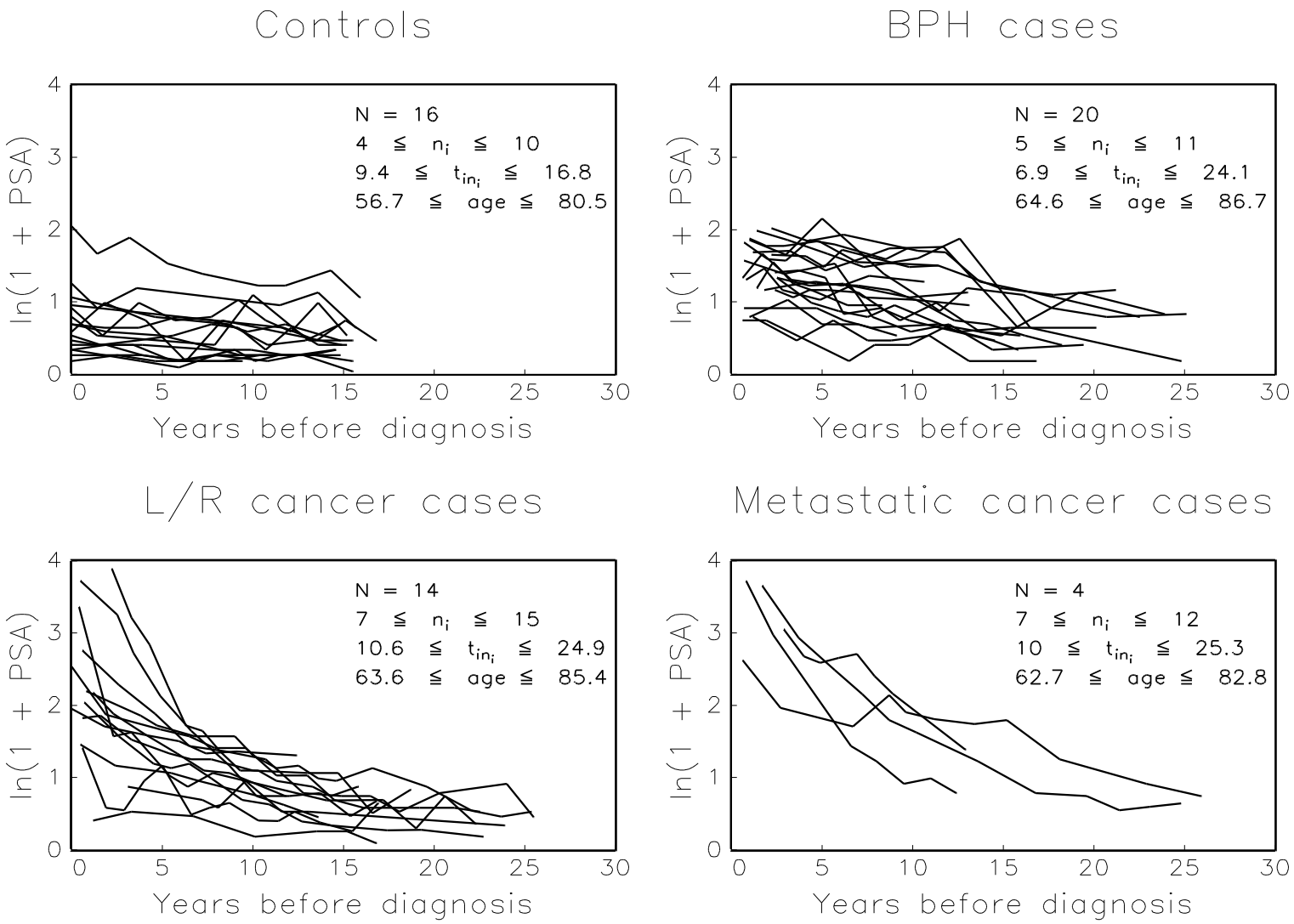
- **P**rostate-**S**pecific **A**ntigen is an enzyme produced by both normal and cancerous prostate cells

- PSA level is related to the volume of prostate tissue.
- Problem: Patients with **B**enign **P**rostatic **H**yperplasia also have an increased PSA level
- Overlap in PSA distribution for cancer and BPH cases seriously complicates the detection of prostate cancer.
- Research question (hypothesis based on clinical practice):

> Can longitudinal PSA profiles be used to detect prostate cancer in an early stage ?

- A retrospective case-control study based on frozen serum samples:
  - ▷ 16 control patients
  - ▷ 20 BPH cases
  - ▷ 14 local cancer cases
  - ▷ 4 metastatic cancer cases
- Complication: No perfect match for age at diagnosis and years of follow-up possible
- Hence, analyses will have to correct for these age differences between the diagnostic groups.

- Individual profiles:



- Remarks:
  - ▷ Much variability between subjects
  - ▷ Little variability within subjects
  - ▷ Highly unbalanced data

Chapter 2
A Model for Longitudinal Data

- Introduction
- The 2-stage model formulation
- Example: Rat data
- The general linear mixed-effects model
- Hierarchical versus marginal model
- Example: Rat data
- Example: Bivariate observations

2.1 Introduction

- In practice: often unbalanced data:
  - ▷ unequal number of measurements per subject
  - ▷ measurements not taken at fixed time points
- Therefore, multivariate regression techniques are often not applicable
- Often, subject-specific longitudinal profiles can be well approximated by linear regression functions
- This leads to a 2-stage model formulation:
  - ▷ **Stage 1:** Linear regression model for each subject separately
  - ▷ **Stage 2:** Explain variability in the subject-specific regression coefficients using known covariates

2.2 A 2-stage Model Formulation

2.2.1 Stage 1

- Response Y_{ij} for ith subject, measured at time t_{ij}, $i = 1, \ldots, N$, $j = 1, \ldots, n_i$
- Response vector $\boldsymbol{Y_i}$ for ith subject: $\boldsymbol{Y_i} = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})'$
- Stage 1 model:

$$\boldsymbol{Y_i} = Z_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon_i}$$

- Z_i is a $(n_i \times q)$ matrix of known covariates
- $\boldsymbol{\beta}_i$ is a q-dimensional vector of subject-specific regression coefficients
- $\boldsymbol{\varepsilon_i} \sim N(\mathbf{0}, \Sigma_i)$, often $\Sigma_i = \sigma^2 I_{n_i}$
- Note that the above model describes the observed variability within subjects

2.2.2 Stage 2

- Between-subject variability can now be studied from relating the $\boldsymbol{\beta}_i$ to known covariates

- Stage 2 model:

$$\boldsymbol{\beta}_i = K_i\boldsymbol{\beta} + \boldsymbol{b_i}$$

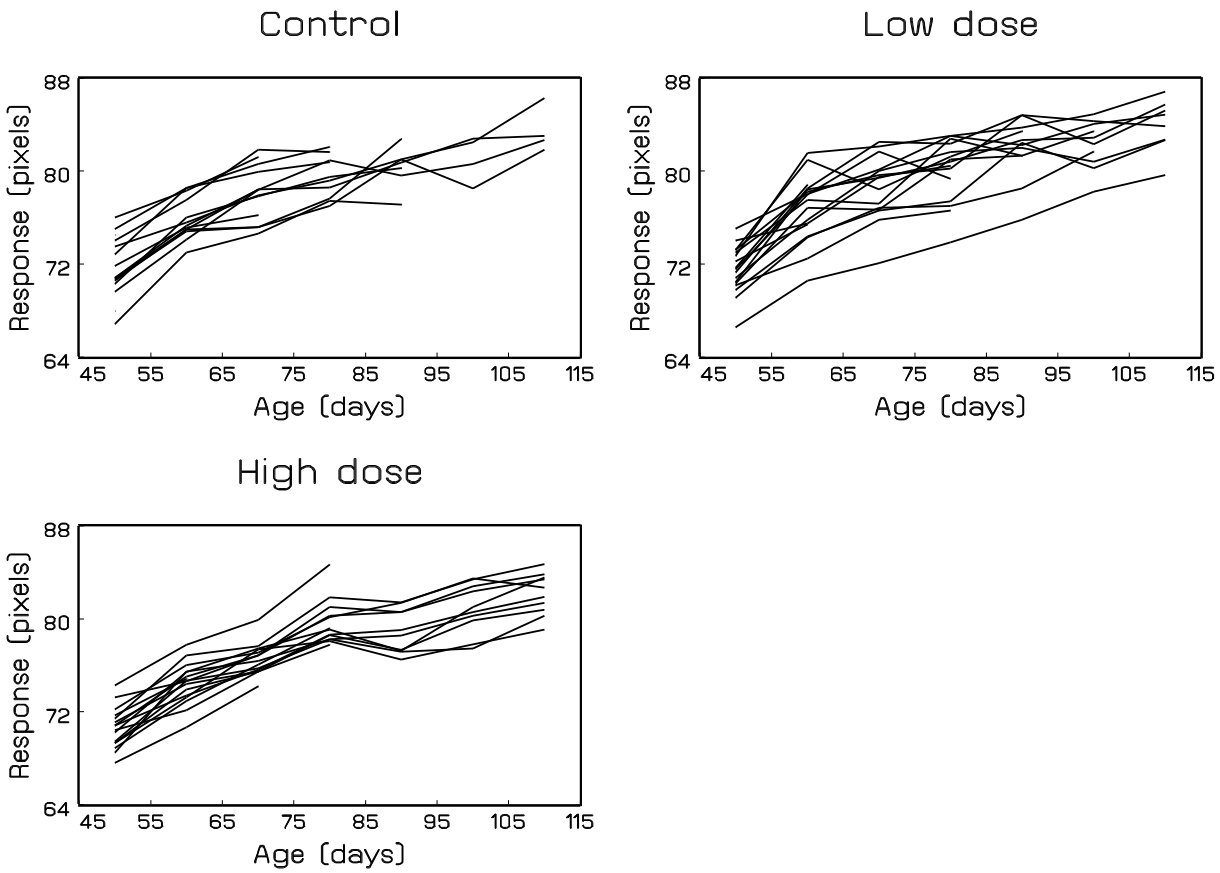
- K_i is a $(q \times p)$ matrix of known covariates

- $\boldsymbol{\beta}$ is a p-dimensional vector of unknown regression parameters

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

2.3 Example: The Rat Data

- Individual profiles:



- Transformation of the time scale to linearize the profiles:

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

- Note that $t = 0$ corresponds to the start of the treatment (moment of randomization)

- Stage 1 model: $Y_{ij} = \beta_{1i} + \beta_{2i} t_{ij} + \varepsilon_{ij}, \quad j = 1, \ldots, n_i$

- Matrix notation:

$$\boldsymbol{Y_i} = Z_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon_i} \quad \text{with} \quad Z_i = \begin{pmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{pmatrix}$$

- In the second stage, the subject-specific intercepts and time effects are related to the treatment of the rats

- Stage 2 model:

$$\begin{cases} \beta_{1i} = \beta_0 + b_{1i}, \\ \\ \beta_{2i} = \beta_1 L_i + \beta_2 H_i + \beta_3 C_i + b_{2i}, \end{cases}$$

- L_i, H_i, and C_i are indicator variables:

$$L_i = \begin{cases} 1 & \text{if low dose} \\ 0 & \text{otherwise} \end{cases} \quad H_i = \begin{cases} 1 & \text{if high dose} \\ 0 & \text{otherwise} \end{cases} \quad C_i = \begin{cases} 1 & \text{if control} \\ 0 & \text{otherwise} \end{cases}$$

- Parameter interpretation:
  - ▷ β_0: average response at the start of the treatment (independent of treatment)
  - ▷ β_1, β_2, and β_3: average time effect for each treatment group

2.4 The General Linear Mixed-effects Model

- A 2-stage approach can be performed explicitly in the analysis

- However, this is just another example of the use of summary statistics:
  - ▷ $\boldsymbol{Y_i}$ is summarized by $\widehat{\boldsymbol{\beta}_i}$
  - ▷ summary statistics $\widehat{\boldsymbol{\beta}_i}$ analysed in second stage

- The associated drawbacks can be avoided by combining the two stages into one model:

$$\begin{cases} \boldsymbol{Y_i} = Z_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon_i} \\ \boldsymbol{\beta}_i = K_i\boldsymbol{\beta} + \boldsymbol{b_i} \end{cases} \Longrightarrow \quad \boldsymbol{Y_i} = \underbrace{Z_iK_i}_{X_i}\boldsymbol{\beta} + Z_i\boldsymbol{b_i} + \boldsymbol{\varepsilon_i} = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b_i} + \boldsymbol{\varepsilon_i}$$

- General linear mixed-effects model:

$$\begin{cases} \boldsymbol{Y_i} = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b_i} + \boldsymbol{\varepsilon_i} \\ \\ \boldsymbol{b_i} \sim N(\mathbf{0}, D), \qquad \boldsymbol{\varepsilon_i} \sim N(\mathbf{0}, \Sigma_i), \\ \\ \boldsymbol{b_1}, \ldots, \boldsymbol{b_N}, \boldsymbol{\varepsilon_1}, \ldots, \boldsymbol{\varepsilon_N} \text{ independent} \end{cases}$$

- Terminology:
  - ▷ Fixed effects: $\boldsymbol{\beta}$
  - ▷ Random effects: $\boldsymbol{b_i}$
  - ▷ Variance components: elements in D and Σ_i

2.5 Hierarchical versus Marginal Model

- The general linear mixed model is given by:

$$\begin{cases} \boldsymbol{Y_i} = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b_i} + \boldsymbol{\varepsilon_i} \\ \\ \boldsymbol{b_i} \sim N(\boldsymbol{0}, D), \qquad \boldsymbol{\varepsilon_i} \sim N(\boldsymbol{0}, \Sigma_i), \\ \\ \boldsymbol{b_1}, \ldots, \boldsymbol{b_N}, \boldsymbol{\varepsilon_1}, \ldots, \boldsymbol{\varepsilon_N} \text{ independent} \end{cases}$$

- It can be rewritten as:

$$\boldsymbol{Y_i}|\boldsymbol{b_i} \sim N(X_i\boldsymbol{\beta} + Z_i\boldsymbol{b_i}, \Sigma_i), \qquad \boldsymbol{b_i} \sim N(\boldsymbol{0}, D)$$

- It is therefore also called a hierarchical model:
  - ▷ A model for $\boldsymbol{Y_i}$ given $\boldsymbol{b_i}$
  - ▷ A model for $\boldsymbol{b_i}$

- Marginally, we have that $\boldsymbol{Y_i}$ is distributed as:

$$\boldsymbol{Y_i} \sim N(X_i\boldsymbol{\beta}, Z_iDZ_i' + \Sigma_i)$$

- Hence, very specific assumptions are made about the dependence of mean and covariance onn the covariates X_i and Z_i:
  - ▷ **Implied mean :** $X_i\boldsymbol{\beta}$
  - ▷ **Implied covariance :** $V_i = Z_iDZ_i' + \Sigma_i$

- Note that the hierarchical model implies the marginal one, **NOT** vice versa

2.6 Example: The Rat Data

- Stage 1 model: $$Y_{ij} = \beta_{1i} + \beta_{2i} t_{ij} + \varepsilon_{ij}, \quad j = 1, \ldots, n_i$$

- Stage 2 model: $$\begin{cases} \beta_{1i} = \beta_0 + b_{1i}, \\ \beta_{2i} = \beta_1 L_i + \beta_2 H_i + \beta_3 C_i + b_{2i}, \end{cases}$$

- Combined: $$Y_{ij} = (\beta_0 + b_{1i}) + (\beta_1 L_i + \beta_2 H_i + \beta_3 C_i + b_{2i}) t_{ij} + \varepsilon_{ij}$$

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

- Implied marginal mean structure:
  - ▷ Linear average evolution in each group
  - ▷ Equal average intercepts
  - ▷ Different average slopes

- Implied marginal covariance structure ($\Sigma_i = \sigma^2 I_{n_i}$):

$$\begin{aligned} \text{Cov}(\boldsymbol{Y_i}(t_1), \boldsymbol{Y_i}(t_2)) &= \begin{pmatrix} 1 & t_1 \end{pmatrix} D \begin{pmatrix} 1 \\ t_2 \end{pmatrix} + \sigma^2 \delta_{\{t_1,t_2\}} \\ &= d_{22} t_1 \, t_2 + d_{12}(t_1 + t_2) + d_{11} + \sigma^2 \delta_{\{t_1,t_2\}}. \end{aligned}$$

- Note that the model implicitly assumes that the variance function is quadratic over time, with positive curvature d_{22}.

- A model which assumes that all variability in subject-specific slopes can be ascribed to treatment differences can be obtained by omitting the random slopes b_{2i} from the above model:

$$Y_{ij} = (\beta_0 + b_{1i}) + (\beta_1 L_i + \beta_2 H_i + \beta_3 C_i) t_{ij} + \varepsilon_{ij}$$

$$= \begin{cases} \beta_0 + b_{1i} + \beta_1 t_{ij} + \varepsilon_{ij}, & \text{if low dose} \\ \beta_0 + b_{1i} + \beta_2 t_{ij} + \varepsilon_{ij}, & \text{if high dose} \\ \beta_0 + b_{1i} + \beta_3 t_{ij} + \varepsilon_{ij}, & \text{if control.} \end{cases}$$

- This is the so-called random-intercepts model

- The same marginal mean structure is obtained as under the model with random slopes

- Implied marginal covariance structure ($\Sigma_i = \sigma^2 I_{n_i}$):

$$\begin{aligned}\text{Cov}(\boldsymbol{Y_i}(t_1), \boldsymbol{Y_i}(t_2)) &= \begin{pmatrix} 1 \end{pmatrix} D \begin{pmatrix} 1 \end{pmatrix} + \sigma^2 \delta_{\{t_1,t_2\}} \\ &= d_{11} + \sigma^2 \delta_{\{t_1,t_2\}}.\end{aligned}$$

- Hence, the implied covariance matrix is compound symmetry:
  - ▷ constant variance $d_{11} + \sigma^2$
  - ▷ constant correlation $\rho_I = d_{11}/(d_{11} + \sigma^2)$ between any two repeated measurements within the same rat

2.7 Example: Bivariate Observations

- Balanced data, two measurements per subject ($n_i = 2$), two models:

Model 1:

Random intercepts

+

heterogeneous errors

$$V = \begin{pmatrix} 1 \\ 1 \end{pmatrix} (d) \begin{pmatrix} 1 & 1 \end{pmatrix} + \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$$

$$= \begin{pmatrix} d + \sigma_1^2 & d \\ d & d + \sigma_2^2 \end{pmatrix}$$

Model 2:

Uncorrelated intercepts and slopes

+

measurement error

$$V = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$$

$$= \begin{pmatrix} d_1 + \sigma^2 & d_1 \\ d_1 & d_1 + d_2 + \sigma^2 \end{pmatrix}$$

- Different hierarchical models can produce the same marginal model
- Hence, a good fit of the marginal model cannot be interpreted as evidence for any of the hierarchical models.
- A satisfactory treatment of the hierarchical model is only possible within a Bayesian context.

Chapter 3
Estimation and Inference in the Marginal Model

- ML and REML estimation
- Fitting linear mixed models in SAS
- Negative variance components
- Inference

3.1 ML and REML Estimation

- Recall that the general linear mixed model equals

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

$$\left.\begin{array}{l} \boldsymbol{b_i} \sim N(\mathbf{0}, D) \\ \boldsymbol{\varepsilon_i} \sim N(\mathbf{0}, \Sigma_i) \end{array}\right\} \text{ independent}$$

- The implied marginal model equals $\quad \boldsymbol{Y_i} \sim N(X_i\boldsymbol{\beta}, Z_iDZ_i' + \Sigma_i)$

- Note that inferences based on the marginal model do not explicitly assume the presence of random effects representing the natural heterogeneity between subjects

- Notation:
  - ▷ $\boldsymbol{\beta}$: vector of fixed effects (as before)
  - ▷ $\boldsymbol{\alpha}$: vector of all variance components in D and Σ_i
  - ▷ $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}')'$: vector of all parameters in marginal model

- Marginal likelihood function:

$$L_{\mathsf{ML}}(\boldsymbol{\theta}) = \prod_{i=1}^{N} \left\{ (2\pi)^{-n_i/2} \left| V_i(\boldsymbol{\alpha}) \right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \left(\boldsymbol{Y_i} - X_i\boldsymbol{\beta} \right)' V_i^{-1}(\boldsymbol{\alpha}) \left(\boldsymbol{Y_i} - X_i\boldsymbol{\beta} \right) \right) \right\}$$

- If $\boldsymbol{\alpha}$ were known, MLE of $\boldsymbol{\beta}$ equals

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \left(\sum_{i=1}^{N} X_i' W_i X_i \right)^{-1} \sum_{i=1}^{N} X_i' W_i \boldsymbol{y_i},$$

  where W_i equals V_i^{-1}.

- In most cases, α is not known, and needs to be replaced by an estimate $\widehat{\alpha}$

- Two frequently used estimation methods for α:
  - ▷ Maximum likelihood
  - ▷ Restricted maximum likelihood

3.2 Fitting Linear Mixed Models in SAS

- A model for the prostate data:

$$\begin{aligned}
&\ln(\mathsf{PSA}_{ij}+1)\\
&= \beta_1\mathsf{Age}_i + \beta_2 C_i + \beta_3 B_i + \beta_4 L_i + \beta_5 M_i\\
&\quad + (\beta_6\mathsf{Age}_i + \beta_7 C_i + \beta_8 B_i + \beta_9 L_i + \beta_{10} M_i)\, t_{ij}\\
&\quad + (\beta_{11}\mathsf{Age}_i + \beta_{12} C_i + \beta_{13} B_i + \beta_{14} L_i + \beta_{15} M_i)\, t_{ij}^2\\
&\quad + b_{1i} + b_{2i} t_{ij} + b_{3i} t_{ij}^2 + \varepsilon_{ij}.
\end{aligned}$$

- SAS program ($\Sigma_i = \sigma^2 I_{n_i}$):

```
proc mixed data=prostate method=reml;
class id group timeclss;
model lnpsa = group age time group*time age*time
              time2 group*time2 age*time2 / solution;
random intercept time time2 / type=un subject=id;
run;
```

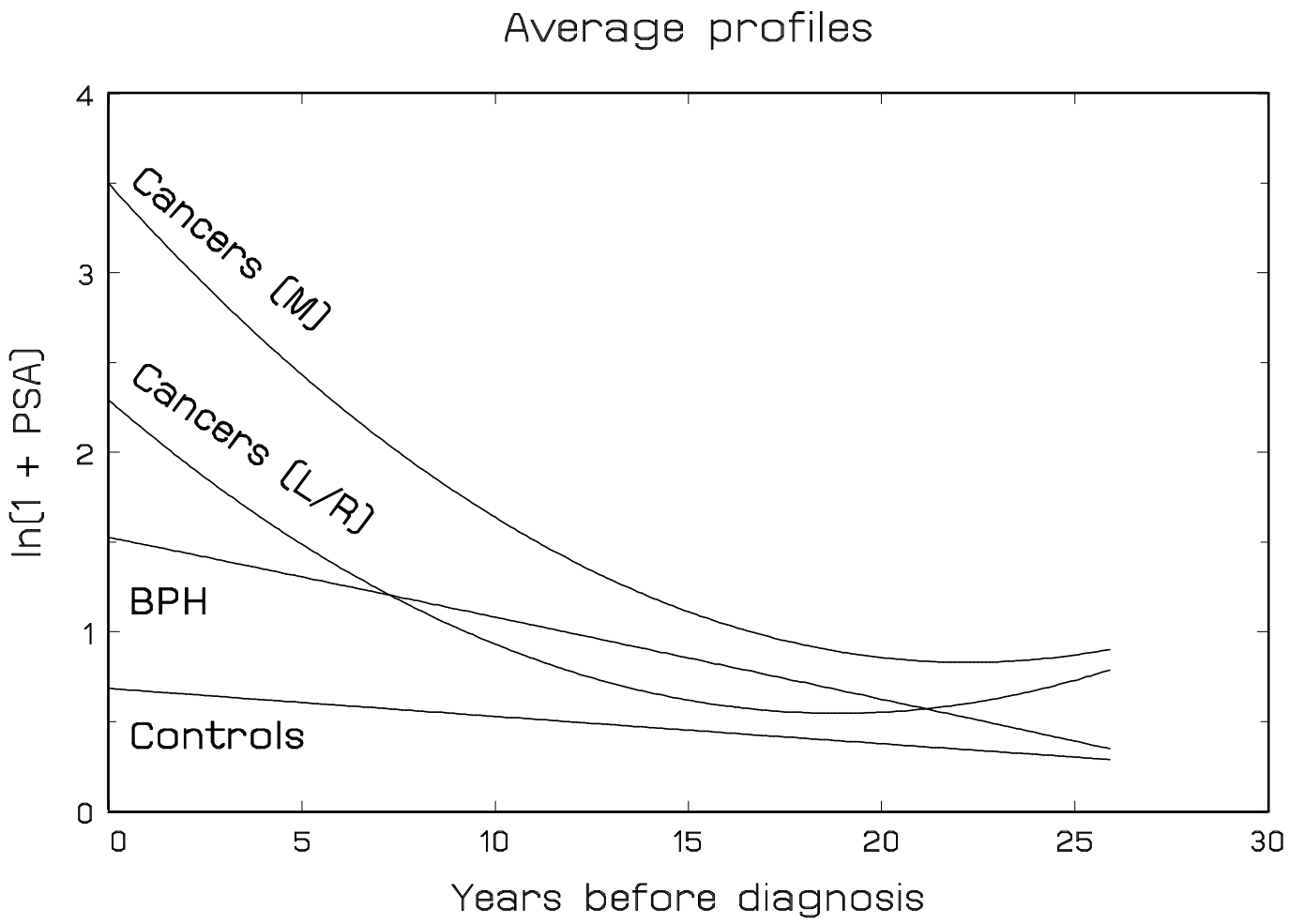
- ML and REML estimates for fixed effects:

Effect	Parameter	MLE (s.e.)	REMLE (s.e.)
Age effect	β_1	0.026 (0.013)	0.027 (0.014)
Intercepts:			
Control	β_2	−1.077 (0.919)	−1.098 (0.976)
BPH	β_3	−0.493 (1.026)	−0.523 (1.090)
L/R cancer	β_4	0.314 (0.997)	0.296 (1.059)
Met. cancer	β_5	1.574 (1.022)	1.549 (1.086)
Age×time effect	β_6	−0.010 (0.020)	−0.011 (0.021)
Time effects:			
Control	β_7	0.511 (1.359)	0.568 (1.473)
BPH	β_8	0.313 (1.511)	0.396 (1.638)
L/R cancer	β_9	−1.072 (1.469)	−1.036 (1.593)
Met. cancer	β_{10}	−1.657 (1.499)	−1.605 (1.626)
Age×time^2 effect	β_{11}	0.002 (0.008)	0.002 (0.009)
Time^2 effects:			
Control	β_{12}	−0.106 (0.549)	−0.130 (0.610)
BPH	β_{13}	−0.119 (0.604)	−0.158 (0.672)
L/R cancer	β_{14}	0.350 (0.590)	0.342 (0.656)
Met. cancer	β_{15}	0.411 (0.598)	0.395 (0.666)

- ML and REML estimates for variance components:

Effect	Parameter	MLE (s.e.)	REMLE (s.e.)
Covariance of $\boldsymbol{b_i}$:			
var(b_{1i})	d_{11}	0.398 (0.083)	0.452 (0.098)
var(b_{2i})	d_{22}	0.768 (0.187)	0.915 (0.230)
var(b_{3i})	d_{33}	0.103 (0.032)	0.131 (0.041)
cov(b_{1i}, b_{2i})	$d_{12} = d_{21}$	−0.443 (0.113)	−0.518 (0.136)
cov(b_{2i}, b_{3i})	$d_{23} = d_{32}$	−0.273 (0.076)	−0.336 (0.095)
cov(b_{3i}, b_{1i})	$d_{13} = d_{31}$	0.133 (0.043)	0.163 (0.053)
Residual variance:			
var(ε_{ij})	σ^2	0.028 (0.002)	0.028 (0.002)
Log-likelihood		−1.788	−31.235

- Fitted average profiles at median age at diagnosis:



3.3 Negative Variance Components

- Reconsider the model for the rat data:

$$Y_{ij} = (\beta_0 + b_{1i}) + (\beta_1 L_i + \beta_2 H_i + \beta_3 C_i + b_{2i}) t_{ij} + \varepsilon_{ij}$$

- REML estimates obtained from SAS procedure MIXED:

Effect	Parameter	REMLE (s.e.)
Intercept	β_0	68.606 (0.325)
Time effects:		
Low dose	β_1	7.503 (0.228)
High dose	β_2	6.877 (0.231)
Control	β_3	7.319 (0.285)
Covariance of $\boldsymbol{b_i}$:		
var(b_{1i})	d_{11}	3.369 (1.123)
var(b_{2i})	d_{22}	0.000 (——)
cov(b_{1i}, b_{2i})	$d_{12} = d_{21}$	0.090 (0.381)
Residual variance:		
var(ε_{ij})	σ^2	1.445 (0.145)
REML log-likelihood		−466.173

- This suggests that the REML likelihood could be further increased by allowing negative estimates for d_{22}

- In SAS, this can be done by adding the option 'nobound' to the PROC MIXED statement.

- Results:

		Parameter restrictions for $\boldsymbol{\alpha}$	
		$d_{ii} \geq 0, \sigma^2 \geq 0$	$d_{ii} \in \mathbb{R}, \sigma^2 \in \mathbb{R}$
Effect	Parameter	REMLE (s.e.)	REMLE (s.e.)
Intercept	β_0	68.606 (0.325)	68.618 (0.313)
Time effects:			
Low dose	β_1	7.503 (0.228)	7.475 (0.198)
High dose	β_2	6.877 (0.231)	6.890 (0.198)
Control	β_3	7.319 (0.285)	7.284 (0.254)
Covariance of $\boldsymbol{b_i}$:			
var(b_{1i})	d_{11}	3.369 (1.123)	2.921 (1.019)
var(b_{2i})	d_{22}	0.000 (——)	-0.287 (0.169)
cov(b_{1i}, b_{2i})	$d_{12} = d_{21}$	0.090 (0.381)	0.462 (0.357)
Residual variance:			
var(ε_{ij})	σ^2	1.445 (0.145)	1.522 (0.165)
REML log-likelihood		-466.173	-465.193

- Note that the REML log-likelihood value has been further increased and a negative estimate for d_{22} is obtained.

- Brown & Prescott (1999, p. 237) :

Negative variance components

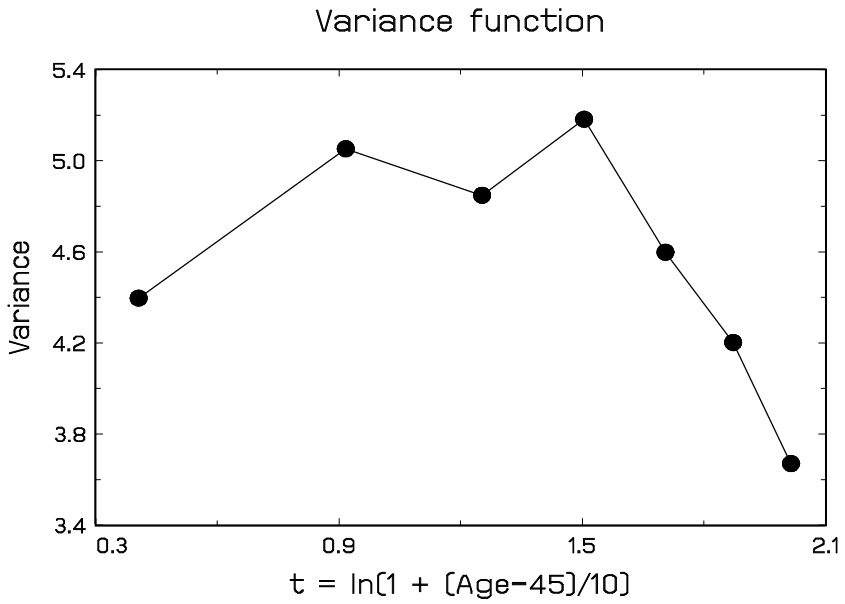
The usual action when a negative variance component estimate is obtained for a random coefficient would be to refit the model with the random coefficient removed. However, the user should be warned that not all packages will produce a negative variance component estimate. For example, in PROC MIXED we have found that non-convergence or a message stating that the **G** matrix is not positive semi-definite are usually indications of a negative variance component. (A matrix, **A**, is positive semi-definite if $\mathbf{x}'\mathbf{A}\mathbf{x}$ is a non-negative number for all vectors, **x**.) The recommended action is then to remove the random coefficients one by one in decreasing order of complexity until all variance components become positive.

- Meaning of negative variance component ?
  - ▷ Fitted variance function:

$$\text{Var}(\boldsymbol{Y}_i(t)) = \begin{pmatrix} 1 & t \end{pmatrix} \widehat{D} \begin{pmatrix} 1 \\ t \end{pmatrix} + \widehat{\sigma}^2$$

$$= \widehat{d_{22}} t^2 + 2\widehat{d_{12}} t + \widehat{d_{11}} + \widehat{\sigma}^2 = \mathbf{-0.287} t^2 + 0.924t + 4.443$$

  - ▷ The suggested negative curvature in the variance function is supported by the sample variance function:



- This again shows that the hierarchical and marginal models are not equivalent:
  - ▷ The marginal model allows negative variance components, as long as the marginal covariances $V_i = Z_i D Z_i' + \sigma^2 I_{n_i}$ are positive definite
  - ▷ The hierarchical interpretation of the model does not allow negative variance components

3.4 Inference for the Marginal Model

- Estimate for $\boldsymbol{\beta}$:

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \left(\sum_{i=1}^{N} X_i' W_i X_i \right)^{-1} \sum_{i=1}^{N} X_i' W_i \boldsymbol{y_i}$$

  with $\boldsymbol{\alpha}$ replaced by its ML or REML estimate

- Conditional on $\boldsymbol{\alpha}$, $\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ is multivariate normal with mean $\boldsymbol{\beta}$ and covariance

$$\begin{aligned} \text{Var}(\widehat{\boldsymbol{\beta}}) &= \left(\sum_{i=1}^{N} X_i' W_i X_i \right)^{-1} \left(\sum_{i=1}^{N} X_i' W_i \text{var}(\boldsymbol{Y_i}) W_i X_i \right) \left(\sum_{i=1}^{N} X_i' W_i X_i \right)^{-1} \\ &= \left(\sum_{i=1}^{N} X_i' W_i X_i \right)^{-1} \end{aligned}$$

- In practice one again replaces $\boldsymbol{\alpha}$ by its ML or REML estimate

- Inference for $\boldsymbol{\beta}$:
  - ▷ Wald tests
  - ▷ t- and F-tests
  - ▷ LR tests (not with REML)
- Inference for $\boldsymbol{\alpha}$:
  - ▷ Wald tests
  - ▷ LR tests (even with REML)
  - ▷ Caution 1: Marginal vs. hierarchical model
  - ▷ Caution 2: Boundary problems

Chapter 4
Inference for the Random Effects

- ▷ Empirical Bayes inference
- ▷ Example: Prostate data
- ▷ Shrinkage
- ▷ Example: Prostate data

4.1 Empirical Bayes Inference

- Random effects $\boldsymbol{b_i}$ reflect how the evolution for the ith subject deviates from the expected evolution $X_i\boldsymbol{\beta}$.

- Estimation of the $\boldsymbol{b_i}$ helpful for detecting outlying profiles

- This is only meaningful under the hierarchical model interpretation:

$$\mathbf{Y_i}|\boldsymbol{b_i} \sim N(X_i\boldsymbol{\beta} + Z_i\boldsymbol{b_i}, \Sigma_i) \qquad \boldsymbol{b_i} \sim N(\mathbf{0}, D)$$

- Since the $\boldsymbol{b_i}$ are random, it is most natural to use Bayesian methods

- Terminology: prior distribution $N(\mathbf{0}, D)$ for $\boldsymbol{b_i}$

- Posterior density:

$$f(\boldsymbol{b}_i|\boldsymbol{y}_i) \equiv f(\boldsymbol{b}_i|\boldsymbol{Y}_i = \boldsymbol{y}_i) = \frac{f(\boldsymbol{y}_i|\boldsymbol{b}_i)\ f(\boldsymbol{b}_i)}{\int\ f(\boldsymbol{y}_i|\boldsymbol{b}_i)\ f(\boldsymbol{b}_i)\ d\boldsymbol{b}_i}$$

$$\propto f(\boldsymbol{y}_i|\boldsymbol{b}_i)\ f(\boldsymbol{b}_i)$$

$$\propto \ldots$$

$$\propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{b_i} - DZ_i'W_i(\boldsymbol{y_i} - X_i\boldsymbol{\beta})\right)'\Lambda_i^{-1}\left(\boldsymbol{b_i} - DZ_i'W_i(\boldsymbol{y_i} - X_i\boldsymbol{\beta})\right)\right\}$$

for some positive definite matrix Λ_i.

- Posterior distribution:

$$\boldsymbol{b_i} \mid \boldsymbol{y_i} \sim N\left(DZ_i'W_i(\boldsymbol{y_i} - X_i\boldsymbol{\beta}), \Lambda_i\right)$$

- Posterior mean as estimate for $\boldsymbol{b_i}$:

$$\widehat{\boldsymbol{b_i}}(\boldsymbol{\theta}) \;=\; E\left[\boldsymbol{b_i} \mid \boldsymbol{Y_i} = \boldsymbol{y_i}\right] \;=\; \int \boldsymbol{b_i}\; f(\boldsymbol{b_i}|\boldsymbol{y_i})\; d\boldsymbol{b_i} \;=\; DZ_i'W_i(\boldsymbol{\alpha})(\boldsymbol{y_i} - X_i\boldsymbol{\beta})$$

- Parameters in $\boldsymbol{\theta}$ are replaced by their ML or REML estimates, obtained from fitting the marginal model.

- $\widehat{\boldsymbol{b_i}} = \widehat{\boldsymbol{b_i}}(\widehat{\boldsymbol{\theta}})$ is called the Empirical Bayes estimate of $\boldsymbol{b_i}$.

- Approximate t- and F-tests to account for the variability introduced by replacing $\boldsymbol{\theta}$ by $\widehat{\boldsymbol{\theta}}$, similar to tests for fixed effects.

4.2 Example: Prostate Data

- We reconsider our linear mixed model:

$$\begin{aligned}
&\ln(\mathsf{PSA}_{ij}+1)\\
&= \beta_1\mathsf{Age}_i + \beta_2 C_i + \beta_3 B_i + \beta_4 L_i + \beta_5 M_i\\
&\quad + (\beta_6\mathsf{Age}_i + \beta_7 C_i + \beta_8 B_i + \beta_9 L_i + \beta_{10} M_i)\, t_{ij}\\
&\quad + (\beta_{11}\mathsf{Age}_i + \beta_{12} C_i + \beta_{13} B_i + \beta_{14} L_i + \beta_{15} M_i)\, t_{ij}^2\\
&\quad + b_{1i} + b_{2i} t_{ij} + b_{3i} t_{ij}^2 + \varepsilon_{ij}.
\end{aligned}$$

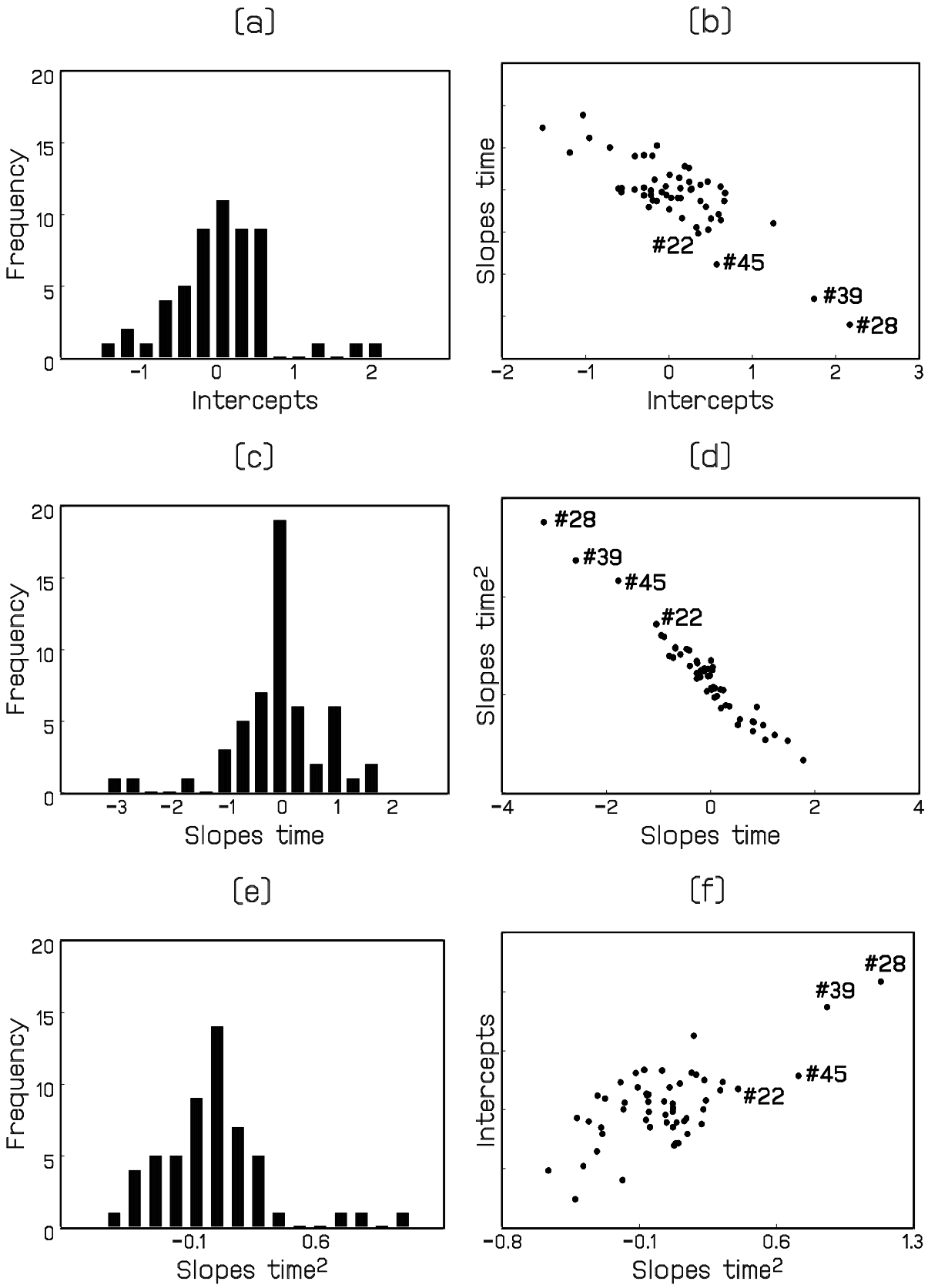
- In SAS the estimates can be obtained from adding the option 'solution' to the random statement:

```
random intercept time time2
       / type=un subject=id solution;

ods listing exclude solutionr;
ods output solutionr=out;
```

- The ODS statements are used to write the EB estimates into a SAS output data set, and to prevent SAS from printing them in the output window.

- In practice, histograms and scatterplots of certain components of $\widehat{\boldsymbol{b_i}}$ are used to detect model deviations or subjects with 'exceptional' evolutions over time


(a)
Frequency
Intercepts
(b)
Slopes time
Intercepts
#22
#45
#39
#28
(c)
Frequency
Slopes time
(d)
Slopes time2
Slopes time
#28
#39
#45
#22
(e)
Frequency
Slopes time2
(f)
Intercepts
Slopes time2
#28
#39
#45
#22

- Strong negative correlations in agreement with correlation matrix corresponding to fitted D:

$$\widehat{D}_{\text{corr}} = \begin{pmatrix} 1.000 & -0.803 & 0.658 \\ -0.803 & 1.000 & -0.968 \\ 0.658 & -0.968 & 1.000 \end{pmatrix}$$

- Histograms and scatterplots show outliers

- Subjects #22, #28, #39, and #45, have highest four slopes for time^2 and smallest four slopes for time, i.e., with the strongest (quadratic) growth.

- Subjects #22, #28 and #39 have been further examined and have been shown to be metastatic cancer cases which were misclassified as local cancer cases.

- Subject #45 is the metastatic cancer case with the strongest growth

Part II

Marginal Models for Non-Gaussian Data

Chapter 5
Generalized Estimating Equations

- ▷ General idea
- ▷ Asymptotic properties
- ▷ Working correlation
- ▷ Special case and application
- ▷ SAS code and output

5.1 General Idea

- Univariate GLM, score function of the form (scalar Y_i):

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{N} \frac{\partial \mu_i}{\partial \boldsymbol{\beta}} v_i^{-1}(y_i - \mu_i) = \mathbf{0}, \quad \text{with } v_i = \text{Var}(Y_i).$$

- In longitudinal setting: $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N)$:

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{N} D_i' \left[\boldsymbol{V}_i(\boldsymbol{\alpha})\right]^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}$$

  where

  ▷ D_i is an $n_i \times p$ matrix with (i,j)th elements $\dfrac{\partial \mu_{ij}}{\partial \boldsymbol{\beta}}$

  ▷ Is V_i $n_i \times n_i$ diagonal?

  ▷ $\boldsymbol{y}_i$ and $\boldsymbol{\mu}_i$ are n_i-vectors with elements y_{ij} and μ_{ij}

- The corresponding matrices $\boldsymbol{V}_i = \text{Var}(\boldsymbol{Y}_i)$ involve a set of nuisance parameters, $\boldsymbol{\alpha}$ say, which determine the covariance structure of $\boldsymbol{Y}_i$.

- Same form as for full likelihood procedure

- We restrict specification to the first moment only

- The second moment is only specified in the variances, not in the correlations.

- Solving these equations:
  - ▷ version of iteratively weighted least squares
  - ▷ Fisher scoring

- Liang and Zeger (1986)

5.2 Large Sample Properties

As $N \to \infty$

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim N(\mathbf{0}, I_0^{-1})$$

where

$$I_0 = \sum_{i=1}^{N} D_i'[V_i(\boldsymbol{\alpha})]^{-1} D_i$$

- **(Unrealistic) Conditions:**
  - ▷ $\boldsymbol{\alpha}$ is known
  - ▷ the parametric form for $\boldsymbol{V}_i(\boldsymbol{\alpha})$ is known
- **Solution: working correlation matrix**

5.3 Unknown Covariance Structure

Keep the score equations

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{N} [D_i]' [V_i(\boldsymbol{\alpha})]^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}$$

BUT

- suppose $\boldsymbol{V}_i(.)$ is not the true variance of $\boldsymbol{Y}_i$ but only a plausible guess, a so-called working correlation matrix
- specify correlations and not covariances, because the variances follow from the mean structure
- the score equations are solved as before

The asymptotic normality results change to

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim N(\mathbf{0}, I_0^{-1} I_1 I_0^{-1})$$

$$I_0 = \sum_{i=1}^{N} D_i' [V_i(\boldsymbol{\alpha})]^{-1} D_i$$

$$I_1 = \sum_{i=1}^{N} D_i' [V_i(\boldsymbol{\alpha})]^{-1} \mathsf{Var}(\boldsymbol{Y}_i) [V_i(\boldsymbol{\alpha})]^{-1} D_i.$$

5.4 The Sandwich Estimator

- This is the so-called **sandwich estimator**:
  - ▷ I_0 is the bread
  - ▷ I_1 is the filling (ham or cheese)
- Correct guess $\Longrightarrow$ likelihood variance
- The estimators $\hat{\boldsymbol{\beta}}$ are consistent even if the working correlation matrix is incorrect
- An estimate is found by replacing the unknown variance matrix $\text{Var}(\boldsymbol{Y}_i)$ by

$$(\boldsymbol{Y}_i - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{Y}_i - \hat{\boldsymbol{\mu}}_i)'.$$

- Even if this estimator is bad for Var($\boldsymbol{Y}_i$) it leads to a good estimate of I_1, provided that:
  - ▷ replication in the data is sufficiently large
  - ▷ same model for $\boldsymbol{\mu}_i$ is fitted to groups of subjects
  - ▷ observation times do not vary too much between subjects
- A bad choice of working correlation matrix **can** affect the efficiency of $\hat{\boldsymbol{\beta}}$

5.5 The Working Correlation Matrix

$$\boxed{V_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \phi A_i^{1/2}(\boldsymbol{\beta}) R_i(\boldsymbol{\alpha}) A_i^{1/2}(\boldsymbol{\beta}).}$$

- Variance function: A_i is $(n_i \times n_i)$ diagonal with elements $v(\mu_{ij})$, the known GLM variance function.
- Working correlation: $R_i(\boldsymbol{\alpha})$ possibly depends on a different set of parameters $\boldsymbol{\alpha}$.
- Overdispersion parameter: ϕ, assumed 1 or estimated from the data.
- The unknown quantities are expressed in terms of the Pearson residuals

$$e_{ij} = \frac{y_{ij} - \mu_{ij}}{\sqrt{v(\mu_{ij})}}.$$

  Note that e_{ij} depends on $\boldsymbol{\beta}$.

5.6 Estimation of Working Correlation

Liang and Zeger (1986) proposed moment-based estimates for the working correlation.

	Corr(Y_{ij}, Y_{ik})	Estimate
Independence	0	—
Exchangeable	α	$\hat{\alpha} = \frac{1}{N}\Sigma_{i=1}^{N} \frac{1}{n_i(n_i-1)} \Sigma_{j \neq k} e_{ij}e_{ik}$
AR(1)	α	$\hat{\alpha} = \frac{1}{N}\Sigma_{i=1}^{N} \frac{1}{n_i-1} \Sigma_{j \leq n_i-1} e_{ij}e_{i,j+1}$
Unstructured	α_{jk}	$\hat{\alpha}_{jk} = \frac{1}{N}\Sigma_{i=1}^{N} e_{ij}e_{ik}$

Dispersion parameter:

$$\hat{\phi} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{n_i} \sum_{j=1}^{n_i} e_{ij}^2.$$

5.7 Special Case: Linear Mixed Models

- Estimate for $\boldsymbol{\beta}$:

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \left(\sum_{i=1}^{N} X_i' W_i X_i \right)^{-1} \sum_{i=1}^{N} X_i' W_i \boldsymbol{Y_i}$$

  with $\boldsymbol{\alpha}$ replaced by its ML or REML estimate

- Conditional on $\boldsymbol{\alpha}$, $\widehat{\boldsymbol{\beta}}$ has mean

$$E\left[\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})\right] = \left(\sum_{i=1}^{N} X_i' W_i X_i \right)^{-1} \sum_{i=1}^{N} X_i' W_i X_i \boldsymbol{\beta} = \boldsymbol{\beta}$$

  provided that $E(\boldsymbol{Y_i}) = X_i \boldsymbol{\beta}$

- Hence, in order for $\widehat{\boldsymbol{\beta}}$ to be unbiased, it is sufficient that the mean of the response is correctly specified.

- Conditional on $\boldsymbol{\alpha}$, $\widehat{\boldsymbol{\beta}}$ has covariance

$$\text{Var}(\widehat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^{N} X_i' W_i X_i\right)^{-1} \left(\sum_{i=1}^{N} X_i' W_i \text{Var}(\boldsymbol{Y_i}) W_i X_i\right) \left(\sum_{i=1}^{N} X_i' W_i X_i\right)^{-1} = \left(\sum_{i=1}^{N} X_i' W_i X_i\right)^{-1}$$

- Note that this **model-based version** assumes that the covariance matrix $\text{Var}(\boldsymbol{Y_i})$ is correctly modelled as $V_i = Z_i D Z_i' + \Sigma_i$.

- An **empirically corrected version** is:

$$\text{Var}(\widehat{\boldsymbol{\beta}}) = \underbrace{\left(\sum_{i=1}^{N} X_i' W_i X_i\right)^{-1}}_{\downarrow \text{ BREAD}} \underbrace{\left(\sum_{i=1}^{N} X_i' W_i \text{Var}(\boldsymbol{Y_i}) W_i X_i\right)}_{\downarrow \text{ MEAT}} \underbrace{\left(\sum_{i=1}^{N} X_i' W_i X_i\right)^{-1}}_{\downarrow \text{ BREAD}}$$

5.8 Application to the Toenail Data

5.8.1 The model

- Consider the model:

$$Y_{ij} \sim \text{Bernoulli}(\mu_{ij}), \qquad \log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \beta_0 + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij}$$

- Y_{ij}: severe infection (yes/no) at occasion j for patient i
- t_{ij}: measurement time for occasion j
- T_i: treatment group

5.8.2 Standard GEE

- **SAS Code:**

```
proc genmod data=test descending;
class idnum timeclss;
model onyresp = treatn time treatn*time
                / dist=binomial;
repeated subject=idnum / withinsubject=timeclss
                         type=exch covb corrw modelse;
run;
```

- SAS statements:
  - ▷ The REPEATED statements defines the GEE character of the model.
  - ▷ 'type=': working correlation specification (UN, AR(1), EXCH, IND,...)
  - ▷ 'modelse': model-based s.e.'s on top of default empirically corrected s.e.'s
  - ▷ 'corrw': printout of working correlation matrix
  - ▷ 'withinsubject=': specification of the ordering within subjects

- **Selected output:**
  - ▷ Regression paramters:

Analysis Of Initial Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-0.5571	0.1090	-0.7708	-0.3433	26.10
treatn	1	0.0240	0.1565	-0.2827	0.3307	0.02
time	1	-0.1769	0.0246	-0.2251	-0.1288	51.91
treatn*time	1	-0.0783	0.0394	-0.1556	-0.0010	3.95
Scale	0	1.0000	0.0000	1.0000	1.0000	

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	-0.5840	0.1734	-0.9238	-0.2441	-3.37	0.0008
treatn	0.0120	0.2613	-0.5001	0.5241	0.05	0.9633
time	-0.1770	0.0311	-0.2380	-0.1161	-5.69	<.0001
treatn*time	-0.0886	0.0571	-0.2006	0.0233	-1.55	0.1208

Analysis Of GEE Parameter Estimates
Model-Based Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	-0.5840	0.1344	-0.8475	-0.3204	-4.34	<.0001
treatn	0.0120	0.1866	-0.3537	0.3777	0.06	0.9486
time	-0.1770	0.0209	-0.2180	-0.1361	-8.47	<.0001
treatn*time	-0.0886	0.0362	-0.1596	-0.0177	-2.45	0.0143

- The working correlation:

Exchangeable Working Correlation

Correlation 0.420259237

Part III

Hierarchical Models for Non-Gaussian Data

Chapter 6
Generalized Linear Mixed Models (GLMM)

- ▷ Introduction: LMM Revisited
- ▷ Generalized Linear Mixed Models (GLMM)
- ▷ Fitting Algorithms
- ▷ Example

6.1 Introduction: LMM Revisited

- We re-consider the linear mixed model:

$$\boldsymbol{Y_i}|\boldsymbol{b_i} \sim N(X_i\boldsymbol{\beta} + Z_i\boldsymbol{b_i}, \Sigma_i), \qquad \boldsymbol{b_i} \sim N(\mathbf{0}, D)$$

- The implied marginal model equals $\boldsymbol{Y_i} \sim N(X_i\boldsymbol{\beta}, Z_iDZ_i' + \Sigma_i)$

- Hence, even under conditional independence, i.e., all Σ_i equal to $\sigma^2 I_{n_i}$, a marginal association structure is implied through the random effects.

- The same ideas can now be applied in the context of GLM's to model association between discrete repeated measures.

6.2 Generalized Linear Mixed Models (GLMM)

- Given a vector $\boldsymbol{b_i}$ of random effects for cluster i, it is assumed that all responses Y_{ij} are independent, with density

$$f(y_{ij}|\theta_{ij},\phi) \;=\; \exp\left\{\phi^{-1}[y_{ij}\theta_{ij} - \psi(\theta_{ij})] + c(y_{ij},\phi)\right\}$$

- θ_{ij} is now modelled as $\theta_{ij} \;=\; \boldsymbol{x_{ij}}'\boldsymbol{\beta} \;+\; \boldsymbol{z_{ij}}'\boldsymbol{b_i}$

- As before, it is assumed that $\boldsymbol{b_i} \;\sim\; N(\mathbf{0}, D)$

- Let $f_{ij}(y_{ij}|\boldsymbol{b_i},\boldsymbol{\beta},\phi)$ denote the conditional density of Y_{ij} given $\boldsymbol{b_i}$, the conditional density of $\boldsymbol{Y_i}$ equals

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

- The marginal distribution of $\boldsymbol{Y_i}$ is given by

$$f_i(\boldsymbol{y_i}|\boldsymbol{\beta}, D, \phi) = \int f_i(\boldsymbol{y_i}|\boldsymbol{b_i}, \boldsymbol{\beta}, \phi)\ f(\boldsymbol{b_i}|D)\ d\boldsymbol{b_i}$$

$$= \int \prod_{j=1}^{n_i} f_{ij}(y_{ij}|\boldsymbol{b_i}, \boldsymbol{\beta}, \phi)\ f(\boldsymbol{b_i}|D)\ d\boldsymbol{b_i}$$

  where $f(\boldsymbol{b_i}|D)$ is the density of the $N(\mathbf{0}, D)$ distribution.

- The likelihood function for $\boldsymbol{\beta}$, D, and ϕ now equals

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

$$= \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} f_{ij}(y_{ij}|\boldsymbol{b_i}, \boldsymbol{\beta}, \phi)\ f(\boldsymbol{b_i}|D)\ d\boldsymbol{b_i}$$

- Under the normal linear model, the integral can be worked out analytically.

- In general, approximations are required:
  - ▷ Approximation of integrand
  - ▷ Approximation of data
  - ▷ Approximation of integral

- Predictions of random effects can be based on the posterior distribution

$$f(\boldsymbol{b}_i|\boldsymbol{Y}_i = \boldsymbol{y}_i)$$

- 'Empirical Bayes (EB) estimate':
  Posterior mode, with unknown parameters replaced by their MLE

6.3 Laplace Approximation of Integrand

- Integrals in $L(\boldsymbol{\beta}, D, \phi)$ 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

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

- 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

6.4 Approximation of Data

6.4.1 General Idea

- Re-write GLMM as:

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

  with variance for errors equal to $\text{Var}(Y_{ij}|\boldsymbol{b_i}) = \phi v(\mu_{ij})$

- 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 $\boldsymbol{b_i} = \boldsymbol{0}$

6.4.2 Penalized quasi-likelihood (PQL)

- Linear Taylor expansion around current $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{b_i}}$:

$$Y_{ij} \approx h(\boldsymbol{x}'_{ij}\widehat{\boldsymbol{\beta}} + \boldsymbol{z}'_{ij}\widehat{\boldsymbol{b_i}}) + h'(\boldsymbol{x}'_{ij}\widehat{\boldsymbol{\beta}} + \boldsymbol{z}'_{ij}\widehat{\boldsymbol{b_i}})\boldsymbol{x}'_{ij}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + h'(\boldsymbol{x}'_{ij}\widehat{\boldsymbol{\beta}} + \boldsymbol{z}'_{ij}\widehat{\boldsymbol{b_i}})\boldsymbol{z}'_{ij}(\boldsymbol{b_i} - \widehat{\boldsymbol{b_i}}) + \varepsilon_{ij}$$

$$\approx \widehat{\mu}_{ij} + v(\widehat{\mu}_{ij})\boldsymbol{x}'_{ij}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + v(\widehat{\mu}_{ij})\boldsymbol{z}'_{ij}(\boldsymbol{b_i} - \widehat{\boldsymbol{b_i}}) + \varepsilon_{ij}$$

- In vector notation: $\boldsymbol{Y_i} \approx \widehat{\boldsymbol{\mu}}_i + \widehat{V}_i X_i(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + \widehat{V}_i Z_i(\boldsymbol{b_i} - \widehat{\boldsymbol{b_i}}) + \boldsymbol{\varepsilon_i}$

- Re-ordering terms yields:

$$\boldsymbol{Y_i}^* \equiv \widehat{V}_i^{-1}(\boldsymbol{Y_i} - \widehat{\boldsymbol{\mu}}_i) + X_i\widehat{\boldsymbol{\beta}} + Z_i\widehat{\boldsymbol{b_i}} \approx X_i\boldsymbol{\beta} + Z_i\boldsymbol{b_i} + \boldsymbol{\varepsilon_i}^*,$$

- Model fitting by iterating between updating the pseudo responses $\boldsymbol{Y_i}^*$ and fitting the above linear mixed model to them.

6.4.3 Marginal quasi-likelihood (MQL)

- Linear Taylor expansion around current $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{b_i} = \mathbf{0}$:

$$Y_{ij} \approx h(\boldsymbol{x}'_{ij}\widehat{\boldsymbol{\beta}}) + h'(\boldsymbol{x}'_{ij}\widehat{\boldsymbol{\beta}})\boldsymbol{x}'_{ij}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + h'(\boldsymbol{x}'_{ij}\widehat{\boldsymbol{\beta}})\boldsymbol{z}'_{ij}\boldsymbol{b_i} + \varepsilon_{ij}$$

$$\approx \widehat{\mu}_{ij} + v(\widehat{\mu}_{ij})\boldsymbol{x}'_{ij}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + v(\widehat{\mu}_{ij})\boldsymbol{z}'_{ij}\boldsymbol{b_i} + \varepsilon_{ij}$$

- In vector notation: $\boldsymbol{Y_i} \approx \widehat{\boldsymbol{\mu}}_i + \widehat{V}_i X_i(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + \widehat{V}_i Z_i \boldsymbol{b_i} + \boldsymbol{\varepsilon_i}$

- Re-ordering terms yields:

$$\boldsymbol{Y_i}^* \equiv \widehat{V}_i^{-1}(\boldsymbol{Y_i} - \widehat{\boldsymbol{\mu}}_i) + X_i\widehat{\boldsymbol{\beta}} \approx X_i\boldsymbol{\beta} + Z_i\boldsymbol{b_i} + \boldsymbol{\varepsilon_i}^*$$

- Model fitting by iterating between updating the pseudo responses $\boldsymbol{Y_i}^*$ and fitting the above linear mixed model to them.

6.4.4 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

6.5 Approximation of Integral

- The likelihood contribution of every subject is of the form

$$\int f(z)\phi(z)dz$$

  where $\phi(z)$ is the density of the (multivariate) normal distribution

- Gaussian quadrature methods replace the integral by a weighted sum:

$$\int f(z)\phi(z)dz \approx \sum_{q=1}^{Q} w_q f(z_q)$$

- Q is the order of the approximation. The higher Q the more accurate the approximation will be

- The nodes (or quadrature points) z_q are solutions to the Qth order Hermite polynomial

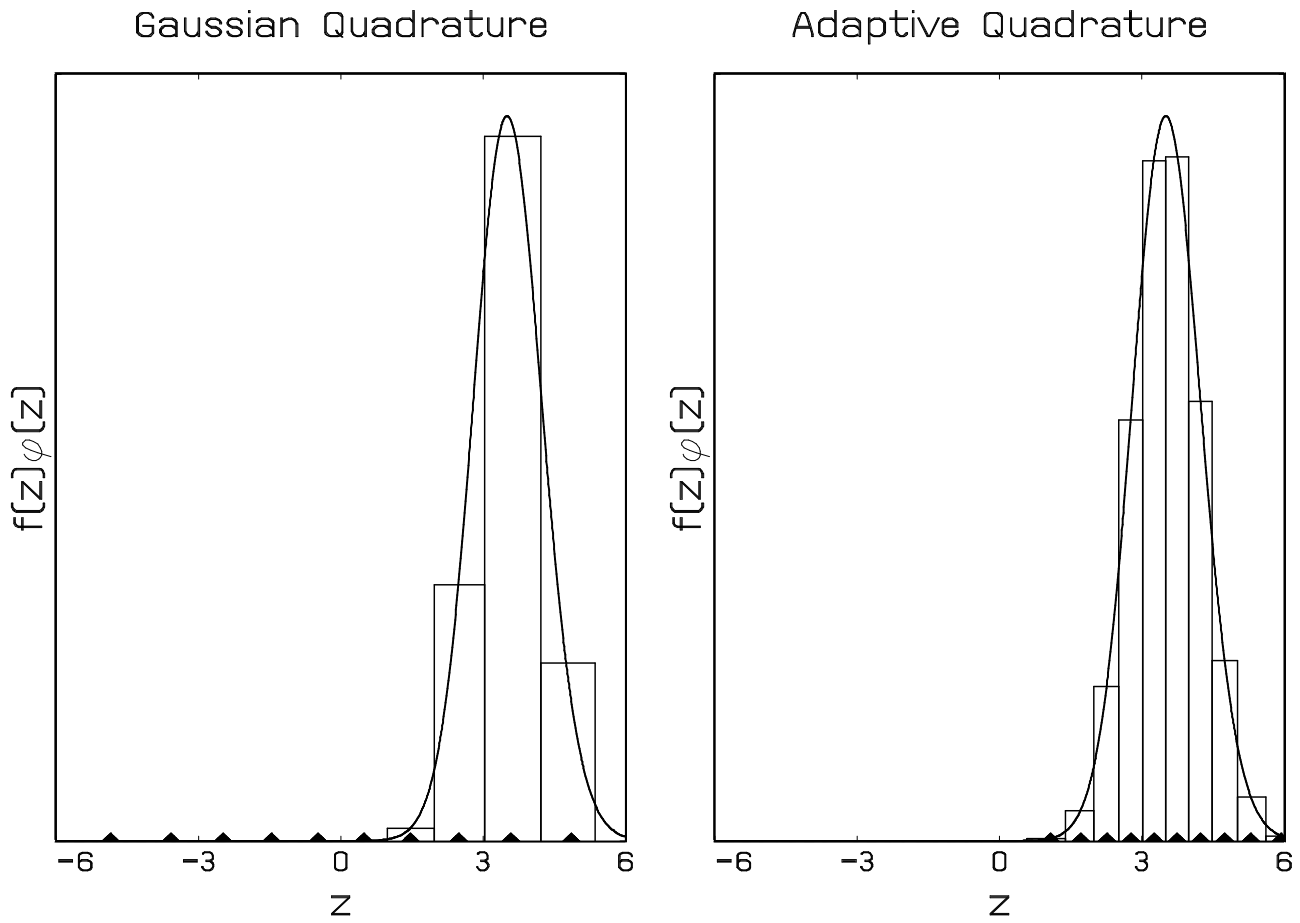
- The w_q are well-chosen weights

- The nodes z_q and weights w_q are reported in tables. Alternatively, an algorithm is available for calculating all z_q and w_q for any value Q.

- With **Gaussian quadrature**, the nodes and weights are fixed, independent of $f(z)\phi(z)$.

- With **adaptive Gaussian quadrature**, the nodes and weights are adapted to the 'support' of $f(z)\phi(z)$.

- Graphically ($Q = 10$):



- 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.

6.6 Example: Toenail Data

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

$$Y_{ij}|b_i \sim \text{Bernoulli}(\pi_{ij}), \qquad \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + b_i + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij}$$

- Notation:
  - ▷ T_i: treatment indicator for subject i
  - ▷ t_{ij}: time point at which jth measurement is taken for ith subject
- Adaptive as well as non-adaptive Gaussian quadrature, for various Q.

- Results:

	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)
σ	2.26 (0.12)	3.09 (0.21)	4.53 (0.39)	3.86 (0.33)	4.04 (0.39)
-2ℓ	1344.1	1259.6	1254.4	1249.6	1247.7
	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)
σ	4.51 (0.62)	3.70 (0.34)	4.07 (0.43)	4.01 (0.38)	4.02 (0.38)
-2ℓ	1259.1	1257.1	1248.2	1247.8	1247.8

- Conclusions:
  - ▷ (Log-)likelihoods are not comparable
  - ▷ 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 fitting algorithms:
  - ▷ Adaptive Gaussian Quadrature, $Q = 50$
  - ▷ MQL and PQL

- Summary of results:

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)

- Severe differences between QUAD (gold standard ?) and MQL/PQL.

- MQL/PQL may yield (very) biased results, especially for binary data.

Chapter 7
Fitting GLMM's in SAS

- ▷ Overview
- ▷ Example: Toenail data

7.1 Overview

- **GLIMMIX:** Laplace, MQL, PQL, adaptive quadrature
- **NLMIXED:** Adaptive and non-adaptive quadrature
  $\longrightarrow$ not discussed here

7.2 Example: Toenail data

- Re-consider logistic model with random intercepts for toenail data
- SAS code (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;
```

- MQL obtained with option 'method=RMPL'
- Laplace obtained with option 'method=LAPLACE'

- Adaptive quadrature with option 'method=QUAD(qpoints=5)'

- Inclusion of random slopes:

```
random intercept time / subject=idnum type=un;
```

Part IV

Marginal Versus Hierarchical Models

Chapter 8
Marginal Versus Random-effects Models

- ▷ Interpretation of GLMM parameters
- ▷ Marginalization of GLMM
- ▷ Example: Toenail data

8.1 Interpretation of GLMM Parameters: Toenail Data

- We compare our GLMM results for the toenail data with those from fitting GEE's (unstructured working correlation):

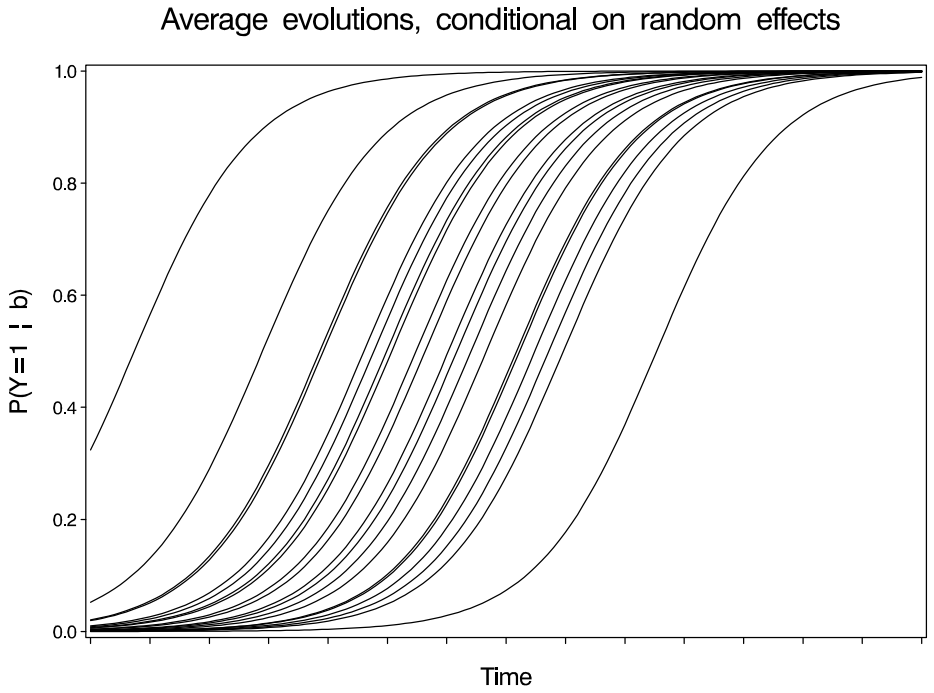
	GLMM	GEE
Parameter	Estimate (s.e.)	Estimate (s.e.)
Intercept group A	−1.6308 (0.4356)	−0.7219 (0.1656)
Intercept group B	−1.7454 (0.4478)	−0.6493 (0.1671)
Slope group A	−0.4043 (0.0460)	−0.1409 (0.0277)
Slope group B	−0.5657 (0.0601)	−0.2548 (0.0380)

- The strong differences can be explained as follows:
  - ▷ Consider the following GLMM:

$$Y_{ij}|b_i \sim \text{Bernoulli}(\pi_{ij}), \qquad \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + b_i + \beta_1 t_{ij}$$

  - ▷ The conditional means $E(Y_{ij}|b_i)$, as functions of t_{ij}, are given by

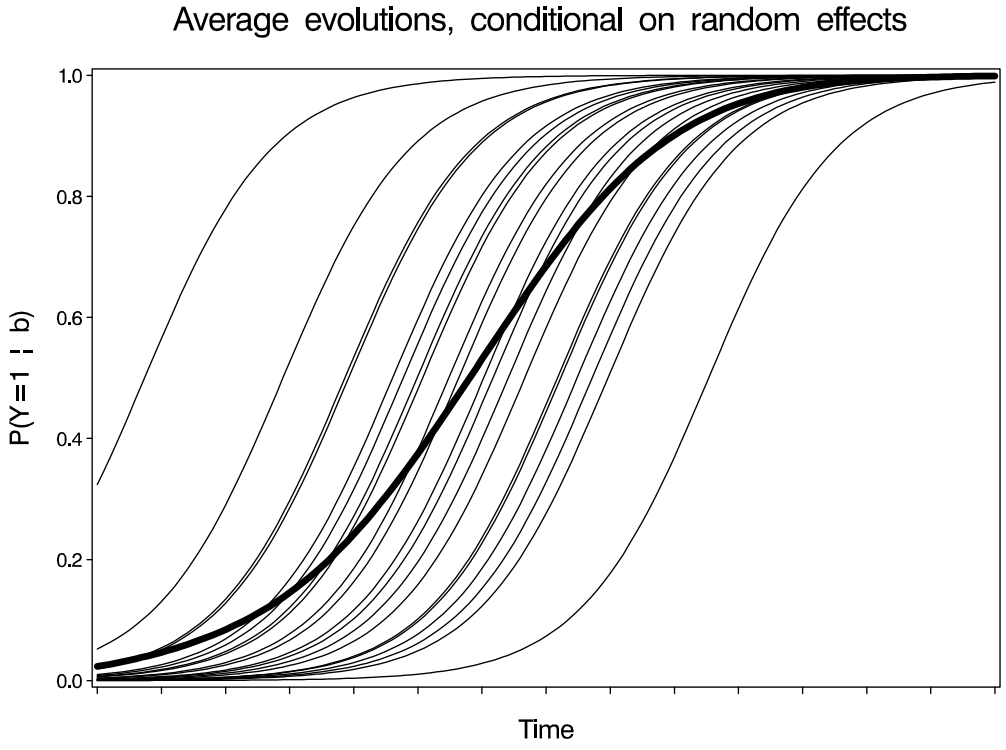
$$E(Y_{ij}|b_i) = \frac{\exp(\beta_0 + b_i + \beta_1 t_{ij})}{1 + \exp(\beta_0 + b_i + \beta_1 t_{ij})}$$



▷ The marginal average evolution is now obtained from averaging over the random effects:

$$E(Y_{ij}) = E[E(Y_{ij}|b_i)] = E\left[\frac{\exp(\beta_0 + b_i + \beta_1 t_{ij})}{1 + \exp(\beta_0 + b_i + \beta_1 t_{ij})}\right]$$

$$\neq \frac{\exp(\beta_0 + \beta_1 t_{ij})}{1 + \exp(\beta_0 + \beta_1 t_{ij})}$$



- Hence, the parameter vector $\boldsymbol{\beta}$ in the GEE model needs to be interpreted completely different from the parameter vector $\boldsymbol{\beta}$ in the GLMM:
  - ▷ GEE: marginal interpretation
  - ▷ GLMM: conditional interpretation, conditionally upon level of random effects

- In general, the model for the marginal average is not of the same parametric form as the conditional average in the GLMM.

- For logistic mixed models, with normally distributed random random intercepts, it can be shown that the marginal model can be well approximated by again a logistic model, but with parameters approximately satisfying

$$\frac{\widehat{\boldsymbol{\beta}}^{\mathsf{RE}}}{\widehat{\boldsymbol{\beta}}^{\mathsf{M}}} = \sqrt{c^2\sigma^2 + 1} > 1, \qquad \sigma^2 = \text{variance random intercepts}$$

$$c = 16\sqrt{3}/(15\pi)$$

- For the toenail application, σ was estimated as 4.0164, such that the ratio equals $\sqrt{c^2\sigma^2 + 1} = 2.5649$.

- The ratio's between the GLMM and GEE estimates are:

	GLMM	GEE	
Parameter	Estimate (s.e.)	Estimate (s.e.)	Ratio
Intercept group A	−1.6308 (0.4356)	−0.7219 (0.1656)	2.2590
Intercept group B	−1.7454 (0.4478)	−0.6493 (0.1671)	2.6881
Slope group A	−0.4043 (0.0460)	−0.1409 (0.0277)	2.8694
Slope group B	−0.5657 (0.0601)	−0.2548 (0.0380)	2.2202

- Note that this problem does not occur in linear mixed models:
  - ▷ Conditional mean: $E(\boldsymbol{Y_i}|\boldsymbol{b}_i) = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i$
  - ▷ Specifically: $E(\boldsymbol{Y_i}|\boldsymbol{b}_i = \boldsymbol{0}) = X_i\boldsymbol{\beta}$
  - ▷ Marginal mean: $E(\boldsymbol{Y_i}) = X_i\boldsymbol{\beta}$

- The problem arises from the fact that, in general,

$$E[g(Y)] \neq g[E(Y)]$$

- So, whenever the random effects enter the conditional mean in a non-linear way, the regression parameters in the marginal model need to be interpreted differently from the regression parameters in the mixed model.

- In practice, the marginal mean can be derived from the GLMM output by integrating out the random effects.

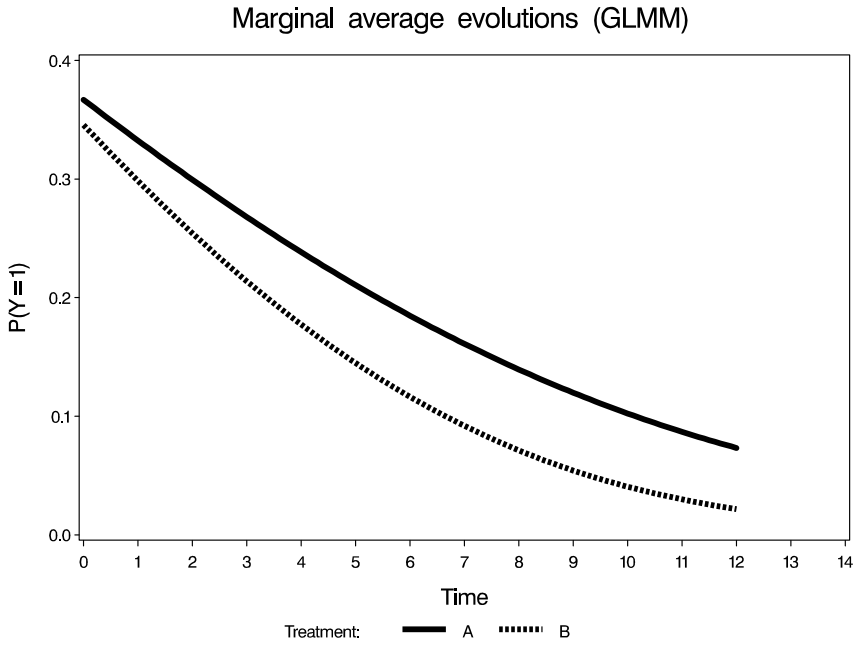
- This can be done numerically via Gaussian quadrature, or based on sampling methods.

8.2 Marginalization of GLMM: Toenail Data

- As an example, we plot the average evolutions based on the GLMM output obtained in the toenail example:

$$P(Y_{ij}=1)$$

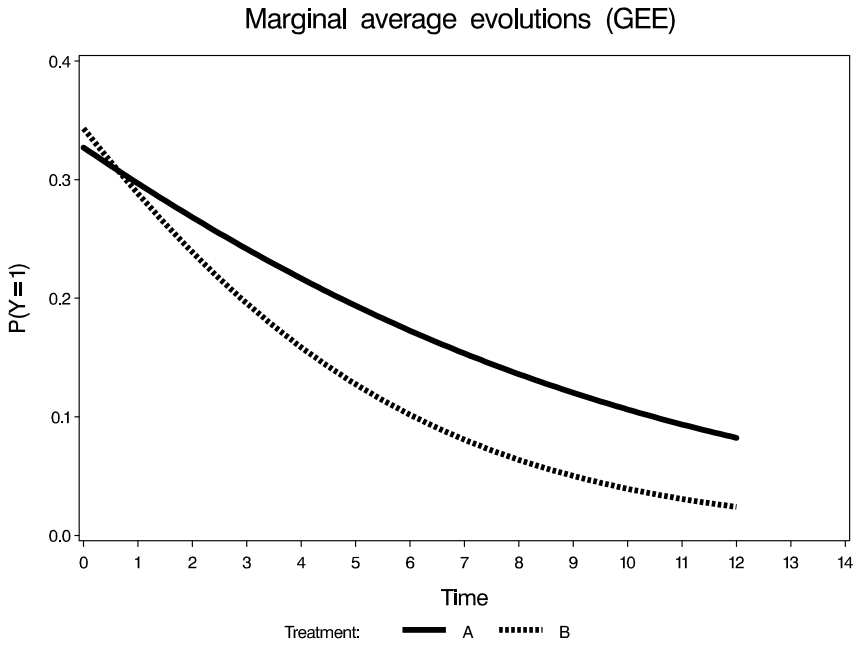
$$= \begin{cases} E\left[\dfrac{\exp(-1.6308+b_i-0.4043t_{ij})}{1+\exp(-1.6308+b_i-0.4043t_{ij})}\right], \\ \\ E\left[\dfrac{\exp(-1.7454+b_i-0.5657t_{ij})}{1+\exp(-1.7454+b_i-0.5657t_{ij})}\right], \end{cases}$$



- Average evolutions obtained from the GEE analyses:

$$P(Y_{ij} = 1)$$

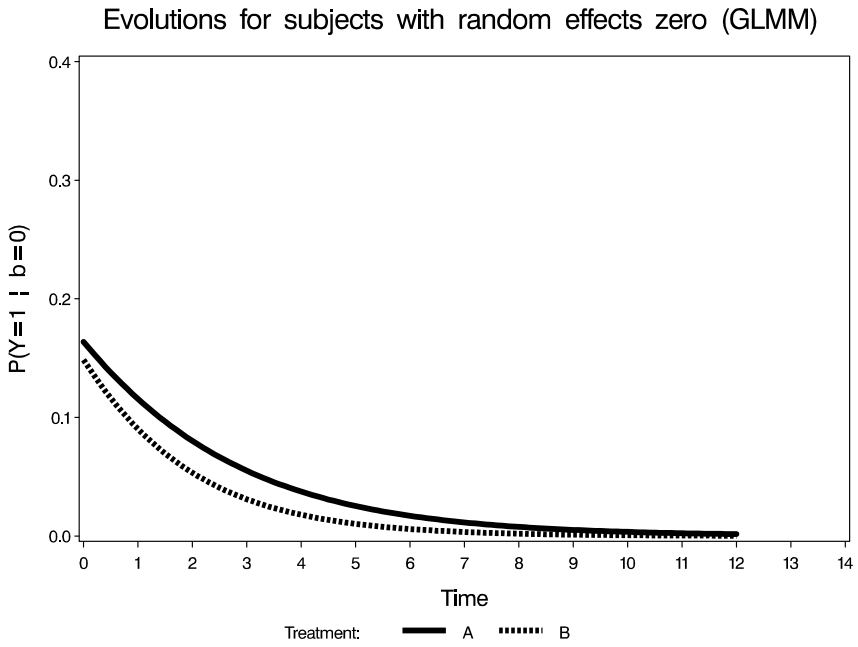
$$= \begin{cases} \dfrac{\exp(-0.7219 - 0.1409 t_{ij})}{1 + \exp(-0.7219 - 0.1409 t_{ij})} \\ \\ \dfrac{\exp(-0.6493 - 0.2548 t_{ij})}{1 + \exp(-0.6493 - 0.2548 t_{ij})} \end{cases}$$



- In a GLMM context, rather than plotting the marginal averages, one can also plot the profile for an 'average' subject, i.e., a subject with random effect $\boldsymbol{b}_i = \mathbf{0}$:

$$P(Y_{ij} = 1 | b_i = 0)$$

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



8.3 Example: Toenail Data Revisited

- Overview of all analyses on toenail data:

Parameter	QUAD	PQL	MQL	GEE
Intercept group A	−1.63 (0.44)	−0.72 (0.24)	−0.56 (0.17)	−0.72 (0.17)
Intercept group B	−1.75 (0.45)	−0.72 (0.24)	−0.53 (0.17)	−0.65 (0.17)
Slope group A	−0.40 (0.05)	−0.29 (0.03)	−0.17 (0.02)	−0.14 (0.03)
Slope group B	−0.57 (0.06)	−0.40 (0.04)	−0.26 (0.03)	−0.25 (0.04)
Var. random intercepts (τ^2)	15.99 (3.02)	4.71 (0.60)	2.49 (0.29)	

- Conclusion:

$$|\text{GEE}| < |\text{MQL}| < |\text{PQL}| < |\text{QUAD}|$$

Part V

Non-linear Models

Chapter 9
Non-Linear Mixed Models

- From linear to non-linear models
- Orange Tree Example
- Song Bird Example

9.1 From Linear to Non-linear Models

- We have studied:
  - ▷ linear models $\longleftrightarrow$ generalized linear models
  - ▷ for Gaussian data $\longleftrightarrow$ non-Gaussian data
  - ▷ for cross-sectional (univariate) data $\longleftrightarrow$ for correlated data (clustered data, multivariate data, longitudinal data)

- In all cases, a certain amount of linearity is preserved:
  - ▷ E.g., in generalized linear models, linearity operates at the level of the linear predictor, describing a transformation of the mean (logit, log, probit,...)

- This implies that all predictor functions are linear in the parameters:

$$\beta_0 + \beta_1 x_{1i} + \ldots \beta_p x_{pi}$$

- This may still be considered a limiting factor, and non-linearity in the parametric functions may be desirable.

- Examples:
  - ▷ Certain growth curves, especially when they include both a growth spurt as well as asymptote behavior towards the end of growth
  - ▷ Dose-response modeling
  - ▷ Pharmacokinetic and pharmacodynamic models

9.2 LMM and GLMM

- In linear mixed models, the mean is modeled as a linear function of regression parameters and random effects:

$$E(Y_{ij}|b_i) = \boldsymbol{x_{ij}}'\boldsymbol{\beta} + \boldsymbol{z_{ij}}'\boldsymbol{b}_i$$

- In generalized linear mixed models, apart from a link function, the mean is again modeled as a linear function of regression parameters and random effects:

$$E(Y_{ij}|b_i) = h(\boldsymbol{x_{ij}}'\boldsymbol{\beta} + \boldsymbol{z_{ij}}'\boldsymbol{b}_i)$$

- In some applications, models are needed, in which the mean is no longer modeled as a function of a linear predictor $\boldsymbol{x_{ij}}'\boldsymbol{\beta} + \boldsymbol{z_{ij}}'\boldsymbol{b}_i$. These are called non-linear mixed models.

9.3 Non-linear Mixed Models

- In non-linear mixed models, it is assumed that the conditional distribution of Y_{ij}, given $\boldsymbol{b}_i$ is belongs to the exponential family (Normal, Binomial, Poisson,...).

- The mean is modeled as:

$$E(Y_{ij}|b_i) \;=\; h(\boldsymbol{x_{ij}}, \boldsymbol{\beta}, \boldsymbol{z_{ij}}, \boldsymbol{b}_i)$$

- As before, the random effects are assumed to be normally distributed, with mean $\mathbf{0}$ and covariance D.

- Let $f_{ij}(y_{ij}|\boldsymbol{b}_i, \boldsymbol{\beta}, \phi)$ be the conditional density of Y_{ij} given $\boldsymbol{b}_i$, and let $f(\boldsymbol{b}_i|D)$ be the density of the $N(\mathbf{0}, D)$ distribution.

- Under conditional independence, the marginal likelihood equals

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

$$= \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} f_{ij}(y_{ij}|\boldsymbol{b}_i, \boldsymbol{\beta}, \phi)\ f(\boldsymbol{b}_i|D)\ d\boldsymbol{b}_i$$

- The above likelihood is of the same form as the likelihood obtained earlier for a generalized linear mixed model.

- As before, numerical quadrature is used to approximate the integral

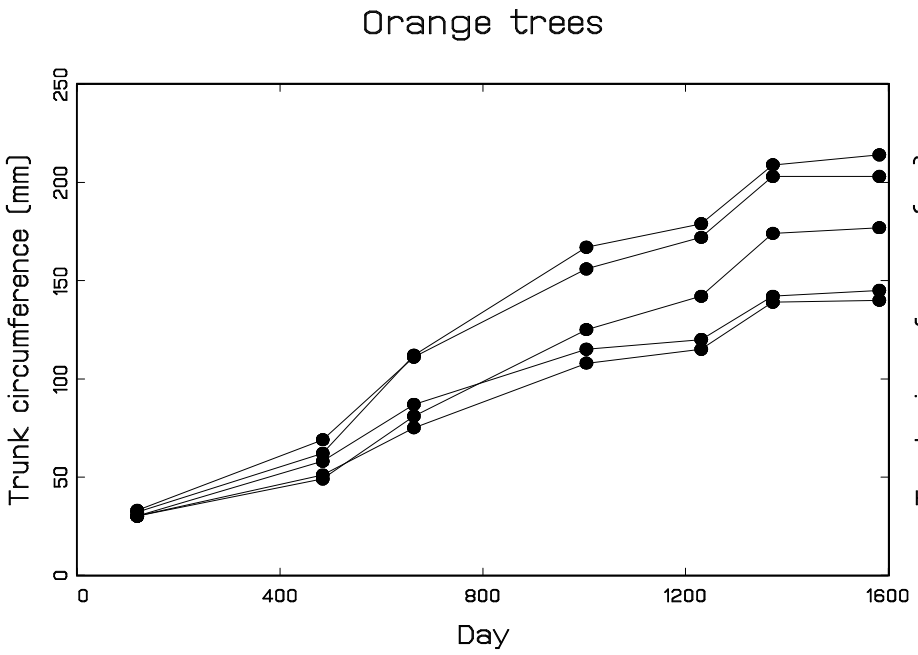
- Non-linear mixed models can also be fitted within the SAS procedure NLMIXED.

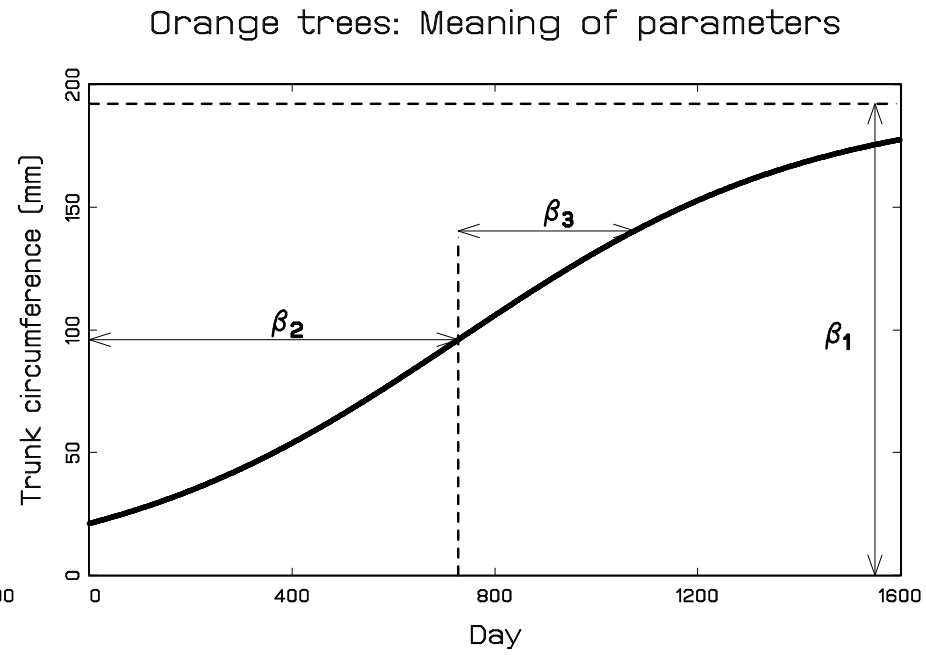
9.4 Example: Orange Trees

- We consider an experiment in which the trunk circumference (in mm) is measured for 5 orange trees, on 7 different occasions.

- Data:

	Response				
Day	Tree 1	Tree 2	Tree 3	Tree 4	Tree 5
118	30	33	30	32	30
484	58	69	51	62	49
664	87	111	75	112	81
1004	115	156	108	167	125
1231	120	172	115	179	142
1372	142	203	139	209	174
1582	145	203	140	214	177





- The following non-linear mixed model has been proposed:

$$Y_{ij} = \frac{\beta_1 + b_i}{1 + \exp[-(t_{ij} - \beta_2)/\beta_3]} + \varepsilon_{ij}$$

$$b_i \sim N(0, \sigma_b^2) \qquad \varepsilon_{ij} \sim N(0, \sigma^2)$$

- In SAS PROC NLMIXED, the model can be fitted as follows:

```
proc nlmixed data=tree;
parms beta1=190 beta2=700 beta3=350
      sigmab=10 sigma=10;
num = b;
ex = exp(-(day-beta2)/beta3);
den = 1 + ex;
model y ~ normal(num/den,sigma**2);
random b ~ normal(beta1,sigmab**2) subject=tree;
run;
```

- An equivalent program is given by

```
proc nlmixed data=tree;
parms beta1=190 beta2=700 beta3=350
      sigmab=10 sigma=10;
num = b + beta1;
ex = exp(-(day-beta2)/beta3);
den = 1 + ex;
model y ~ normal(num/den,sigma**2);
random b ~ normal(0,sigmab**2) subject=tree;
run;
```

- Selected output:

Specifications

Data Set	WORK.TREE
Dependent Variable	y
Distribution for Dependent Variable	Normal
Random Effects	b
Distribution for Random Effects	Normal
Subject Variable	tree
Optimization Technique	Dual Quasi-Newton
Integration Method	Adaptive Gaussian Quadrature

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|
beta1	192.05	15.6577	4	12.27	0.0003
beta2	727.91	35.2487	4	20.65	<.0001
beta3	348.07	27.0798	4	12.85	0.0002
sigmab	31.6463	10.2614	4	3.08	0.0368
sigma	7.8430	1.0125	4	7.75	0.0015

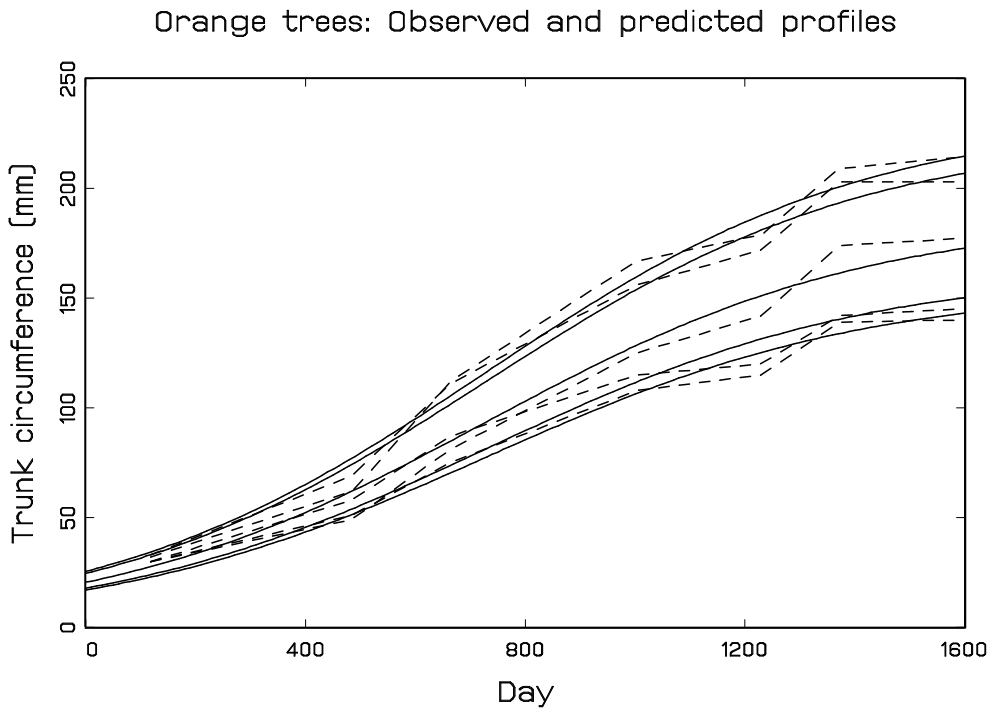
- Note that the number of quadrature points was selected adaptively to be equal to only one. Refitting the model with 50 quadrature points yielded identical results.

- Empirical Bayes estimates, and subject-specific predictions can be obtained as follows:

```
proc nlmixed data=tree;
parms beta1=190 beta2=700 beta3=350
      sigmab=10 sigma=10;
num = b + beta1;
ex = exp(-(day-beta2)/beta3);
den = 1 + ex;
ratio=num/den;
model y ~ normal(ratio,sigma**2);
random b ~ normal(0,sigmab**2) subject=tree out=eb;
predict ratio out=ratio;
run;
```

- We can now compare the observed data to the subject-specific predictions

$$\widehat{y}_{ij} = \frac{\widehat{\beta}_1 + \widehat{b}_i}{1 + \exp[-(t_{ij} - \widehat{\beta}_2)/\widehat{\beta}_3]}$$



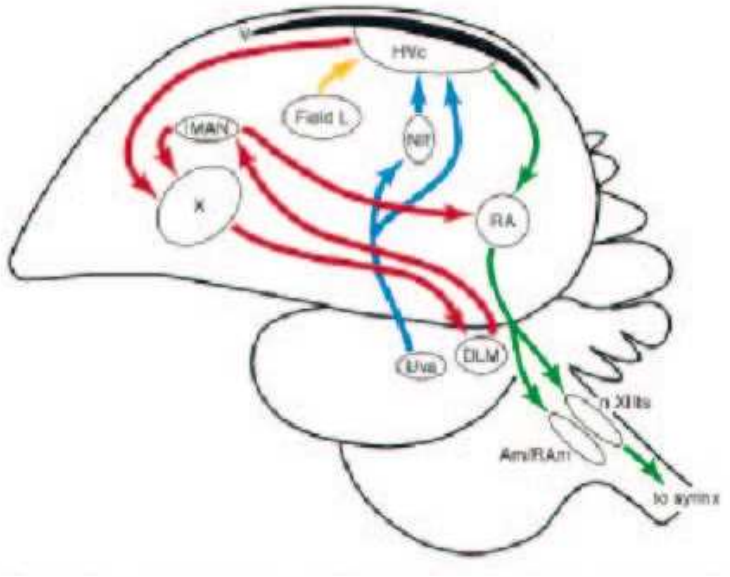
9.5 Example: Song Birds

- At the University of Antwerp, a novel in-vivo MRI approach to discern the functional characteristics of specific neuronal populations in a strongly connected brain circuitry has been established.

- Of particular interest: the song control system (SCS) in songbird brain.

- The high vocal center (HVC), one of the major nuclei in this circuit, contains interneurons and two distinct types of neurons projecting respectively to the nucleus robustus archistriatalis, RA or to area X.

- Schematically,



9.5.1 The MRI Data

- T1-weighted multi slice SE images were acquired (Van Meir *et al* 2003).
- After acquisition of a set of control images, $MnCl_2$ was injected in the cannulated HVC.
- Two sets of 5 coronal slices (one through HVC and RA, one through area X) were acquired every 15 min for up to 6–7 hours after injection.
- This results in 30 data sets of 10 slices of each bird (5 controls and 5 testosterone treated).
- The change of relative SI of each time point is calculated from the mean signal intensity determined within the defined regions of interest (area X and RA) and in an adjacent control area.

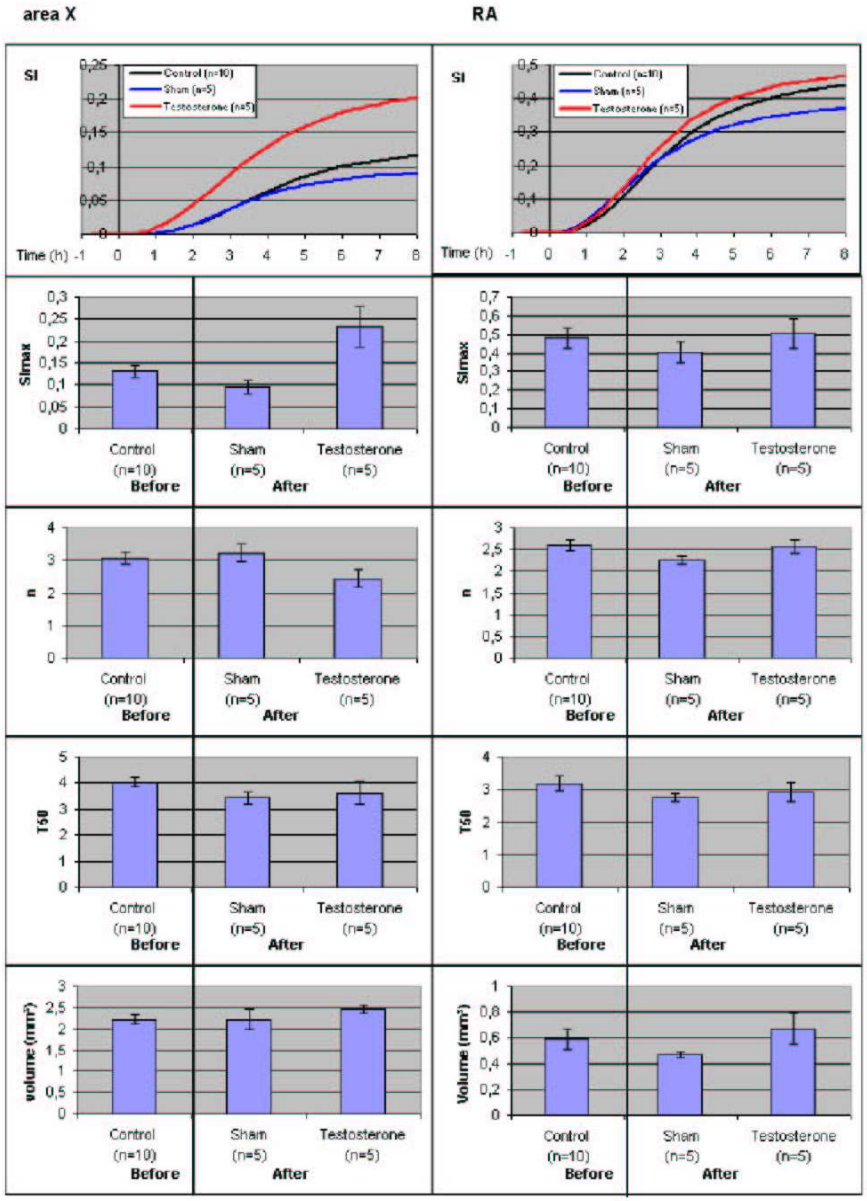
9.5.2 Initial Study

- The effect of testosterone (T) on the dynamics of Mn^{2+} accumulation in RA and area X of female starling has been established.

- This has been done with dynamic ME-MRI in individual birds injected with Manganese in their HVC.

- This was done in a **2-stage approach:** The individual SI data, determined as the mean intensity of a region of interest, were submitted to a sigmoid curve fitting providing curve parameters which revealed upon a two-way repeated ANOVA analysis that neurons projecting to RA and those to area X were affected differently by testosterone treatment.

- However, this approach could be less reliable if the fit-parameters were mutually dependent.


area X
RA
SI
Time (h)
Control (n=10)
Sham (n=5)
Testosterone (n=5)
Smax
n
T50
volume (mm³)
Volume (mm³)
Control (n=10)
Sham (n=5)
Testosterone (n=5)
Before
After

- Thus: an integrated non-linear modeling approach is necessary: the MRI signal intensities (SI) need to be fitted to a non-linear mixed model assuming that the SI follow a pre-specified distribution depending on a covariate indicating the time after $MnCl_2$ injection and parameterized through fixed and random (bird-specific) effects.

- An initial model needs to be proposed and then subsequently simplified.

9.5.3 A Model for RA in the Second Period

- Let RA_{ij} be the measurement at time j for bird i.
- The following initial non-linear model is assumed:

$$RA_{ij} = \frac{(\phi_0 + \phi_1 G_i + f_i) T_{ij}^{\eta_0+\eta_1 G_i+n_i}}{(\tau_0 + \tau_1 G_i + t_i)^{\eta_0+\eta_1 G_i+n_i} + T_{ij}^{\eta_0+\eta_1 G_i+n_i}} + \gamma_0 + \gamma_1 G_i + \varepsilon_{ij}$$

- The following conventions are used:
  - ▷ G_i is an indicator for group membership (0 for control, 1 for active)
  - ▷ T_{ij} is a covariate indicating the measurement time
  - ▷ The fixed effects parameters:

  - ∗ the "intercept" parameters ϕ_0, η_0, τ_0, and γ_0
  - ∗ the group effect parameters ϕ_1, η_1, τ_1, and γ_1
  - ∗ If the latter are simultaneously zero, there is no difference between both groups. If at least one of them is (significantly) different from zero, then the model indicates a difference between both groups.

- ▷ There are three bird-specific or random effects, f_i, n_i, and t_i, following a trivariate zero-mean normal and general covariance matrix D
- ▷ The residual error terms ε_{ij} are assumed to be mutually independent and independent from the random effects, and to follow a $N(0, \sigma^2)$ distribution.

- Thus, the general form of the model has 8 fixed-effects parameters, and 7 variance components (3 variances in D, 3 covariances in D, and σ^2).

- SAS program:

```
data hulp2;
set m.vincent03;
if time <= 0 then delete;
if periode = 1 then delete;
run;

proc nlmixed data=hulp2 qpoints=3;
parms phim=0.64 phimdiff=0 eta=1.88 etadiff=0  tau=3.68
      taudiff=0 gamma=-0.01 gdiff=0
      d11=0.01 sigma2=0.01 d22=0.01 d33=0.01;
teller = (phim + phimdiff * groep + vm ) *
         (time ** (eta + etadiff * groep + n ));
noemer = ((tau + t + taudiff * groep )
         ** (eta + etadiff * groep +n) )
         + (time ** (eta + etadiff * groep + n));
gemid = teller/noemer + gamma + gdiff * groep;
model si_ra ~ normal(gemid,sigma2);
random vm t n ~ normal([0, 0, 0],[d11,0,d22, 0, 0, d33])
       subject=vogel out=m.eb;
run;
```

- Gaussian quadrature is used.

- The covariances in D are set equal to zero, due to computational difficulty.

- Fitting the model produces the following output.

- First, the bookkeeping information is provided:

Specifications

Data Set	WORK.HULP2
Dependent Variable	SI_RA
Distribution for Dependent Variable	Normal
Random Effects	vm t n
Distribution for Random Effects	Normal
Subject Variable	vogel
Optimization Technique	Dual Quasi-Newton
Integration Method	Adaptive Gaussian Quadrature

Dimensions

Observations Used	262
Observations Not Used	58
Total Observations	320
Subjects	10
Max Obs Per Subject	28
Parameters	12
Quadrature Points	3

- Next, starting values and the corresponding likelihood are provided:

Parameters

phim	phimdiff	eta	etadiff	tau	taudiff	gamma	gdiff	d11
0.64	0	1.88	0	3.68	0	-0.01	0	0.01

sigma2	d22	d33	NegLogLike
0.01	0.01	0.01	-325.62539

- The iteration history tells convergence is not straightforward:

Iteration History

Iter	Calls	NegLogLike	Diff	MaxGrad	Slope
1	14	-571.60061	245.9752	346172	-1341908
2	18	-590.08646	18.48585	25374.59	-3290.57
3	19	-617.20333	27.11687	23321.39	-61.3569
4	21	-630.54411	13.34079	248131.3	-15.4346
5	22	-640.39285	9.848737	7211.013	-21.8031
6	23	-656.41057	16.01772	12295.64	-10.1471
7	25	-666.68301	10.27244	96575.61	-11.8592
8	26	-670.60017	3.917164	100677.4	-5.38138
9	27	-676.45628	5.856102	30099.22	-6.11815
10	30	-677.36448	0.908204	123312.7	-1.98302
20	53	-688.60175	4.468022	32724.37	-4.48668

```
30        74    -695.15573    0.716814    43575.75    -0.18929

40        93    -697.90828    0.303986    1094.634    -0.51907

50       112    -699.51124    0.002001    166.3506    -0.00338

60       137    -700.79832    0.002657    710.6714    -0.00044

70       157    -701.17648    0.000123    20.20774    -0.00018
71       159    -701.17688    0.000396    125.3658    -0.00006
72       162    -701.21656    0.039676    1426.738     -0.0008
73       164    -701.31217    0.095616     136.553    -0.05932
74       166    -701.31463    0.002454    98.78744    -0.00443
75       168     -701.3147    0.000071    3.915711    -0.00015
76       170     -701.3147    9.862E-7    1.290999    -2.05E-6

        NOTE: GCONV convergence criterion satisfied.
```

- The fit statistics can be used for model comparison, as always:

Fit Statistics

-2 Log Likelihood	-1403
AIC (smaller is better)	-1379
AICC (smaller is better)	-1377
BIC (smaller is better)	-1375

- Finally, parameter estimates are provided:

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
phim	0.4085	0.06554	7	6.23	0.0004	0.05	0.2535	0.5634	0.000428
phimdiff	0.08167	0.09233	7	0.88	0.4058	0.05	-0.1366	0.3000	0.000591
eta	2.2117	0.1394	7	15.87	<.0001	0.05	1.8821	2.5412	-9.87E-6
etadiff	0.4390	0.1865	7	2.35	0.0508	0.05	-0.00206	0.8801	0.000717
tau	2.8006	0.2346	7	11.94	<.0001	0.05	2.2457	3.3554	-0.00028
taudiff	0.07546	0.3250	7	0.23	0.8231	0.05	-0.6932	0.8441	0.000046
gamma	0.000237	0.004391	7	0.05	0.9585	0.05	-0.01015	0.01062	0.008095
gdiff	0.001919	0.005923	7	0.32	0.7554	0.05	-0.01209	0.01592	-0.00244
d11	0.02059	0.009283	7	2.22	0.0620	0.05	-0.00136	0.04254	-0.00547
sigma2	0.000180	0.000017	7	10.66	<.0001	0.05	0.000140	0.000221	-1.291
d22	0.2400	0.1169	7	2.05	0.0791	0.05	-0.03633	0.5163	-0.00053
d33	0.02420	0.03488	7	0.69	0.5101	0.05	-0.05829	0.1067	-0.00059

- Next, a backward selection is conducted, using likelihood ratio tests.

- First, the variance d_{33} of n_i was removed. The corresponding test statistic has a $\chi^2_{0:1}$ null distribution ($p = 0.4387$).

- Next, fixed-effect parameters γ_0, γ_1, τ_1, and ϕ_1 are removed.

- The final model is:

$$RA_{ij} = \frac{(\phi_0 + f_i)T_{ij}^{\eta_0+\eta_1 G_i}}{(\tau_0 + t_i)^{\eta_0+\eta_1 G_i} + T_{ij}^{\eta_0+\eta_1 G_i}} + \varepsilon_{ij}$$

- Overview of parameter estimates and standard errors for initial and final model:

		Estimate (s.e.)	
Effect	Parameter	Initial	Final
	ϕ_0	0.4085(0.0655)	0.4526(0.0478)
	ϕ_1	0.0817(0.0923)	
	η_0	2.2117(0.1394)	2.1826(0.0802)
	η_1	0.4390(0.1865)	0.4285(0.1060)
	τ_0	2.8006(0.2346)	2.8480(0.1761)
	τ_1	0.0755(0.3250)	
	γ_0	0.00024(0.0044)	
	γ_1	0.0019(0.0059)	
Var(f_i)	d_{11}	0.0206(0.0092)	0.0225(0.0101)
Var(t_i)	d_{22}	0.2400(0.1169)	0.2881(0.1338)
Var(n_i)	d_{33}	0.0242(0.0349)	
Var(ε_{ij})	σ^2	0.00018(0.000017)	0.00019(0.000017)

9.5.4 AREA.X at the Second Period

- The same initial model will be fitted to $AREA_{ij}$.

- In this case, a fully general model can be fitted, including also the covariances between the random effects.

- Code to do so:

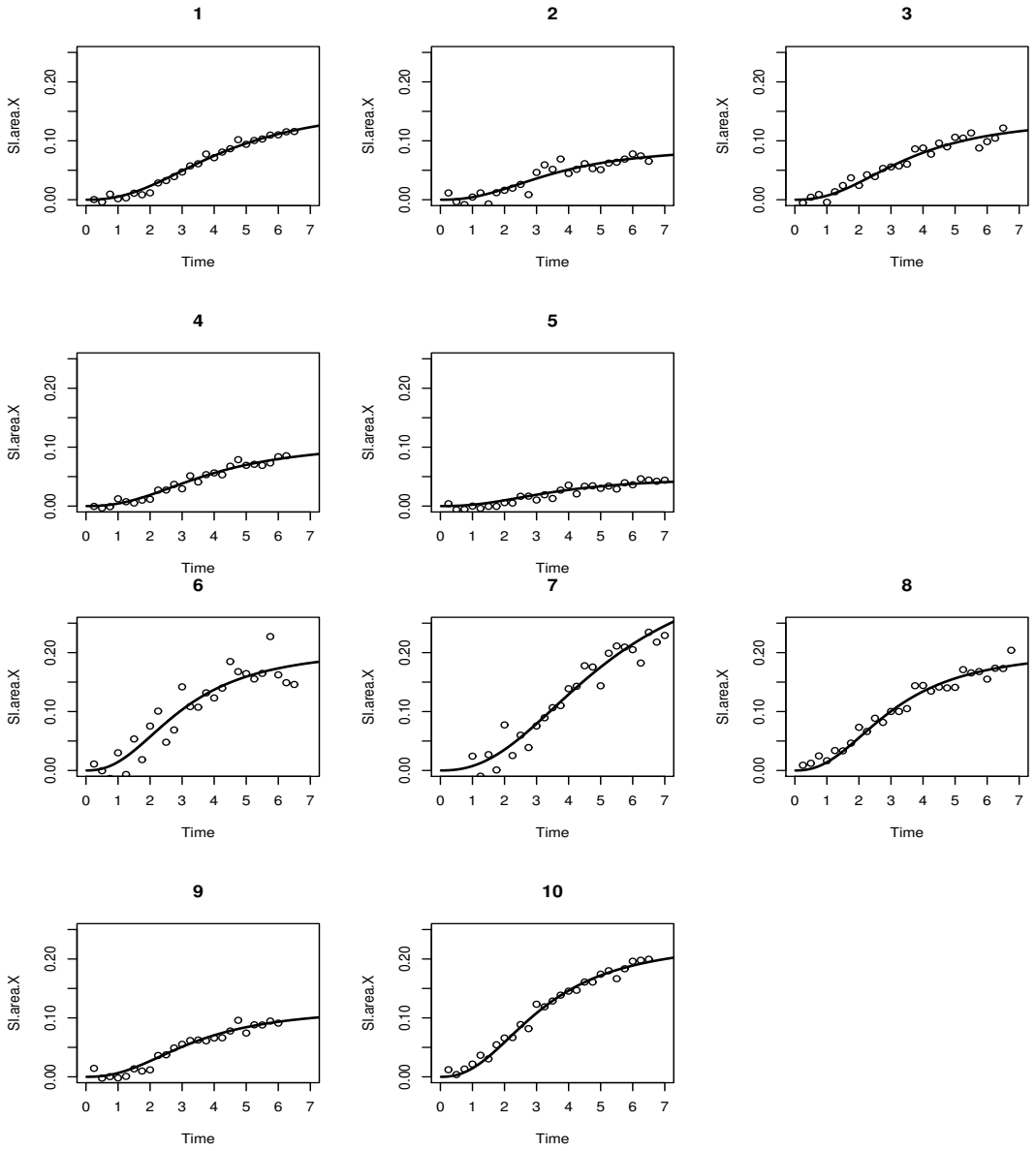
```
proc nlmixed data=hulp2 qpoints=3;
parms phim=0.1124 phimdiff=0.1200 eta=2.4158 etadiff=-0.1582 tau=3.8297
      taudiff=-0.05259 gamma=-0.00187 gdiff=0.002502 sigma2=0.000156
      d11=0.002667  d22=0.2793 d33=0.08505 d12=0 d13=0 d23=0;
teller = (phim + phimdiff * groep + vm)
         * (time ** (eta + etadiff * groep + n ));
noemer = ((tau + taudiff * groep +t )
         ** (eta + etadiff * groep + n) )
         + (time ** (eta + etadiff * groep +n));
gemid = teller/noemer + gamma + gdiff * groep;
model si_area_X ~ normal(gemid,sigma2);
random vm t n ~ normal([0, 0, 0],[d11, d12, d22, d13, d23, d33])
       subject=vogel out=m.eb2;
run;
```

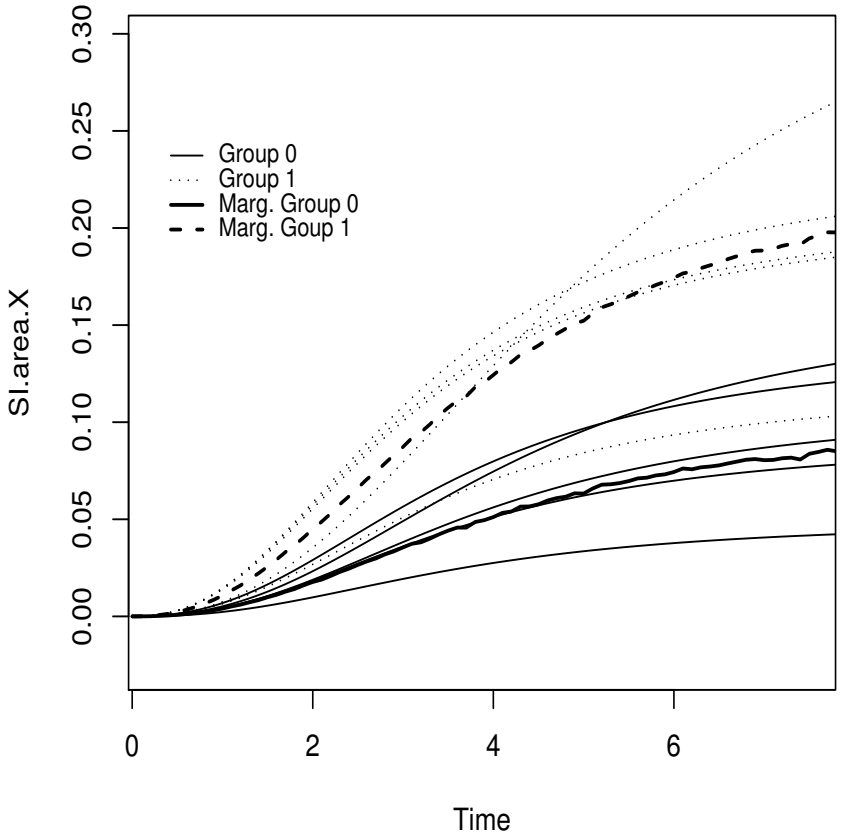
- Model simplification:
  - ▷ First, the random n_i effect is removed (implying the removal of d_{11}, d_{12}, and d_{13}), using a likelihood ratio test statistic with value 4.08 and null distribution $\chi^2_{2:3}$. The corresponding $p = 0.1914$.
  - ▷ Removal of the random t_i effect is not possible since the likelihood ratio equals 54.95 on 1:2 degrees of freedom ($p < 0.0001$).
  - ▷ Removal of the covariance between the random t_i and f_i effects is not possible ($G^2 = 4.35$ on 1 d.f., $p = 0.0371$).
  - ▷ Next, the following fixed-effect parameters were removed: γ_0, γ_1, η_1 and τ_1.
  - ▷ The fixed-effect ϕ_1 was found to be highly significant and therefore could not be removed from the model ($G^2 = 10.5773$ on 1 d.f., $p = 0.0011$). This indicates a significant difference between the two groups.
- The resulting final model is:

$$AREA_{ij} = \frac{(\phi_0 + \phi_1 G_i + f_i) T_{ij}^{\eta_0}}{(\tau_0 + t_i)^{\eta_0} + T_{ij}^{\eta_0}} + \varepsilon_{ij}.$$

Effect	Parameter	Estimate (s.e.)	
		Initial	Final
	ϕ_0	0.1118 (0.0333)	0.1035 (0.0261)
	ϕ_1	0.1116 (0.0458)	0.1331 (0.0312)
	η_0	2.4940 (0.5390)	2.3462 (0.1498)
	η_1	-0.0623 (0.5631)	
	τ_0	3.6614 (0.5662)	3.7264 (0.3262)
	τ_1	-0.1303 (0.6226)	
	γ_0	-0.0021 (0.0032)	
	γ_1	0.0029 (0.0048)	
Var(f_i)	d_{11}	0.0038 (0.0020)	0.004271 (0.0022)
Var(t_i)	d_{22}	0.2953 (0.2365)	0.5054 (0.2881)
Var(n_i)	d_{33}	0.1315 (0.1858)	
Cov(f_i, t_i)	d_{12}	0.0284 (0.0205)	0.03442 (0.0229)
Cov(f_i, n_i)	d_{13}	-0.0116 (0.0159)	
Cov(t_i, n_i)	d_{23}	-0.0095 (0.1615)	
Var(ε_{ij})	σ^2	0.00016 (0.000014)	0.00016 (0.000014)

Individual curves, showing data points as well as empirical Bayes predictions for the bird-specific curves, show that the sigmoidal curves describe the data quite well.





- We can also explore all individual as well as marginal average fitted curves per group. The marginal effect was obtained using the sampling-based method.

- It is clear that Area.X is higher for most treated birds (group 1) compared to the untreated birds (group 0).

Part VI

Incomplete Data

Chapter 10
Setting The Scene

- Orthodontic growth data
- Depression trial
- Age-related macular degeneration trial
- Notation
- Taxonomy

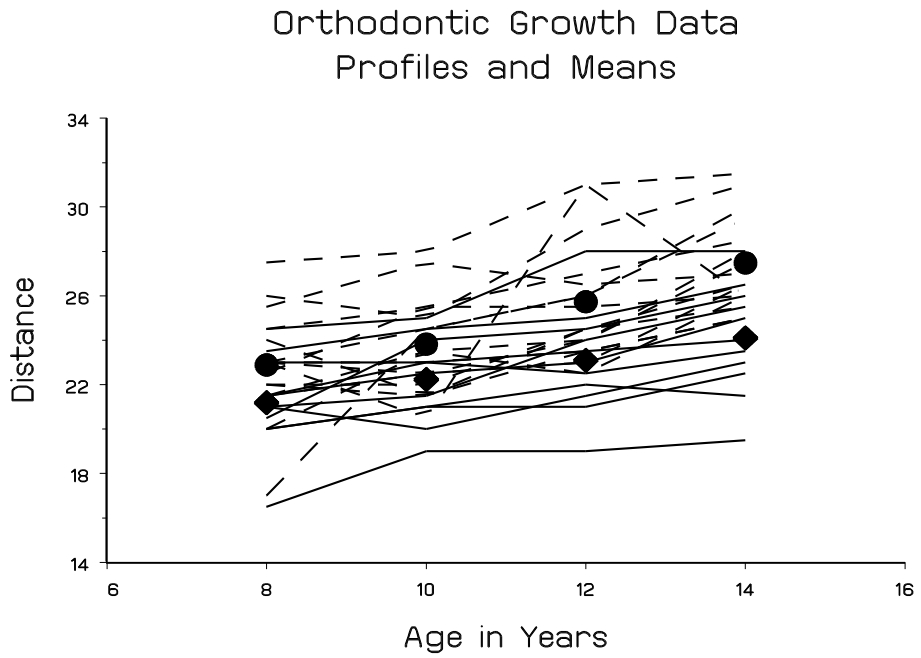
10.1 Growth Data

- Taken from Potthoff and Roy, Biometrika (1964)
- Research question:

> **Is dental growth related to gender ?**

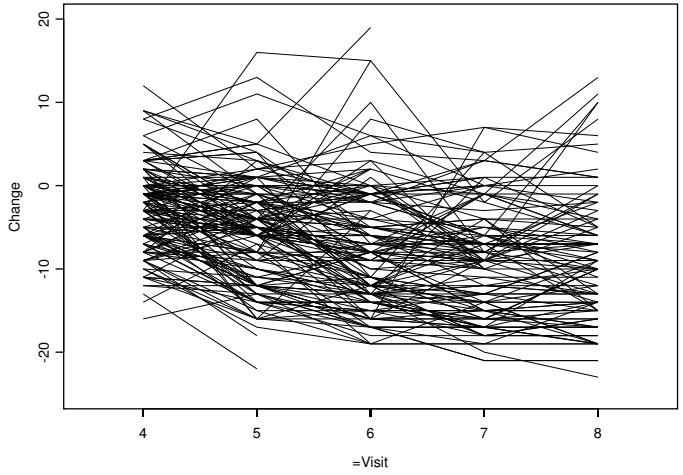
- The distance from the center of the pituitary to the maxillary fissure was recorded at ages 8, 10, 12, and 14, for 11 girls and 16 boys

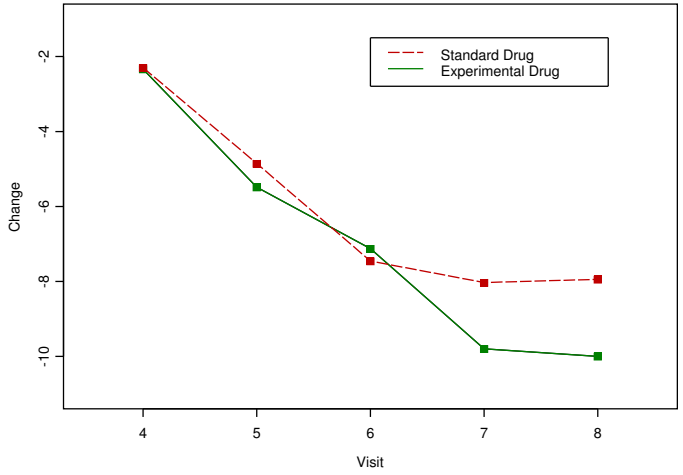
- Individual profiles:
  - ▷ Much variability between girls / boys
  - ▷ Considerable variability within girls / boys
  - ▷ Fixed number of measurements per subject
  - ▷ Measurements taken at fixed time points



10.2 The Depression Trial

- Clinical trial: experimental drug *versus* standard drug
- 170 patients
- Response: change versus baseline in $HAMD_{17}$ score
- 5 post-baseline visits: 4–8





10.3 Age-related Macular Degeneration Trial

- Pharmacological Therapy for Macular Degeneration Study Group (1997)
- An occular pressure disease which makes patients progressively lose vision
- 240 patients enrolled in a multi-center trial (190 completers)
- **Treatment:** Interferon-α (6 million units) versus placebo
- **Visits:** baseline and follow-up at 4, 12, 24, and 52 weeks
- **Continuous outcome: visual acuity:** # letters correctly read on a vision chart
- **Binary outcome:** visual acuity versus baseline ≥ 0 or ≤ 0

- Missingness:

Measurement occasion					
4 wks	12 wks	24 wks	52 wks	Number	%
Completers					
O	O	O	O	188	78.33
Dropouts					
O	O	O	M	24	10.00
O	O	M	M	8	3.33
O	M	M	M	6	2.50
M	M	M	M	6	2.50
Non-monotone missingness					
O	O	M	O	4	1.67
O	M	M	O	1	0.42
M	O	O	O	2	0.83
M	O	M	M	1	0.42

CRF	TRT	VISUAL0	VISUAL4	VISUAL12	VISUAL24	VISUAL52	lesion
1002	4	59	55	45	.	.	3
1003	4	65	70	65	65	55	1
1006	1	40	40	37	17	.	4
1007	1	67	64	64	64	68	2
1010	4	70	.	.	.	.	1
1110	4	59	53	52	53	42	3
1111	1	64	68	74	72	65	1
1112	1	39	37	43	37	37	3
1115	4	59	58	49	54	58	2
1803	1	49	51	71	71	.	1
1805	4	58	50	.	.	.	1

...

10.4 Notation

- Subject i at occasion (time) $j = 1, \ldots, n_i$
- **Measurement** Y_{ij}
- **Missingness indicator** $R_{ij} = \begin{cases} 1 & \text{if } Y_{ij} \text{ is observed,} \\ 0 & \text{otherwise.} \end{cases}$
- Group Y_{ij} into a vector $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})' = (\boldsymbol{Y}_i^o, \boldsymbol{Y}_i^m)$

  $\begin{cases} \boldsymbol{Y}_i^o & \text{contains } Y_{ij} \text{ for which } R_{ij} = 1, \\ \boldsymbol{Y}_i^m & \text{contains } Y_{ij} \text{ for which } R_{ij} = 0. \end{cases}$
- Group R_{ij} into a vector $\boldsymbol{R}_i = (R_{i1}, \ldots, R_{in_i})'$
- D_i: time of dropout: $D_i = 1 + \Sigma_{j=1}^{n_i} R_{ij}$

10.5 Direct Likelihood/Bayesian Inference: Ignorability

$$\boxed{\text{MAR}}: \; f(\boldsymbol{Y}_i^o | \boldsymbol{X}_i, \boldsymbol{\theta}) \; \cancel{f(\boldsymbol{r}_i | \boldsymbol{X}_i, \boldsymbol{Y}_i^o, \boldsymbol{\psi})}$$

Mechanism is MAR

$\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ distinct

Interest in $\boldsymbol{\theta}$

(Use observed information matrix)

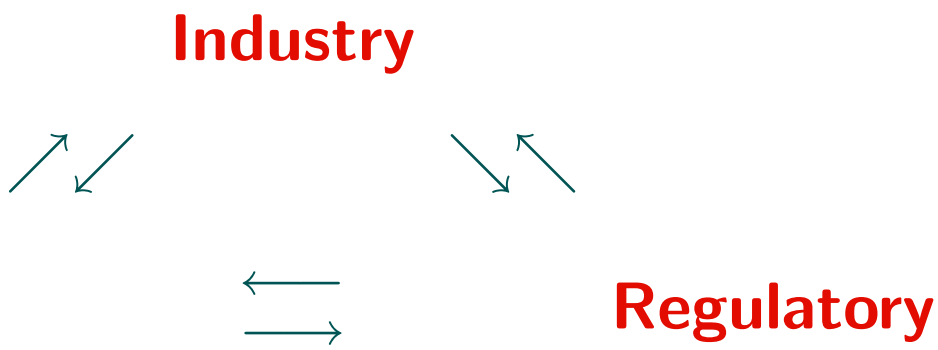
$\Longrightarrow$ Lik./Bayes inference valid

Outcome type	**Modeling strategy**	**Software**
Gaussian	Linear mixed model	SAS MIXED
Non-Gaussian	Gen./Non-linear mixed model	SAS GLIMMIX, NLMIXED

10.6 Rubin, 1976

- Ignorability: Rubin (Biometrika, 1976): 35 years ago!
- Little and Rubin (1976, 2002)
- Why did it take so long?

10.7 A Vicious Triangle



- **Academe:** The R^2 principle
- **Regulatory:** Rigid procedures $\longleftrightarrow$ scientific developments
- **Industry:** We **cannot / do not want to** apply new methods

10.8 Terminology & Confusion

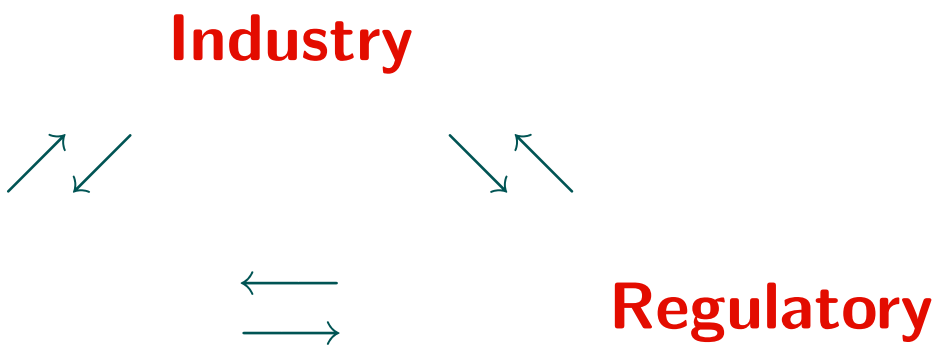
- The Ministry of Disinformation:

$\longleftarrow$ **All directions**

Other directions $\longrightarrow$

- **MCAR, MAR, MNAR:** "What do the terms mean?"
- **MAR, random dropout, informative missingness, ignorable, censoring,. . .**
- **Dropout from the study, dropout from treatment, lost to follow up,. . .**
- **"Under MAR patients dropping out and patients not dropping out are similar."**

10.9 A Virtuous Triangle



- FDA/Industry Workshops
- DIA/EMA Meetings
- **The NAS Experience**

10.10 The NAS Experience: A Wholesome Product

- **FDA $\longrightarrow$ NAS $\longrightarrow$ the working group**
- **Composition**
- **Encompassing:**
  - ▷ terminology/taxonomy/concepts
  - ▷ prevention
  - ▷ treatment

10.11 Taxonomy

- **Missingness pattern:** complete — monotone — non-monotone
- Dropout pattern: complete — dropout — intermittent
- **Model framework:** SEM — PMM — SPM
- **Missingness mechanism:** MCAR — MAR — MNAR
- **Ignorability:** ignorable — non-ignorable
- **Inference paradigm:** frequentist — likelihood — Bayes

10.12 The NAS Panel

Name	Specialty	Affiliation
Rod Little	biostat	U Michigan
Ralph D'Agostino	biostat	Boston U
Kay Dickerson	epi	Johns Hopkins
Scott Emerson	biostat	U Washington
John Farrar	epi	U Penn
Constantine Frangakis	biostat	Johns Hopkins
Joseph Hogan	biostat	Brown U
Geert Molenberghs	biostat	U Hasselt & K.U.Leuven
Susan Murphy	stat	U Michigan
James Neaton	biostat	U Minnesota
Andrea Rotnitzky	stat	Buenos Aires & Harvard
Dan Scharfstein	biostat	Johns Hopkins
Joseph Shih	biostat	New Jersey SPH
Jay Siegel	biostast	J&J
Hal Stern	stat	UC at Irvine

10.13 Modeling Frameworks & Missing Data Mechanisms

$$f(\boldsymbol{y}_i, \boldsymbol{r}_i | X_i, \boldsymbol{\theta}, \boldsymbol{\psi})$$

Selection Models: $f(\boldsymbol{y}_i | X_i, \boldsymbol{\theta}) \boxed{f(\boldsymbol{r}_i | X_i, \boldsymbol{y}_i^o, \boldsymbol{y}_i^m, \boldsymbol{\psi})}$

MCAR $\longrightarrow$ **MAR** $\longrightarrow$ **MNAR**

$f(\boldsymbol{r}_i | X_i, \boldsymbol{\psi})$ $\qquad$ $f(\boldsymbol{r}_i | X_i, \boldsymbol{y}_i^o, \boldsymbol{\psi})$ $\qquad$ $f(\boldsymbol{r}_i | X_i, \boldsymbol{y}_i^o, \boldsymbol{y}_i^m, \boldsymbol{\psi})$

Pattern-mixture Models: $f(\boldsymbol{y}_i | X_i, \boldsymbol{r}_i, \boldsymbol{\theta})\, f(\boldsymbol{r}_i | X_i, \boldsymbol{\psi})$

Shared-parameter Models: $f(\boldsymbol{y}_i | X_i, \boldsymbol{b}_i, \boldsymbol{\theta})\, f(\boldsymbol{r}_i | X_i, \boldsymbol{b}_i, \boldsymbol{\psi})$

10.14 Frameworks and Their Methods

$$f(\boldsymbol{y}_i, \boldsymbol{r}_i | X_i, \boldsymbol{\theta}, \boldsymbol{\psi})$$

Selection Models: $f(\boldsymbol{y}_i | X_i, \boldsymbol{\theta}) \boxed{f(\boldsymbol{r}_i | X_i, \boldsymbol{y}_i^o, \boldsymbol{y}_i^m, \boldsymbol{\psi})}$

MCAR/simple	$\longrightarrow$	**MAR**	$\longrightarrow$	**MNAR**
CC?		**direct likelihood!**		joint model!?
LOCF?		**direct Bayesian!**		sensitivity analysis?!
single imputation?		**multiple imputation (MI)!**		
⋮		**IPW $\supset$ W-GEE!**		
		d.l. + IPW = double robustness! (consensus)		

10.15 Frameworks and Their Methods: Start

$$f(\boldsymbol{y}_i, \boldsymbol{r}_i | X_i, \boldsymbol{\theta}, \boldsymbol{\psi})$$

Selection Models: $f(\boldsymbol{y}_i | X_i, \boldsymbol{\theta}) \boxed{f(\boldsymbol{r}_i | X_i, \boldsymbol{y}_i^o, \boldsymbol{y}_i^m, \boldsymbol{\psi})}$

MCAR/simple $\longrightarrow$ **MAR** $\longrightarrow$ **MNAR**

direct likelihood!

direct Bayesian!

multiple imputation (MI)!

IPW $\supset$ **W-GEE!**

d.l. + IPW = double robustness!

10.16 Frameworks and Their Methods: Next

$$f(\boldsymbol{y}_i, \boldsymbol{r}_i | X_i, \boldsymbol{\theta}, \boldsymbol{\psi})$$

Selection Models: $f(\boldsymbol{y}_i | X_i, \boldsymbol{\theta}) \boxed{f(\boldsymbol{r}_i | X_i, \boldsymbol{y}_i^o, \boldsymbol{y}_i^m, \boldsymbol{\psi})}$

MCAR/simple $\longrightarrow$ **MAR** $\longrightarrow$ **MNAR**

~~joint model!?~~

sensitivity analysis!

PMM

MI (MGK, J&J)

local influence

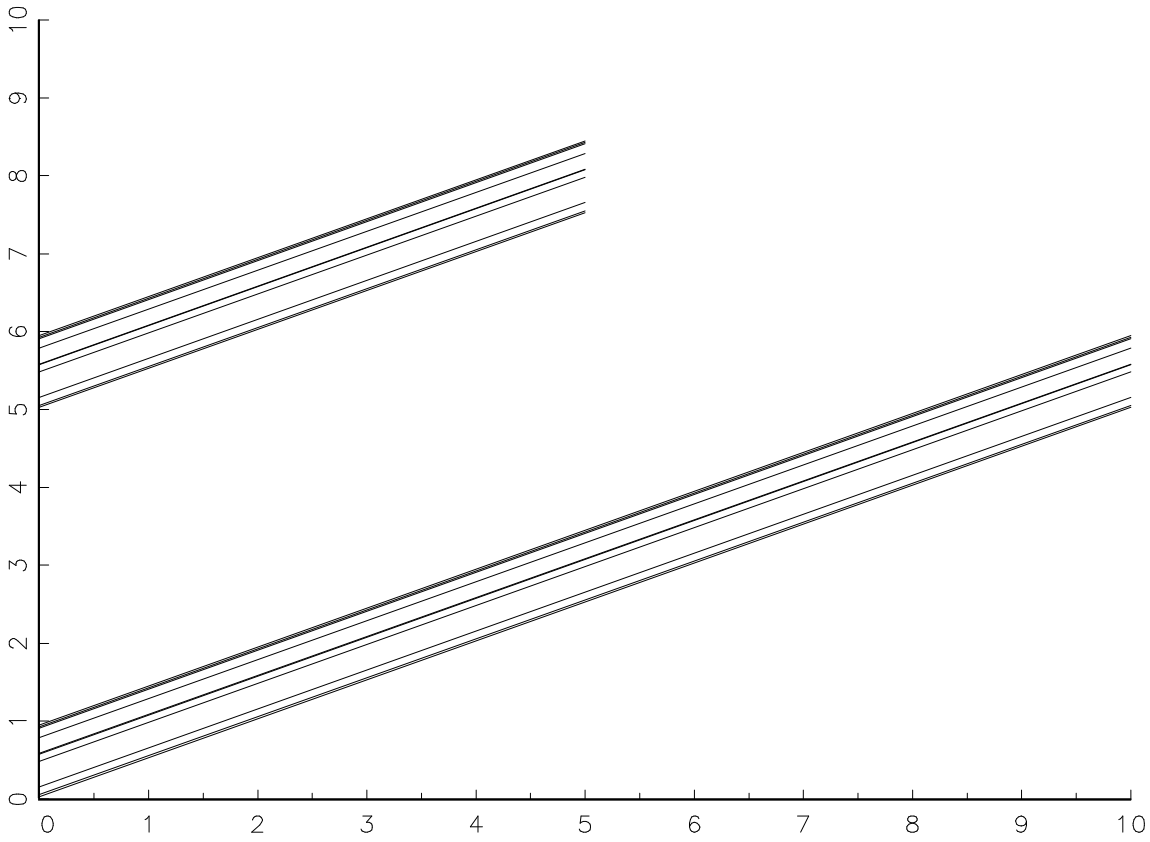
interval ignorance

IPW based

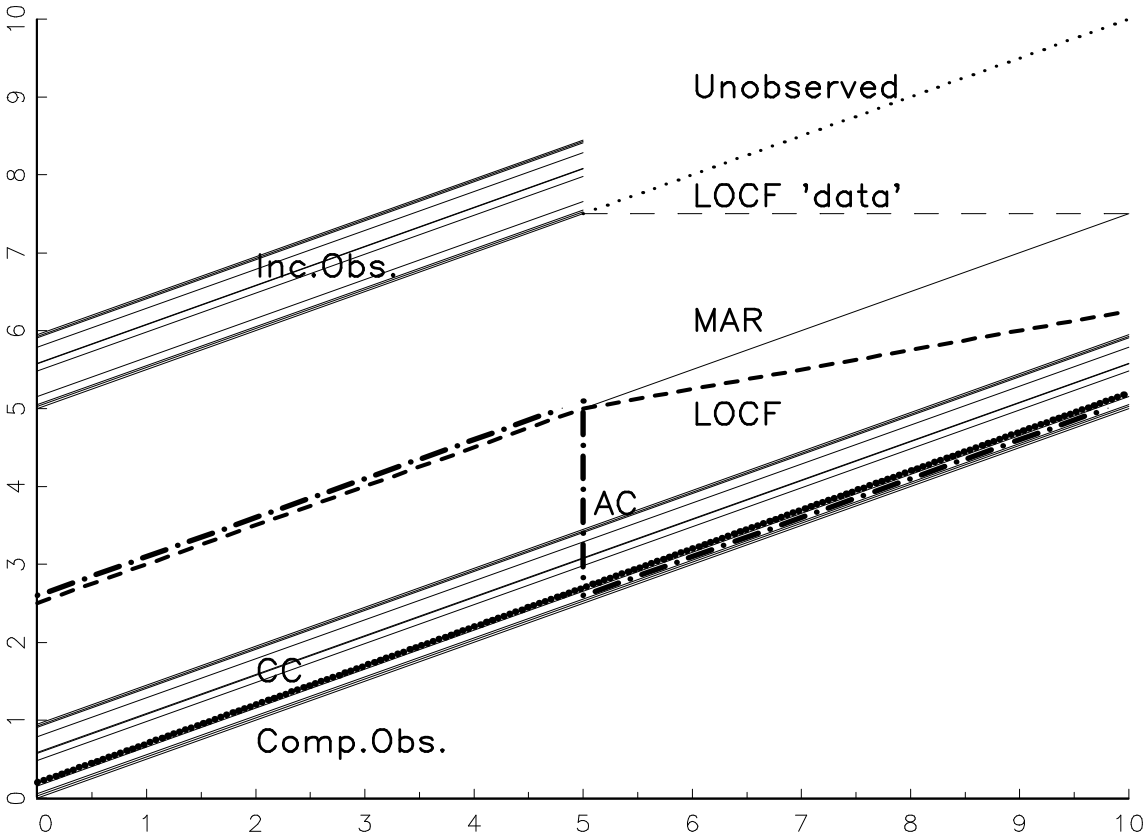
10.17 Overview

MCAR/simple	CC LOCF	biased inefficient not simpler than MAR methods
MAR	**direct likelihood** **direct Bayes** **weighted GEE** **MI**	easy to conduct Gaussian & non-Gaussian
MNAR	variety of methods	strong, untestable assumptions most useful in **sensitivity analysis**

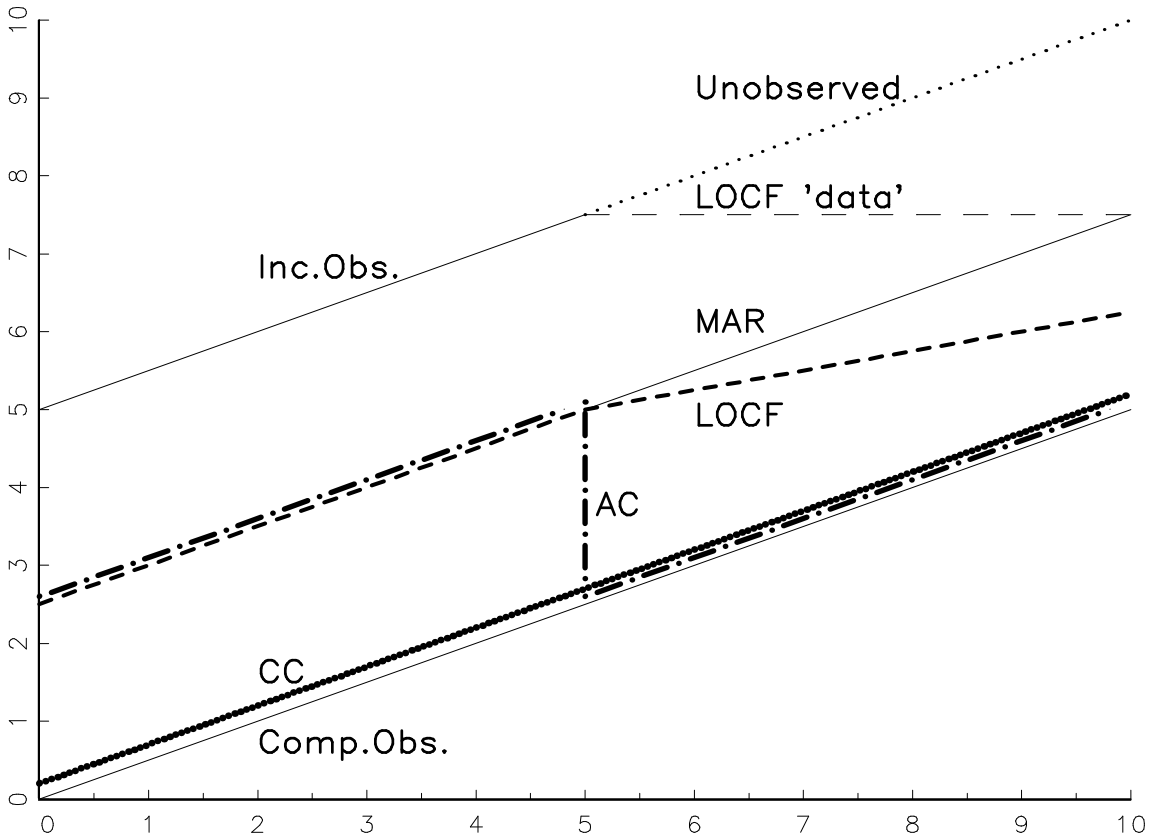
10.18 Incomplete Longitudinal Data



Data and Modeling Strategies

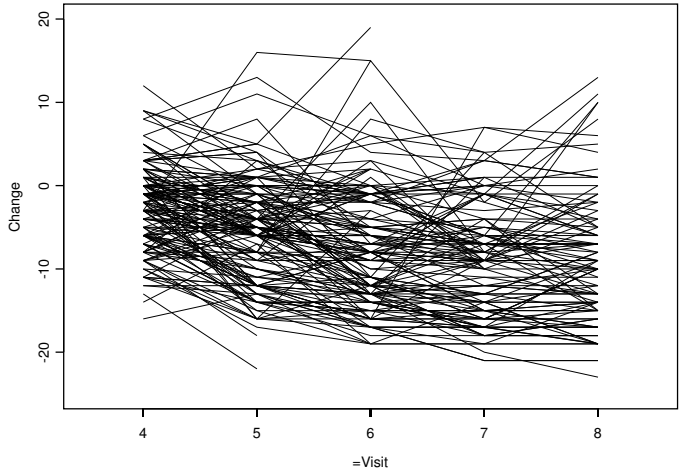


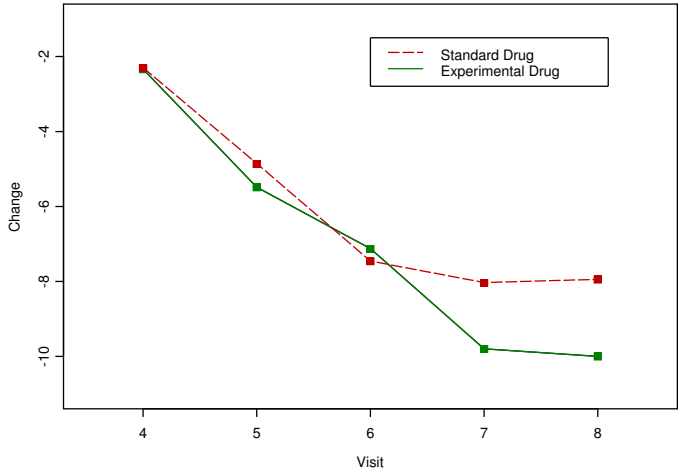
Modeling Strategies



10.19 The Depression Trial

- Clinical trial: experimental drug *versus* standard drug
- 170 patients
- Response: change versus baseline in $HAMD_{17}$ score
- 5 post-baseline visits: 4–8





10.20 Analysis of the Depression Trial

- Treatment effect at visit 8 (last follow-up measurement):

Method	**Estimate**	**(s.e.)**	p**-value**
CC	-1.94	(1.17)	0.0995
LOCF	-1.63	(1.08)	0.1322
MAR	-2.38	(1.16)	0.0419

Observe the slightly significant p-value under the MAR model

Chapter 11
Direct Likelihood / Ignorable Likelihood

- Simple methods
- Direct likelihood / ignorability

11.1 Simple Methods

MCAR

Complete case analysis:

$\Rightarrow$ **delete** incomplete subjects

- Standard statistical software
- Loss of information
- Impact on precision and power
- Missingness $\neq$ MCAR $\Rightarrow$ bias

Last observation carried forward:

$\Rightarrow$ **impute** missing values

- Standard statistical software
- Increase of information
- Constant profile after dropout: unrealistic
- Usually bias

11.2 Ignorability

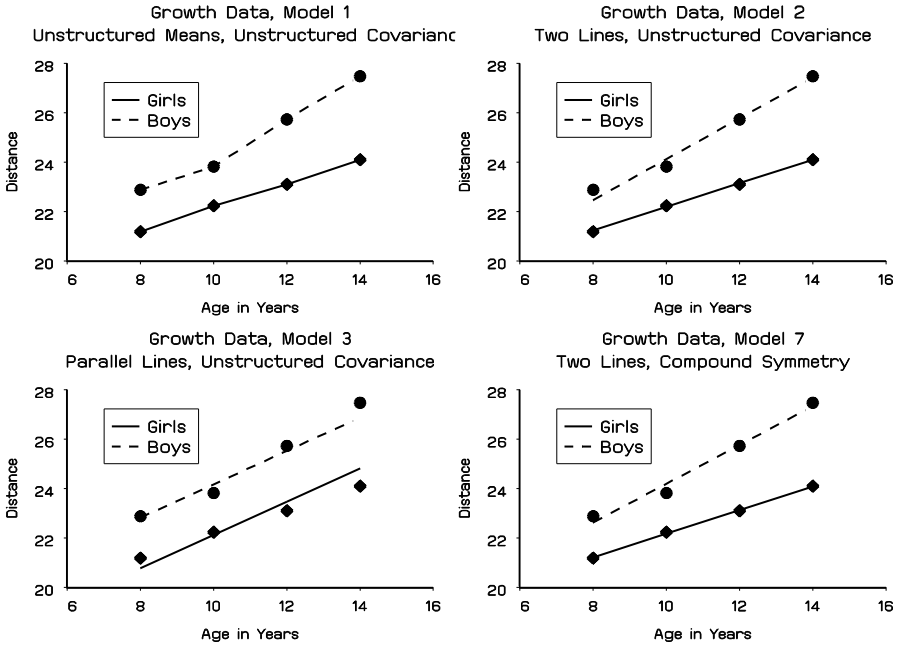
Likelihood/Bayesian + MAR

&

Frequentist + MCAR

11.3 Original, Complete Orthodontic Growth Data

	Mean	Covar	# par
1	unstructured	unstructured	18
2	$\neq$ slopes	unstructured	14
3	$=$ slopes	unstructured	13
7	**$\neq$ slopes**	**CS**	**6**

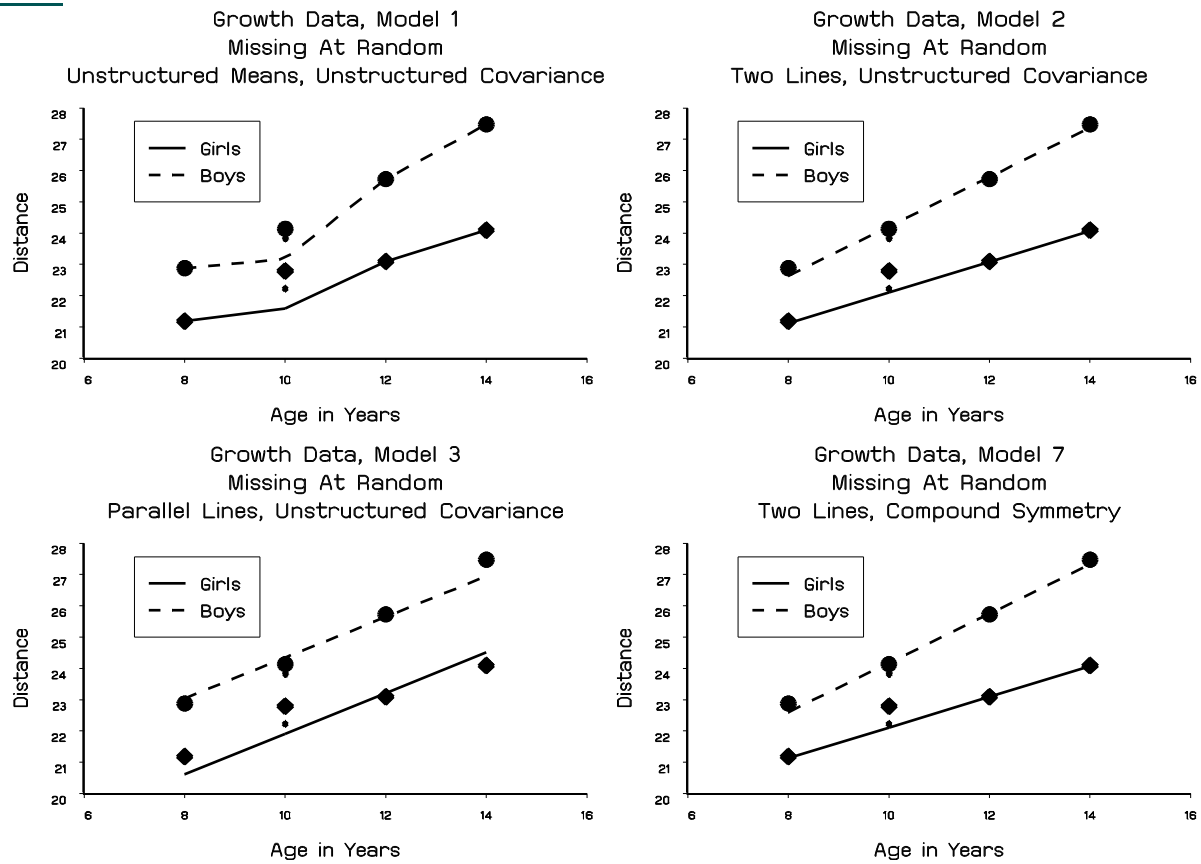


11.4 Trimmed Growth Data: Simple Methods

Method	Model	Mean	Covar	# par
Complete case	7a	= slopes	CS	5
LOCF	2a	quadratic	unstructured	16
Unconditional mean	7a	= slopes	CS	5
Conditional mean	1	unstructured	unstructured	18

11.5 Trimmed Growth Data: Direct Likelihood

	Mean	Covar	# par
7	$\neq$ slopes	CS	6



11.6 Illustration

Data:

20	30
10	40

75
25

LOCF:

20	30
10	40

75	0
0	25

$\Longrightarrow$

95	30
10	65

$\Longrightarrow \quad \hat{\theta} = \frac{95}{200} = 0.475 \ [0.406; 0.544]$ (biased & too narrow)

CC:

20	30
10	40

0	0
0	0

$\Longrightarrow$

20	30
10	40

$\Longrightarrow \quad \hat{\theta} = \frac{20}{100} = 0.200 \ [0.122; 0.278]$ (biased & too wide)

MAR:

20	30
10	40

30	45
5	20

$\Longrightarrow$

50	75
15	60

$\Longrightarrow \quad \hat{\theta} = \frac{50}{200} = 0.250 \ [0.163; 0.337]$

11.6.1 Growth Data: SAS Code for Model 1

IDNR	AGE	SEX	MEASURE
1	8	2	21.0
1	10	2	20.0
1	12	2	21.5
1	14	2	23.0
...			
3	8	2	20.5
3	12	2	24.5
3	14	2	26.0
...			

- SAS code:

```
proc mixed data = growth method = ml;
   class sex idnr age;
   model measure = sex age*sex / s;
   repeated age / type = un
                  subject = idnr;
   run;
```

▷ Subjects in terms of IDNR blocks

▷ age ensures proper ordering of observations within subjects!

11.6.2 Growth Data: SAS Code for Model 2

IDNR	AGE	SEX	MEASURE
1	8	2	21.0
1	10	2	20.0
1	12	2	21.5
1	14	2	23.0
...			
3	8	2	20.5
3	12	2	24.5
3	14	2	26.0
...			

- SAS code:

```
data help;
   set growth;
   agecat = age;
   run;

proc mixed data = growth method = ml;
   class sex idnr agecat;
   model measure = sex age*sex / s;
   repeated agecat / type = un
                     subject = idnr;
   run;
```

  - ▷ Time ordering variable needs to be categorical

11.7 Analysis of the Depression Trial

- **Complete case analysis**:

```
%cc(data=depression, id=patient, time=visit, response=change, out={cc});
```

  $\Rightarrow$ performs analysis on CC data set

- **LOCF analysis**:

```
%locf(data=depression, id=patient, time=visit, response=change, out={locf});
```

  $\Rightarrow$ performs analysis on LOCF data

- **Direct-likelihood analysis:** $\Rightarrow$ fit linear mixed model to incomplete data

- Treatment effect at visit 8 (last follow-up measurement):

Method	**Estimate**	**(s.e.)**	p**-value**
CC	-1.94	(1.17)	0.0995
LOCF	-1.63	(1.08)	0.1322
MAR	-2.38	(1.16)	0.0419

Observe the slightly significant p-value under the MAR model

Chapter 12
Analysis of the ARMD Trial

- Model for continuous outcomes:

$$Y_{ij} = \beta_{j1} + \beta_{j2} T_i + \varepsilon_{ij}$$

  with:

  - $T_i = 0$ for placebo and $T_i = 1$ for interferon-α
  - t_j $(j = 1, \ldots, 4)$ refers to the four follow-up measurements
  - unstructured variance-covariance matrix

- Marginal mean for GEE:

$$\text{logit}[P(Y_{ij} = 1|T_i, t_j)] = \beta_{j1} + \beta_{j2}T_i$$

- Model for GLMM with random interecept:

$$\text{logit}[P(Y_{ij} = 1|T_i, t_j, b_i)] = \beta_{j1} + b_i + \beta_{j2}T_i$$

  with

  ▷ $b_i \sim N(0, \tau^2)$

- **Complete case analysis preparation (continuous outcome):**

```
%cc(data=armd155,id=subject,time=time,response=diff,
    out=armdcc2);
```

- **Complete case analysis preparation (discrete outcome):**

```
%cc(data=armd111,id=subject,time=time,response=bindif,
    out=armdcc);
```

- **Preparing for LOCF analysis (continuous outcome):**

```
%locf(data=armd155,id=subject,time=time,response=diff,
      out=armdlocf2);
```

- **Preparing for LOCF analysis (discrete outcome):**

```
%locf(data=armd111,id=subject,time=time,response=bindif,
      out=armdlocf);
```

- **Program for linear mixed model (PROC MIXED):**

```
proc mixed data=armdcc2 method=ml;
class time treat subject;
model diff = time treat*time / noint solution ddfm=kr;
repeated time / subject=subject type=un;
run;
```

- **Program for GEE (PROC GENMOD):**

```
proc genmod data=armdcc;
class time treat subject;
model bindif = time treat*time / noint dist=binomial;
repeated subject=subject / withinsubject=time type=exch modelse;
run;
```

- **Program for GEE (PROC GEE):**

```
proc gee data=armdcc;
class time treat subject;
model bindif = time treat*time / noint dist=binomial;
repeated subject=subject / withinsubject=time type=exch modelse;
run;
```

- **Program for GLMM (PROC GLIMMIX):**

```
proc glimmix data=armdcc method=gauss(q=20);
nloptions maxiter=50 technique=newrap;
class time treat subject;
model bindif = time treat*time / noint solution dist=binary;
random intercept  / subject=subject type=un g gcorr;
run;
```

- **Program for GLMM (PROC NLMIXED):**

```
data help; set armdcc;
time1=0; if time=1 then time1=1;
time2=0; if time=2 then time2=1;
time3=0; if time=3 then time3=1;
time4=0; if time=4 then time4=1;
run;

proc nlmixed data=help  qpoints=20 maxiter=100 technique=newrap;
title 'CC - mixed - numerical integration';
eta = beta11*time1+beta12*time2+beta13*time3+beta14*time4
      +b
      +(beta21*time1+beta22*time2+beta23*time3+beta24*time4)
      *(2-treat);
p = exp(eta)/(1+exp(eta));
model bindif ~ binary(p);
random b ~ normal(0,tau*tau) subject=subject;
estimate 'tau^2' tau*tau;
run;
```

Effect	Parameter	CC	LOCF	direct lik.
Parameter estimates (standard errors) for linear mixed model				
Intercept 4	β_{11}	-3.24(0.77)	-3.48(0.77)	-3.48(0.77)
Intercept 12	β_{21}	-4.66(1.14)	-5.72(1.09)	-5.85(1.11)
Intercept 24	β_{31}	-8.33(1.39)	-8.34(1.30)	-9.05(1.36)
Intercept 52	β_{41}	-15.13(1.73)	-14.16(1.53)	-16.21(1.67)
Treatm. eff. 4	β_{12}	2.32(1.05)	2.20(1.08)	2.20(1.08)
Treatm. eff. 12	β_{22}	2.35(1.55)	3.38(1.53)	3.51(1.55)
Treatm. eff. 24	β_{32}	2.73(1.88)	2.41(1.83)	3.03(1.89)
Treatm. eff. 52	β_{42}	4.17(2.35)	3.43(2.15)	4.86(2.31)
		p-**values**		
Treatm. eff. 4	β_{12}	0.0282	0.0432	0.0435
Treatm. eff. 12	β_{22}	0.1312	0.0287	0.0246
Treatm. eff. 24	β_{32}	0.1491	0.1891	0.1096
Treatm. eff. 52	β_{42}	0.0772	0.1119	0.0366
Treatm. eff. (overall)		0.1914	0.1699	0.1234

Effect	Parameter	CC	LOCF	direct lik.
		PQL		
Int.4	β_{11}	-1.19(0.31)	-1.05(0.28)	-1.00(0.26)
Int.12	β_{21}	-1.05(0.31)	-1.18(0.28)	-1.19(0.28)
Int.24	β_{31}	-1.35(0.32)	-1.30(0.28)	-1.26(0.29)
Int.52	β_{41}	-1.97(0.36)	-1.89(0.31)	-2.02(0.35)
Trt.4	β_{12}	0.45(0.42)	0.24(0.39)	0.22(0.37)
Trt.12	β_{22}	0.58(0.41)	0.68(0.38)	0.71(0.37)
Trt.24	β_{32}	0.55(0.42)	0.50(0.39)	0.49(0.39)
Trt.52	β_{42}	0.44(0.47)	0.39(0.42)	0.46(0.46)
R.I. s.d.	τ	1.42(0.14)	1.53(0.13)	1.40(0.13)
R.I. var.	τ^2	2.03(0.39)	2.34(0.39)	1.95(0.35)
		Numerical integration		
Int.4	β_{11}	-1.73(0.42)	-1.63(0.39)	-1.50(0.36)
Int.12	β_{21}	-1.53(0.41)	-1.80(0.39)	-1.73(0.37)
Int.24	β_{31}	-1.93(0.43)	-1.96(0.40)	-1.83(0.39)
Int.52	β_{41}	-2.74(0.48)	-2.76(0.44)	-2.85(0.47)
Trt.4	β_{12}	0.64(0.54)	0.38(0.52)	0.34(0.48)
Trt.12	β_{22}	0.81(0.53)	0.98(0.52)	1.00(0.49)
Trt.24	β_{32}	0.77(0.55)	0.74(0.52)	0.69(0.50)
Trt.52	β_{42}	0.60(0.59)	0.57(0.56)	0.64(0.58)
R.I. s.d.	τ	2.19(0.27)	2.47(0.27)	2.20(0.25)
R.I. var.	τ^2	4.80(1.17)	6.08(1.32)	4.83(1.11)

Chapter 13
Weighted Generalized Estimating Equations

- ▷ General Principle
- ▷ Analysis of the analgesic trial
- ▷ Analysis of the ARMD trial
- ▷ Analysis of the depression trial

13.1 General Principle

MAR and non-ignorable !

- Standard GEE inference correct only under MCAR
- Under MAR: *weighted* GEE

  Robins, Rotnitzky & Zhao (JASA, 1995)

  Fitzmaurice, Molenberghs & Lipsitz (JRSSB, 1995)

- Decompose dropout time $D_i = (R_{i1}, \ldots, R_{in}) = (1, \ldots, 1, 0, \ldots, 0)$

- Weigh a contribution by inverse dropout probability

$$\nu_{id_i} \equiv P[D_i = d_i] \;=\; \prod_{k=2}^{d_i-1} (1 - P[R_{ik} = 0 | R_{i2} = \ldots = R_{i,k-1} = 1]) \times$$

$$P[R_{id_i} = 0 | R_{i2} = \ldots = R_{i,d_i-1} = 1]^{I\{d_i \leq T\}}$$

- Adjust estimating equations

$$\sum_{i=1}^{N} \frac{1}{\nu_{id_i}} \cdot \frac{\partial \boldsymbol{\mu_i}}{\partial \boldsymbol{\beta}'} V_i^{-1} (\boldsymbol{y_i} - \boldsymbol{\mu_i}) \;=\; \mathbf{0}$$

13.2 Computing the Weights

- Predicted values from (PROC GENMOD) output

- The weights are now defined at the individual measurement level:
  - ▷ At the first occasion, the weight is $w = 1$
  - ▷ At other than the last ocassion, the weight is the already accumulated weight, multiplied by $1-$the predicted probability
  - ▷ At the last occasion *within a sequence where dropout occurs* the weight is multiplied by the predicted probability
  - ▷ At the end of the process, the weight is inverted

13.3 The Analgesic Trial

- single-arm trial with 530 patients recruited (491 selected for analysis)
- analgesic treatment for pain caused by chronic nonmalignant disease
- treatment was to be administered for 12 months
- we will focus on Global Satisfaction Assessment (GSA)
- GSA scale goes from 1=very good to 5=very bad
- GSA was rated by each subject 4 times during the trial, at months 3, 6, 9, and 12.

- Research questions:
  - ▷ Evolution over time
  - ▷ Relation with baseline covariates: age, sex, duration of the pain, type of pain, disease progression, Pain Control Assessment (PCA), ...
  - ▷ Investigation of dropout

- Frequencies:

GSA	Month 3		Month 6		Month 9		Month 12	
1	55	14.3%	38	12.6%	40	17.6%	30	13.5%
2	112	29.1%	84	27.8%	67	29.5%	66	29.6%
3	151	39.2%	115	38.1%	76	33.5%	97	43.5%
4	52	13.5%	51	16.9%	33	14.5%	27	12.1%
5	15	3.9%	14	4.6%	11	4.9%	3	1.4%
Tot	385		302		227		223	

- Missingness:

Measurement occasion					
Month 3	Month 6	Month 9	Month 12	Number	%
Completers' pattern					
O	O	O	O	163	41.2
Dropout patterns					
O	O	O	M	51	12.91
O	O	M	M	51	12.91
O	M	M	M	63	15.95
Non-monotone patterns					
O	O	M	O	30	7.59
O	M	O	O	7	1.77
O	M	O	M	2	0.51
O	M	M	O	18	4.56
M	O	O	O	2	0.51
M	O	O	M	1	0.25
M	O	M	O	1	0.25
M	O	M	M	3	0.76

13.4 Analysis of the Analgesic Trial

- A logistic regression for the dropout indicator:

$$\begin{aligned}\text{logit}[P(D_i = j | D_i \geq j, \cdot)] = \; & \psi_0 + \psi_{11} I(\text{GSA}_{i,j-1} = 1) + \psi_{12} I(\text{GSA}_{i,j-1} = 2) \\ & + \psi_{13} I(\text{GSA}_{i,j-1} = 3) + \psi_{14} I(\text{GSA}_{i,j-1} = 4) \\ & + \psi_2 \text{PCA0}_i + \psi_3 \text{PF}_i + \psi_4 \text{GD}_i\end{aligned}$$

  with

  - ▷ $\text{GSA}_{i,j-1}$ the 5-point outcome at the previous time
  - ▷ $I(\cdot)$ is an indicator function
  - ▷ PCA0_i is pain control assessment at baseline
  - ▷ PF_i is physical functioning at baseline
  - ▷ GD_i is genetic disorder at baseline are used)

Effect	Par.	Estimate (s.e.)
Intercept	ψ_0	-1.80 (0.49)
Previous GSA$=1$	ψ_{11}	-1.02 (0.41)
Previous GSA$=2$	ψ_{12}	-1.04 (0.38)
Previous GSA$=3$	ψ_{13}	-1.34 (0.37)
Previous GSA$=4$	ψ_{14}	-0.26 (0.38)
Basel. PCA	ψ_2	0.25 (0.10)
Phys. func.	ψ_3	0.009 (0.004)
Genetic disfunc.	ψ_4	0.59 (0.24)

- There is some evidence for MAR: $P(D_i = j | D_i \geq j)$ depends on previous GSA.
- Furthermore: baseline PCA, physical functioning and genetic/congenital disorder.

- GEE and WGEE:

$$\text{logit}[P(Y_{ij} = 1 | t_j, \text{PCA0}_i)] = \beta_1 + \beta_2 t_j + \beta_3 t_j^2 + \beta_4 \text{PCA0}_i$$

Effect	**Parameter**	**GEE**	**WGEE**
Intercept	β_1	2.95 (0.47)	2.17 (0.69)
Time	β_2	-0.84 (0.33)	-0.44 (0.44)
Time^2	β_3	0.18 (0.07)	0.12 (0.09)
Basel. PCA	β_4	-0.24 (0.10)	-0.16 (0.13)

- A hint of potentially important differences between both

13.5 Analgesic Trial: Steps for WGEE in SAS

1. **Preparatory data manipulation:**

```
%dropout(...)
```

2. **Logistic regression for weight model:**

```
proc genmod data=gsac;
   class prevgsa;
   model dropout = prevgsa pca0 physfct gendis / pred dist=b;
   ods output obstats=pred;
   run;
```

3. **Conversion of predicted values to weights:**

```
...
%dropwgt(...)
```

4. Weighted GEE analysis:

```
proc genmod data=repbin.gsaw;
   scwgt wi;
   class patid timecls;
   model gsabin = time|time pca0 / dist=b;
   repeated subject=patid / type=un corrw within=timecls;
   run;
```

13.6 Analgesic Trial: Steps for WGEE in SAS, Using PROC GEE

- Available since SAS 9.4 (SAS/STAT 13.2)
- **Preparation:**

```
data gsaw;
   set gsaw;
   by patid;
   prevgsa = lag(gsa);
   if first.id then prevgsa = 1;
   time = time-1;
   timeclss = time;
run;
```

- **Weighted GEE analysis:**

```
ods graphics on;
proc gee data=gsaw plots=histogram;
   class patid timecls prevgsa;
   model gsabin = time|time pca0 / dist=bin;
   repeated subject=patid / within=timecls corr=un;
   missmodel prevgsa pca0 physfunt gendist / type=obslevel;
run;
```

13.7 Analysis of the ARMD Trial

- Model for the weights:

$$\text{logit}[P(D_i = j | D_i \geq j)] = \psi_0 + \psi_1 y_{i,j-1} + \psi_2 T_i + \psi_{31} L_{1i} + \psi_{32} L_{2i} + \psi_{34} L_{3i} + \psi_{41} I(t_j = 2) + \psi_{42} I(t_j = 3)$$

  with

  - $y_{i,j-1}$ the binary outcome at the previous time $t_{i,j-1} = t_{j-1}$ (since time is common to all subjects)
  - $T_i = 1$ for interferon-α and $T_i = 0$ for placebo
  - $L_{ki} = 1$ if the patient's eye lesion is of level $k = 1, \ldots, 4$ (since one dummy variable is redundant, only three are used)
  - $I(\cdot)$ is an indicator function

- Results for the weights model:

Effect	**Parameter**	**Estimate (s.e.)**
Intercept	ψ_0	0.14 (0.49)
Previous outcome	ψ_1	0.04 (0.38)
Treatment	ψ_2	-0.86 (0.37)
Lesion level 1	ψ_{31}	-1.85 (0.49)
Lesion level 2	ψ_{32}	-1.91 (0.52)
Lesion level 3	ψ_{33}	-2.80 (0.72)
Time 2	ψ_{41}	-1.75 (0.49)
Time 3	ψ_{42}	-1.38 (0.44)

- GEE:

$$\text{logit}[P(Y_{ij} = 1|T_i, t_j)] = \beta_{j1} + \beta_{j2} T_i$$

  with

  - $T_i = 0$ for placebo and $T_i = 1$ for interferon-α
  - t_j $(j = 1, \ldots, 4)$ refers to the four follow-up measurements
  - Comparison between CC, LOCF, and GEE analyses

- SAS code: Molenberghs and Verbeke (2005, Section 32.5)

- Results:

				Observed data	
Effect	**Par.**	**CC**	**LOCF**	**Unweighted**	**WGEE**
Int.4	β_{11}	-1.01(0.24;0.24)	-0.87(0.20;0.21)	-0.87(0.21;0.21)	-0.98(0.10;0.44)
Int.12	β_{21}	-0.89(0.24;0.24)	-0.97(0.21;0.21)	-1.01(0.21;0.21)	-1.78(0.15;0.38)
Int.24	β_{31}	-1.13(0.25;0.25)	-1.05(0.21;0.21)	-1.07(0.22;0.22)	-1.11(0.15;0.33)
Int.52	β_{41}	-1.64(0.29;0.29)	-1.51(0.24;0.24)	-1.71(0.29;0.29)	-1.72(0.25;0.39)
Tr.4	β_{12}	0.40(0.32;0.32)	0.22(0.28;0.28)	0.22(0.28;0.28)	0.80(0.15;0.67)
Tr.12	β_{22}	0.49(0.31;0.31)	0.55(0.28;0.28)	0.61(0.29;0.29)	1.87(0.19;0.61)
Tr.24	β_{32}	0.48(0.33;0.33)	0.42(0.29;0.29)	0.44(0.30;0.30)	0.73(0.20;0.52)
Tr.52	β_{42}	0.40(0.38;0.38)	0.34(0.32;0.32)	0.44(0.37;0.37)	0.74(0.31;0.52)
Corr.	ρ	0.39	0.44	0.39	0.33

13.8 WGEE for the ARMD Trial in SAS: PROC GENMOD

1. **Data manipulation prior to estimating the parameters in the weight model:**

```
%dropout(data=armd111,id=subject,time=time,response=bindif,
         out=armdhlp);
```

2. **Code for the weight model:**

```
proc genmod data=armdhlp descending;
class trt prev lesion time;
model dropout = prev trt lesion time / pred dist=binomial;
ods output obstats=pred;
run;
```

3. Data manipulation to prepare for WGEE:

```
data pred;
set pred;
keep observation pred;
run;

data armdhlp;
merge pred armdhlp;
run;

%dropwgt(data=armdhlp,id=subject,time=time,pred=pred,
         dropout=dropout,out=armdwgee);
```

4. WGEE using PROC GENMOD:

```
proc genmod data=armdcc;
weight wi;
class time treat subject;
model bindif = time treat*time / noint dist=binomial;
repeated subject=subject / withinsubject=time type=exch modelse;
run;
```

13.9 ARMD Trial Analyzed With PROC GEE

		Weights	
Effect	**Parameter**	**observation**	**subject**
Int.4	β_{11}	-0.95 (0.20)	-0.98 (0.35)
Int.12	β_{21}	-1.03 (0.22)	-1.77 (0.30)
Int.24	β_{31}	-1.03 (0.23)	-1.11 (0.29)
Int.52	β_{41}	-1.52 (0.30)	-1.72 (0.37)
Tr.4	β_{12}	0.32 (0.28)	0.78 (0.56)
Tr.12	β_{22}	0.65 (0.29)	1.83 (0.47)
Tr.24	β_{32}	0.39 (0.30)	0.71 (0.49)
Tr.52	β_{42}	0.30 (0.39)	0.72 (0.47)
Corr.	ρ	0.38	0.33

- **Subject-level weights:**
  - ▷ A single weight is calculated for the entire subject
  - ▷ The original proposal of the method
  - ▷ Slightly easier to calculate
  - ▷ Less precise

- **Observation-level weights:**
  - ▷ A separate weight is calculated for every measurement within a subject
  - ▷ A later modification
  - ▷ Increased precision
  - ▷ Default in SAS

13.10 WGEE for the ARMD Trial in SAS: PROC GEE

1. **Preparatory steps (creation of the variables needed in the MISSMODEL statement):**

```
data help;
set armdwgee;
by subject;
prevbindif=lag(bindif);
if first.id then prevbindif=1;
time2=0;
if time=2 then time2=1;
time3=0;
if time=3 then time3=1;
run;
```

2. **PROC GEE code:**

```
proc gee data=help;
class time treat subject lesion;
model bindif = time treat*time / noint dist=binomial;
repeated subject=subject / withinsubject=time type=exch corrw
                           modelse;
missmodel prevbindif treat lesion time2 time3 / type=obslevel;
run;
```

13.11 Analysis of the Depression Trial

- Response: create binary indicator ybin for $HAMD_{17} > 7$
- Model for dropout:

$$\text{logit}[P(D_i = j | D_i \geq j)] = \psi_0 + \psi_1 y_{i,j-1} + \gamma T_i$$

with

  - $y_{i,j-1}$: the binary indicator at the previous occasion
  - T_i: treatment indicator for patient i

- SAS code:
  - ▷ Preparing the dataset:

    ```
    %dropout(data=depression,id=patient,time=visit,response=ybin,out=dropout);
    ```

    producing:
    - ▷ dropout indicates whether missingness at a given time occurs
    - ▷ prev contains outcome at the previous occasion
  - ▷ The logistic model for dropout:

    ```
    proc genmod data=dropout descending;
       class trt;
       model dropout = prev trt / pred dist=b;
       output out=pred p=pred;
       run;
    ```

  - ▷ The weights can now be included in the GENMOD program which specifies the GEE, through the WEIGHT or SCWGT statements:

```
proc genmod data=study descending;
   weight wi;
   class patient visitclass trt;
   model ybin = trt visit trt*visit basval basval*visit / dist=bin;
   repeated subject=patient / withinsubject=visitclass type=cs corrw;
   run;
```

- Results:

	WGEE			GEE		
Effect	**est.**	**(s.e.)**	**p-value**	**est.**	**(s.e.)**	**p-value**
Treatment at visit 4	-1.57	(0.99)	0.11	-0.24	(0.57)	0.67
Treatment at visit 5	-0.67	(0.65)	0.30	0.09	(0.40)	0.82
Treatment at visit 6	0.62	(0.56)	0.27	0.17	(0.34)	0.62
Treatment at visit 7	-0.57	(0.37)	0.12	-0.43	(0.35)	0.22
Treatment at visit 8	-0.84	(0.39)	0.03	-0.71	(0.38)	0.06

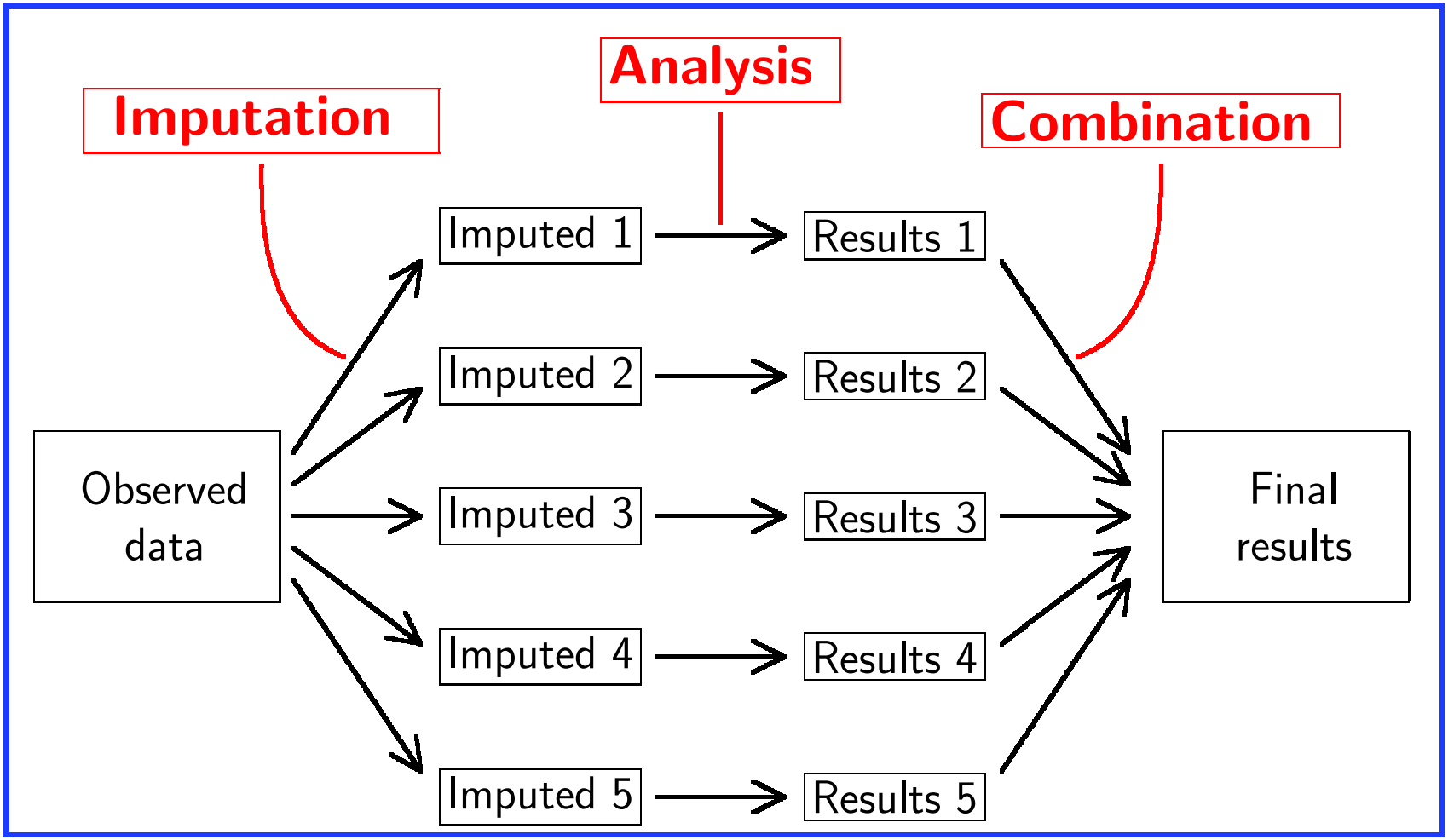
Chapter 14
Multiple Imputation

- General idea
- Estimation
- Hypothesis testing
- Use of MI in practice
- Analysis of the growth data
- Analysis of the ARMD trial
- Creating monotone missingness

14.1 General Principles

- Valid under MAR
- An alternative to direct likelihood and WGEE
- Three steps:
  1. The missing values are filled in M times $\Longrightarrow$ M complete data sets
  2. The M complete data sets are analyzed by using standard procedures
  3. The results from the M analyses are combined into a single inference
- Rubin (1987), Rubin and Schenker (1986), Little and Rubin (1987)
- Molenberghs and Kenward (2007), van Buuren (2012), Carpenter and Kenward (2013)

- Multiple imputation ($M = 5$ imputations):



14.2 Use of MI in Practice

- Many analyses of the same incomplete set of data
- A combination of missing outcomes and missing covariates
- As an alternative to WGEE: MI can be combined with classical GEE
- MI in SAS:

Imputation Task: **PROC MI**

↓

Analysis Task: **PROC "MYFAVORITE"**

↓

Inference Task: **PROC MIANALYZE**

14.3 MI Analysis of the ARMD Trial

- $M = 10$ imputations
- GEE:

$$\text{logit}[P(Y_{ij} = 1|T_i, t_j)] = \beta_{j1} + \beta_{j2}T_i$$

- GLMM:

$$\text{logit}[P(Y_{ij} = 1|T_i, t_j, b_i)] = \beta_{j1} + b_i + \beta_{j2}T_i, \qquad b_i \sim N(0, \tau^2)$$

- $T_i = 0$ for placebo and $T_i = 1$ for interferon-α
- t_j $(j = 1, \ldots, 4)$ refers to the four follow-up measurements
- Imputation based on the continuous outcome

- Results:

Effect	Par.	GEE	GLMM
Int.4	β_{11}	-0.84(0.20)	-1.46(0.36)
Int.12	β_{21}	-1.02(0.22)	-1.75(0.38)
Int.24	β_{31}	-1.07(0.23)	-1.83(0.38)
Int.52	β_{41}	-1.61(0.27)	-2.69(0.45)
Trt.4	β_{12}	0.21(0.28)	0.32(0.48)
Trt.12	β_{22}	0.60(0.29)	0.99(0.49)
Trt.24	β_{32}	0.43(0.30)	0.67(0.51)
Trt.52	β_{42}	0.37(0.35)	0.52(0.56)
R.I. s.d.	τ		2.20(0.26)
R.I. var.	τ^2		4.85(1.13)

14.4 SAS Tools for MI

MONOTONE: Input data **must** be monotone only!

There are several options:

reg: every 'later' variable is regressed on the earlier ones (normal model)

logistic: every 'later' variable is regressed on the earlier ones (logistic model)

discrim: version for categorical data where discriminant functions are used

regpmm: predictive mean matching (a choice is made from observations with value similar to the predicted mean)

propensity: Some detail:

- The propensity score is the conditional probability of assignment to a particular treatment given a vector of observed covariates.
- A propensity score is generated for each variable with missing values to indicate the probability of that observations being missing.

- The observations are then grouped based on these propensity scores, and an approximate Bayesian bootstrap imputation (Rubin 1987) is applied to each group.
- Less suitable when also relations betwen variables (association, regression) is of interest.
- Hence of limited use.

MCMC: can be used with monotone and non-monotone missingness:

- a multivariate normal model is considered
- Monte Carlo draws are taken from it
- Default in SAS

FCS: a extension of MONOTONE that also allows for non-monotone missingness:

- ▷ Fill in 'starting values' for the missing values (e.g., from a predictive model based on completers)
- ▷ Cycle repeatedly through the following two steps:
  - ∗ Build a predictive model for the jth variable given all the others (recall that missing values are 'temporarily' filled in)
  - ∗ Based on the updated model, draw values for the missing ones
- ▷ Same options as MONOTONE, except for propensity scores

14.5 SAS Code for MI

1. Preparatory data analysis so that there is one line per subject

2. **The imputation task (default MCMC method):**

```
proc mi data=armd13 seed=486048 out=armd13a simple nimpute=10 round=0.1;
   var lesion diff4 diff12 diff24 diff52;
   by treat;
   run;
```

   Note that the imputation task is conducted on the continuous outcome 'diff·', indicating the difference in number of letters versus baseline

3. Then, data manipulation takes place to define the binary indicators and to create a longitudinal version of the dataset

4. The imputation task using FCS:

```
proc mi data=m.armd13 seed=486048 simple out=m.armd13fcs nimpute=30 round=0.01;
fcs reg(diff4=lesion);
fcs reg(diff12=lesion diff4);
fcs reg(diff24=lesion diff4 diff12);
fcs reg(diff52=lesion diff4 diff12 diff24);
var lesion diff4 diff12 diff24 diff52;
by treat;
run;
```

5. Details on intermediate steps:

▷ **Dichotomization of imputed data:**

```
proc sort data=m.armd13a;
by _imputation_ subject;
run;

data m.armd13a;
set m.armd13a;
bindif4=0;  if diff4 <= 0   then bindif4=1;
bindif12=0; if diff12 <= 0  then bindif12=1;
bindif24=0; if diff24 <= 0  then bindif24=1;
bindif52=0; if diff52 <= 0  then bindif52=1;
if diff4=. then bindif4=.;
if diff12=. then bindif12=.;
if diff24=. then bindif24=.;
if diff52=. then bindif52=.;
run;
```

▷ Transforming horizontal dataset in vertical dataset:

```
data m.armd13b;
set m.armd13a;
array x (4) bindif4 bindif12 bindif24 bindif52;
array y (4) diff4 diff12 diff24 diff52;
do j=1 to 4;
   bindif=x(j);
   diff=y(j);
   time=j;
   output;
end;
run;
```

▷ Creating dummies:

```
data m.armd13c;
set m.armd13b;
time1=0;
time2=0;
time3=0;
time4=0;
```

```
trttime1=0;
trttime2=0;
trttime3=0;
trttime4=0;
if time=1 then time1=1;
if time=2 then time2=1;
if time=3 then time3=1;
if time=4 then time4=1;
if (time=1 & treat=1) then trttime1=1;
if (time=2 & treat=1) then trttime2=1;
if (time=3 & treat=1) then trttime3=1;
if (time=4 & treat=1) then trttime4=1;
run;

proc sort data=m.armd13cs;
by _imputation_ subject time;
run;
```

6. The analysis task (GEE):

```
proc gee data=armd13c;
class time subject;
by _imputation_;
model bindif = time1 time2 time3 time4 trttime1 trttime2 trttime3 trttime4
      / noint dist=binomial covb;
repeated subject=subject / withinsubject=time type=exch modelse;
ods output GEEEmpPEst=gmparms parminfo=gmpinfo CovB=gmcovb;
run;
```

Deletion of redundant parameter information:

```
data gmpinfo;
set gmpinfo;
if parameter='Prm1' then delete;
run;
```

7. The analysis task (GLMM):

```
proc nlmixed data=armd13c qpoints=20 maxiter=100 technique=newrap cov ecov;
by _imputation_;
eta = beta11*time1+beta12*time2+beta13*time3+beta14*time4+b
     +beta21*trttime1+beta22*trttime2+beta23*trttime3+beta24*trttime4;
p = exp(eta)/(1+exp(eta));
model bindif ~ binary(p);
random b ~ normal(0,tau*tau) subject=subject;
estimate 'tau2' tau*tau;
ods output ParameterEstimates=nlparms
           CovMatParmEst=nlcovb
           AdditionalEstimates=nlparmsa
           CovMatAddEst=nlcovba;
run;
```

8. The inference task (GEE):

```
proc mianalyze parms=gmparms covb=gmcovb parminfo=gmpinfo wcov bcov tcov;
modeleffects time1 time2 time3 time4 trttime1 trttime2 trttime3 trttime4;
run;
```

9. The inference task (GLMM):

```
proc mianalyze parms=nlparms covb=nlcovb wcov bcov tcov;
modeleffects beta11 beta12 beta13 beta14 beta21 beta22 beta23 beta24;
run;
```

Chapter 15
Creating Monotone Missingness

- When missingness is non-monotone, one might think of several mechanisms operating simultaneously:
  - ▷ A simple (MCAR or MAR) mechanism for the intermittent missing values
  - ▷ A more complex (MNAR) mechanism for the missing data past the moment of dropout
- Analyzing such data are complicated, especially with methods that apply to dropout only

- Solution:
  - ▷ Generate multiple imputations that render the datasets monotone missing, by including into the MI procedure:

```
mcmc impute=monotone;
```

  - ▷ Apply method of choice to the so-completed multiple sets of data

- Note: this is different from the monotone method in PROC MI, intended to fully complete already monotone sets of data:
  - The MONOTONE statement takes monotone patterns as input and returns completed data.
  - The MCMC statement with `impute=monotone` option takes data with also non-monotone patterns as input and returns monotonized data.

15.1 Example: Creating Monotone Missingness to Then Apply Weighted GEE

- Consider again the analgesic trial
- Multiple imputation to create monotone missingness:

```
proc mi data=m.gsa4 seed=459864 simple nimpute=10
        round=0.1 out=m.gsaimput;
title 'Monotone multiple imputation';
mcmc impute = monotone;
var pca0 physfct gsa1 gsa2 gsa3 gsa4;
run;
```

- Preparation of the data in vertical format, so that the data can be used in ordinary GEE:

```
data m.gsaw;
set m.gsa4;
array y (4) gsa1 gsa2 gsa3 gsa4;
do j=1 to 4;
   gsa=y(j);
   time=j;
   timecls=time;
   gsabin=.;
   if gsa=1 then gsabin=1;
   if gsa=2 then gsabin=1;
   if gsa=3 then gsabin=1;
   if gsa=4 then gsabin=0;
   if gsa=5 then gsabin=0;
   output;
end;
run;
```

- Standard GEE:

```
proc gee data=m.gsaw plots=histogram;
   title 'Standard GEE for GSA data';
   class patid timecls;
   model gsabin = time|time pca0 / dist=bin;
   repeated subject=patid / within=timecls corr=un;
run;
```

- Steps to prepare the data for weighted GEE, including definition of the 'previous' outcome:

```
data m.gsaimput02;
set m.gsaimput;
array y (4) gsa1 gsa2 gsa3 gsa4;
do j=1 to 4;
   gsa=y(j);
   time=j;
   timecls=time;
```

```
   patid2=1000*_imputation_+patid;
   output;
end;
run;

proc sort data=m.gsaimput02;
by _imputation_ patid2;
run;

data m.gsaimput03;
   set m.gsaimput02;
   by patid2;
   prevgsa = lag(gsa);
   if time=1 then prevgsa = 1;
   timeclss = time;
run;
```

```
data m.gsaimput03;
   set m.gsaimput03;
   if gsa<=3.5 then gsabin=1;
   if gsa>3.5 then gsabin=0;
   gsabin=gsabin+gsa-gsa;
   run;
```

- Weighted GEE, where weights are created at observation level:

```
ods graphics on;
proc gee data=m.gsaimput03 plots=histogram;
   title 'Weighted GEE for GSA Data Based on Multiple
          Imputation to Monotonize - OBSLEVEL';
   by _imputation_;
   class patid timecls;
   model gsabin = time|time pca0 / dist=bin covb;
   repeated subject=patid / within=timecls corr=un ecovb;
   missmodel prevgsa pca0 physfct  / type=obslevel;
   ods output GEEEmpPEst=gmparms parminfo=gmpinfo
              modelinfo=modelinfo GEERCov=gmcovb;
run;

proc mianalyze parms=gmparms  parminfo=gmpinfo covb=gmcovb;
title 'Multiple Imputation Analysis After Weighted GEE for GSA Data';
modeleffects intercept time time*time pca0;
run;
```

- To use weights at subject rather than observation level:

```
missmodel prevgsa pca0 physfct  / type=sublevel;
```

- Evidently, using these monotonized data, also standard GEE can be used:

```
ods graphics on;
proc gee data=m.gsaimput03 plots=histogram;
   title 'Standard GEE for GSA Data Based on
          Multiple Imputation to Monotonize';
   by _imputation_;
   class patid timecls;
   model gsabin = time|time pca0 / dist=bin covb;
   repeated subject=patid / within=timecls corr=un ecovb;
   ods output GEEEmpPEst=gmparms parminfo=gmpinfo
              modelinfo=modelinfo GEERCov=gmcovb;
run;
```

- Files:
  - ▷ `analg11(met-proc-gee).sas`
  - ▷ `analg11(met-proc-gee).lst`
- Overview of results:

Effect	Par.	Est.(s.e.)	p-value	Est.(s.e.)	p-value
		Standard GEE			
		Without MI		After MI	
Intercept	β_0	2.90(0.46)		2.87(0.45)	
Time	β_1	-0.81(0.32)	0.0124	-0.83(0.32)	0.0087
Time^2	β_2	0.17(0.07)	0.0083	0.18(0.06)	0.0058
PCA$_0$	β_3	-0.23(0.10)	0.0178	-0.21(0.10)	0.0253
		Weighted GEE (after MI)			
		Observation level		Subject level	
Intercept	β_0	2.74(0.46)		2.62(0.60)	
Time	β_1	-0.76(0.33)	0.0231	-0.71(0.40)	0.0747
Time^2	β_2	0.17(0.07)	0.0155	0.16(0.08)	0.0444
PCA$_0$	β_3	-0.19(0.10)	0.0384	-0.21(0.12)	0.0853

- The dropout model is similar but slightly different than the one used with PROC GENMOD.

- Weighted GEE leads to increased standard errors, as observed before.

- This effect is less pronounced when weigths are constructed at observation level, rather than at subject level.

- A typical output for one of the imputed datasets takes the form (first imputation out of ten; with weights at observation level):

Parameter Estimates for Response Model
with Empirical Standard Error

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	2.4299	0.5890	1.2755	3.5843	4.13	<.0001
TIME	-0.5881	0.3912	-1.3548	0.1787	-1.50	0.1328
TIME*TIME	0.1392	0.0794	-0.0165	0.2949	1.75	0.0796
PCA0	-0.1797	0.1173	-0.4096	0.0501	-1.53	0.1254

Parameter Estimates for Missingness Model

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	3.1335	0.4060	2.3377	3.9293	7.72	<.0001
prevgsa	-0.1974	0.0822	-0.3585	-0.0363	-2.40	0.0163
PCA0	-0.2495	0.0956	-0.4370	-0.0621	-2.61	0.0091
PHYSFCT	-0.0079	0.0037	-0.0151	-0.0007	-2.16	0.0311

Chapter 16
Sensitivity Analysis Based on Multiple Imputation

- Multiple imputation in its basic form: MAR

- Various ways to deviate from this:
  - ▷ Apply shift and/or inflation factor to imputed data (in some groups)
  - ▷ Apply inflation factor to imputed data (in some groups)
  - ▷ Use a 'placebo' rather than an 'active' predictive distributions
  - ▷ More generally, base predictive distribution on any subset of your choice
  - ▷ Use pattern-mixture models with NCMV, CCMV,...

- Implemented in PROC MI using the MNAR statement.

- SAS code for a **shift** to the treatment group in the ARMD data:

```
proc mi data=m.armd13 seed=486048 simple out=m.armd13as1
        nimpute=10 round=0.1;
   title 'Shift multiple imputation';
   class treat;
   var lesion diff4 diff12 diff24 diff52;
   fcs reg;
   mnar adjust (diff12 / shift=10 adjustobs=(treat='2'));
   mnar adjust (diff24 / shift=15 adjustobs=(treat='2'));
   mnar adjust (diff52 / shift=20 adjustobs=(treat='2'));
   by treat;
run;
```

- SAS code for a **subgroup adjustment** in the ARMD data:

```
proc mi data=m.armd13 seed=486048 simple out=m.armd13as2 nimpute=10;
title 'Model multiple imputation';
class treat;
var lesion diff4 diff12 diff24 diff52;
fcs reg;
mnar model (diff4  / modelobs= (treat='1'));
mnar model (diff12 / modelobs= (treat='1'));
mnar model (diff24 / modelobs= (treat='1'));
mnar model (diff52 / modelobs= (treat='1'));
run;
```

- Suppose we want to undertake NCMV adjustment:
  - ▷ The method can be applied to monotone data only
  - ▷ First MI (10 imputations): Start by making the data monotone (under MAR)
  - ▷ Second MI (1 imputation): Then apply NCMV to the monotonized data
  - ▷ The end result is 10 imputations, as we want

- SAS code for **NCMV** in the ARMD data:

```
proc mi data=m.armd13 seed=486048 simple out=m.armd13as3 nimpute=10;
title 'Montone imputation';
var lesion diff4 diff12 diff24 diff52;
mcmc impute=monotone;
by treat;
run;

proc mi data=m.armd13as3 seed=486048 simple out=m.armd13as4 nimpute=1;
title 'Model multiple imputation';
var lesion diff4 diff12 diff24 diff52;
monotone reg;
mnar model (diff4 diff12 diff24 diff52 / modelobs=ncmv);
by treat;
run;
```

- In the latter case, SAS prints what predictive distributions have been used:

```
Observations Used for Imputation Models Under MNAR Assumption

Imputed
Variable    Observations

diff4       Nonmissing lesion, diff4; Missing diff12, ..., diff52
diff12      Nonmissing lesion, ..., diff12; Missing diff24, diff52
diff24      Nonmissing lesion, ..., diff24; Missing diff52
diff52      Complete Cases
```

- Results:
  - ▷ GEE
  - ▷ GLMM

Effect	Par.	MAR	shift	placebo	NCMV
Generalized estimating equations					
Int.4	β_{11}	-0.82(0.20)	-0.73(0.20)	-0.81(0.20)	-0.83(0.21)
Int.12	β_{21}	-0.97(0.22)	-0.71(0.19)	-0.98(0.22)	-1.06(0.21)
Int.24	β_{31}	-1.07(0.23)	-0.56(0.19)	-1.05(0.22)	-1.00(0.22)
Int.52	β_{41}	-1.66(0.27)	-0.82(0.20)	-1.58(0.29)	-1.59(0.27)
Trt.4	β_{12}	0.17(0.29)	0.07(0.28)	0.17(0.28)	0.17(0.29)
Trt.12	β_{22}	0.56(0.29)	0.29(0.27)	0.56(0.29)	0.67(0.28)
Trt.24	β_{32}	0.41(0.30)	-0.10(0.27)	0.39(0.29)	0.34(0.29)
Trt.52	β_{42}	0.41(0.35)	-0.43(0.30)	0.32(0.35)	0.32(0.35)
Generalized linear mixed models					
Int.4	β_{11}	-1.46(0.36)	-1.32(0.36)	-1.39(0.35)	-1.42(0.35)
Int.12	β_{21}	-1.75(0.38)	-1.27(0.35)	-1.67(0.37)	-1.80(0.36)
Int.24	β_{31}	-1.83(0.38)	-1.01(0.34)	-1.78(0.38)	-1.70(0.38)
Int.52	β_{41}	-2.71(0.45)	-1.47(0.36)	-2.62(0.46)	-2.64(0.44)
Trt.4	β_{12}	0.32(0.48)	0.12(0.50)	0.25(0.48)	0.24(0.48)

Chapter 17
Overview

MCAR/simple	CC LOCF	biased inefficient not simpler than MAR methods
MAR	**direct likelihood** **direct Bayes** **weighted GEE** **MI**	easy to conduct Gaussian & non-Gaussian
MNAR	variety of methods	strong, untestable assumptions most useful in **sensitivity analysis**