

Introduction to Longitudinal Data Analysis

Geert Verbeke

Geert Molenberghs

Interuniversity Institute for Biostatistics and statistical Bioinformatics
K.U.Leuven and U.Hasselt

geert.verbeke@med.kuleuven.be

http://perswww.kuleuven.be/geert_verbeke

geert.molenberghs@uhasselt.be

www.censtat.uhasselt.be

Master of Statistics

Contents

1	Related References	1
I	Continuous Longitudinal Data	7
2	Introduction	8
3	Cross-sectional versus Longitudinal Data	39
4	Simple Methods	52
5	The Multivariate Regression Model	62
6	A Model for Longitudinal Data	90

7	Exploratory Data Analysis	121
8	Estimation of the Marginal Model	153
9	Inference for the Marginal Model	179
10	Inference for the Random Effects	234
11	General Guidelines for Model Building	258
12	Power Analyses under Linear Mixed Models	291
II	Marginal Models for Non-Gaussian Longitudinal Data	310
13	The Toenail Data	311
14	The Analgesic Trial	314
15	The National Toxicology Program (NTP) Data	317
16	Generalized Linear Models	320
17	Parametric Modeling Families	338

18	Conditional Models	342
19	Full Marginal Models	358
20	Generalized Estimating Equations	370
21	A Family of GEE Methods	382
III	Generalized Linear Mixed Models for Non-Gaussian Longitudinal Data	402
22	The Beta-binomial Model	403
23	Generalized Linear Mixed Models (GLMM)	409
24	Fitting GLMM's in SAS	427
IV	Marginal Versus Random-effects Models and Case Studies	437
25	Marginal Versus Random-effects Models	438
26	Case Study: The NTP Data	451
27	Case Study: Binary Analysis of Analgesic Trial	473

28	Case Study: Ordinal Analysis of Analgesic Trial	487
29	Count Data: The Epilepsy Study	502
V	Incomplete Data	528
30	Setting The Scene	529
31	Proper Analysis of Incomplete Data	542
32	Weighted Generalized Estimating Equations	565
33	Multiple Imputation	583
34	Creating Monotone Missingness	598
VI	Topics in Methods and Sensitivity Analysis for Incomplete Data	600
35	An MNAR Selection Model and Local Influence	601
36	Mechanism for Growth Data	615
37	Interval of Ignorance	618

38	Pattern-mixture Models	643
39	Concluding Remarks	658

Chapter 1

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.
- 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., Laird, N.M., and Ware, J.H. (2004). *Applied Longitudinal Analysis*. New York: John Wiley & Sons.
- 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.
- 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.
- Longford, N.T. (1993). *Random Coefficient Models*. Oxford: Oxford University Press.
- McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models* (second edition). London: Chapman & Hall.
- Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer-Verlag.
- Pinheiro, J.C. and Bates D.M. (2000). *Mixed effects models in S and S-Plus*. New York: Springer.
- 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.
- 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.

Part I

Continuous Longitudinal Data

Chapter 2

Introduction

- ▷ Repeated Measures / Longitudinal data
- ▷ Examples

2.1 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**

2.2 Captopril Data

- Taken from Hand, Daly, Lunn, McConway, & Ostrowski (1994)
- 15 patients with hypertension
- The response of interest is the supine blood pressure, before and after treatment with CAPTOPRIL

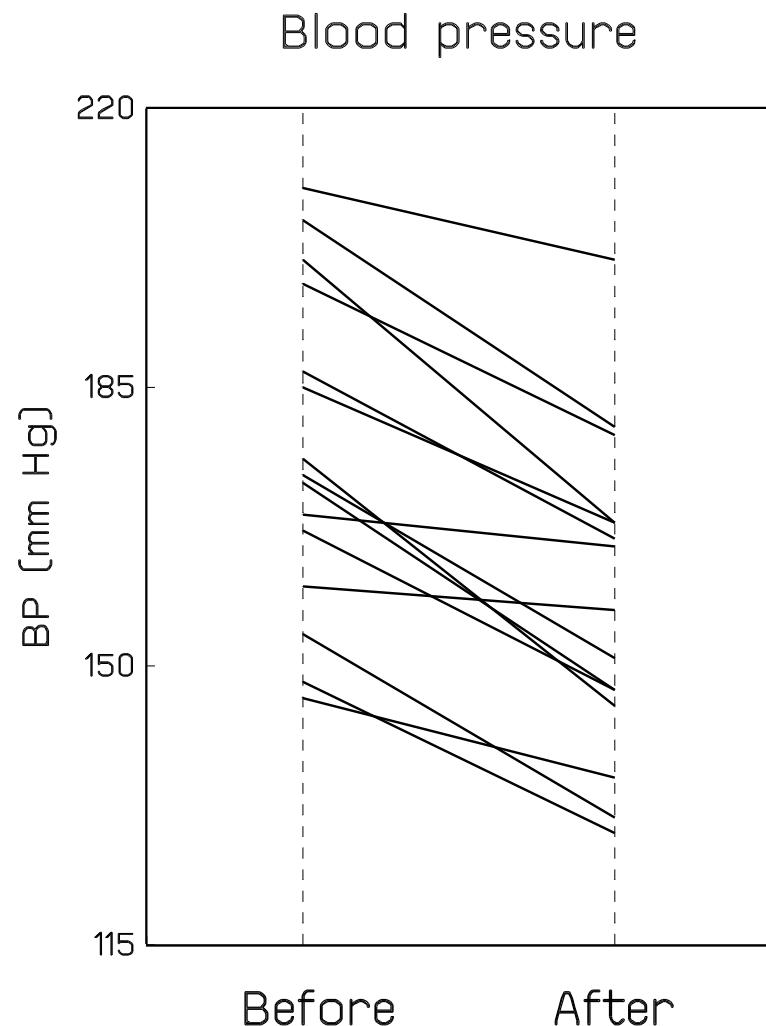
Patiënt	Before		After	
	SBP	DBP	SBP	DBP
1	210	130	201	125
2	169	122	165	121
3	187	124	166	121
4	160	104	157	106
5	167	112	147	101
6	176	101	145	85
7	185	121	168	98
8	206	124	180	105
9	173	115	147	103
10	146	102	136	98
11	174	98	151	90
12	201	119	168	98
13	198	106	179	110
14	148	107	129	103
15	154	100	131	82

- Research question:

How does treatment affect BP ?

- Remarks:

- ▷ Paired observations:
Most simple example of longitudinal data
- ▷ Much variability between subjects



2.3 Growth Curves

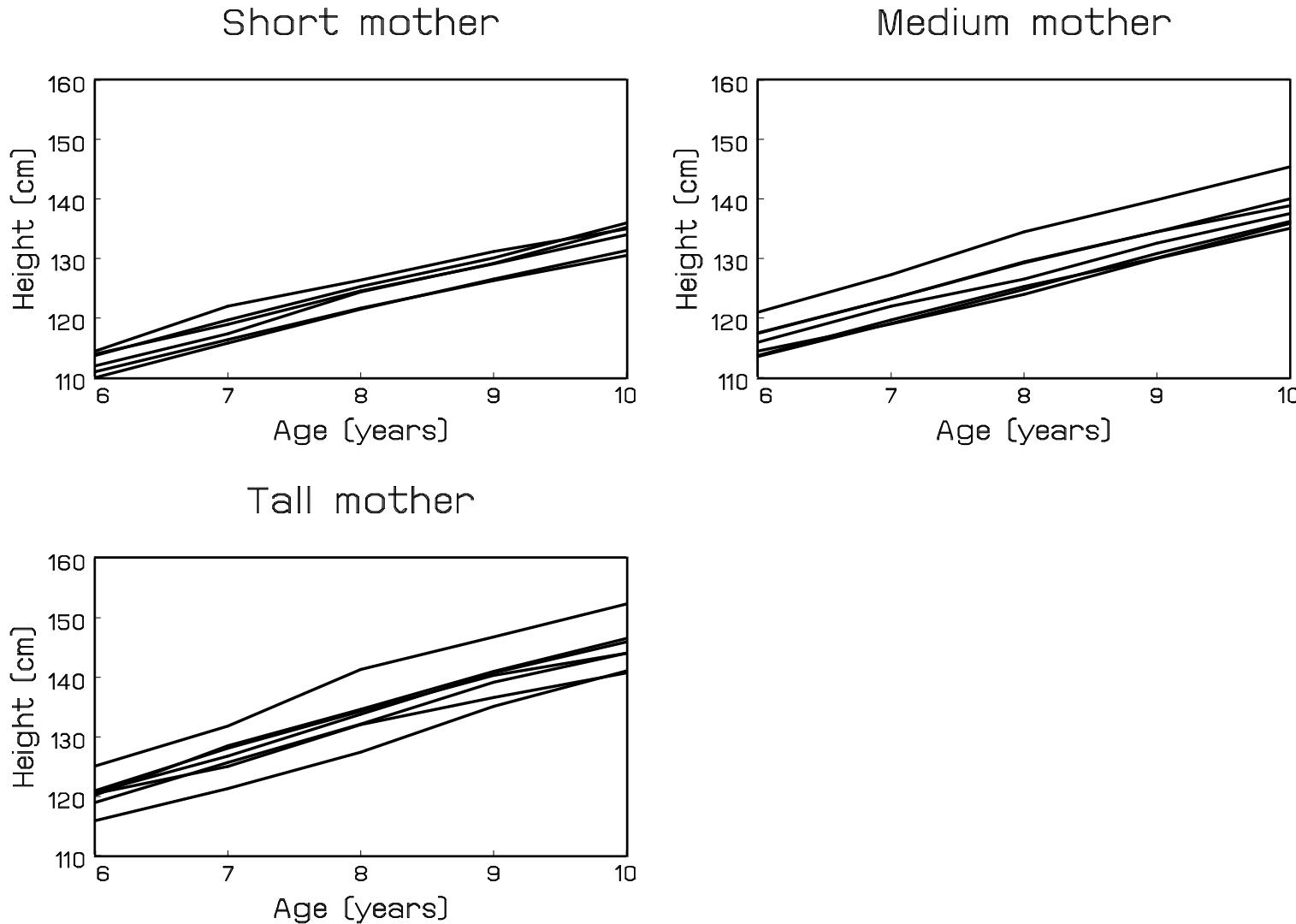
- Taken from Goldstein 1979
- The height of 20 schoolgirls, with small, medium, or tall mothers, was measured over a 4-year period:

	Mothers height	Children numbers
Small mothers	< 155 cm	1 → 6
Medium mothers	[155cm; 164cm]	7 → 13
Tall mothers	> 164 cm	14 → 20

- Research question:

Is growth related to height of mother ?

- Individual profiles:



- Remarks:

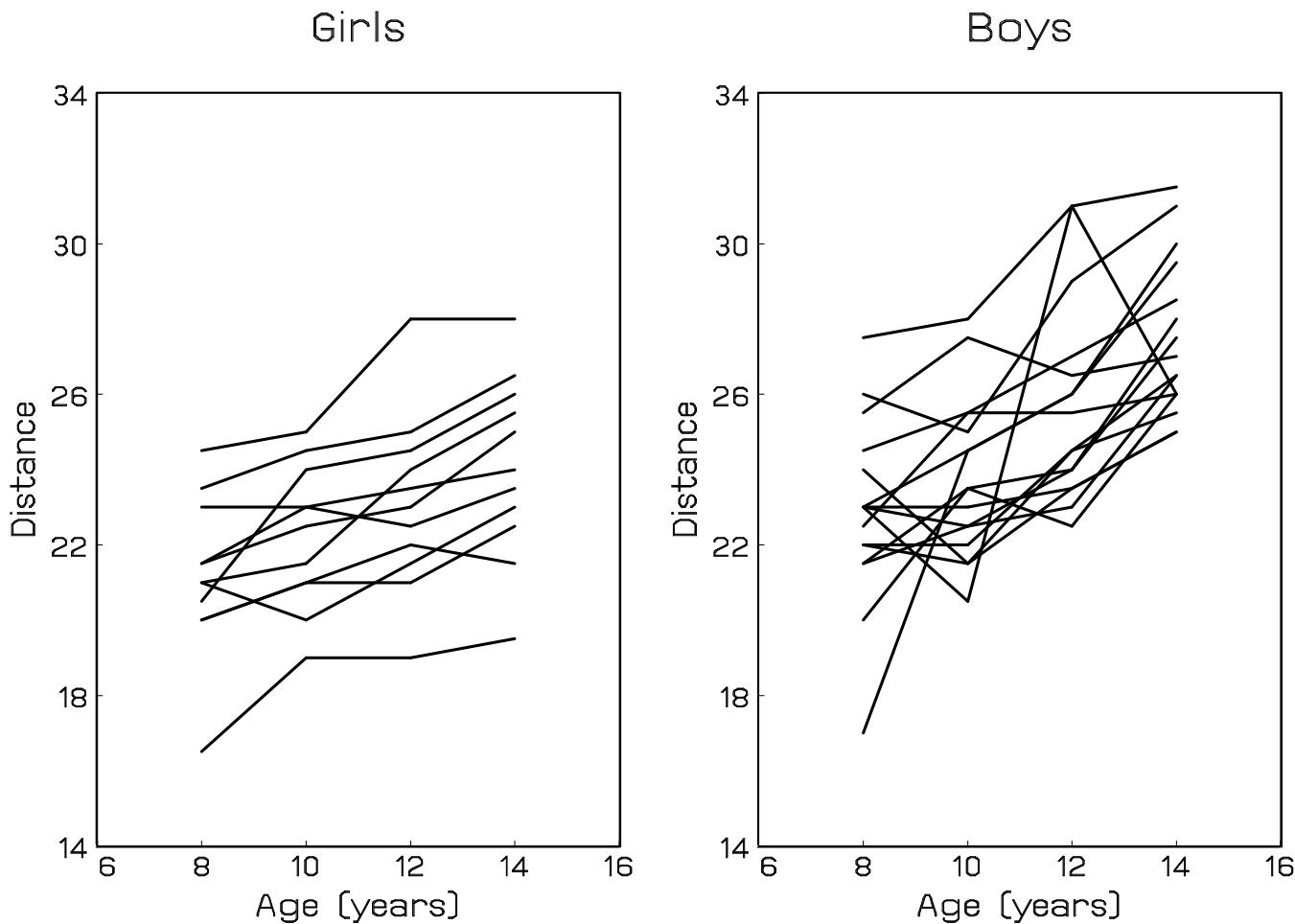
- ▷ Almost perfect linear relation between Age and Height
- ▷ Much variability between girls
- ▷ Little variability within girls
- ▷ Fixed number of measurements per subject
- ▷ Measurements taken at fixed time points

2.4 Growth Data

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

Is dental growth related to gender ?

- Individual profiles:



- Remarks:

- ▷ Much variability between children
- ▷ Considerable variability within children
- ▷ Fixed number of measurements per subject
- ▷ Measurements taken at fixed time points

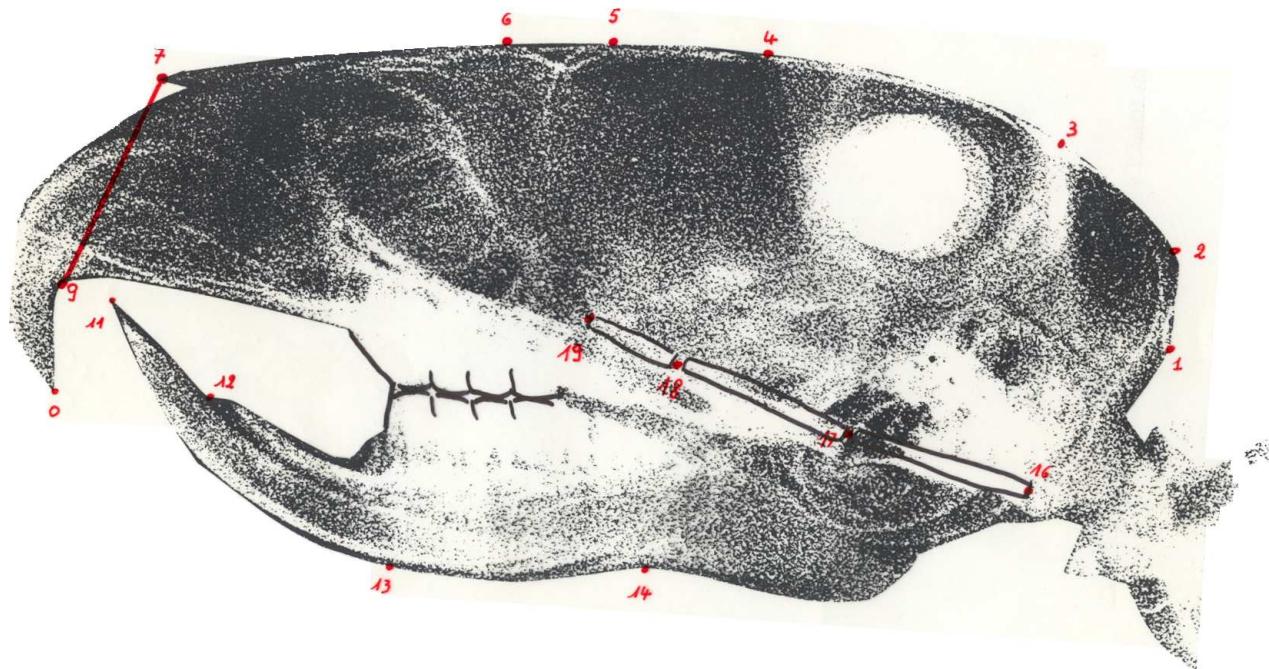
2.5 Rat Data

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

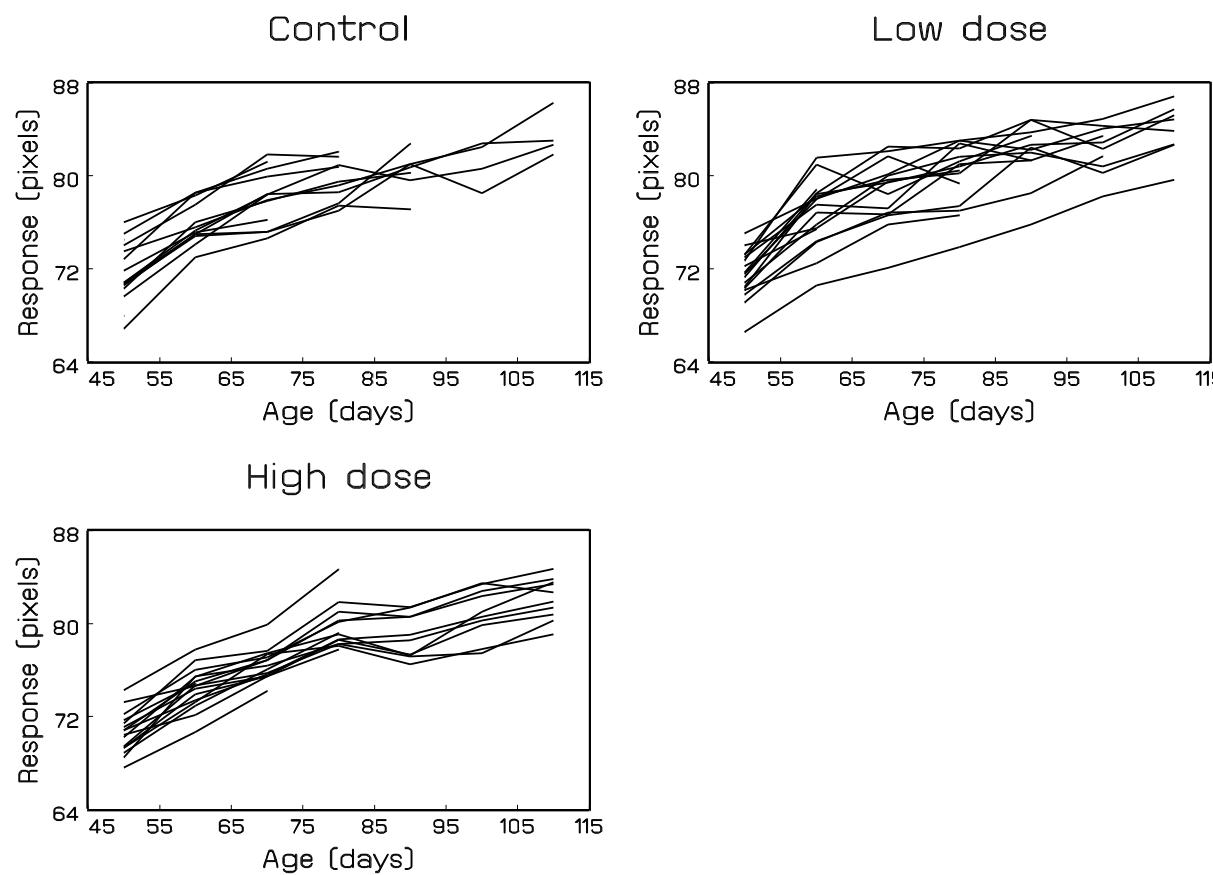
How does craniofacial growth depend on
testosteron production ?

- 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%):

Age (days)	# Observations				Total
	Control	Low	High		
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

2.6 Toenail Data

- Reference: De Backer, De Keyser, De Vroey, Lesaffre, British Journal of Dermatology (1996).
- **Toenail Dermatophyte Onychomycosis:** Common toenail infection, difficult to treat, affecting more than 2% of population.
- Classical treatments with antifungal compounds need to be administered until the whole nail has grown out healthy.
- New compounds have been developed which reduce treatment to 3 months
- Randomized, double-blind, parallel group, multicenter study for the comparison of two such new compounds (*A* and *B*) for oral treatment.

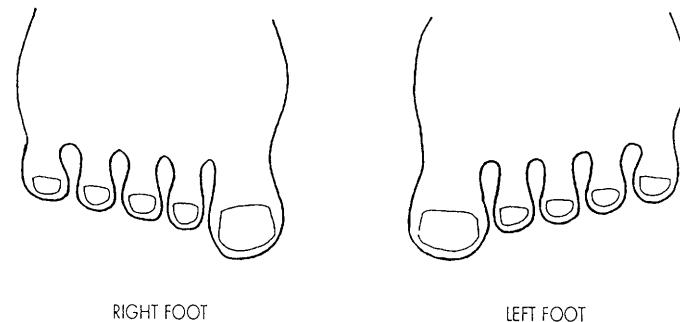
- Research question:

Are both treatments equally effective for
the treatment of TDO ?

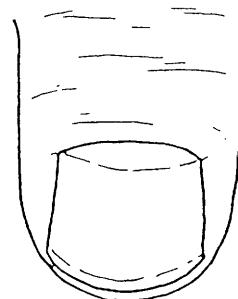
- 2×189 patients randomized, 36 centers
- 48 weeks of total follow up (12 months)
- 12 weeks of treatment (3 months)
- Measurements at months 0, 1, 2, 3, 6, 9, 12.

- Response considered here: Unaffected nail length (mm):

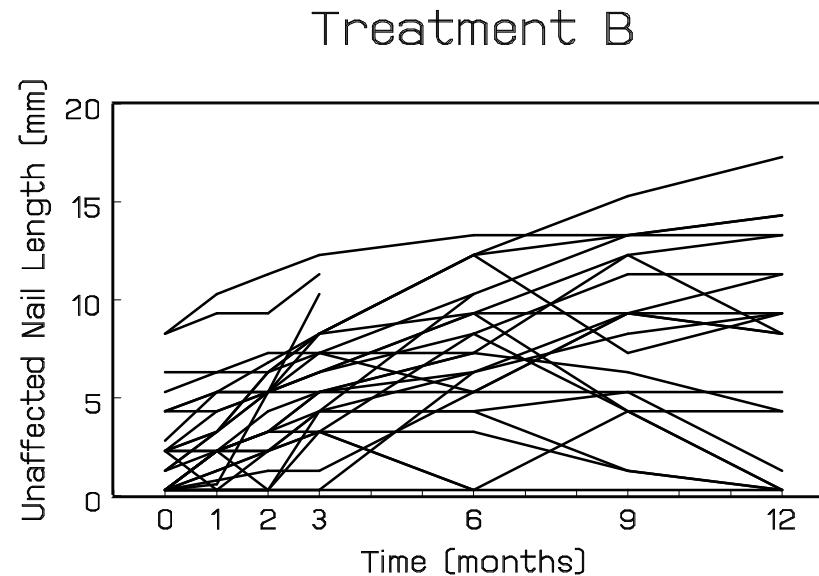
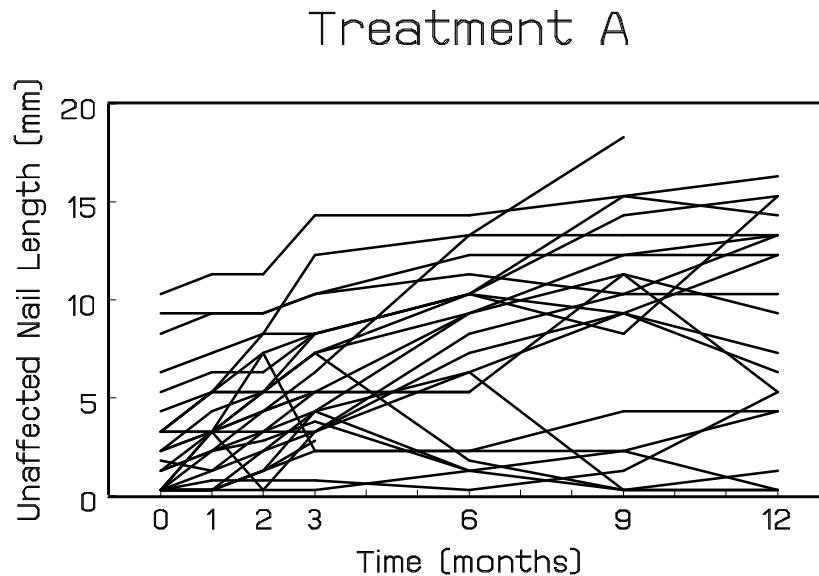
Please mark each infected nail with a "X"
Indicate the TARGET nail, which you selected at baseline, for assessment with a circle.



Please indicate on the diagram the margin of the unaffected target nail.



- As response is related to toe size, we restrict to patients with big toenail as target nail \Rightarrow 150 and 148 subjects.
- 30 randomly selected profiles, in each group:



- Complication: Dropout (24%):

Time (months)	# Observations		
	Treatment A	Treatment B	Total
0	150	148	298
1	149	142	291
2	146	138	284
3	140	131	271
6	131	124	255
9	120	109	229
12	118	108	226

- Remarks:

- ▷ Much variability between subjects
- ▷ Much variability within subjects
- ▷ Fixed number of measurements scheduled per subject, but not all measurements available due to dropout, for unknown reason.
- ▷ Measurements taken at fixed time points

2.7 Mastitis in Dairy Cattle

- Taken from Diggle & Kenward, Applied statistics (1994)
- Mastitis : Infectious disease, typically reducing milk yields
- Research question:

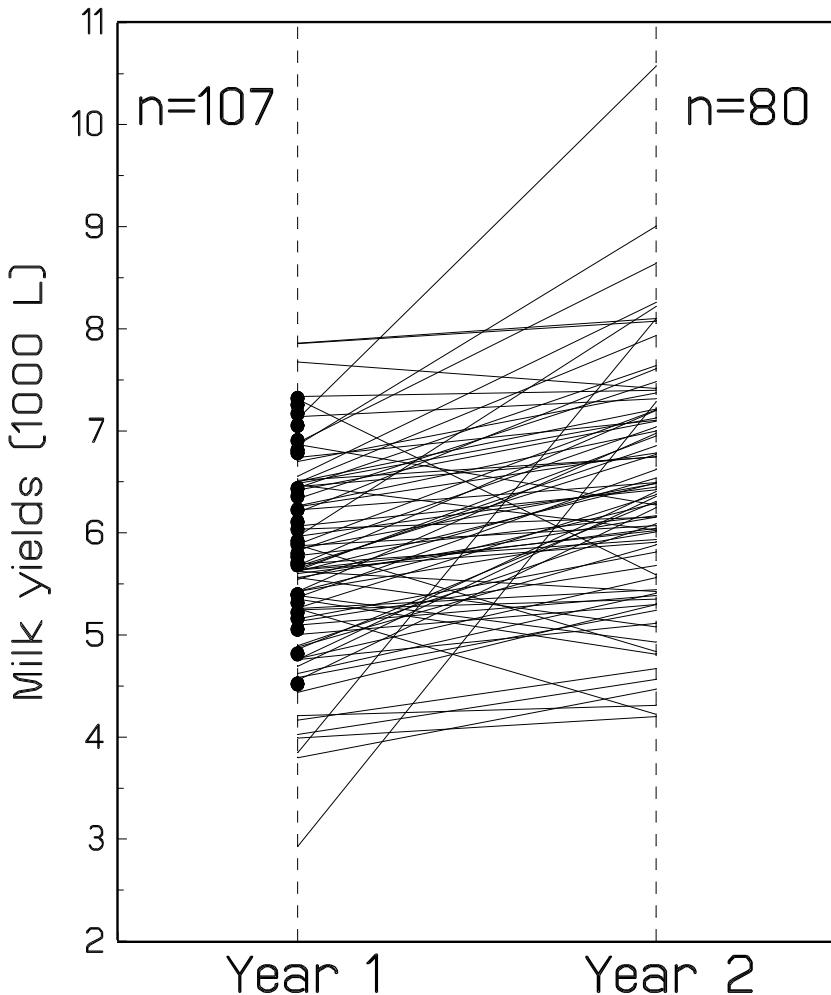
Are high yielding cows more susceptible ?

- Hence, is the probability of occurrence of mastitis related to the yield that would have been observed had mastitis not occurred ?
- Hypothesis cannot be tested directly since ‘covariate is missing for all events’

- Individual profiles:

- Remarks:

- ▷ Paired observations: Most simple example of longitudinal data
- ▷ Much variability between cows
- ▷ Missingness process itself is of interest



2.8 The Baltimore Longitudinal Study of Aging (BLSA)

- Reference: Shock, Greulich, Andres, Arenberg, Costa, Lakatta, & Tobin, National Institutes of Health Publication, Washington, DC: National Institutes of Health (1984).
- BLSA: Ongoing, multidisciplinary observational study, started in 1958, with the study of normal human aging as primary objective
- Participants:
 - ▷ volunteers, predominantly white, well educated, and financially comfortable
 - ▷ return approximately every 2 years for 3 days of biomedical and psychological examinations
 - ▷ at first only males (over 1500 by now), later also females
 - ▷ an average of almost 7 visits and 16 years of follow-up

- The BLSA is a unique resource for rapidly evaluating longitudinal hypotheses:
 - ▷ data from repeated clinical examinations
 - ▷ a bank of frozen blood and urine samples
- Drawbacks of such observational studies:
 - ▷ More complicated analyses needed (see later)
 - ▷ Observed evolutions may be highly influenced by many covariates which may or may not be recorded in the study

2.8.1 Prostate Data

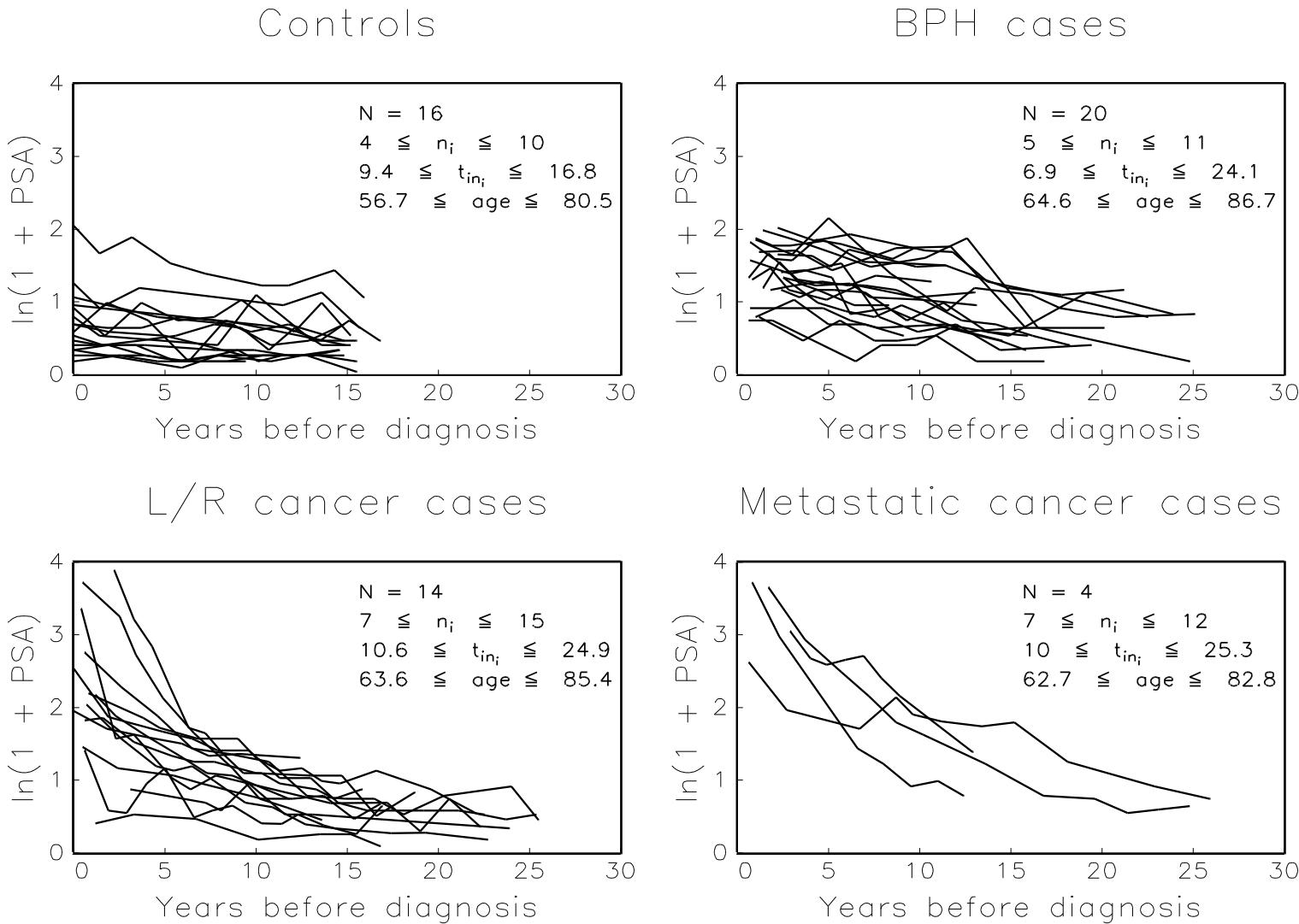
- References:
 - ▷ Carter *et al* (1992, Cancer Research).
 - ▷ Carter *et al* (1992, Journal of the American Medical Association).
 - ▷ Morrell *et al* (1995, Journal of the American Statistical Association).
 - ▷ Pearson *et al* (1994, Statistics in Medicine).
- Prostate disease is one of the most common and most costly medical problems in the United States
- Important to look for markers which can detect the disease at an early stage
- **Prostate-Specific Antigen** is an enzyme produced by both normal and cancerous prostate cells

- PSA level is related to the volume of prostate tissue.
- Problem: Patients with **Benign Prostatic Hyperplasia** 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

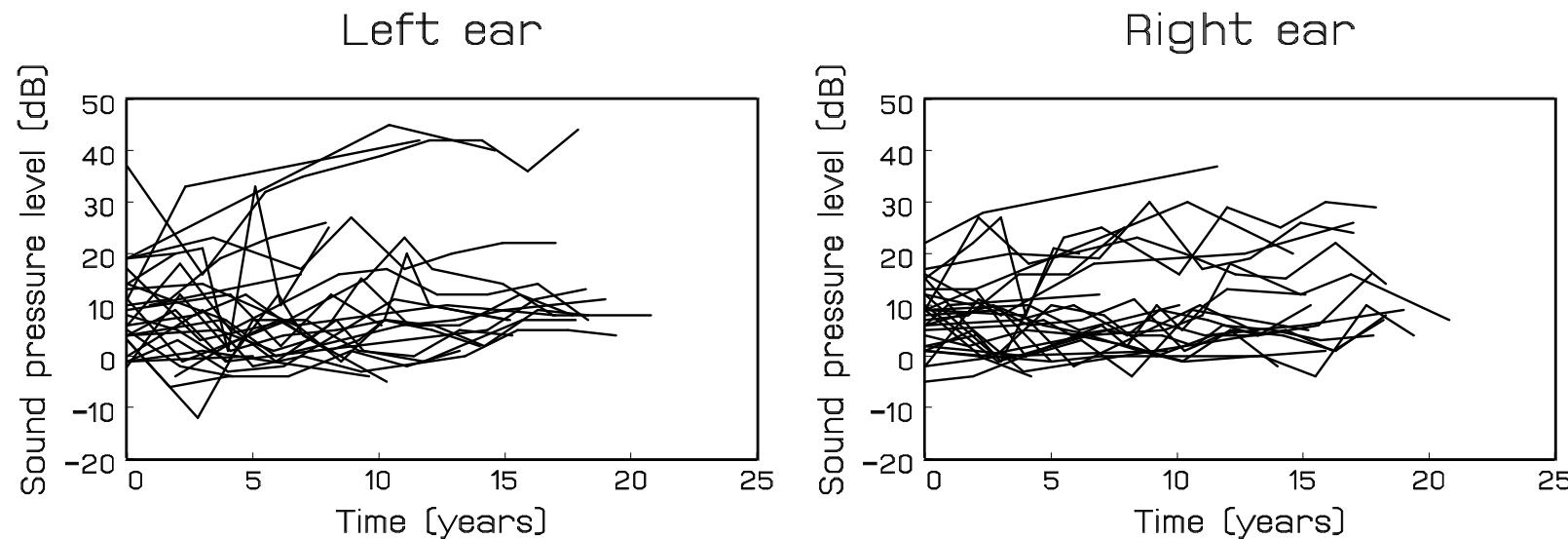
2.8.2 Hearing Data

- References:
 - ▷ Brant & Fozard, Journal of the Acoustic Society of America (1990).
 - ▷ Morrell & Brant, Statistics in Medicine (1991).
- Hearing thresholds, by means of sound proof chamber and Bekesy audiometer
- 11 frequencies : $125 \rightarrow 8000$ Hz, both ears
- Research question:

How does hearing depend on aging ?

- Data considered here:
 - ▷ 500 Hz
 - ▷ 6170 observations (3089 left ear, 3081 right ear) from 681 males without any otologic disease
 - ▷ followed for up to 22 years, with a maximum of 15 measurements/subject

- 30 randomly selected profiles, for each ear:



- Remarks:

- ▷ Much variability between subjects
- ▷ Much variability within subjects
- ▷ Highly unbalanced data

Chapter 3

Cross-sectional versus Longitudinal Data

- ▷ Introduction
- ▷ Paired verus unpaired t -test
- ▷ Cross-sectional versus longitudinal data

3.1 Introduction

- The examples have illustrated several aspects of longitudinal data structures:
 - ▷ Experimental and observational
 - ▷ Balanced and unbalanced
 - ▷ With or without missing data (dropout)
- Often, there is far more variability between subjects than within subjects.
- This is also reflected in correlation within units

- For example, for the growth curves, the correlation matrix of the 5 repeated measurements equals

$$\begin{pmatrix} 1.00 & 0.95 & 0.96 & 0.93 & 0.87 \\ 0.95 & 1.00 & 0.97 & 0.96 & 0.89 \\ 0.96 & 0.97 & 1.00 & 0.98 & 0.94 \\ 0.93 & 0.96 & 0.98 & 1.00 & 0.98 \\ 0.87 & 0.89 & 0.94 & 0.98 & 1.00 \end{pmatrix}$$

- This correlation structure cannot be ignored in the analyses ([Section 3.2](#))
- The advantage however is that longitudinal data allow to study changes within subjects ([Section 3.3](#)).

3.2 Paired versus Unpaired t -test

3.2.1 Paired t -test

- The simplest case of longitudinal data are paired data
- We re-consider the diastolic blood pressures from the Captopril data
- The data can be summarized as:

Descriptive Statistics (capto.sta)					
<u>Continue...</u>	Valid N	Mean	Minimum	Maximum	Std. Dev.
DIA_VOR	15	112.3333	98.00000	130.0000	10.47219
DIA_NA	15	103.0667	82.00000	125.0000	12.55540

- There is an average decrease of more than 9 mmHG
- The classical analysis of paired data is based on comparisons within subjects:

$$\Delta_i = Y_{i1} - Y_{i2}, \quad i = 1, \dots, 15$$

- A positive Δ_i corresponds to a decrease of the BP, while a negative Δ_i is equivalent to an increase.
- Testing for treatment effect is now equivalent to testing whether the average difference μ_Δ equals zero.

- Statistica output:

<u>Continue...</u>	Marked differences are significant at p < .05000							
Variable	Mean	Std. Dv.	N	Diff.	Std. Dv. Diff.	t	df	p
DIA_VOOR	112.33	10.472						
DIA_NA	103.07	12.555	15	9.2667	8.6145	4.1662	14	.0010

- Hence, the average change in BP is statistically, significantly different from zero ($p = 0.001$).

3.2.2 Unpaired, Two-sample, t -test

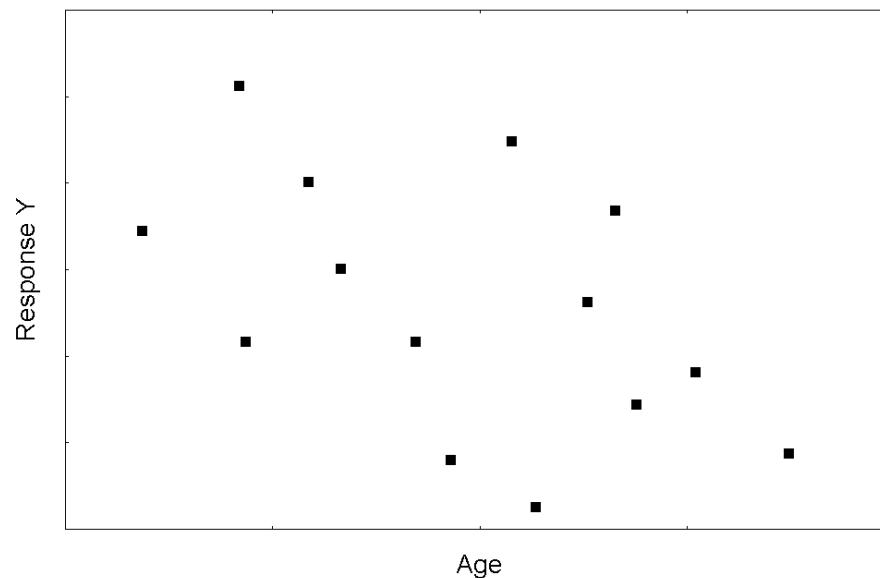
- What if we had ignored the paired nature of the data ?
- We then could have used a two-sample (unpaired) t -test to compare the average BP of untreated patients (controls) with treated patients.
- We would still have found a significant difference ($p = 0.0366$), but the p -value would have been more than $30 \times$ larger compared to the one obtained using the paired t -test ($p = 0.001$).
- Conclusion:

$$15 \times 2 \neq 30 \times 1$$

- The two-sample t -test does not take into account the fact that the 30 measurements are not independent observations.
- This illustrates that classical statistical models which assume independent observations will not be valid for the analysis of longitudinal data

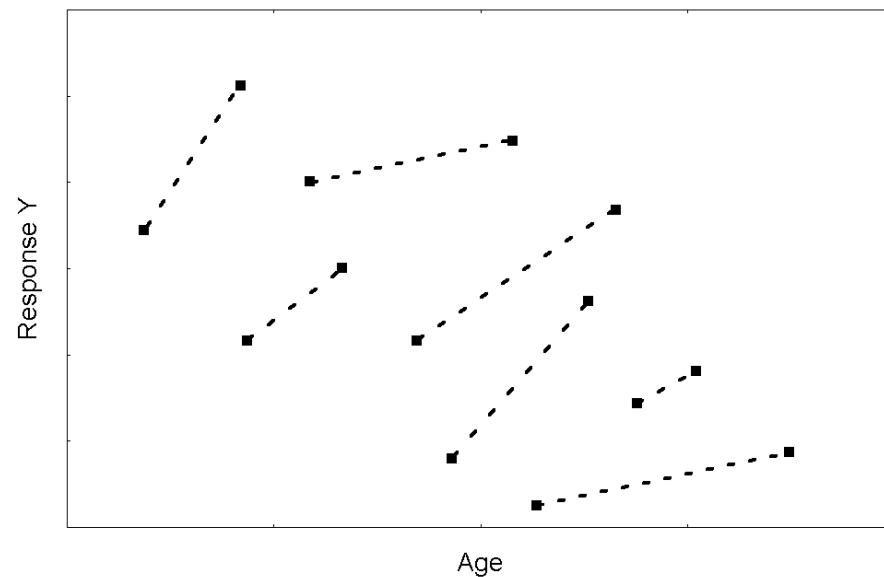
3.3 Cross-sectional versus Longitudinal Data

- Suppose it is of interest to study the relation between some response Y and age
- A cross-sectional study yields the following data:



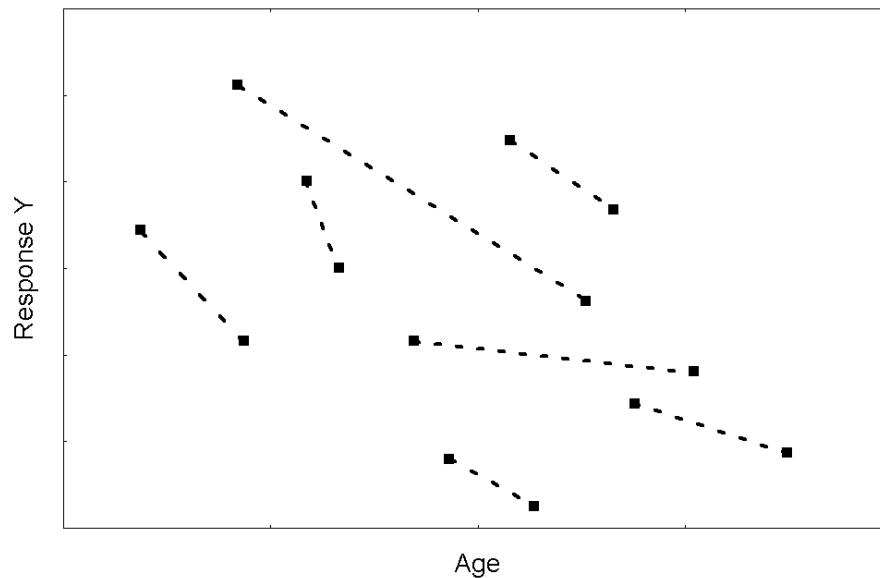
- The graph suggests a negative relation between Y and age.

- Exactly the same observations could also have been obtained in a longitudinal study, with 2 measurements per subject.
- First case:



Are we now still inclined to conclude that there is a negative relation between Y and Age ?

- The graph suggests a negative cross-sectional relation but a positive longitudinal trend.
- Second case:



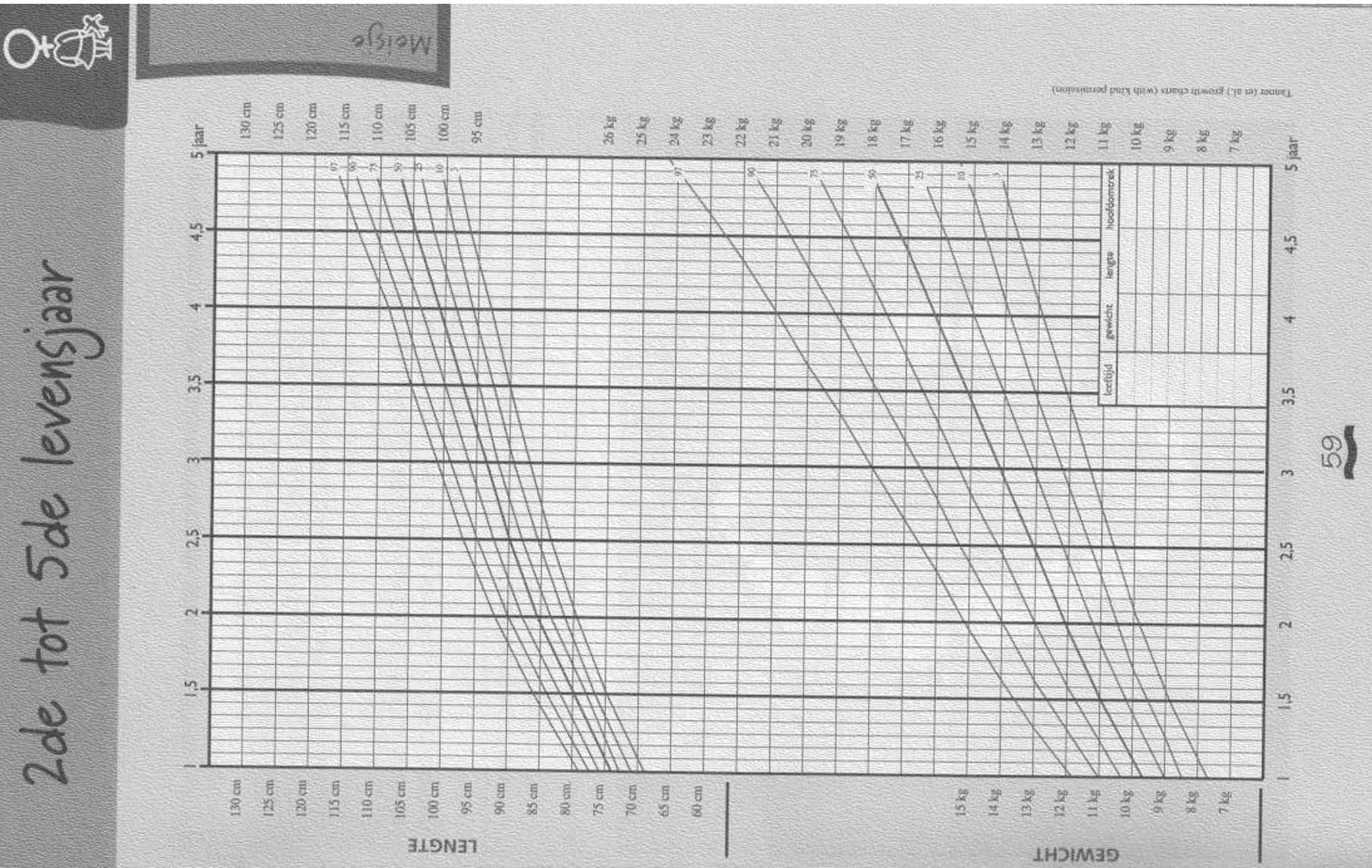
- The graph now suggests the cross-sectional as well as longitudinal trend to be negative.

- Conclusion:

Longitudinal data allow to distinguish differences between subjects from changes within subjects

- Application: Growth curves for babies (next page)

2de tot 5de levensjaar



Chapter 4

Simple Methods

- ▷ Introduction
- ▷ Overview of frequently used methods
- ▷ Summary statistics

4.1 Introduction

- The reason why classical statistical techniques fail in the context of longitudinal data is that observations within subjects are correlated.
- In many cases the correlation between two repeated measurements decreases as the time span between those measurements increases.
- A correct analysis should account for this
- The paired t -test accounts for this by considering subject-specific differences $\Delta_i = Y_{i1} - Y_{i2}$.
- This reduces the number of measurements to just one per subject, which implies that classical techniques can be applied again.

- In the case of more than 2 measurements per subject, similar simple techniques are often applied to reduce the number of measurements for the i th subject, from n_i to 1.
- Some examples:
 - ▷ Analysis at each time point separately
 - ▷ Analysis of Area Under the Curve (AUC)
 - ▷ Analysis of endpoints
 - ▷ Analysis of increments
 - ▷ Analysis of covariance

4.2 Overview of Frequently Used Methods

4.2.1 Analysis at Each Time Point

- The data are analysed at each occasion separately.

- Advantages:
 - ▷ Simple to interpret
 - ▷ Uses all available data

- Disadvantages:
 - ▷ Does not consider ‘overall’ differences
 - ▷ Does not allow to study evolution differences
 - ▷ Problem of multiple testing

4.2.2 Analysis of Area Under the Curve

- For each subject, the area under its curve is calculated :

$$AUC_i = (t_{i2} - t_{i1}) \times (y_{i1} + y_{i2})/2 + (t_{i3} - t_{i2}) \times (y_{i2} + y_{i3})/2 + \dots$$

- Afterwards, these AUC_i are analyzed.
- Advantages:
 - ▷ No problems of multiple testing
 - ▷ Does not explicitly assume balanced data
 - ▷ Compares ‘overall’ differences
- Disadvantage: Uses only partial information : AUC_i

4.2.3 Analysis of Endpoints

- In randomized studies, there are no systematic differences at baseline.
- Hence, ‘treatment’ effects can be assessed by only comparing the measurements at the last occasion.
- Advantages:
 - ▷ No problems of multiple testing
 - ▷ Does not explicitly assume balanced data
- Disadvantages:
 - ▷ Uses only partial information : y_{in_i}
 - ▷ Only valid for large data sets

4.2.4 Analysis of Increments

- A simple method to compare evolutions between subjects, correcting for differences at baseline, is to analyze the subject-specific changes $y_{in_i} - y_{i1}$.
- Advantages:
 - ▷ No problems of multiple testing
 - ▷ Does not explicitly assume balanced data
- Disadvantage: Uses only partial information : $y_{in_i} - y_{i1}$

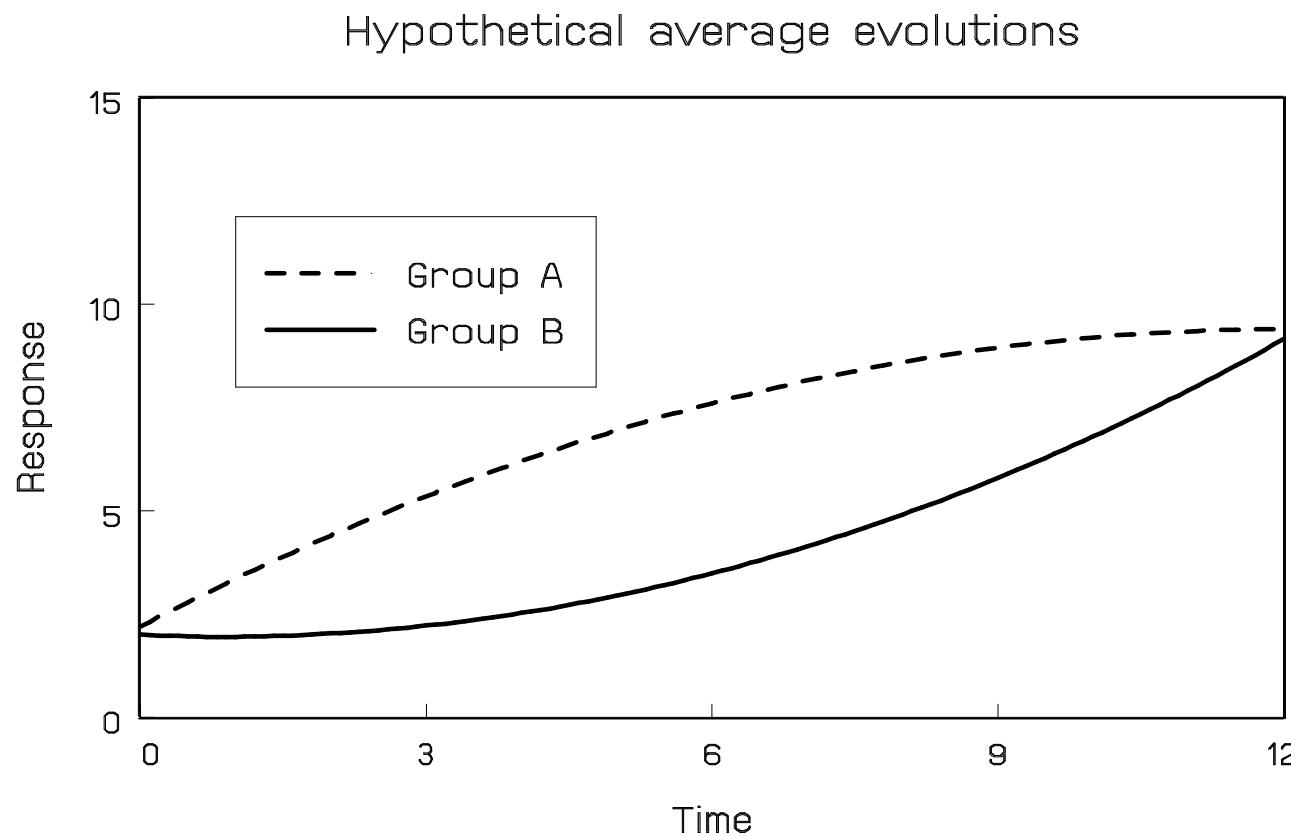
4.2.5 Analysis of Covariance

- Another way to analyse endpoints, correcting for differences at baseline, is to use analysis of covariance techniques, where the first measurement is included as covariate in the model.
- Advantages:
 - ▷ No problems of multiple testing
 - ▷ Does not explicitly assume balanced data
- Disadvantages:
 - ▷ Uses only partial information : y_{i1} and y_{in_i}
 - ▷ Does not take into account the variability of y_{i1}

4.3 Summary Statistics

- The AUC, endpoints and increments are examples of summary statistics
- Such summary statistics summarize the vector of repeated measurements for each subject separately.
- This leads to the following general procedure :
 - ▷ **Step 1** : Summarize data of each subject into one statistic, a summary statistic
 - ▷ **Step 2** : Analyze the summary statistics, e.g. analysis of covariance to compare groups after correction for important covariates
- This way, the analysis of longitudinal data is reduced to the analysis of independent observations, for which classical statistical procedures are available.

- However, all these methods have the disadvantage that (lots of) information is lost
- Further, they often do not allow to draw conclusions about the way the endpoint has been reached:



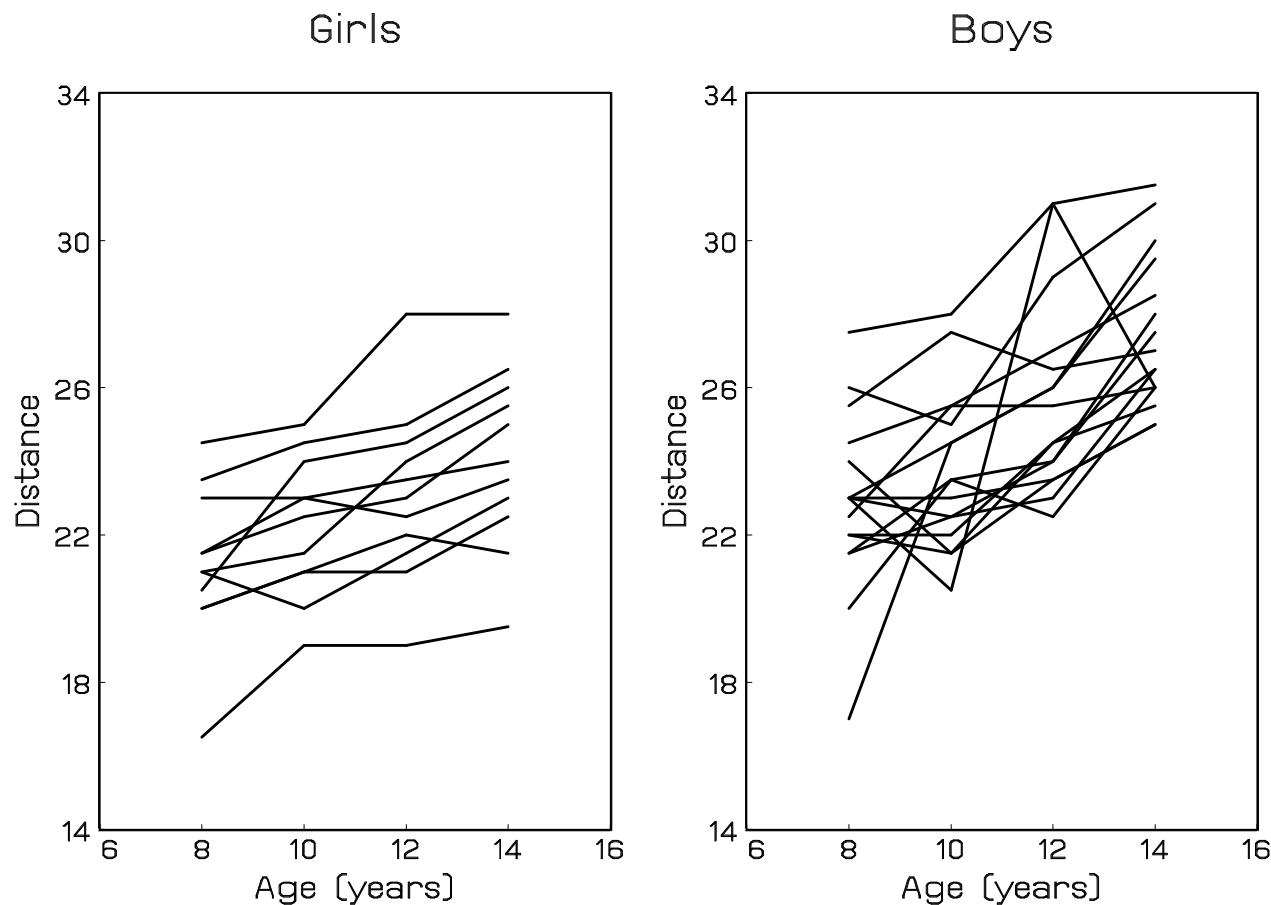
Chapter 5

The Multivariate Regression Model

- ▷ The general multivariate model
- ▷ Model fitting with SAS
- ▷ Model reduction
- ▷ Remarks

5.1 The General Multivariate Model

- We re-consider the growth data:



- This is a completely balanced data set:
 - ▷ 4 measurements for all subjects
 - ▷ measurements taken at exactly the same time points
- Let \mathbf{Y}_i be the vector of n repeated measurements for the i th subject :

$$\mathbf{Y}_i = \left(Y_{i1} \ Y_{i2} \ \dots \ Y_{in} \right)'$$

- The general multivariate model assumes that \mathbf{Y}_i satisfies a regression model

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i \quad \text{with} \quad \begin{cases} \mathbf{X}_i : \text{matrix of covariates} \\ \boldsymbol{\beta} : \text{vector of regression parameters} \\ \boldsymbol{\varepsilon}_i : \text{vector of error components, } \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \Sigma) \end{cases}$$

- We then have the following distribution for \mathbf{Y}_i : $\mathbf{Y}_i \sim N(X_i\boldsymbol{\beta}, \Sigma)$
- The mean structure $X_i\boldsymbol{\beta}$ is modelled as in classical linear regression and ANOVA models
- Usually, Σ is just a general $(n \times n)$ covariance matrix.
However, special structures for Σ can be assumed (see later).
- Assuming independence across individuals, $\boldsymbol{\beta}$ and the parameters in Σ can be estimated by maximizing

$$L_{ML} = \prod_{i=1}^N \left\{ (2\pi)^{-n/2} |\Sigma|^{-\frac{1}{2}} \exp \left(-\frac{1}{2} (\mathbf{y}_i - X_i\boldsymbol{\beta})' \Sigma^{-1} (\mathbf{y}_i - X_i\boldsymbol{\beta}) \right) \right\}$$

- Inference is based on classical maximum likelihood theory:
 - ▷ LR tests
 - ▷ Asymptotic WALD tests
- More details on inference will be discussed later

5.2 Model Fitting With SAS

5.2.1 Model Parameterization

- As an example, we fit a model with unstructured mean and unstructured covariance matrix to the growth data (Model 1).
- Let x_i be equal to 0 for a boy, and equal to 1 for a girl
- One possible parameterization of the model is

$$Y_{i1} = \beta_{0,8}(1 - x_i) + \beta_{1,8}x_i + \varepsilon_{i1}$$

$$Y_{i2} = \beta_{0,10}(1 - x_i) + \beta_{1,10}x_i + \varepsilon_{i2}$$

$$Y_{i3} = \beta_{0,12}(1 - x_i) + \beta_{1,12}x_i + \varepsilon_{i3}$$

$$Y_{i4} = \beta_{0,14}(1 - x_i) + \beta_{1,14}x_i + \varepsilon_{i4}$$

- In matrix notation:

$$\mathbf{Y}_i = X_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i,$$

with

$$X_i = \begin{pmatrix} (1-x_i) & 0 & 0 & 0 & x_i & 0 & 0 & 0 \\ 0 & (1-x_i) & 0 & 0 & 0 & x_i & 0 & 0 \\ 0 & 0 & (1-x_i) & 0 & 0 & 0 & x_i & 0 \\ 0 & 0 & 0 & (1-x_i) & 0 & 0 & 0 & x_i \end{pmatrix}$$

and with

$$\boldsymbol{\beta} = (\beta_{0,8}, \beta_{0,10}, \beta_{0,12}, \beta_{0,14}, \beta_{1,8}, \beta_{1,10}, \beta_{1,12}, \beta_{1,14})'$$

5.2.2 SAS Program

- SAS syntax:

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

- Data structure:
one record per observation:

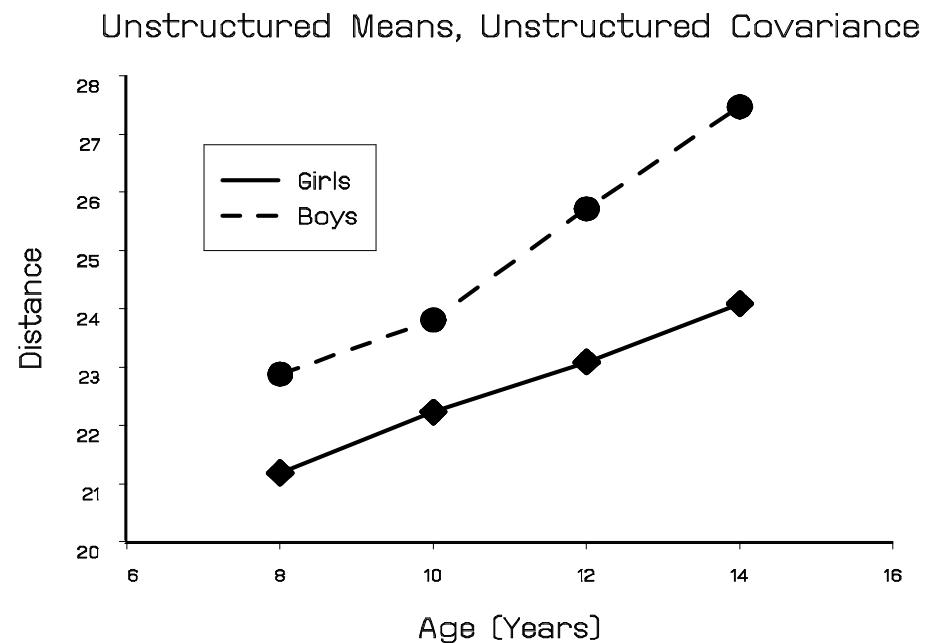
idnr	age	sex	measure
1.0000	8.0000	1.0000	21.000
1.0000	10.000	1.0000	20.000
1.0000	12.000	1.0000	21.500
1.0000	14.000	1.0000	23.000
2.0000	8.0000	1.0000	21.000
2.0000	10.000	1.0000	21.500
.....
26.000	12.000	0.0000	26.000
26.000	14.000	0.0000	30.000
27.000	8.0000	0.0000	22.000
27.000	10.000	0.0000	21.500
27.000	12.000	0.0000	23.500
27.000	14.000	0.0000	25.000

- The mean is modeled in the MODEL statement, as in other SAS procedures for linear models
- The covariance matrix is modeled in the REPEATED statement:
 - ▷ option ‘type=’ specifies covariance structure
 - ▷ option ‘subject=idnr’ specifies the clusters in the data set
 - ▷ the variable ‘age’ is used to order measurements within clusters

5.2.3 Results

- Maximized log-likelihood value: $\ell = -208.25$
- Estimates for parameters in mean structure, and implied fitted averages:

Parameter	MLE	(s.e.)
$\beta_{0,8}$	22.8750	(0.5598)
$\beta_{0,10}$	23.8125	(0.4921)
$\beta_{0,12}$	25.7188	(0.6112)
$\beta_{0,14}$	27.4688	(0.5371)
$\beta_{1,8}$	21.1818	(0.6752)
$\beta_{1,10}$	22.2273	(0.5935)
$\beta_{1,12}$	23.0909	(0.7372)
$\beta_{1,14}$	24.0909	(0.6478)



- Fitted covariance and correlation matrices:

$$\hat{\Sigma} = \begin{pmatrix} 5.0143 & 2.5156 & 3.6206 & 2.5095 \\ 2.5156 & 3.8748 & 2.7103 & 3.0714 \\ 3.6206 & 2.7103 & 5.9775 & 3.8248 \\ 2.5095 & 3.0714 & 3.8248 & 4.6164 \end{pmatrix} \implies \begin{pmatrix} 1.0000 & 0.5707 & 0.6613 & 0.5216 \\ 0.5707 & 1.0000 & 0.5632 & 0.7262 \\ 0.6613 & 0.5632 & 1.0000 & 0.7281 \\ 0.5216 & 0.7262 & 0.7281 & 1.0000 \end{pmatrix}$$

5.3 Model Reduction

- In many circumstances, one will be interested in reducing the model.
- For the growth data for example, one may be interested in finding out whether the fitted average profiles can be well described by straight lines.
- Also, the covariance matrix contained 10 parameters, not even of interest. If this can be reduced, one may gain efficiency for the mean structure.
- In practice, one therefore usually tries to reduce the mean and covariance structures, yielding more parsimonious models
- This is now illustrated using the growth data

5.3.1 Reduction of the Mean Structure

Model 2: Linear Average Trends

- Linear average trend within each group, unstructured 4×4 covariance matrix Σ
- Model 2 is given by ($x_i = 1$ for girls):

$$Y_{ij} = \beta_0 + \beta_{01}x_i + \beta_{10}t_j(1 - x_i) + \beta_{11}t_jx_i + \varepsilon_{ij},$$

- In matrix notation, this equals $\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$, with design matrix

$$\mathbf{X}_i = \begin{pmatrix} 1 & x_i & 8(1 - x_i) & 8x_i \\ 1 & x_i & 10(1 - x_i) & 10x_i \\ 1 & x_i & 12(1 - x_i) & 12x_i \\ 1 & x_i & 14(1 - x_i) & 14x_i \end{pmatrix}.$$

- Parameterization $\beta = (\beta_0, \beta_{01}, \beta_{10}, \beta_{11})'$:

- ▷ β_0 : intercept for boys
 - ▷ $\beta_0 + \beta_{01}$: intercept for girls
 - ▷ β_{10} : slope for boys
 - ▷ β_{11} : slope for girls

- SAS program :

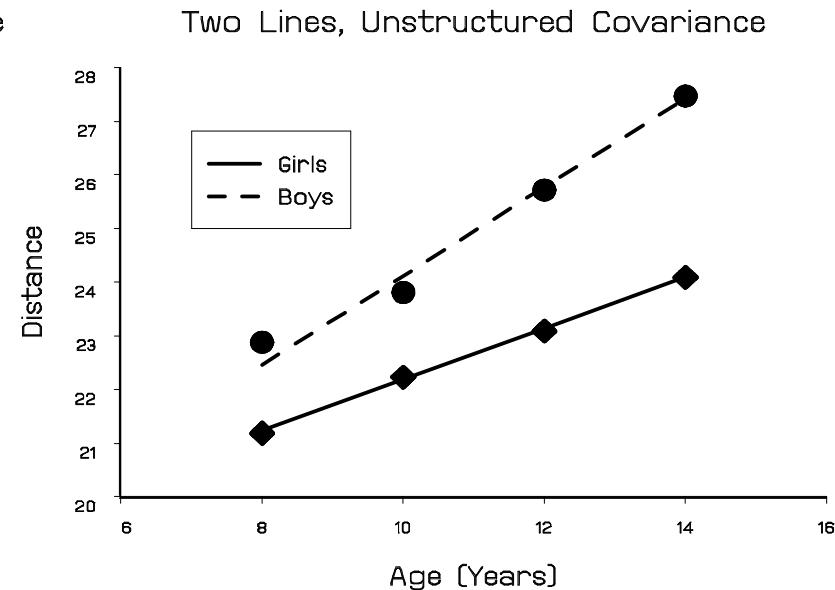
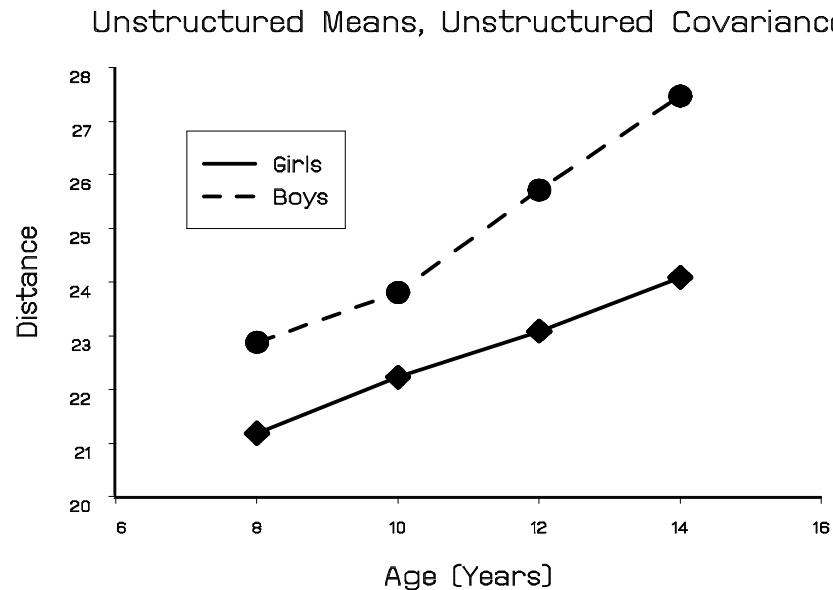
```
proc mixed data = growth method = ml;
  class idnr sex ageclss;
  model measure = sex age*sex / s;
  repeated ageclss / type = un subject = idnr;
  run;
```

- The variable ageclss is a copy of the original variable age

- LR test Model 2 versus Model 1:

	Mean	Covar	par	-2ℓ	Ref	G^2	df	p
1	unstr.	unstr.	18	416.509				
2	\neq slopes	unstr.	14	419.477	1	2.968	4	0.5632

- Predicted trends: girls : $\hat{Y}_j = 17.43 + 0.4764t_j$ boys : $\hat{Y}_j = 15.84 + 0.8268t_j$



Model 3 : Parallel Average Profiles

- Linear average trend within each sex group, the same slope for both groups
- Unstructured 4×4 covariance matrix Σ
- Model 3 is given by:

$$Y_{ij} = \beta_0 + \beta_{01}x_i + \beta_1t_j + \varepsilon_{ij}.$$

- In matrix notation, this equals $\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$, with design matrix

$$\mathbf{X}_i = \begin{pmatrix} 1 & x_i & 8 \\ 1 & x_i & 10 \\ 1 & x_i & 12 \\ 1 & x_i & 14 \end{pmatrix}$$

- Parameterization $\beta = (\beta_0, \beta_{01}, \beta_1)'$:
 - ▷ β_0 : intercept for boys
 - ▷ $\beta_0 + \beta_{01}$: intercept for girls
 - ▷ β_1 : common slope for boys and girls

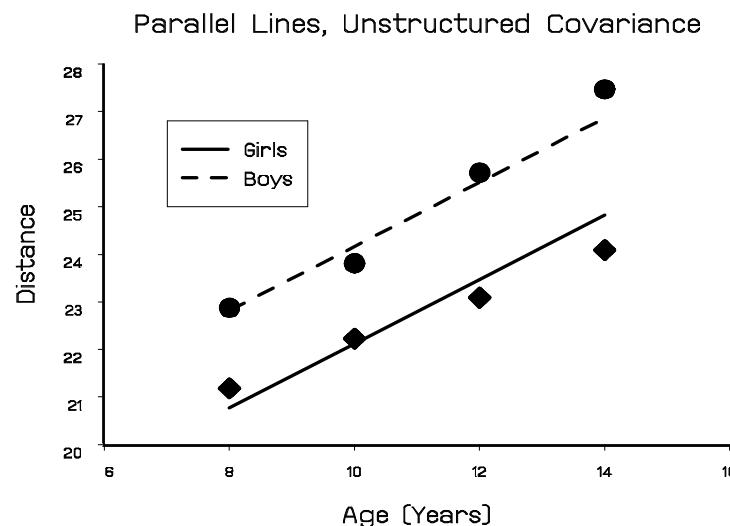
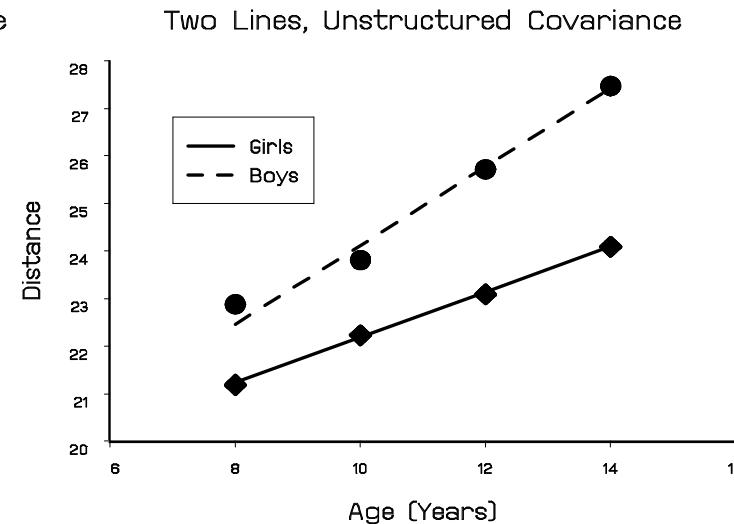
- SAS program :

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

- LR test:

	Mean	Covar	par	-2ℓ	Ref	G^2	df	p
1	unstr.	unstr.	18	416.509				
2	\neq slopes	unstr.	14	419.477	1	2.968	4	0.5632
3	= slopes	unstr.	13	426.153	2	6.676	1	0.0098

- Predicted trends: girls : $\hat{Y}_j = 15.37 + 0.6747t_j$ boys : $\hat{Y}_j = 17.42 + 0.6747t_j$



5.3.2 Reduction of the Covariance Structure

- In order to reduce the number of parameters in the covariance structure, we can now fit models with more parsimonious structures
- This often leads to more efficient inferences for the mean parameters.
- This is particularly useful when many repeated measurements are taken per subject.
- SAS includes a large variety of covariance structures (see SAS help function)

- Some examples:

Structure	Example	Structure	Example
Unstructured type=UN	$\begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{pmatrix}$	Toeplitz type=TOEP	$\begin{pmatrix} \sigma^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma^2 & \sigma_{12} \\ \sigma_{13} & \sigma_{12} & \sigma^2 \end{pmatrix}$
Simple type=SIMPLE	$\begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix}$	Toeplitz (1) type=Toep(1)	$\begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix}$
Compound symmetry type=CS	$\begin{pmatrix} \sigma_1^2 + \sigma^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 + \sigma^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 + \sigma^2 \end{pmatrix}$	Heterogeneous compound symmetry type=CSH	$\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho\sigma_1\sigma_3 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho\sigma_2\sigma_3 \\ \rho\sigma_1\sigma_3 & \rho\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix}$
Banded type=UN(2)	$\begin{pmatrix} \sigma_1^2 & \sigma_{12} & 0 \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ 0 & \sigma_{23} & \sigma_3^2 \end{pmatrix}$	Heterogeneous first-order autoregressive type=ARH(1)	$\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho^2\sigma_1\sigma_3 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho\sigma_2\sigma_3 \\ \rho^2\sigma_1\sigma_3 & \rho\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix}$
First-order autoregressive type=AR(1)	$\begin{pmatrix} \sigma^2 & \rho\sigma^2 & \rho^2\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ \rho^2\sigma^2 & \rho\sigma^2 & \sigma^2 \end{pmatrix}$	Heterogeneous Toeplitz type=TOEPh	$\begin{pmatrix} \sigma_1^2 & \rho_1\sigma_1\sigma_2 & \rho_2\sigma_1\sigma_3 \\ \rho_1\sigma_1\sigma_2 & \sigma_2^2 & \rho_1\sigma_2\sigma_3 \\ \rho_2\sigma_1\sigma_3 & \rho_1\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix}$

Model 4: Toeplitz Covariance Structure

- Linear average trend within each sex group
- The estimated covariance matrix (s.e.) of the unstructured covariance matrix under Model 2 equals:

$$\begin{pmatrix} 5.12(1.42) & 2.44(0.98) & 3.61(1.28) & 2.52(1.06) \\ 2.44(0.98) & 3.93(1.08) & 2.72(1.07) & 3.06(1.01) \\ 3.61(1.28) & 2.72(1.07) & 5.98(1.63) & 3.82(1.25) \\ 2.52(1.06) & 3.06(1.01) & 3.82(1.25) & 4.62(1.26) \end{pmatrix}$$

- This suggests that a possible model reduction could consist of assuming equal variances, and banded covariances.

- This is the so-called Toeplitz covariance matrix Σ , with elements of the form $\Sigma_{ij} = \alpha_{|i-j|}$:

$$\Sigma = \begin{pmatrix} \alpha_0 & \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_1 & \alpha_0 & \alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_1 & \alpha_0 & \alpha_1 \\ \alpha_3 & \alpha_2 & \alpha_1 & \alpha_0 \end{pmatrix}$$

- Note that this is only really meaningful when the time points at which measurements are taken are equally spaced, as in the current example.
- SAS program :

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

- LR test Model 4 versus Model 2:

	Mean	Covar	par	-2ℓ	Ref	G^2	df	p
1	unstr.	unstr.	18	416.509				
2	\neq slopes	unstr.	14	419.477	1	2.968	4	0.5632
4	\neq slopes	banded	8	424.643	2	5.166	6	0.5227

- Fitted covariance and correlation matrices:

$$\hat{\Sigma} = \begin{pmatrix} 4.9439 & 3.0507 & 3.4054 & 2.3421 \\ 3.0507 & 4.9439 & 3.0507 & 3.4054 \\ 3.4054 & 3.0507 & 4.9439 & 3.0507 \\ 2.3421 & 3.4054 & 3.0507 & 4.9439 \end{pmatrix} \implies \begin{pmatrix} 1.0000 & 0.6171 & 0.6888 & 0.4737 \\ 0.6171 & 1.0000 & 0.6171 & 0.6888 \\ 0.6888 & 0.6171 & 1.0000 & 0.6171 \\ 0.4737 & 0.6888 & 0.6171 & 1.0000 \end{pmatrix}$$

Model 5: AR(1) Covariance Structure

- Linear average trend within each sex group
- The AR(1) covariance structure assumes exponentially decaying correlations, i.e., elements of Σ of the form $\Sigma_{ij} = \sigma^2 \rho^{|i-j|}$:

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}$$

- Note that this is also only really meaningful when the time points at which measurements are taken are equally spaced.

- SAS program:

```
proc mixed data = growth method = ml;
class sex idnr ageclss;
model measure = sex age*sex / s;
repeated ageclss / type = AR(1) subject = idnr;
run;
```

- LR test Model 5 versus Models 2 and 4 :

	Mean	Covar	par	-2ℓ	Ref	G^2	df	p
1	unstr.	unstr.	18	416.509				
2	\neq slopes	unstr.	14	419.477	1	2.968	4	0.5632
4	\neq slopes	banded	8	424.643	2	5.166	6	0.5227
5	\neq slopes	AR(1)	6	440.681	2	21.204	8	0.0066
					4	16.038	2	0.0003

- Fitted covariance and correlation matrices:

$$\hat{\Sigma} = \begin{pmatrix} 4.8903 & 2.9687 & 1.8021 & 1.0940 \\ 2.9687 & 4.8903 & 2.9687 & 1.8021 \\ 1.8021 & 2.9687 & 4.8903 & 2.9687 \\ 1.0940 & 1.8021 & 2.9687 & 4.8903 \end{pmatrix} \implies \begin{pmatrix} 1.0000 & 0.6070 & 0.3685 & 0.2237 \\ 0.6070 & 1.0000 & 0.6070 & 0.3685 \\ 0.3685 & 0.6070 & 1.0000 & 0.6070 \\ 0.2237 & 0.3685 & 0.6070 & 1.0000 \end{pmatrix}$$

5.4 Remarks

- The multivariate regression model is primarily suitable when measurements are taken at a relatively small number of fixed time points
- Even if some measurements are missing, the multivariate regression model can be applied, as long as the software allows for unequal numbers of measurements per subject.
- In the SAS procedure MIXED, this is taken care of in the REPEATED statement

```
repeated ageclss /      ;
```

from which it can be derived which outcomes have been observed, and which ones are missing.

- In case of large numbers of repeated measurements:
 - ▷ Multivariate regression models can only be applied under very specific mean and covariance structures, even in case of complete balance.
 - ▷ For example, unstructured means and/or unstructured covariances require estimation of very many parameters
- In case of highly unbalanced data:
 - ▷ Multivariate regression models can only be applied under very specific mean and covariance structures.
 - ▷ For example, Toeplitz and AR(1) covariances are not meaningful since time points are not equally spaced.
 - ▷ For example, compound symmetric covariances are meaningful, but based on very strong assumptions.

Chapter 6

A Model for Longitudinal Data

- ▷ Introduction
- ▷ The 2-stage model formulation
- ▷ Examples: Rat and prostate data
- ▷ The general linear mixed-effects model
- ▷ Hierarchical versus marginal model
- ▷ Examples: Rat and prostate data
- ▷ A model for the residual covariance structure

6.1 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

6.2 A 2-stage Model Formulation

6.2.1 Stage 1

- Response Y_{ij} for i th subject, measured at time t_{ij} , $i = 1, \dots, N$, $j = 1, \dots, n_i$
- Response vector \mathbf{Y}_i for i th subject: $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{in_i})'$
- Stage 1 model:

$$\mathbf{Y}_i = Z_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i$$

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

6.2.2 Stage 2

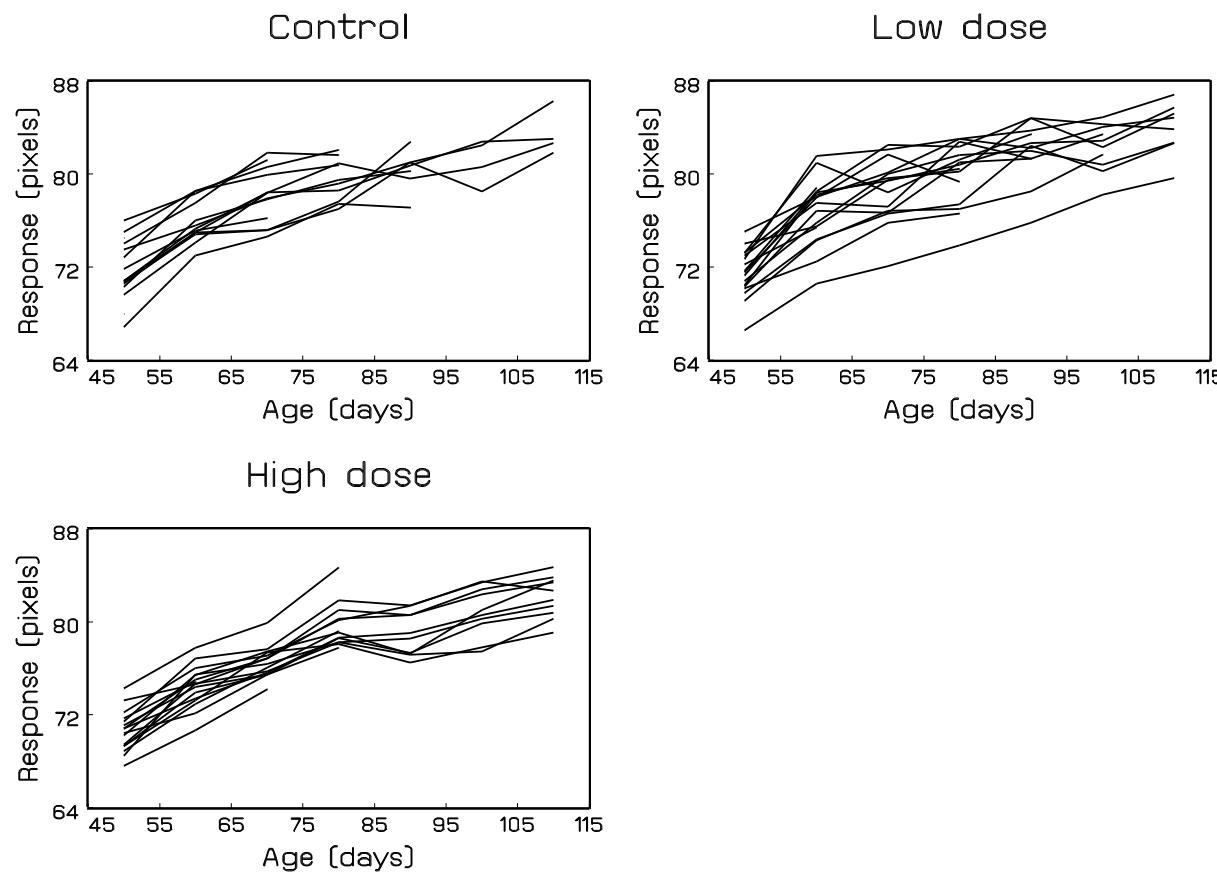
- Between-subject variability can now be studied from relating the β_i to known covariates
- Stage 2 model:

$$\beta_i = K_i \beta + b_i$$

- K_i is a $(q \times p)$ matrix of known covariates
- β is a p -dimensional vector of unknown regression parameters
- $b_i \sim N(\mathbf{0}, D)$

6.3 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: $Y_{ij} = \beta_{1i} + \beta_{2i}t_{ij} + \varepsilon_{ij}, \quad j = 1, \dots, n_i$
- Matrix notation:

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

- In the second stage, the subject-specific intercepts and time effects are related to the treatment of the rats
- Stage 2 model:

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

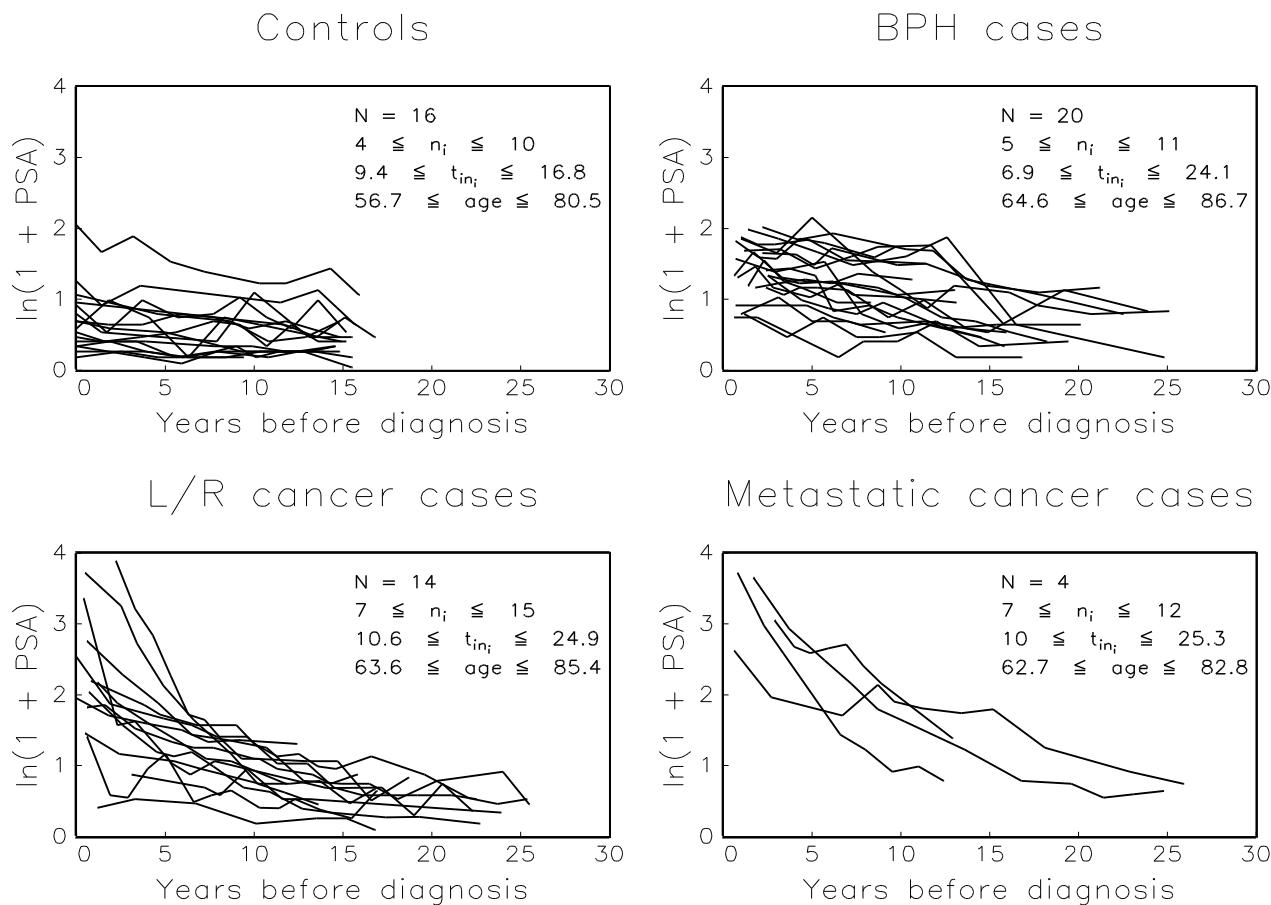
- L_i , H_i , and C_i are indicator variables:

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

- 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

6.4 Example: The Prostate Data

- Individual profiles:



- Transformation of the response:

$$\text{PSA}_{ij} \longrightarrow Y_{ij} = \ln(\text{PSA}_{ij} + 1)$$

- Stage 1 model: $Y_{ij} = \beta_{1i} + \beta_{2i}t_{ij} + \beta_{3i}t_{ij}^2 + \varepsilon_{ij}, \quad j = 1, \dots, n_i$

- Matrix notation:

$$\mathbf{Y}_i = Z_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i \quad \text{with} \quad Z_i = \begin{pmatrix} 1 & t_{i1} & t_{i1}^2 \\ 1 & t_{i2} & t_{i2}^2 \\ \vdots & \vdots & \vdots \\ 1 & t_{in_i} & t_{in_i}^2 \end{pmatrix}$$

- In the second stage, the subject-specific intercepts and time effects are related to the age (at diagnosis) and disease status

- Stage 2 model:

$$\begin{cases} \beta_{1i} = \beta_1 \text{Age}_i + \beta_2 C_i + \beta_3 B_i + \beta_4 L_i + \beta_5 M_i + b_{1i}, \\ \beta_{2i} = \beta_6 \text{Age}_i + \beta_7 C_i + \beta_8 B_i + \beta_9 L_i + \beta_{10} M_i + b_{2i}, \\ \beta_{3i} = \beta_{11} \text{Age}_i + \beta_{12} C_i + \beta_{13} B_i + \beta_{14} L_i + \beta_{15} M_i + b_{3i} \end{cases}$$

- C_i , B_i , L_i and M_i are indicator variables:

$$C_i = \begin{cases} 1 & \text{if Control} \\ 0 & \text{otherwise} \end{cases}$$

$$L_i = \begin{cases} 1 & \text{if L/R cancer case} \\ 0 & \text{otherwise} \end{cases}$$

$$B_i = \begin{cases} 1 & \text{if BPH case} \\ 0 & \text{otherwise} \end{cases}$$

$$M_i = \begin{cases} 1 & \text{if Metastatic cancer case} \\ 0 & \text{otherwise} \end{cases}$$

- Parameter interpretation:
 - ▷ β_2 , β_3 , β_4 , and β_5 : average intercepts after correction for age
 - ▷ β_7 , β_8 , β_9 , and β_{10} : average linear time effects after correction for age.
 - ▷ β_{12} , β_{13} , β_{14} , and β_{15} : average quadratic time effects after correction for age.

6.5 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:
 - ▷ \mathbf{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:

$$\begin{cases} \mathbf{Y}_i = Z_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i \\ \boldsymbol{\beta}_i = K_i \boldsymbol{\beta} + \mathbf{b}_i \end{cases} \implies \mathbf{Y}_i = \underbrace{Z_i K_i}_{X_i} \boldsymbol{\beta} + Z_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i = X_i \boldsymbol{\beta} + Z_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i$$

- General linear mixed-effects model:

$$\left\{ \begin{array}{l} Y_i = X_i \beta + Z_i b_i + \varepsilon_i \\ \\ b_i \sim N(\mathbf{0}, D), \quad \varepsilon_i \sim N(\mathbf{0}, \Sigma_i), \\ \\ b_1, \dots, b_N, \varepsilon_1, \dots, \varepsilon_N \text{ independent} \end{array} \right.$$

- Terminology:
 - ▷ Fixed effects: β
 - ▷ Random effects: b_i
 - ▷ Variance components: elements in D and Σ_i

6.6 Hierarchical versus Marginal Model

- The general linear mixed model is given by:

$$\begin{cases} \mathbf{Y}_i = X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i \\ \\ \mathbf{b}_i \sim N(\mathbf{0}, D), \quad \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \Sigma_i), \\ \\ \mathbf{b}_1, \dots, \mathbf{b}_N, \boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_N \text{ independent} \end{cases}$$

- It can be rewritten as:

$$\mathbf{Y}_i | \mathbf{b}_i \sim N(X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i, \Sigma_i), \quad \mathbf{b}_i \sim N(\mathbf{0}, D)$$

- It is therefore also called a hierarchical model:

- ▷ A model for \mathbf{Y}_i given \mathbf{b}_i
 - ▷ A model for \mathbf{b}_i

- Marginally, we have that \mathbf{Y}_i is distributed as:

$$\mathbf{Y}_i \sim N(X_i\boldsymbol{\beta}, Z_i D Z_i' + \Sigma_i)$$

- Hence, very specific assumptions are made about the dependence of mean and covariance on the covariates X_i and Z_i :
 - ▷ **Implied mean** : $X_i\boldsymbol{\beta}$
 - ▷ **Implied covariance** : $V_i = Z_i D Z_i' + \Sigma_i$
- Note that the hierarchical model implies the marginal one, **NOT** vice versa

6.7 Example: The Rat Data

- Stage 1 model:
$$Y_{ij} = \beta_{1i} + \beta_{2i}t_{ij} + \varepsilon_{ij}, \quad j = 1, \dots, n_i$$
- 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 ($\Sigma_i = \sigma^2 I_{n_i}$):

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

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

$$\begin{aligned}\text{Cov}(\mathbf{Y}_i(t_1), \mathbf{Y}_i(t_2)) &= \left(\begin{matrix} 1 \end{matrix} \right) D \left(\begin{matrix} 1 \end{matrix} \right) + \sigma^2 \delta_{\{t_1, t_2\}} \\ &= d_{11} + \sigma^2 \delta_{\{t_1, t_2\}}.\end{aligned}$$

- 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

6.8 Example: The Prostate Data

- Stage 1 model:
$$Y_{ij} = \beta_{1i} + \beta_{2i}t_{ij} + \beta_{3i}t_{ij}^2 + \varepsilon_{ij}, \quad j = 1, \dots, n_i$$
- Stage 2 model:
$$\begin{cases} \beta_{1i} = \beta_1 \text{Age}_i + \beta_2 C_i + \beta_3 B_i + \beta_4 L_i + \beta_5 M_i + b_{1i}, \\ \beta_{2i} = \beta_6 \text{Age}_i + \beta_7 C_i + \beta_8 B_i + \beta_9 L_i + \beta_{10} M_i + b_{2i}, \\ \beta_{3i} = \beta_{11} \text{Age}_i + \beta_{12} C_i + \beta_{13} B_i + \beta_{14} L_i + \beta_{15} M_i + b_{3i}, \end{cases}$$
- Combined:
$$\begin{aligned} Y_{ij} = & \beta_1 \text{Age}_i + \beta_2 C_i + \beta_3 B_i + \beta_4 L_i + \beta_5 M_i \\ & + (\beta_6 \text{Age}_i + \beta_7 C_i + \beta_8 B_i + \beta_9 L_i + \beta_{10} M_i) t_{ij} \\ & + (\beta_{11} \text{Age}_i + \beta_{12} C_i + \beta_{13} B_i + \beta_{14} L_i + \beta_{15} M_i) t_{ij}^2 \\ & + b_{1i} + b_{2i} t_{ij} + b_{3i} t_{ij}^2 + \varepsilon_{ij}. \end{aligned}$$

- Implied marginal mean structure:
 - ▷ Quadratic average evolution in each group
 - ▷ Average intercept and linear as well as quadratic slopes corrected for age differences

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

$$\begin{aligned} \text{Cov}(\mathbf{Y}_i(t_1), \mathbf{Y}_i(t_2)) &= \begin{pmatrix} 1 & t_1 & t_1^2 \end{pmatrix} D \begin{pmatrix} 1 \\ t_2 \\ t_2^2 \end{pmatrix} + \sigma^2 \delta_{\{t_1, t_2\}} \\ &= d_{33} t_1^2 t_2^2 + d_{23}(t_1^2 t_2 + t_1 t_2^2) + d_{22} t_1 t_2 \\ &\quad + d_{13}(t_1^2 + t_2^2) + d_{12}(t_1 + t_2) + d_{11} + \sigma^2 \delta_{\{t_1, t_2\}}. \end{aligned}$$

- The implied variance function is now a four-degree polynomial over time.

6.9 Example: Bivariate Observations

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

Model 1:

Random intercepts

+

heterogeneous errors

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

Model 2:

Uncorrelated intercepts and slopes

+

measurement error

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

- 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.

6.10 A Model for the Residual Covariance Structure

- Often, Σ_i is taken equal to $\sigma^2 I_{n_i}$
- We then obtain conditional independence:
Conditional on b_i , the elements in Y_i are independent
- In the presence of no, or little, random effects, conditional independence is often unrealistic
- For example, the random intercepts model not only implies constant variance, it also implicitly assumes constant correlation between any two measurements within subjects.

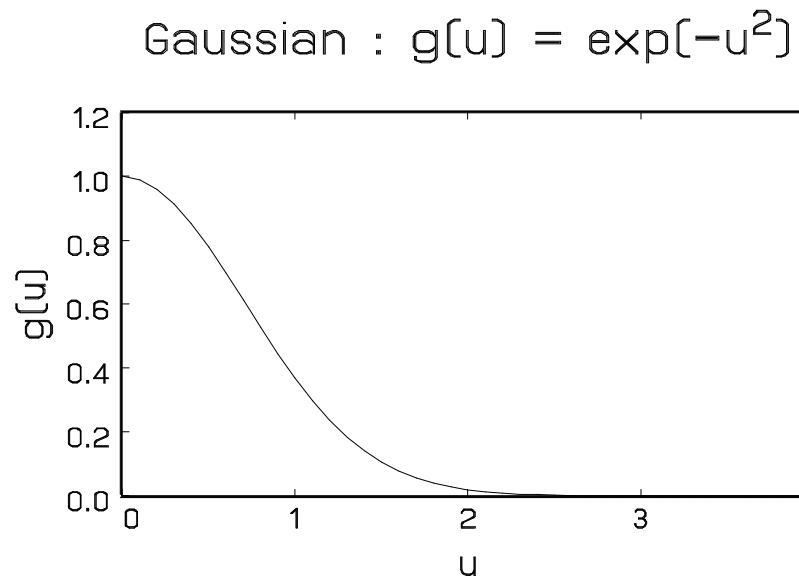
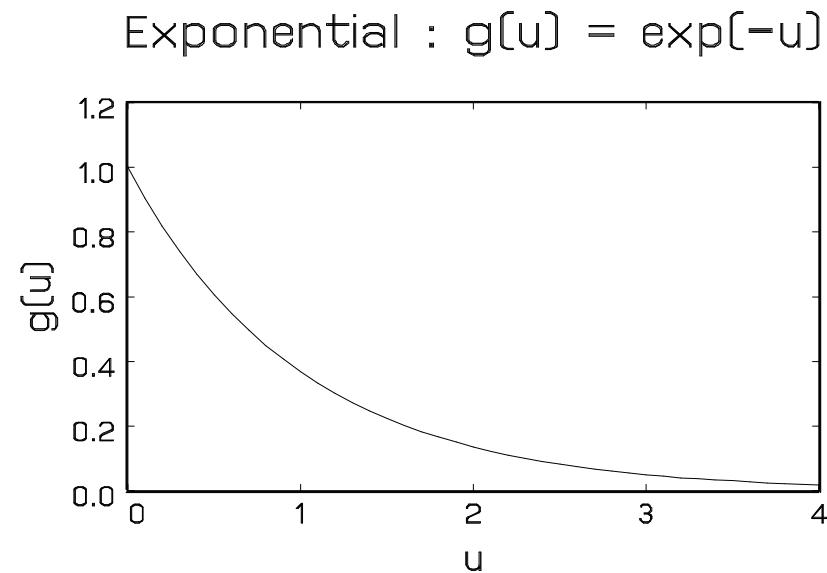
- Hence, when there is no evidence for (additional) random effects, or if they would have no substantive meaning, the correlation structure in the data can be accounted for in an appropriate model for Σ_i

- Frequently used model:
$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \underbrace{\boldsymbol{\varepsilon}_{(1)i} + \boldsymbol{\varepsilon}_{(2)i}}_{\downarrow} \mathbf{\varepsilon}_i$$

- 3 stochastic components:
 - ▷ \mathbf{b}_i : between-subject variability
 - ▷ $\boldsymbol{\varepsilon}_{(1)i}$: measurement error
 - ▷ $\boldsymbol{\varepsilon}_{(2)i}$: serial correlation component

- $\varepsilon_{(2)i}$ represents the belief that part of an individual's observed profile is a response to time-varying stochastic processes operating within that individual.
- This results in a correlation between serial measurements, which is usually a decreasing function of the time separation between these measurements.
- The correlation matrix H_i of $\varepsilon_{(2)i}$ is assumed to have (j, k) element of the form $h_{ijk} = g(|t_{ij} - t_{ik}|)$ for some decreasing function $g(\cdot)$ with $g(0) = 1$
- Frequently used functions $g(\cdot)$:
 - ▷ Exponential serial correlation: $g(u) = \exp(-\phi u)$
 - ▷ Gaussian serial correlation: $g(u) = \exp(-\phi u^2)$

- Graphically, for $\phi = 1$:



- Extreme cases:
 - ▷ $\phi = +\infty$: components in $\varepsilon_{(2)i}$ independent
 - ▷ $\phi = 0$: components in $\varepsilon_{(2)i}$ perfectly correlated

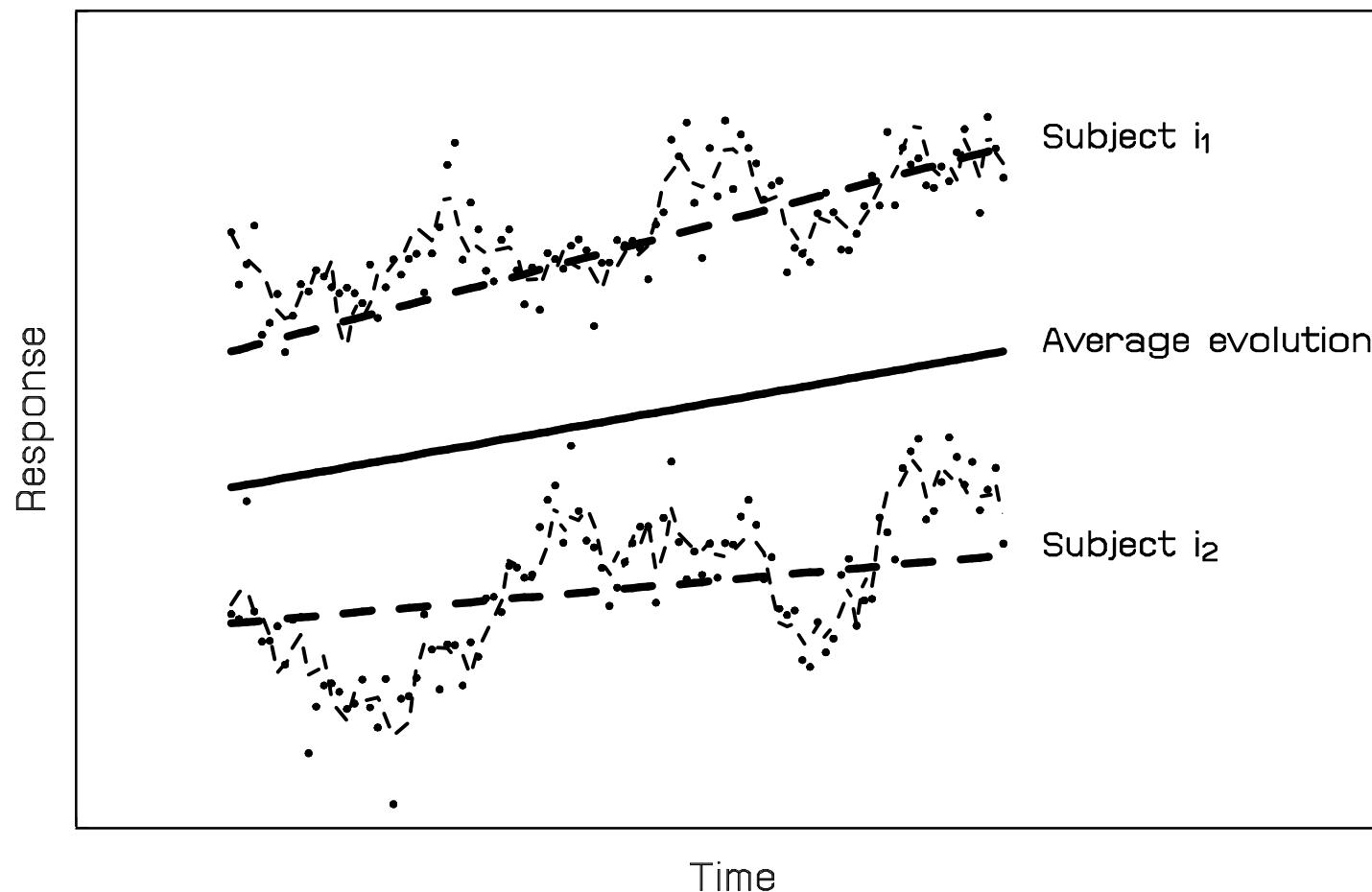
- In general, the smaller ϕ , the stronger is the serial correlation.
- Resulting final linear mixed model:

$$\mathbf{Y}_i = X_i \boldsymbol{\beta} + Z_i \mathbf{b}_i + \boldsymbol{\varepsilon}_{(1)i} + \boldsymbol{\varepsilon}_{(2)i}$$

$$\left. \begin{array}{l} \mathbf{b}_i \sim N(\mathbf{0}, D) \\ \boldsymbol{\varepsilon}_{(1)i} \sim N(\mathbf{0}, \sigma^2 I_{n_i}) \\ \boldsymbol{\varepsilon}_{(2)i} \sim N(\mathbf{0}, \tau^2 H_i) \end{array} \right\} \text{independent}$$

- Graphical representation of all 4 components in the model:

Stochastic components in general linear mixed model



Chapter 7

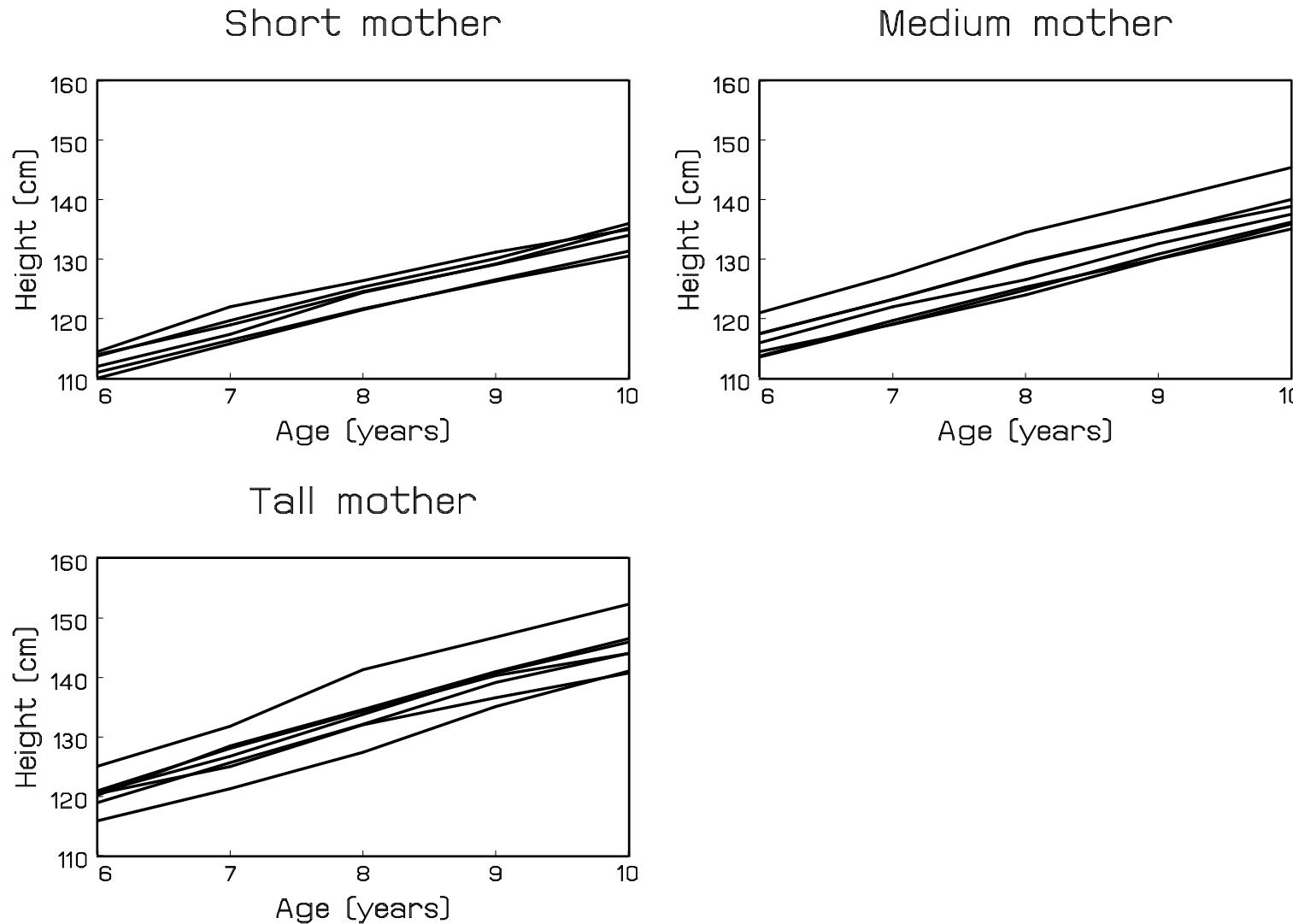
Exploratory Data Analysis

- ▷ Introduction
- ▷ Mean structure
- ▷ Variance function
- ▷ Correlation structure
- ▷ Individual profiles

7.1 Introduction

- A linear mixed model makes assumptions about:
 - ▷ mean structure: (non-)linear, covariates, ...
 - ▷ variance function: constant, quadratic, ...
 - ▷ correlation structure: constant, serial, ...
 - ▷ subject-specific profiles: linear, quadratic, ...
- In practice, linear mixed models are often obtained from a two-stage model formulation
- However, this may or may not imply a valid marginal model

- As an example, reconsider the growth curves:



- The individual profiles support a random-intercepts model
- However, the estimated covariance matrix suggests non-constant variance function:

$$\begin{pmatrix} 6.11 & 6.88 & 8.26 & 7.44 & 7.18 \\ 6.88 & 8.53 & 9.78 & 9.01 & 8.70 \\ 8.26 & 9.78 & 12.04 & 10.99 & 10.96 \\ 7.44 & 9.01 & 10.99 & 10.42 & 10.56 \\ 7.18 & 8.70 & 10.96 & 10.56 & 11.24 \end{pmatrix}.$$

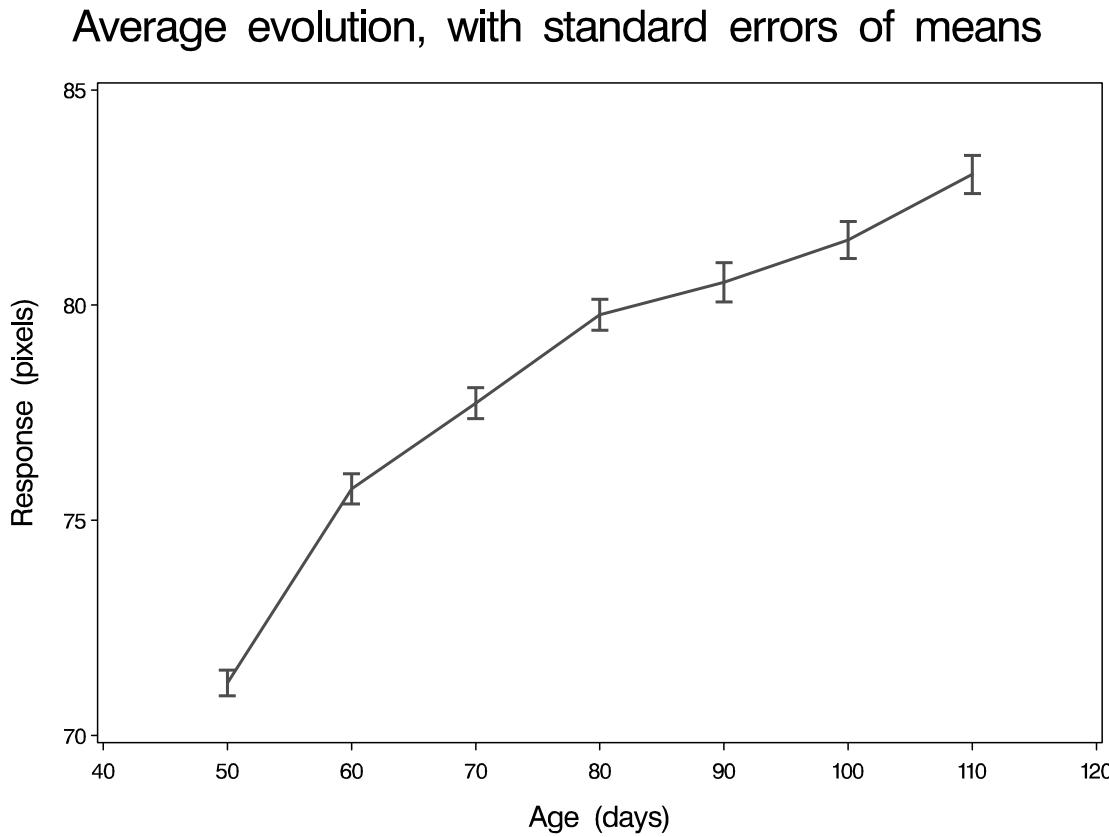
- Data exploration is therefore extremely helpful as additional tool in the selection of appropriate models

7.2 Exploring the Mean Structure

- For balanced data, averages can be calculated for each occasion separately, and standard errors for the means can be added
- Example: rat data:
 - ▷ SAS program:

```
filename fig1 'd:\path\file.eps';
goptions reset=all ftext=swiss device=psepsf gsfname=fig1 gsfmode=replace
      rotate=landscape;
proc gplot data=test;
plot y*age / haxis=axis1 vaxis=axis2;
symbol c=red i=std1mjt w=2 mode=include;
axis1 label=(h=2 'Age (days)') value=(h=1.5) order=(40 to 120 by 10) minor=none;
axis2 label=(h=2 A=90 'Response (pixels)') value=(h=1.5) order=(70 to 85 by 5)
      minor=none;
title h=3 'Average evolution, with standard errors of means';
run;quit;
```

▷ SAS output:



▷ Conclusion: non-linear average trend, increasing standard errors due to dropout

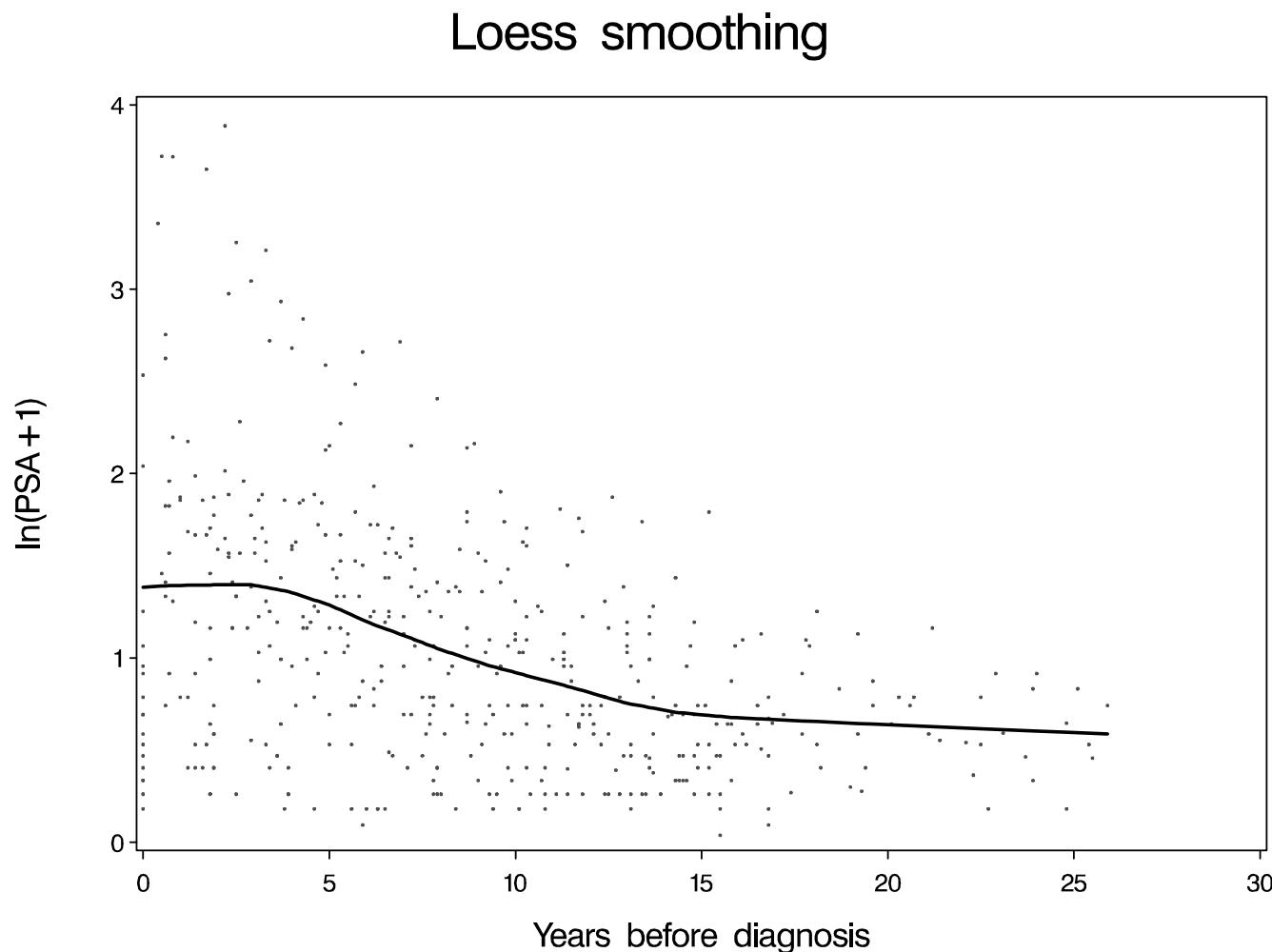
- For unbalanced data:
 - ▷ Discretize the time scale and use simple averaging within intervals
 - ▷ Smoothing techniques to estimate the average evolution nonparametrically
- Example: prostate data:
 - ▷ SAS program for loess smoothing:

```

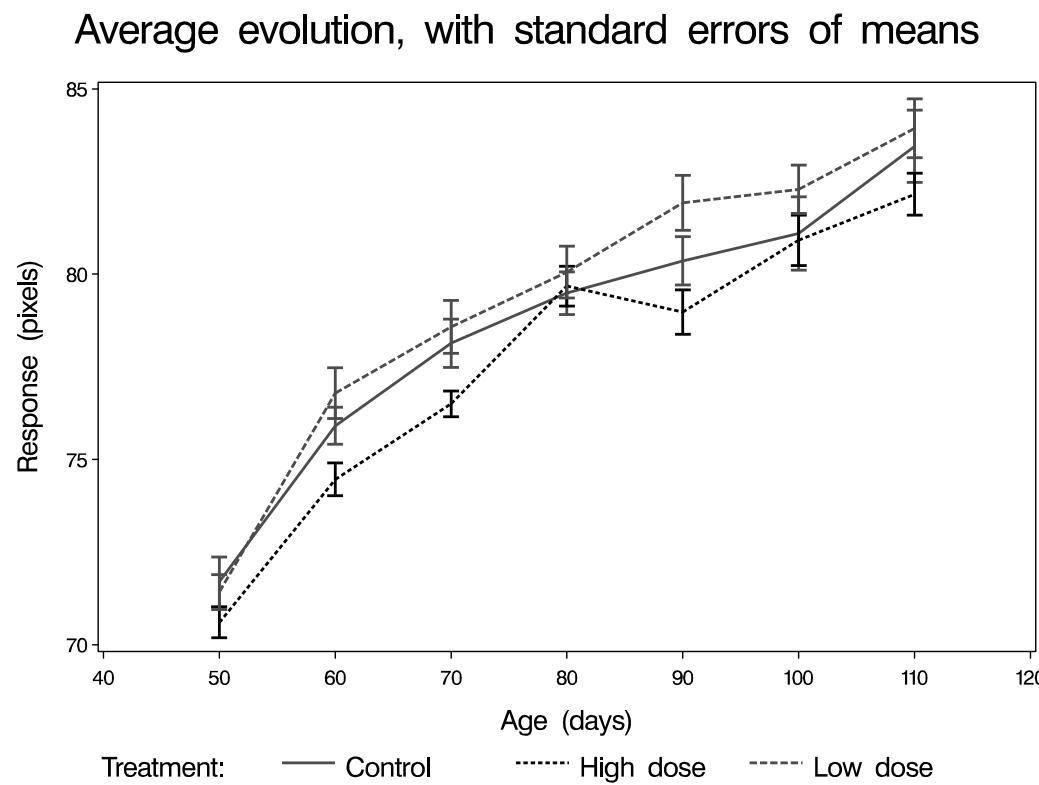
proc loess data=test;           filename fig1 'd:\path\file.eps';
ods output scoreresults=out;   options reset=all ftext=swiss device=psepsf
model lnpsa=time;             gsfname=fig1 gsfmode=replace rotate=landscape;
score data=test;               proc gplot data=out;
run;                           plot lnpsa*time=1 p_lnpsa*time=2
                               / overlay haxis=axis1 vaxis=axis2;
proc sort data=out;            symbol1 c=red v=dot h=0.2 mode=include;
by time;                      symbol2 c=black i=join w=2 mode=include;
run;                           axis1 label=(h=2 'Years before diagnosis')
                               value=(h=1.5) order=(0 to 30 by 5) minor=none;
                               axis2 label=(h=2 A=90 'ln(PSA+1)') value=(h=1.5)
                               order=(0 to 4 by 1) minor=none;
                               title h=3 'Loess smoothing';
                               run;quit;

```

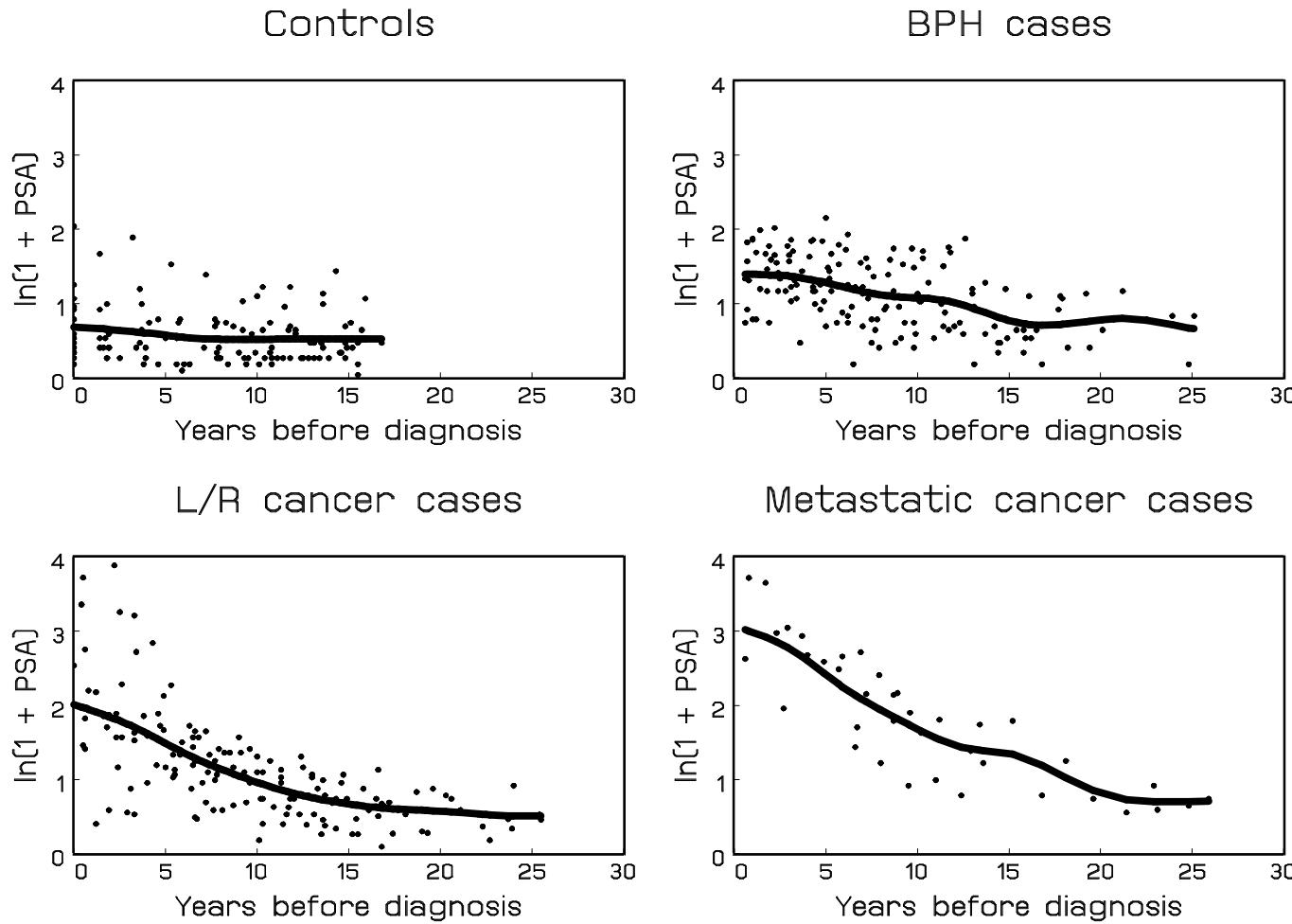
▷ SAS output:



- If (important) covariates or factors are known, similar plots can be constructed for subgroups with different values for these covariates or factors.
- Example for the rat data:



- Example for the prostate data:



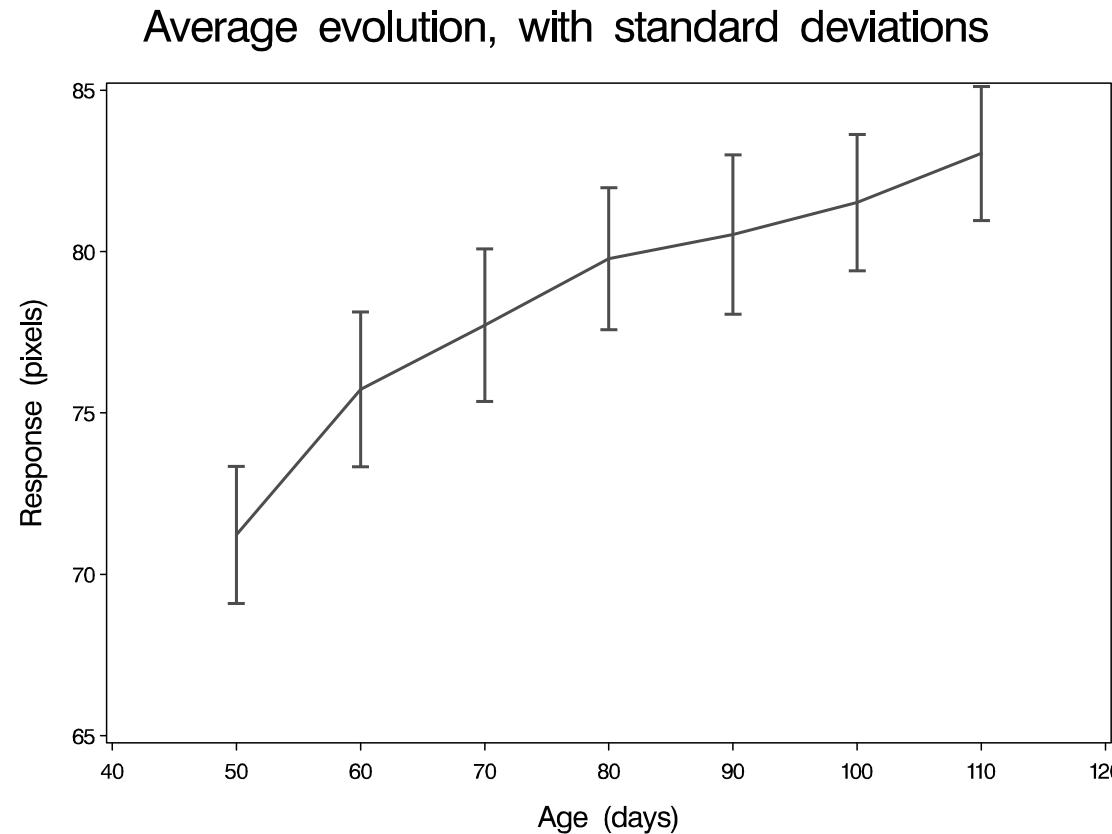
7.3 Exploring the Variance Function

- The variance function equals

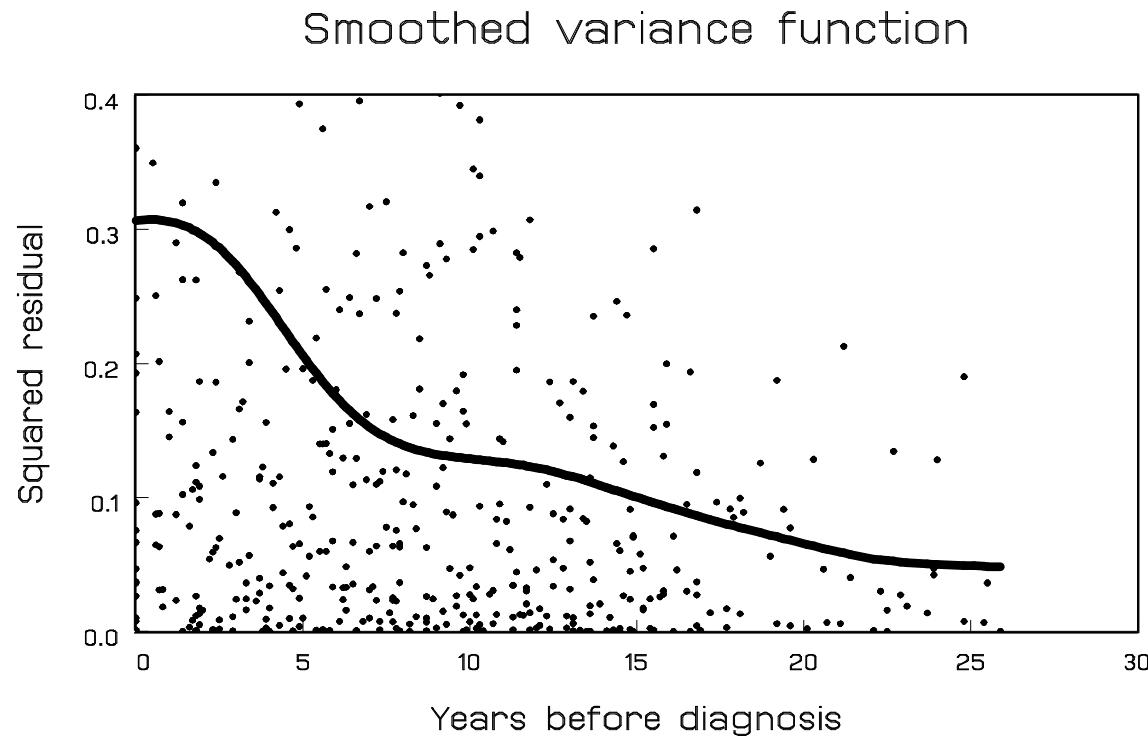
$$\sigma^2(t) = E[Y(t) - \mu(t)]^2$$

- Hence, an estimate for $\sigma^2(t)$ can be obtained from applying any of the techniques described for exploring the mean structure to squared residuals r_{ij}^2

- Example for the rat data (averages with standard deviations):



- Example for the prostate data (based on group-specific smoothing of averages):



7.4 Exploring the Correlation Structure

7.4.1 Scatterplot and Correlation Matrix

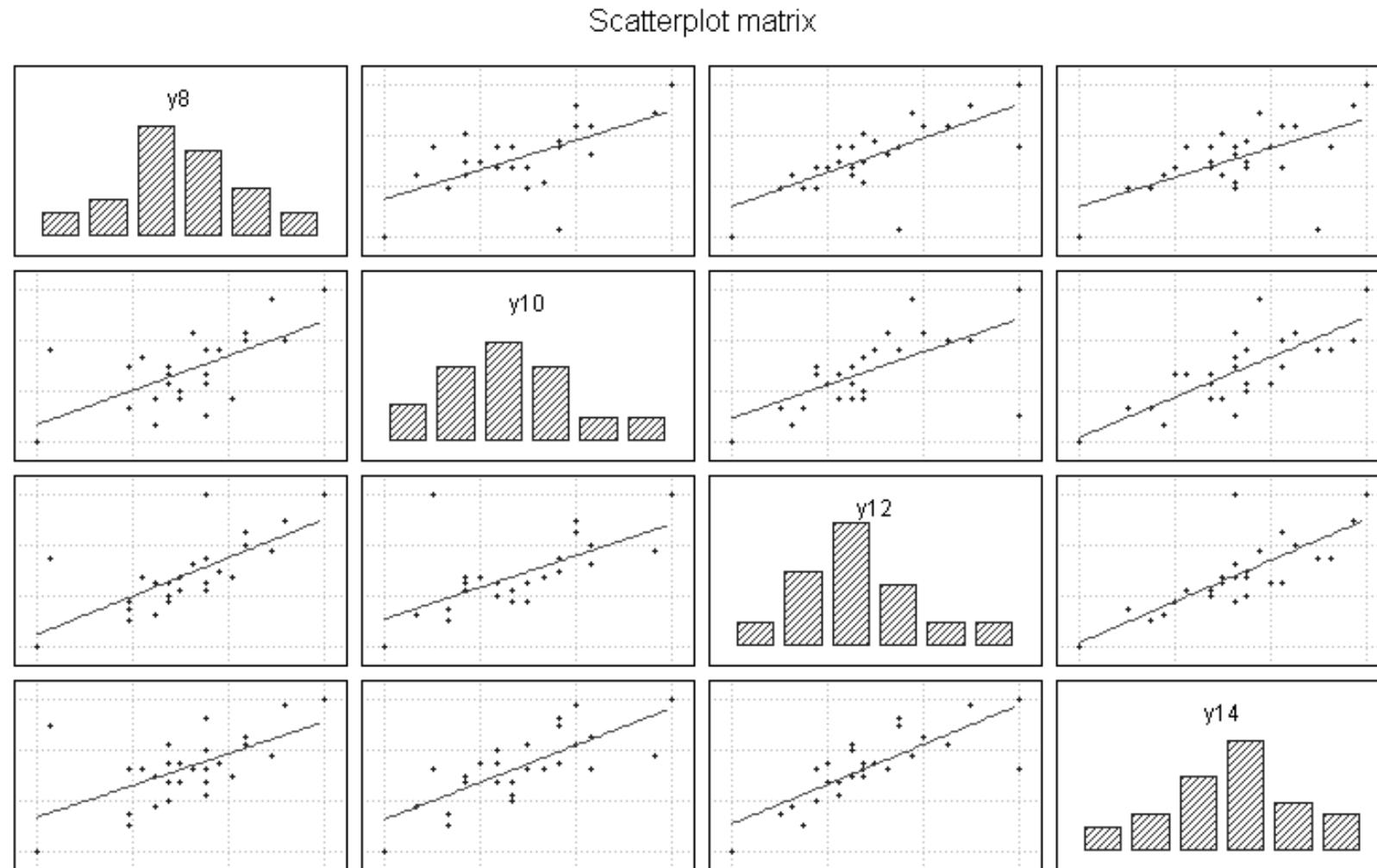
- For balanced longitudinal data, the correlation structure can be studied through the correlation matrix, or a scatterplot matrix

- Correlation matrix for the growth data:

$$\begin{pmatrix} 1.00 & 0.63 & 0.71 & 0.60 \\ 0.63 & 1.00 & 0.63 & 0.76 \\ 0.71 & 0.63 & 1.00 & 0.80 \\ 0.60 & 0.76 & 0.80 & 1.00 \end{pmatrix}.$$

- Graphically, pairwise scatterplots can be used for exploring the correlation between any two repeated measurements

- Scatterplot matrix for the growth data:



7.4.2 Semi-variogram

- For unbalanced data, the same approach can be used, after discretizing the time scale.
- An alternative method, in case the variance function suggests constant variance is the semi-variogram
- Re-consider the general linear mixed model:

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_{(1)i} + \boldsymbol{\varepsilon}_{(2)i}$$

$$\left. \begin{array}{l} \mathbf{b}_i \sim N(\mathbf{0}, D) \\ \boldsymbol{\varepsilon}_{(1)i} \sim N(\mathbf{0}, \sigma^2 I_{n_i}) \\ \boldsymbol{\varepsilon}_{(2)i} \sim N(\mathbf{0}, \tau^2 H_i) \end{array} \right\} \text{independent}$$

- Based on a mean function exploration, residuals $r_{ij} = y_{ij} - \mu(t_{ij})$ can be obtained
- These residuals are assumed to follow the model: $\mathbf{r}_i = Z_i \mathbf{b}_i + \boldsymbol{\varepsilon}_{(1)i} + \boldsymbol{\varepsilon}_{(2)i}$
- The semi-variogram assumes constant variance, which implies that the only random effects in the model will at most be intercepts, i.e., $Z_i = \begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix}'$.
- We will denote the variance of the random intercepts by ν^2
- The covariance matrix is then of the form

$$V_i = \text{Var}(\mathbf{Y}_i) = \text{Var}(\mathbf{r}_i) = \nu^2 Z_i Z_i' + \sigma^2 I_{n_i} + \tau^2 H_i$$

- The residuals r_{ij} have constant variance $\nu^2 + \sigma^2 + \tau^2$

- The correlation between any two residuals r_{ij} and r_{ik} from the same subject i is given by

$$\rho(|t_{ij} - t_{ik}|) = \frac{\nu^2 + \tau^2 g(|t_{ij} - t_{ik}|)}{\nu^2 + \sigma^2 + \tau^2}.$$

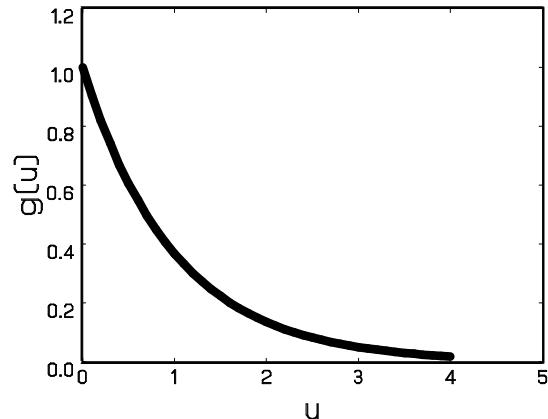
- One can show that, for $j \neq k$,

$$\begin{aligned}\frac{1}{2} E (r_{ij} - r_{ik})^2 &= \sigma^2 + \tau^2 (1 - g(|t_{ij} - t_{ik}|)) \\ &= v(u_{ijk})\end{aligned}$$

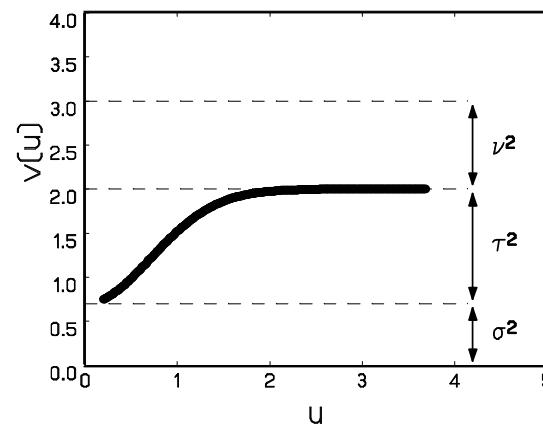
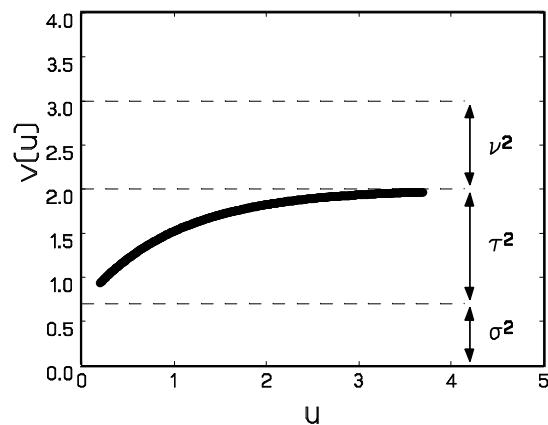
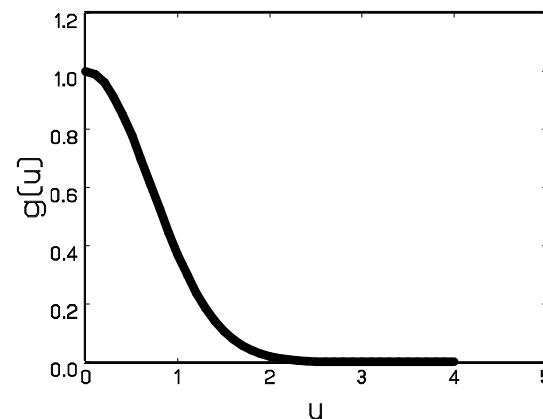
- The function $v(u)$ is called the semi-variogram, and it only depends on the time points t_{ij} through the time lags $u_{ijk} = |t_{ij} - t_{ik}|$.
- Decreasing serial correlation functions $g(\cdot)$ yield increasing semi-variograms $v(u)$, with $v(0) = \sigma^2$, which converge to $\sigma^2 + \tau^2$ as u grows to infinity.

- Semi-variograms for exponential and Gaussian serial correlation functions $g(\cdot)$, $\sigma^2 = 0.7$, $\tau^2 = 1.3$, and $\nu^2 = 1$, $\phi = 1$:

Exponential : $g(u) = \exp(-u)$



Gaussian : $g(u) = \exp(-u^2)$



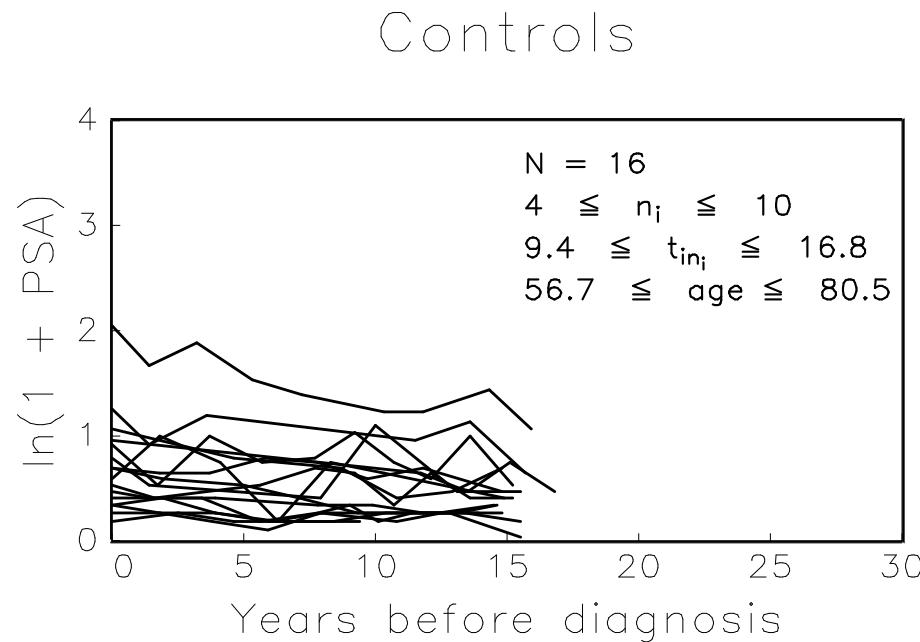
- Obviously, an estimate of $v(u)$ can be used to explore the relative importance of the stochastic components b_i , $\varepsilon_{(1)i}$, and $\varepsilon_{(2)i}$, as well as the nature of the serial correlation function $g(\cdot)$.
- An estimate of $v(u)$ is obtained from smoothing the scatter plot of the $\sum_{i=1}^N n_i(n_i - 1)/2$ half-squared differences $v_{ijk} = (r_{ij} - r_{ik})^2/2$ between pairs of residuals within subjects versus the corresponding time lags $u_{ijk} = |t_{ij} - t_{ik}|$.
- One can also show that, for $i \neq k$: $\frac{1}{2}E[r_{ij} - r_{kl}]^2 = \sigma^2 + \tau^2 + \nu^2$
- Hence, the total variability in the data (assumed constant) can be estimated by

$$\widehat{\sigma}^2 + \widehat{\tau}^2 + \widehat{\nu}^2 = \frac{1}{2N^*} \sum_{i \neq k} \sum_{j=1}^{n_i} \sum_{l=1}^{n_l} (r_{ij} - r_{kl})^2,$$

where N^* is the number of terms in the sum.

- Example: prostate data

 - ▷ We now consider the control group only:



 - ▷ Assuming constant variability, the variogram can be constructed to explore the 3 stochastic components.

▷ SAS program for loess smoothing:

```
/* Calculation of residuals, linear average trend */
proc glm data=prostate;
model lnpsa=time;
output out=out r=residual;
run;

/* Calculation of the variogram */
proc variogram data=out outpair=out;
coordinates xc=time yc=id;
compute robust novariogram;
var residual;
run;

data variogram;set out;
if y1=y2;vario=(v1-v2)**2/2; run;
data variance;set out;
if y1<y2; vario=(v1-v2)**2/2; run;

/* Calculation of the total variance (=0.148) */
proc means data=variance mean;
var vario;
run;
```

```

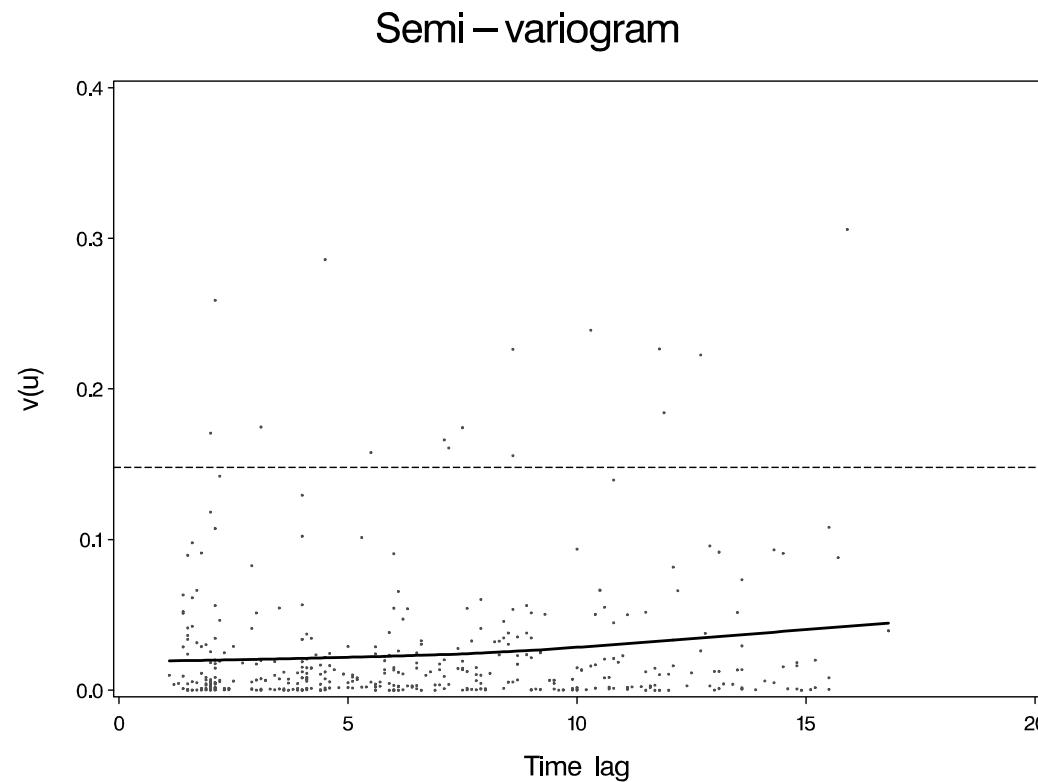
/* Loess smoothing of the variogram */
proc loess data=variogram;
ods output scoreresults=out;
model vario=distance;
score data=variogram;
run;

proc sort data=out;by distance;run;

filename fig1 'd:\path\file.eps';
goptions reset=all ftext=swiss device=psepsf gsfname=fig1
      gsfmode=replace rotate=landscape;
proc gplot data=out;
plot vario*distance=1 p_vario*distance=2
      / overlay haxis=axis1 vaxis=axis2 vref=0.148 lvref=3;
symbol1 c=red v=dot h=0.2 mode=include;
symbol2 c=black i=join w=2 mode=include;
axis1 label=(h=2 'Time lag') value=(h=1.5)
      order=(0 to 20 by 5) minor=none;
axis2 label=(h=2 A=90 'v(u)') value=(h=1.5)
      order=(0 to 0.4 by 0.1) minor=none;
title h=3 'Semi-variogram';
run;quit;

```

- ▷ SAS output:



- ▷ The total variability is estimated to be 0.148
- ▷ Random intercepts represent most of the variability, while there is very little evidence for the presence of serial correlation.

7.5 Exploring the Individual Profiles

7.5.1 Introduction

- As discussed before, linear mixed models are often obtained from a two-stage model formulation
- This is based on a good approximation of the subject-specific profiles by linear regression models
- This requires methods for the exploration of longitudinal profiles

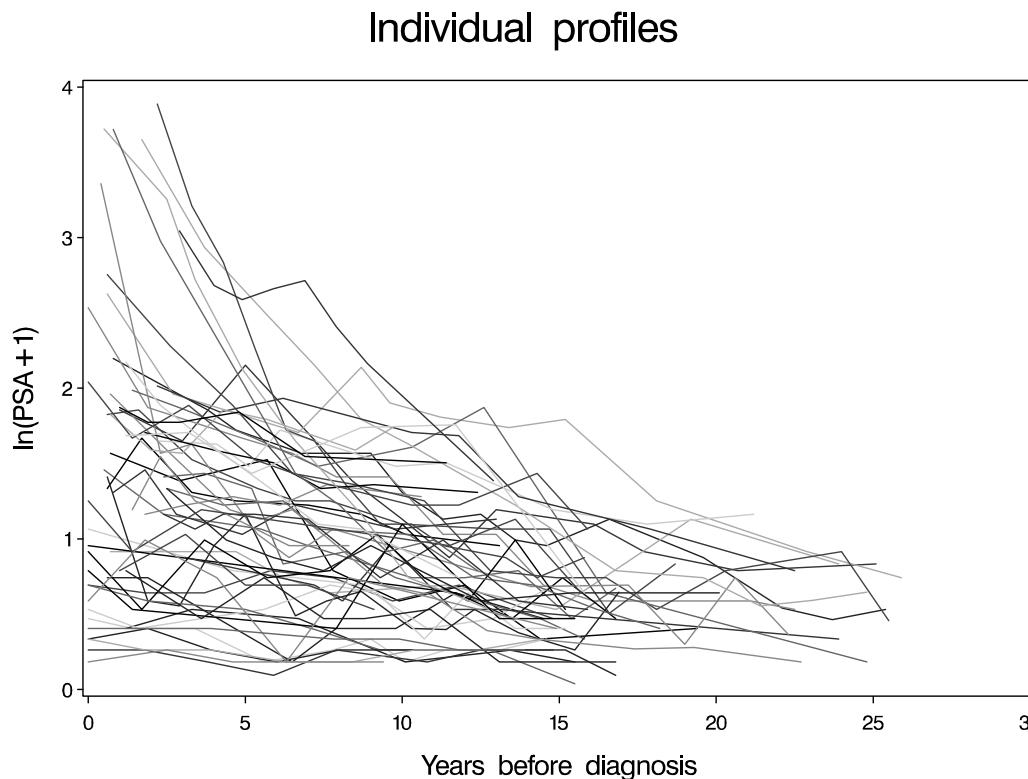
7.5.2 Graphical Exploration

- A natural way to explore longitudinal profiles is by plotting them
- Example: Prostate data:
 - ▷ SAS program:

```
proc sort data=prostate;
by id time;
run;

filename fig1 'd:\path\file.eps';
goptions reset=all ftext=swiss device=psepsf gsfname=fig1
      gsfmode=replace rotate=landscape i=join;
proc gplot data=test;
plot lnpsa*time=id / haxis=axis1 vaxis=axis2 nolegend;
axis1 label=(h=2 'Years before diagnosis') value=(h=1.5)
      order=(0 to 30 by 5) minor=none;
axis2 label=(h=2 A=90 'ln(PSA+1)') value=(h=1.5)
      order=(0 to 4 by 1) minor=none;
title h=3 'Individual profiles';
run;quit;
```

▷ SAS output:



- In case of large data sets:
 - ▷ Randomly select some profiles
 - ▷ Order subjects according to a specific profile characteristic (mean, variability,...) and plot profiles for some profiles

7.5.3 Exploring Subject-specific Regression Model

- Some ad hoc statistical procedures for checking the linear regression models

$$\mathbf{Y}_i = \mathbf{Z}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i$$

used in the first stage of the model formulation.

- Extensions of classical linear regression techniques:
 - ▷ Coefficient R^2 of multiple determination
 - ▷ Formal test for the need of a model extension

Coefficients of Multiple Determination

- In linear regression: $R^2 = \frac{SSTO - SSE}{SSTO}$
- Subject-specific coefficients: $R_i^2 = \frac{SSTO_i - SSE_i}{SSTO_i}$
- Histogram of R_i^2 or scatterplot of R_i^2 versus n_i
- Overall R^2 :
$$R_{\text{meta}}^2 = \frac{\sum_{i=1}^N (SSTO_i - SSE_i)}{\sum_{i=1}^N SSTO_i},$$
- SAS macro available

Test for Model Extension

- Test for the need to extend the linear regression model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with additional covariates in \mathbf{X}^* :

$$F = \frac{(\text{SSE}(R) - \text{SSE}(F))/p^*}{\text{SSE}(F)/(N - p - p^*)}$$

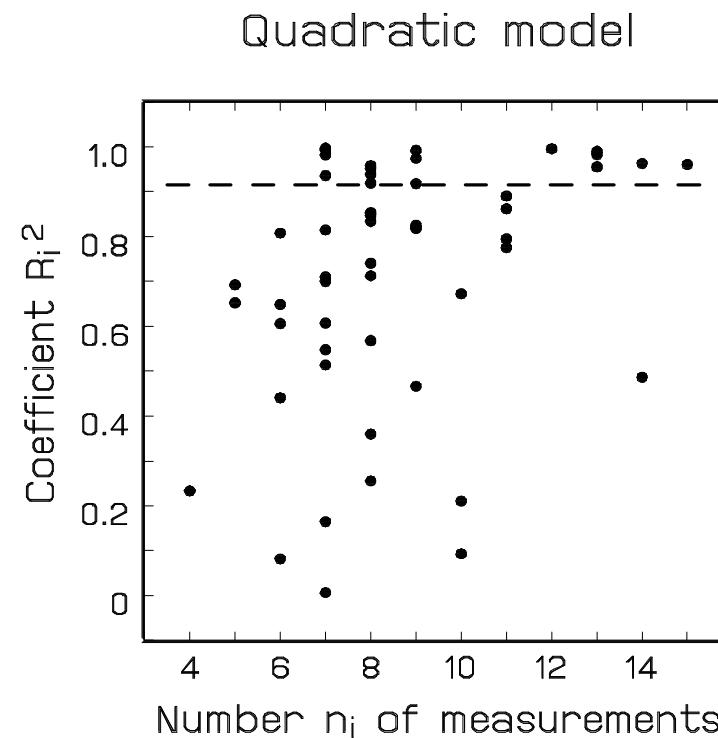
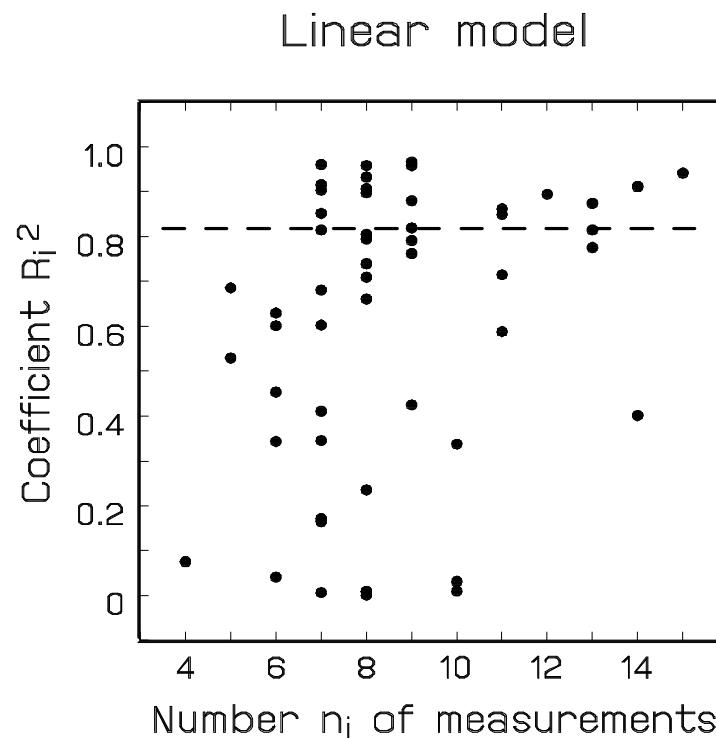
- Overall test for the need to extend the stage 1 model:

$$F_{\text{meta}} = \frac{\left\{ \sum_{\{i:n_i \geq p+p^*\}} (\text{SSE}_i(R) - \text{SSE}_i(F)) \right\} / \left\{ \sum_{\{i:n_i \geq p+p^*\}} p^* \right\}}{\left\{ \sum_{\{i:n_i \geq p+p^*\}} \text{SSE}_i(F) \right\} / \left\{ \sum_{\{i:n_i \geq p+p^*\}} (n_i - p - p^*) \right\}}$$

- Null-distribution is F with $\sum_{\{i:n_i \geq p+p^*\}} p^*$ and $\sum_{\{i:n_i \geq p+p^*\}} (n_i - p - p^*)$ degrees of freedom
- SAS macro available

Example: Prostate Data

- Scatterplots of R_i^2 under linear and quadratic model:



- Linear model:

- ▷ $R_{\text{meta}}^2 = 0.8188$

- ▷ F -test linear vs. quadratic: $F_{54,301} = 6.2181$ ($p < 0.0001$)

- Quadratic model:

- ▷ $R_{\text{meta}}^2 = 0.9143$

- ▷ F -test quadratic vs. cubic: $F_{54,247} = 1.2310$ ($p = 0.1484$)

Chapter 8

Estimation of the Marginal Model

- ▷ Introduction
- ▷ Maximum likelihood estimation
- ▷ Restricted maximum likelihood estimation
- ▷ Fitting linear mixed models in SAS
- ▷ Negative variance components

8.1 Introduction

- Recall that the general linear mixed model equals

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i$$

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

- The implied marginal model equals $\mathbf{Y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{Z}_i D \mathbf{Z}'_i + \Sigma_i)$
- Note that inferences based on the marginal model do not explicitly assume the presence of random effects representing the natural heterogeneity between subjects

- Notation:

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

- Marginal likelihood function:

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

- If α were known, MLE of β equals

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

where W_i equals V_i^{-1} .

- In most cases, α is not known, and needs to be replaced by an estimate $\hat{\alpha}$
- Two frequently used estimation methods for α :
 - ▷ Maximum likelihood
 - ▷ Restricted maximum likelihood

8.2 Maximum Likelihood Estimation (ML)

- $\widehat{\alpha}_{\text{ML}}$ obtained from maximizing

$$L_{\text{ML}}(\boldsymbol{\alpha}, \widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}))$$

with respect to $\boldsymbol{\alpha}$

- The resulting estimate $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}}_{\text{ML}})$ for $\boldsymbol{\beta}$ will be denoted by $\widehat{\boldsymbol{\beta}}_{\text{ML}}$
- $\widehat{\boldsymbol{\alpha}}_{\text{ML}}$ and $\widehat{\boldsymbol{\beta}}_{\text{ML}}$ can also be obtained from maximizing $L_{\text{ML}}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, i.e., with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ simultaneously.

8.3 Restricted Maximum Likelihood Estimation (REML)

8.3.1 Variance Estimation in Normal Populations

- Consider a sample of N observations Y_1, \dots, Y_N from $N(\mu, \sigma^2)$
- For known μ , MLE of σ^2 equals: $\hat{\sigma}^2 = \sum_i (Y_i - \mu)^2 / N$
- $\hat{\sigma}^2$ is unbiased for σ^2
- When μ is not known, MLE of σ^2 equals: $\hat{\sigma}^2 = \sum_i (Y_i - \bar{Y})^2 / N$
- Note that $\hat{\sigma}^2$ is biased for σ^2 : $E(\hat{\sigma}^2) = \frac{N-1}{N} \sigma^2$

- The bias expression tells us how to derive an unbiased estimate:

$$S^2 = \sum_i (Y_i - \bar{Y})^2 / (N - 1)$$

- Apparently, having to estimate μ introduces bias in MLE of σ^2
- How to estimate σ^2 , without estimating μ first ?
- The model for all data simultaneously:

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix} \sim N \left(\begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix}, \sigma^2 I_N \right)$$

- We transform \mathbf{Y} such that μ vanishes from the likelihood:

$$\mathbf{U} = \begin{pmatrix} Y_1 - Y_2 \\ Y_2 - Y_3 \\ \vdots \\ Y_{N-2} - Y_{N-1} \\ Y_{N-1} - Y_N \end{pmatrix} = A' \mathbf{Y} \sim N(\mathbf{0}, \sigma^2 A' A)$$

- MLE of σ^2 , based on \mathbf{U} , equals: $S^2 = \frac{1}{N-1} \sum_i (Y_i - \bar{Y})^2$
- A defines a set of $N-1$ linearly independent ‘error contrasts’
- S^2 is called the REML estimate of σ^2 , and S^2 is independent of A

8.3.2 Estimation of Residual Variance in Linear Regression Model

- Consider a sample of N observations Y_1, \dots, Y_N from a linear regression model:

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix} \sim N(X\boldsymbol{\beta}, \sigma^2 I)$$

- MLE of σ^2 :

$$\hat{\sigma}^2 = (\mathbf{Y} - X\hat{\boldsymbol{\beta}})'(\mathbf{Y} - X\hat{\boldsymbol{\beta}})/N,$$

- Note that $\hat{\sigma}^2$ is biased for σ^2 :

$$E(\hat{\sigma}^2) = \frac{N-p}{N} \sigma^2$$

- The bias expression tells us how to derive an unbiased estimate:

$$\text{MSE} = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})/(N - p),$$

- The MSE can also be obtained from transforming the data orthogonal to \mathbf{X} :

$$\mathbf{U} = \mathbf{A}'\mathbf{Y} \sim N(\mathbf{0}, \sigma^2 \mathbf{A}'\mathbf{A})$$

- The MLE of σ^2 , based on \mathbf{U} , now equals the mean squared error, MSE
- The MSE is again called the REML estimate of σ^2

8.3.3 REML for the Linear Mixed Model

- We first combine all models

$$\mathbf{Y}_i \sim N(X_i\boldsymbol{\beta}, V_i)$$

into one model

$$\mathbf{Y} \sim N(X\boldsymbol{\beta}, V)$$

in which

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_N \end{pmatrix}, \quad X = \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix}, \quad V(\boldsymbol{\alpha}) = \begin{pmatrix} V_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & V_N \end{pmatrix}$$

- Again, the data are transformed orthogonal to X :

$$\mathbf{U} = A'\mathbf{Y} \sim N(\mathbf{0}, A'V(\boldsymbol{\alpha})A)$$

- The MLE of α , based on U is called the REML estimate, and is denoted by $\widehat{\alpha}_{\text{REML}}$
- The resulting estimate $\widehat{\beta}(\widehat{\alpha}_{\text{REML}})$ for β will be denoted by $\widehat{\beta}_{\text{REML}}$
- $\widehat{\alpha}_{\text{REML}}$ and $\widehat{\beta}_{\text{REML}}$ can also be obtained from maximizing

$$L_{\text{REML}}(\boldsymbol{\theta}) = \left| \sum_{i=1}^N X_i' W_i(\boldsymbol{\alpha}) X_i \right|^{-\frac{1}{2}} L_{\text{ML}}(\boldsymbol{\theta})$$

with respect to $\boldsymbol{\theta}$, i.e., with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ simultaneously.

- $L_{\text{REML}}(\boldsymbol{\alpha}, \widehat{\beta}(\boldsymbol{\alpha}))$ is the likelihood of the error contrasts U , and is often called the REML likelihood function.
- Note that $L_{\text{REML}}(\boldsymbol{\theta})$ is **NOT** the likelihood for our original data Y

8.4 Fitting Linear Mixed Models in SAS

- Reconsider the model for the prostate data:

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

- Factor *group* defined by :

- ▷ control : $group = 1$
- ▷ BPH : $group = 2$
- ▷ local cancer : $group = 3$
- ▷ metastatic cancer : $group = 4$

- We will assume $\Sigma_i = \sigma^2 I_{n_i}$
- *time* and *timeclss* are time, expressed in decades before diagnosis
- *age* is age at the time of diagnosis
- $lnpsa = \ln(PSA + 1)$
- SAS program:

```
proc mixed data=prostate method=reml;
class id group timeclss;
model lnpsa = group age group*time age*time group*time2 age*time2 / noint solution;
random intercept time time2 / type=un subject=id g gcorr v vcorr;
repeated timeclss / type=simple subject=id r rcorr;
run;
```

- PROC MIXED statement:
 - ▷ calls procedure MIXED
 - ▷ specifies data-set (records correspond to occasions)
 - ▷ estimation method: ML, REML (default), ...
- CLASS statement: definition of the factors in the model
- MODEL statement:
 - ▷ response variable
 - ▷ fixed effects
 - ▷ options similar to SAS regression procedures

- RANDOM statement:
 - ▷ definition of random effects (including intercepts !)
 - ▷ identification of the ‘subjects’ : independence accross subjects
 - ▷ type of random-effects covariance matrix D
 - ▷ options ‘g’ and ‘gcorr’ to print out D and corresponding correlation matrix
 - ▷ options ‘v’ and ‘vcorr’ to print out V_i and corresponding correlation matrix
- REPEATED statement :
 - ▷ ordering of measurements within subjects
 - ▷ the effect(s) specified must be of the factor-type
 - ▷ identification of the ‘subjects’ : independence accross subjects
 - ▷ type of residual covariance matrix Σ_i
 - ▷ options ‘r’ and ‘rcorr’ to print out Σ_i and corresponding correlation matrix

- Some frequently used covariance structures available in RANDOM and REPEATED statements:

Structure	Example	Structure	Example
Unstructured type=UN	$\begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{pmatrix}$	Toeplitz type=TOEP	$\begin{pmatrix} \sigma^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma^2 & \sigma_{12} \\ \sigma_{13} & \sigma_{12} & \sigma^2 \end{pmatrix}$
Simple type=SIMPLE	$\begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix}$	Toeplitz (1) type=Toep(1)	$\begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix}$
Compound symmetry type=CS	$\begin{pmatrix} \sigma_1^2 + \sigma^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 + \sigma^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 + \sigma^2 \end{pmatrix}$	Heterogeneous compound symmetry type=CSH	$\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho\sigma_1\sigma_3 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho\sigma_2\sigma_3 \\ \rho\sigma_1\sigma_3 & \rho\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix}$
Banded type=UN(2)	$\begin{pmatrix} \sigma_1^2 & \sigma_{12} & 0 \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ 0 & \sigma_{23} & \sigma_3^2 \end{pmatrix}$	Heterogeneous first-order autoregressive type=ARH(1)	$\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho^2\sigma_1\sigma_3 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho\sigma_2\sigma_3 \\ \rho^2\sigma_1\sigma_3 & \rho\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix}$
First-order autoregressive type=AR(1)	$\begin{pmatrix} \sigma^2 & \rho\sigma^2 & \rho^2\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ \rho^2\sigma^2 & \rho\sigma^2 & \sigma^2 \end{pmatrix}$	Heterogeneous Toeplitz type=TOEPh	$\begin{pmatrix} \sigma_1^2 & \rho_1\sigma_1\sigma_2 & \rho_2\sigma_1\sigma_3 \\ \rho_1\sigma_1\sigma_2 & \sigma_2^2 & \rho_1\sigma_2\sigma_3 \\ \rho_2\sigma_1\sigma_3 & \rho_1\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix}$

- When serial correlation is to be fitted, it should be specified in the REPEATED statement, and the option ‘local’ can then be added to also include measurement error, if required.
- Some frequently used serial correlation structures available in RANDOM and REPEATED statements:

Structure	Example
Power type=SP(POW)(list)	$\sigma^2 \begin{pmatrix} 1 & \rho^{d_{12}} & \rho^{d_{13}} \\ \rho^{d_{12}} & 1 & \rho^{d_{23}} \\ \rho^{d_{13}} & \rho^{d_{23}} & 1 \end{pmatrix}$
Exponential type=SP(EXP)(list)	$\sigma^2 \begin{pmatrix} 1 & \exp(-d_{12}/\rho) & \exp(-d_{13}/\rho) \\ \exp(-d_{12}/\rho) & 1 & \exp(-d_{23}/\rho) \\ \exp(-d_{13}/\rho) & \exp(-d_{23}/\rho) & 1 \end{pmatrix}$
Gaussian type=SP(GAU)(list)	$\sigma^2 \begin{pmatrix} 1 & \exp(-d_{12}^2/\rho^2) & \exp(-d_{13}^2/\rho^2) \\ \exp(-d_{12}^2/\rho^2) & 1 & \exp(-d_{23}^2/\rho^2) \\ \exp(-d_{13}^2/\rho^2) & \exp(-d_{23}^2/\rho^2) & 1 \end{pmatrix}$

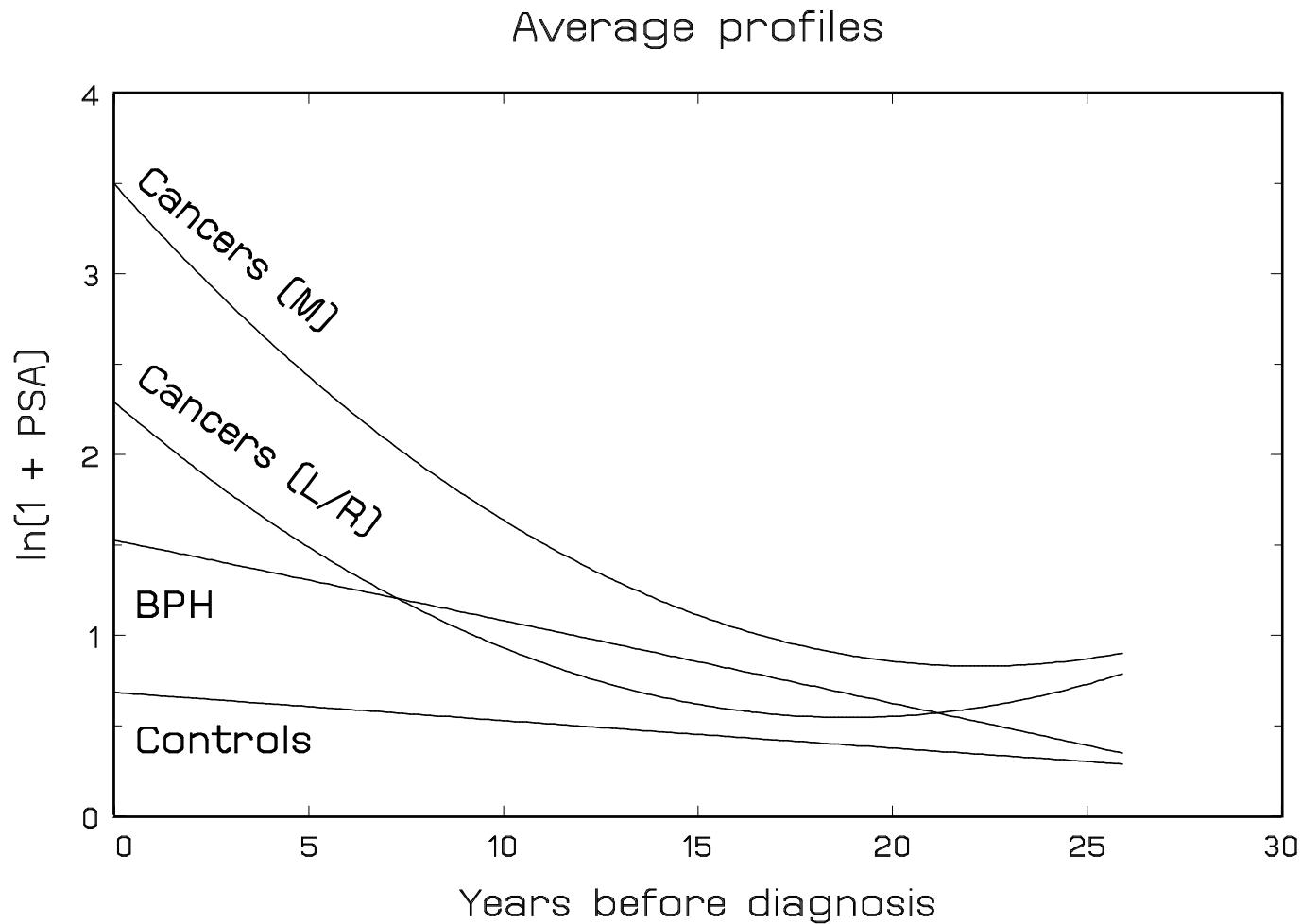
- ML and REML estimates for fixed effects:

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

- Fitted average profiles at median age at diagnosis:



8.5 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 b_i :		
$\text{var}(b_{1i})$	d_{11}	3.369 (1.123)
$\text{var}(b_{2i})$	d_{22}	0.000 (—)
$\text{cov}(b_{1i}, b_{2i})$	$d_{12} = d_{21}$	0.090 (0.381)
Residual variance:		
$\text{var}(\varepsilon_{ij})$	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:

Effect	Parameter	Parameter restrictions for α	
		$d_{ii} \geq 0, \sigma^2 \geq 0$ REMLE (s.e.)	$d_{ii} \in \mathbb{R}, \sigma^2 \in \mathbb{R}$ 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 b_i :			
$\text{var}(b_{1i})$	d_{11}	3.369 (1.123)	2.921 (1.019)
$\text{var}(b_{2i})$	d_{22}	0.000 (—)	-0.287 (0.169)
$\text{cov}(b_{1i}, b_{2i})$	$d_{12} = d_{21}$	0.090 (0.381)	0.462 (0.357)
Residual variance:			
$\text{var}(\varepsilon_{ij})$	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

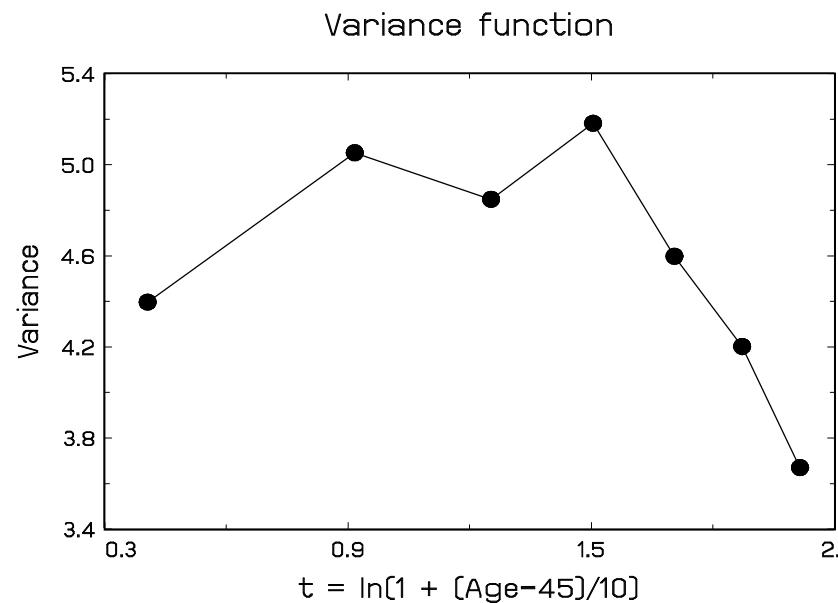
The usual action when a negative variance component estimate is obtained for a random coefficient would be to refit the model with the random coefficient removed. However, the user should be warned that not all packages will produce a negative variance component estimate. For example, in PROC MIXED we 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, \mathbf{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:

$$\begin{aligned}\text{Var}(\mathbf{Y}_i(t)) &= \begin{pmatrix} 1 & t \end{pmatrix} \widehat{\mathbf{D}} \begin{pmatrix} 1 \\ t \end{pmatrix} + \widehat{\sigma}^2 \\ &= \widehat{d}_{22}t^2 + 2\widehat{d}_{12}t + \widehat{d}_{11} + \widehat{\sigma}^2 = -0.287t^2 + 0.924t + 4.443\end{aligned}$$

▷ 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

Chapter 9

Inference for the Marginal Model

- ▷ Inference for fixed effects:
 - * Wald test
 - * t -test and F -test
 - * Robust inference
 - * LR test
- ▷ Inference for variance components:
 - * Wald test
 - * LR test
- ▷ Information criteria

9.1 Inference for the Fixed Effects

- Estimate for β :

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

with α replaced by its ML or REML estimate

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

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

- In practice one again replaces α by its ML or REML estimate

9.1.1 Approximate Wald Test

- For any known matrix L , consider testing

$$H_0 : L\beta = 0, \text{ versus } H_A : L\beta \neq 0$$

- Wald test statistic:

$$G = \widehat{\beta}' L' \left[L \left(\sum_{i=1}^N X_i' V_i^{-1}(\widehat{\alpha}) X_i \right)^{-1} L' \right]^{-1} L \widehat{\beta}$$

- Asymptotic null distribution of G is χ^2 with $\text{rank}(L)$ degrees of freedom

9.1.2 Approximate t -test and F -test

- Wald test based on

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

- Variability introduced from replacing α by some estimate is not taken into account in Wald tests
- Therefore, Wald tests will only provide valid inferences in sufficiently large samples
- In practice, this is often resolved by replacing the χ^2 distribution by an appropriate F -distribution (are the normal by a t).
- For any known matrix L , consider testing

$$H_0 : L\beta = 0, \quad \text{versus} \quad H_A : L\beta \neq 0$$

- F test statistic:

$$F = \frac{\widehat{\beta}' L' \left[L \left(\sum_{i=1}^N X_i' V_i^{-1}(\widehat{\alpha}) X_i \right)^{-1} L' \right]^{-1} L \widehat{\beta}}{\text{rank}(L)}.$$

- Approximate null-distribution of F is F with numerator degrees of freedom equal to $\text{rank}(L)$
- Denominator degrees of freedom to be estimated from the data:
 - ▷ Containment method
 - ▷ Satterthwaite approximation
 - ▷ Kenward and Roger approximation
 - ▷ ...

- In the context of longitudinal data, all methods typically lead to large numbers of degrees of freedom, and therefore also to very similar p -values.
- For univariate hypotheses ($\text{rank}(L) = 1$) the F -test reduces to a t -test

9.1.3 Example: The Prostate Data

- Linear hypotheses of the form

$$H_0 : L\beta = 0, \quad \text{versus} \quad H_A : L\beta \neq 0$$

can be tested in SAS using a CONTRAST statement

- As an example, reconsider the model for the prostate data:

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

- We now test whether the local cancer cases evolve different from the metastatic cancer cases.

- The null-hypothesis is specified by

$$H_0 : \begin{cases} \beta_4 = \beta_5 \\ \beta_9 = \beta_{10} \\ \beta_{14} = \beta_{15}, \end{cases}$$

- This is equivalent with testing

$$H_0 : \begin{pmatrix} 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix} \boldsymbol{\beta} = \mathbf{0},$$

which is of the form $L\boldsymbol{\beta} = \mathbf{0}$

- Related statements in SAS:

```
model lnpsa = group age group*time age*time  
           group*time2 age*time2  
           / noint ddfm=satterth;  
  
contrast 'L/R can = Met can' group 0 0 1 -1,  
           group*time 0 0 1 -1,  
           group*time2 0 0 1 -1 / chisq;
```

- Remarks:

- ▷ The Satterthwaite approximation is used for the denominator degrees of freedom
- ▷ The option ‘chisq’ in CONTRAST statement is needed in order also to obtain a Wald test

- Additional table in the output :

CONTRAST Statement Results							
Source	NDF	DDF	ChiSq	F	Pr > ChiSq	Pr > F	
L/R can = Met can	3	24.4	17.57	5.86	0.0005	0.0037	

- Several CONTRAST statements can now be used to reduce the model, in a stepwise procedure
- This leads to the following simplifications :
 - ▷ no interaction $age \times time^2$
 - ▷ no interaction $age \times time$
 - ▷ quadratic time effect the same for both cancer groups
 - ▷ the quadratic time effect is not significant for the non-cancer groups
 - ▷ the linear time effect is not significant for the controls

- Simultaneous testing of all these hypotheses is testing the null hypothesis

$$H_0 : \begin{cases} \beta_6 = 0 & (\text{no age by time interaction}) \\ \beta_7 = 0 & (\text{no linear time effect for controls}) \\ \beta_{11} = 0 & (\text{no age} \times \text{time}^2 \text{ interaction}) \\ \beta_{12} = 0 & (\text{no quadratic time effect for controls}) \\ \beta_{13} = 0 & (\text{no quadratic time effect for BPH}) \\ \beta_{14} = \beta_{15} & (\text{equal quadratic time effect for both cancer groups}). \end{cases}$$

- This hypothesis is of the form

$$H_0 : \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 \end{pmatrix} \boldsymbol{\beta} = 0$$

- The hypothesis can be tested with the following statements:

```
model lnpsa = group age group*time age*time group*time2 age*time2  
      / noint ddfm=satterth;  
  
contrast 'Final model' age*time 1,  
        group*time 1 0 0 0,  
        age*time2 1,  
        group*time2 1 0 0 0,  
        group*time2 0 1 0 0,  
        group*time2 0 0 1 -1 / chisq;
```

- This results in the following table in the output (Satterthwaite approximation):

CONTRAST Statement Results							
Source	NDF	DDF	ChiSq	F	Pr > ChiSq	Pr > F	
Final model	6	46.7	3.39	0.56	0.7587	0.7561	

- The simplified model is now given by:

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

- SAS procedure MIXED also allows using an ESTIMATE statement to estimate and test linear combinations of the elements of β
- Using similar arguments as for the approximate Wald-test, t -test, and F -test, approximate confidence intervals can be obtained for such linear combinations, also implemented in the ESTIMATE statement.
- Specification of L remains the same as for the CONTRAST statement, but L can now only contain one row.

9.1.4 Robust Inference

- Estimate for β :

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

with α replaced by its ML or REML estimate

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

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

provided that $E(Y_i) = X_i \beta$

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

- Conditional on α , $\hat{\beta}$ has covariance

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

- Note that this assumes that the covariance matrix $\text{Var}(Y_i)$ is correctly modelled as $V_i = Z_i D Z_i' + \Sigma_i$
- This covariance estimate is therefore often called the ‘naive’ estimate.
- The so-called ‘robust’ estimate for $\text{Var}(\hat{\beta})$, which does not assume the covariance matrix to be correctly specified is obtained from replacing $\text{Var}(Y_i)$ by $[Y_i - X_i \hat{\beta}] [Y_i - X_i \hat{\beta}]'$ rather than V_i

- The only condition for $[\mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}][\mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}]'$ to be unbiased for $\text{Var}(\mathbf{Y}_i)$ is that the mean is again correctly specified.
- The so-obtained estimate is called the ‘robust’ variance estimate, also called the sandwich estimate:

$$\text{Var}(\hat{\boldsymbol{\beta}}) = \underbrace{\left(\sum_{i=1}^N \mathbf{X}_i' W_i \mathbf{X}_i \right)^{-1}}_{\downarrow \text{BREAD}} \underbrace{\left(\sum_{i=1}^N \mathbf{X}_i' W_i \text{Var}(\mathbf{Y}_i) W_i \mathbf{X}_i \right)}_{\downarrow \text{MEAT}} \underbrace{\left(\sum_{i=1}^N \mathbf{X}_i' W_i \mathbf{X}_i \right)^{-1}}_{\downarrow \text{BREAD}}$$

- Based on this sandwich estimate, robust versions of the Wald test as well as of the approximate t -test and F -test can be obtained.

- Note that this suggests that as long as interest is only in inferences for the mean structure, little effort should be spent in modeling the covariance structure, provided that the data set is sufficiently large
- Extreme point of view: OLS with robust standard errors
- Appropriate covariance modeling may still be of interest:
 - ▷ for the interpretation of random variation in data
 - ▷ for gaining efficiency
 - ▷ in presence of missing data, robust inference only valid under very severe assumptions about the underlying missingness process (see later)

9.1.5 Example: Prostate Data

- We reconsider the reduced model for the prostate data:

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

- Robust inferences for the fixed effects can be obtained from adding the option ‘empirical’ to the PROC MIXED statement:

```
proc mixed data=prostate method=reml empirical;
```

- Comparison of naive and robust standard errors (**only fixed effects !**):

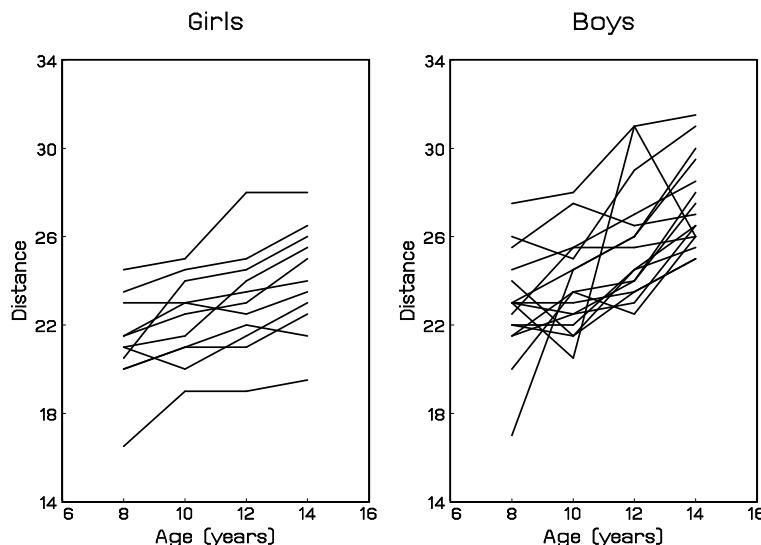
Effect	Parameter	Estimate (s.e. ⁽¹⁾ , s.e. ⁽²⁾)
Age effect	β_1	0.016 (0.006;0.006)
Intercepts:		
Control	β_2	-0.564 (0.428;0.404)
BPH	β_3	0.275 (0.488;0.486)
L/R cancer	β_4	1.099 (0.486;0.499)
Met. cancer	β_5	2.284 (0.531;0.507)
Time effects:		
BPH	β_8	-0.410 (0.068;0.067)
L/R cancer	β_9	-1.870 (0.233;0.360)
Met. cancer	β_{10}	-2.303 (0.262;0.391)
Time ² effects:		
Cancer	$\beta_{14} = \beta_{15}$	0.510 (0.088;0.128)

s.e.⁽¹⁾: Naive, s.e.⁽²⁾: Robust

- For some parameters, the robust standard error is smaller than the naive, model-based one. For other parameters, the opposite is true.

9.1.6 Example: Growth Data

- Comparison of naive and robust standard errors under Model 1 (unstructured mean as well as covariance), for the orthodontic growth data:



Parameter	MLE	(naive s.e.)	(robust s.e.)
$\beta_{0,8}$	22.8750	(0.5598)	(0.5938)
$\beta_{0,10}$	23.8125	(0.4921)	(0.5170)
$\beta_{0,12}$	25.7188	(0.6112)	(0.6419)
$\beta_{0,14}$	27.4688	(0.5371)	(0.5048)
$\beta_{1,8}$	21.1818	(0.6752)	(0.6108)
$\beta_{1,10}$	22.2273	(0.5935)	(0.5468)
$\beta_{1,12}$	23.0909	(0.7372)	(0.6797)
$\beta_{1,14}$	24.0909	(0.6478)	(0.7007)

- How could the covariance structure be improved ?

- We fit a model with a separate covariance structure for each group (Model 0)
- SAS program:

```
proc mixed data=test method=ml ;
  class idnr sex age;
  model measure = age*sex / noint s;
  repeated age / type=un subject=idnr r rcorr group=sex;
  run;
```

- LR test for Model 1 versus Model 0 : $p = 0.0082$
- The fixed-effects estimates remain unchanged.
- The naive standard errors under Model 0 are exactly the same as the sandwich estimated standard errors under Model 1.

9.1.7 Likelihood Ratio Test

- Comparison of nested models with different mean structures, but equal covariance structure
- Null hypothesis of interest equals $H_0 : \beta \in \Theta_{\beta,0}$, for some subspace $\Theta_{\beta,0}$ of the parameter space Θ_β of the fixed effects β .
- Notation:
 - ▷ L_{ML} : ML likelihood function
 - ▷ $\widehat{\theta}_{ML,0}$: MLE under H_0
 - ▷ $\widehat{\theta}_{ML}$: MLE under general model

- Test statistic:

$$-2 \ln \lambda_N = -2 \ln \left[\frac{L_{\text{ML}}(\hat{\boldsymbol{\theta}}_{\text{ML},0})}{L_{\text{ML}}(\hat{\boldsymbol{\theta}}_{\text{ML}})} \right]$$

- Asymptotic null distribution: χ^2 with d.f. equal to the difference in dimension of Θ_β and $\Theta_{\beta,0}$.

9.1.8 Example: Prostate Data

- We reconsider the reduced model:

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

- Testing for the need of age correction, i.e., $H_0 : \beta_1 = 0$
- Results under ML estimation:

ML estimation	
Under $\beta_1 \in \mathbb{R}$	$L_{\text{ML}} = -3.575$
Under $H_0 : \beta_1 = 0$	$L_{\text{ML}} = -6.876$
$-2 \ln \lambda_N$	6.602
degrees of freedom	1
p -value	0.010

- Results under REML estimation:

REML estimation	
Under $\beta_1 \in I\!\!R$	$L_{\text{REML}} = -20.165$
Under $H_0 : \beta_1 = 0$	$L_{\text{REML}} = -19.003$
$-2 \ln \lambda_N$	-2.324
degrees of freedom	—
p -value	—

Negative LR test statistic !

9.1.9 LR Test for Fixed Effects Under REML

- How can the negative LR test statistic be explained ?
- Under REML, the response \mathbf{Y} is transformed into error contrasts $\mathbf{U} = \mathbf{A}'\mathbf{Y}$, for some matrix \mathbf{A} with $\mathbf{A}'\mathbf{X} = 0$.
- Afterwards, ML estimation is performed based on the error contrasts
- The reported likelihood value, $L_{\text{REML}}(\hat{\boldsymbol{\theta}})$ is the likelihood at maximum for the error contrasts \mathbf{U}
- Models with different mean structures lead to different sets of error contrasts
- Hence, the corresponding REML likelihoods are based on different observations, which makes them no longer comparable

- Conclusion:

LR tests for the mean structure are not valid under REML

9.2 Inference for the Variance Components

- Inference for the mean structure is usually of primary interest.
- However, inferences for the covariance structure is of interest as well:
 - ▷ interpretation of the random variation in the data
 - ▷ overparameterized covariance structures lead to inefficient inferences for mean
 - ▷ too restrictive models invalidate inferences for the mean structure

9.2.1 Approximate Wald Test

- Asymptotically, ML and REML estimates of α are normally distributed with correct mean and inverse Fisher information matrix as covariance
- Hence approximate s.e.'s and Wald tests can easily be obtained

9.2.2 Example: Prostate Data

- We reconsider the reduced model:

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

- Standard errors and approximate Wald tests for variance components can be obtained in PROC MIXED from adding the option ‘covtest’ to the PROC MIXED statement:

```
proc mixed data=prostate method=reml covtest;
```

- Related output:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	XRAY	0.4432	0.09349	4.74	<.0001
UN(2,1)	XRAY	-0.4903	0.1239	-3.96	<.0001
UN(2,2)	XRAY	0.8416	0.2033	4.14	<.0001
UN(3,1)	XRAY	0.1480	0.04702	3.15	0.0017
UN(3,2)	XRAY	-0.3000	0.08195	-3.66	0.0003
UN(3,3)	XRAY	0.1142	0.03454	3.31	0.0005
timeclss	XRAY	0.02837	0.002276	12.47	<.0001

- The reported p -values often do not test meaningful hypotheses
- The reported p -values are often wrong

9.2.3 Caution with Wald Tests for Variance Components

Marginal versus Hierarchical Model

- One of the Wald tests for the variance components in the reduced model for the prostate data was

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(3,3)	XRAY	0.1142	0.03454	3.31	0.0005

- This presents a Wald test for $H_0 : d_{33} = 0$
- However, under the hierarchical model interpretation, this null-hypothesis is not of any interest, as d_{23} and d_{13} should also equal zero whenever $d_{33} = 0$.
- Hence, the test is meaningful under the marginal model only, i.e., when no underlying random effects structure is believed to describe the data.

Boundary Problems

- The quality of the normal approximation for the ML or REML estimates strongly depends on the true value α
- Poor normal approximation if α is relatively close to the boundary of the parameter space
- If α is a boundary value, the normal approximation completely fails
- One of the Wald tests for the variance components in the reduced model for the prostate data was

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(3,3)	XRAY	0.1142	0.03454	3.31	0.0005

- This presents a Wald test for $H_0 : d_{33} = 0$

- Under the hierarchical model interpretation, $d_{33} = 0$ is a boundary value, implying the calculation of the above p -value is based on an incorrect null-distribution for the Wald test statistic.
- Indeed, how could ever, under H_0 , \widehat{d}_{33} be normally distributed with mean 0, if d_{33} is estimated under the restriction $d_{33} \geq 0$?
- Hence, the test is only correct, when the null-hypothesis is not a boundary value (e.g., $H_0 : d_{33} = 0.1$).
- Note that, even under the hierarchical model interpretation, a classical Wald test is valid for testing $H_0 : d_{23} = 0$.

9.2.4 Likelihood Ratio Test

- Comparison of nested models with equal mean structures, but different covariance structure
- Null hypothesis of interest equals $H_0 : \alpha \in \Theta_{\alpha,0}$, for some subspace $\Theta_{\alpha,0}$ of the parameter space Θ_α of the variance components α .
- Notation:
 - ▷ L_{ML} : ML likelihood function
 - ▷ $\widehat{\boldsymbol{\theta}}_{\text{ML},0}$: MLE under H_0
 - ▷ $\widehat{\boldsymbol{\theta}}_{\text{ML}}$: MLE under general model
- Test statistic:
$$-2 \ln \lambda_N = -2 \ln \left[\frac{L_{\text{ML}}(\widehat{\boldsymbol{\theta}}_{\text{ML},0})}{L_{\text{ML}}(\widehat{\boldsymbol{\theta}}_{\text{ML}})} \right]$$

- Asymptotic null distribution: χ^2 with d.f. equal to the difference in dimension of Θ_α and $\Theta_{\alpha,0}$.
- Note that, as long as models are compared with the same mean structure, a valid LR test can be obtained under REML as well.
- Indeed, both models can be fitted using the same error contrasts, making the likelihoods comparable.
- Note that, if H_0 is a boundary value, the classical χ^2 approximation may not be valid.
- For some very specific null-hypotheses on the boundary, the correct asymptotic null-distribution has been derived

9.2.5 Marginal Testing for the Need of Random Effects

- Under a hierarchical model interpretation, the asymptotic null-distribution for the LR test statistic for testing significance of all variance components related to one or multiple random effects, can be derived.
- Example: for the prostate model, testing whether the variance components associated to the quadratic random time effect are equal to zero, is equivalent to testing

$$H_0 : d_{13} = d_{23} = d_{33} = 0$$

- Note that, under the hierarchical interpretation of the model, H_0 is on the boundary of the parameter space

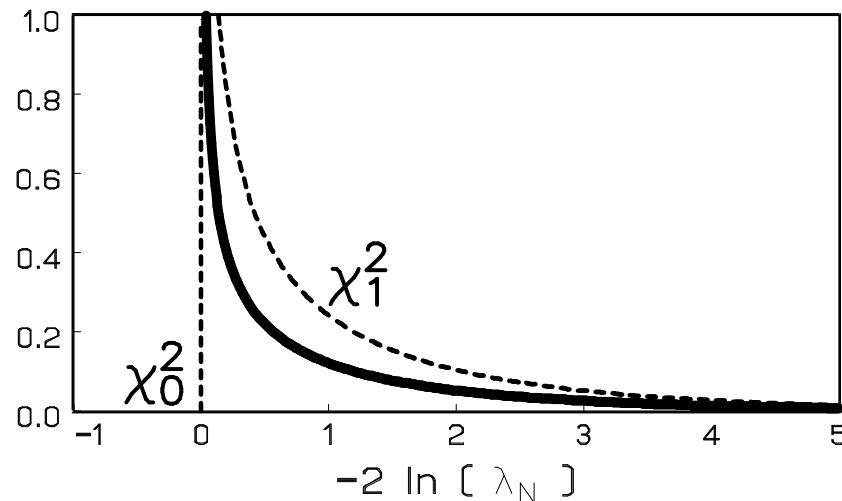
Case 1: No Random Effects versus one Random Effect

- Hypothesis of interest:

$$H_0 : D = 0 \quad \text{versus} \quad H_A : D = d_{11}$$

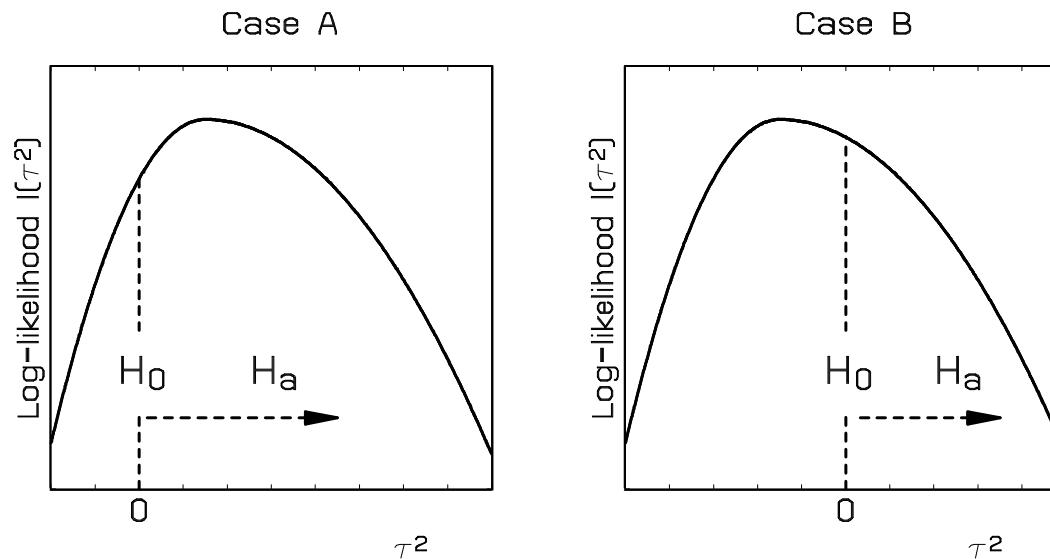
for some non-negative scalar d_{11}

- Asymptotic null-distribution equals $-2 \ln \lambda_N \longrightarrow \chi^2_{0:1}$, the mixture of χ^2_0 and χ^2_1 with equal weights 0.5:



- Under H_0 , $-2 \ln \lambda_N$ equals 0 in 50% of the cases
- Intuitive explanation:
 - ▷ consider the extended parameter space \mathbb{R} for d_{11}
 - ▷ under H_0 , \widehat{d}_{11} will be negative in 50% of the cases
 - ▷ under the restriction $d_{11} \geq 0$, these cases lead to $\widehat{d}_{11} = 0$
 - ▷ hence, $L_{\text{ML}}(\widehat{\boldsymbol{\theta}}_{\text{ML},0}) = L_{\text{ML}}(\widehat{\boldsymbol{\theta}}_{\text{ML}})$ in 50% of the cases

- Graphically ($\tau^2 = d_{11}$):



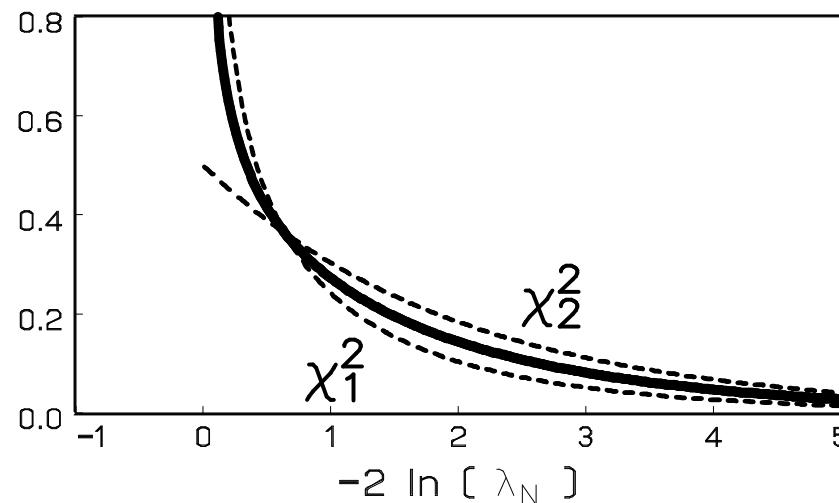
Case 2: One versus two Random Effects

- Hypothesis of interest:

$$H_0 : D = \begin{pmatrix} d_{11} & 0 \\ 0 & 0 \end{pmatrix},$$

for $d_{11} > 0$, versus H_A that D is (2×2) positive semidefinite

- Asymptotic null-distribution: $-2 \ln \lambda_N \longrightarrow \chi_{1:2}^2$, the mixture of χ_1^2 and χ_2^2 with equal weights 0.5:



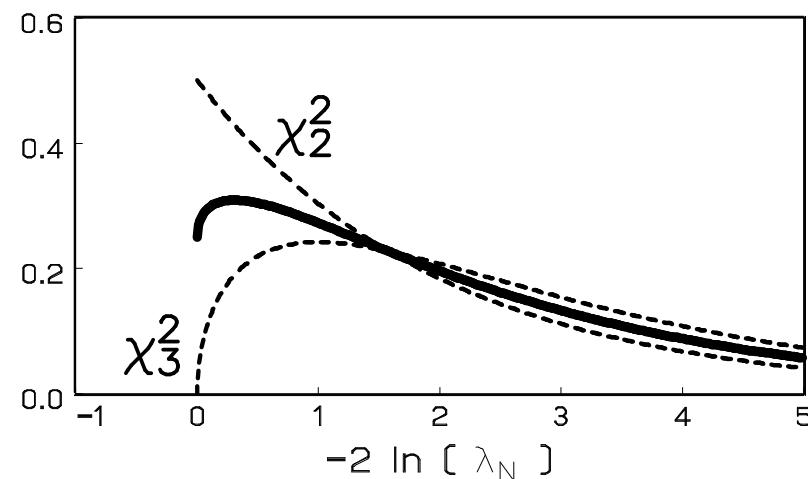
Case 3: q versus $q + 1$ Random Effects

- Hypothesis of interest:

$$H_0 : D = \begin{pmatrix} D_{11} & \mathbf{0} \\ \mathbf{0}' & 0 \end{pmatrix},$$

for D_{11} ($q \times q$) positive definite, versus H_A that D is $((q + 1) \times (q + 1))$ positive semidefinite.

- Asymptotic null-distribution: $-2 \ln \lambda_N \longrightarrow \chi^2_{q:q+1}$, the mixture of χ^2_q and χ^2_{q+1} with equal weights 0.5.



Case 4: q versus $q + k$ Random Effects

- Hypothesis of interest:

$$H_0 : D = \begin{pmatrix} D_{11} & 0 \\ 0 & 0 \end{pmatrix},$$

for D_{11} ($q \times q$) positive definite, versus H_A that D is $((q + k) \times (q + k))$ positive semidefinite.

- Simulations needed to derive asymptotic null distribution

Conclusions

- Correcting for the boundary problem reduces p -values
- Thus, ignoring the boundary problem too often leads to over-simplified covariance structures
- Hence, ignoring the boundary problem may invalidate inferences, even for the mean structure

9.2.6 Example: Rat Data

- We reconsider the model with random intercepts and slopes for the rat data:

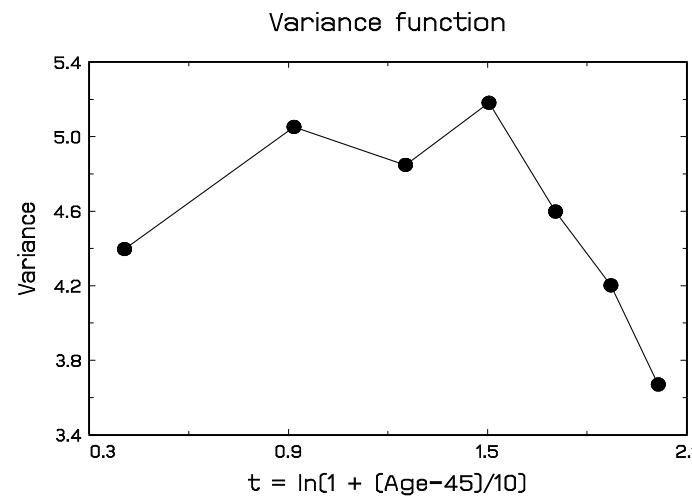
$$Y_{ij} = (\beta_0 + b_{1i}) + (\beta_1 L_i + \beta_2 H_i + \beta_3 C_i + b_{2i})t_{ij} + \varepsilon_{ij}$$

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

- The marginal model assumes linear average trends with common intercept for the 3 groups, and covariance structure:

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

- Exploring the variance function yields:



- This suggested earlier that the above random-effects model might not be valid, as it does not allow negative curvature in the variance function
- It is therefore of interest to test whether the random slopes b_{2i} may be left out of the model.
- Interpretation:
 - ▷ On hierarchical level: all rats receiving the same treatment have the same slope
 - ▷ On marginal level: constant variance, constant correlation

- Null-hypothesis to be tested: $H_0 : d_{12} = d_{22} = 0$
- REML estimates under hierarchical and marginal interpretation, as well as under H_0 :

Effect	Parameter	Parameter restrictions for α		Under H_0 REMLE (s.e.)
		$d_{ii} \geq 0, \sigma^2 \geq 0$	$d_{ii} \in \mathbb{R}, \sigma^2 \in \mathbb{R}$	
Intercept	β_0	68.606 (0.325)	68.618 (0.313)	68.607 (0.331)
Time effects:				
Low dose	β_1	7.503 (0.228)	7.475 (0.198)	7.507 (0.225)
High dose	β_2	6.877 (0.231)	6.890 (0.198)	6.871 (0.228)
Control	β_3	7.319 (0.285)	7.284 (0.254)	7.507 (0.225)
Covariance of b_i :				
$\text{var}(b_{1i})$	d_{11}	3.369 (1.123)	2.921 (1.019)	3.565 (0.808)
$\text{var}(b_{2i})$	d_{22}	0.000 (—)	-0.287 (0.169)	(—)
$\text{cov}(b_{1i}, b_{2i})$	$d_{12} = d_{21}$	0.090 (0.381)	0.462 (0.357)	(—)
Residual variance:				
$\text{var}(\varepsilon_{ij})$	σ^2	1.445 (0.145)	1.522 (0.165)	1.445 (0.145)
REML log-likelihood		-466.173	-465.193	-466.202

Test Under Marginal Interpretation

- Unrestricted parameter space for α , no boundary problem

- Wald test:

▷ Test statistic:

$$\begin{aligned} & \begin{pmatrix} \widehat{d}_{12} & \widehat{d}_{22} \end{pmatrix} \begin{pmatrix} \widehat{\text{Var}}(\widehat{d}_{12}) & \widehat{\text{Cov}}(\widehat{d}_{12}, \widehat{d}_{22}) \\ \widehat{\text{Cov}}(\widehat{d}_{12}, \widehat{d}_{22}) & \widehat{\text{Var}}(\widehat{d}_{22}) \end{pmatrix}^{-1} \begin{pmatrix} \widehat{d}_{12} \\ \widehat{d}_{22} \end{pmatrix} \\ & = \begin{pmatrix} 0.462 & -0.287 \end{pmatrix} \begin{pmatrix} 0.127 & -0.038 \\ -0.038 & 0.029 \end{pmatrix}^{-1} \begin{pmatrix} 0.462 \\ -0.287 \end{pmatrix} = 2.936, \end{aligned}$$

▷ p -value:

$$P(\chi^2_2 \geq 2.936 \mid H_0) = 0.2304$$

- LR test:

- ▷ Test statistic:

$$-2 \ln \lambda_N = -2(-466.202 + 465.193) = 2.018$$

- ▷ *p*-value:

$$P(\chi_2^2 \geq 2.018 \mid H_0) = 0.3646$$

Test Under Hierarchical Interpretation

- Restricted parameter space for α (positive semi-definite D), boundary problem !
- LR test statistic:

$$-2 \ln \lambda_N = -2(-466.202 + 466.173) = 0.058$$

- p -value:

$$P(\chi^2_{1:2} \geq 0.058 \mid H_0)$$

$$= 0.5 P(\chi^2_1 \geq 0.058 \mid H_0) + 0.5 P(\chi^2_2 \geq 0.058 \mid H_0) = 0.8906$$

- Note that the naive p -value, obtained from ignoring the boundary problem is indeed larger:

$$P(\chi^2_2 \geq 0.058 \mid H_0) = 0.9714$$

Reduced Model

- Under both model interpretations, H_0 was accepted, leading to the reduced model:

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

- Marginal interpretation:

- ▷ linear average trends with common intercept for the 3 groups
 - ▷ constant variance estimated to be

$$\widehat{d}_{11} + \widehat{\sigma}^2 = 3.565 + 1.445 = 5.010$$

- ▷ constant (intraclass) correlation

$$\widehat{\rho}_I = \frac{\widehat{d}_{11}}{\widehat{d}_{11} + \widehat{\sigma}^2} = 0.712$$

- The hierarchical interpretation, possible since $\widehat{d}_{11} = 3.565 > 0$, is that heterogeneity between rats is restricted to differences in starting values, not slopes.

9.3 Information Criteria

9.3.1 Definition of Information Criteria

- LR tests can only be used to compare nested models
- How to compare non-nested models ?
- The general idea behind the LR test for comparing model A to a more extensive model B is to select model A if the increase in likelihood under model B is small compared to increase in complexity
- A similar argument can be used to compare non-nested models A and B

- One then selects the model with the largest (log-)likelihood provided it is not (too) complex
- The model is selected with the highest **penalized** log-likelihood $\ell - \mathcal{F}(\#\boldsymbol{\theta})$ for some function $\mathcal{F}(\cdot)$ of the number $\#\boldsymbol{\theta}$ of parameters in the model.
- Different functions $\mathcal{F}(\cdot)$ lead to different criteria:

Criterion	Definition of $\mathcal{F}(\cdot)^*$
Akaike (AIC)	$\mathcal{F}(\#\boldsymbol{\theta}) = \#\boldsymbol{\theta}$
Schwarz (SBC)	$\mathcal{F}(\#\boldsymbol{\theta}) = (\#\boldsymbol{\theta} \ln n^*)/2$
Hannan and Quinn (HQIC)	$\mathcal{F}(\#\boldsymbol{\theta}) = \#\boldsymbol{\theta} \ln(\ln n^*)$
Bozdogan (CAIC)	$\mathcal{F}(\#\boldsymbol{\theta}) = \#\boldsymbol{\theta} (\ln n^* + 1)/2$

: $n^ = n = \sum_{i=1}^N n_i$ under ML

: $n^ = n - p$ under REML

- Information criteria are no formal testing procedures !
- For the comparison of models with different mean structures, information criteria should be based on ML rather than REML, as otherwise the likelihood values would be based on different sets of error contrasts, and therefore would no longer be comparable.

9.3.2 Example: Rat Data

- Consider the random-intercepts model for the rat data:

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

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

- We now want to compare this model with a model which assumes common average slope for the 3 treatments.
- Information criteria can be obtained in SAS from adding the option ‘ic’ to the PROC MIXED statement:

```
proc mixed data=rats method=ml ic;
```

- Summary of results:

Mean structure	ℓ_{ML}	# θ	AIC	SBC
Separate average slopes	-464.326	6	-470.326	-480.914
Common average slope	-466.622	4	-470.622	-477.681

- Selected models:
 - ▷ AIC: model with separate slopes
 - ▷ SBC: model with common slopes
- Based on Wald test, the average slopes are found not to be significantly different from each other ($p = 0.0987$)

Chapter 10

Inference for the Random Effects

- ▷ Empirical Bayes inference
- ▷ Best linear unbiased prediction
- ▷ Example: Prostate data
- ▷ Shrinkage
- ▷ Example: Random-intercepts model
- ▷ Example: Prostate data
- ▷ Normality assumption for random effects

10.1 Empirical Bayes Inference

- Random effects \mathbf{b}_i reflect how the evolution for the i th subject deviates from the expected evolution $X_i\boldsymbol{\beta}$.
- Estimation of the \mathbf{b}_i helpful for detecting outlying profiles
- This is only meaningful under the hierarchical model interpretation:

$$\mathbf{Y}_i | \mathbf{b}_i \sim N(X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i, \Sigma_i) \quad \mathbf{b}_i \sim N(\mathbf{0}, D)$$

- Since the \mathbf{b}_i are random, it is most natural to use Bayesian methods
- Terminology: prior distribution $N(\mathbf{0}, D)$ for \mathbf{b}_i

- Posterior density:

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

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

$$\propto \dots$$

$$\propto \exp \left\{ -\frac{1}{2} (\mathbf{b}_i - DZ'_i W_i (\mathbf{y}_i - X_i \boldsymbol{\beta}))' \Lambda_i^{-1} (\mathbf{b}_i - DZ'_i W_i (\mathbf{y}_i - X_i \boldsymbol{\beta})) \right\}$$

for some positive definite matrix Λ_i .

- Posterior distribution:

$$\mathbf{b}_i | \mathbf{y}_i \sim N(DZ'_i W_i (\mathbf{y}_i - X_i \boldsymbol{\beta}), \Lambda_i)$$

- Posterior mean as estimate for \boldsymbol{b}_i :

$$\widehat{\boldsymbol{b}_i}(\boldsymbol{\theta}) = E[\boldsymbol{b}_i \mid \boldsymbol{Y}_i = \boldsymbol{y}_i] = \int \boldsymbol{b}_i f(\boldsymbol{b}_i | \boldsymbol{y}_i) d\boldsymbol{b}_i = DZ'_i W_i(\boldsymbol{\alpha})(\boldsymbol{y}_i - X_i \boldsymbol{\beta})$$

- $\widehat{\boldsymbol{b}_i}(\boldsymbol{\theta})$ is normally distributed with covariance matrix

$$\text{var}(\widehat{\boldsymbol{b}_i}(\boldsymbol{\theta})) = DZ'_i \left\{ W_i - W_i X_i \left(\sum_{i=1}^N X'_i W_i X_i \right)^{-1} X'_i W_i \right\} Z_i D$$

- Note that inference for \boldsymbol{b}_i should account for the variability in \boldsymbol{b}_i
- Therefore, inference for \boldsymbol{b}_i is usually based on

$$\text{var}(\widehat{\boldsymbol{b}_i}(\boldsymbol{\theta}) - \boldsymbol{b}_i) = D - \text{var}(\widehat{\boldsymbol{b}_i}(\boldsymbol{\theta}))$$

- Wald tests can be derived
- Parameters in θ are replaced by their ML or REML estimates, obtained from fitting the marginal model.
- $\widehat{b}_i = \widehat{b}_i(\widehat{\theta})$ is called the Empirical Bayes estimate of b_i .
- Approximate t - and F -tests to account for the variability introduced by replacing θ by $\widehat{\theta}$, similar to tests for fixed effects.

10.2 Best Linear Unbiased Prediction (BLUP)

- Often, parameters of interest are linear combinations of fixed effects in β and random effects in b_i
- For example, a subject-specific slope is the sum of the average slope for subjects with the same covariate values, and the subject-specific random slope for that subject.
- In general, suppose $u = \lambda'_\beta \beta + \lambda'_b b_i$ is of interest
- Conditionally on α , $\hat{u} = \lambda'_\beta \hat{\beta} + \lambda'_b \hat{b}_i$ is BLUP:
 - ▷ linear in the observations Y_i
 - ▷ unbiased for u
 - ▷ minimum variance among all unbiased linear estimators

10.3 Example: Prostate Data

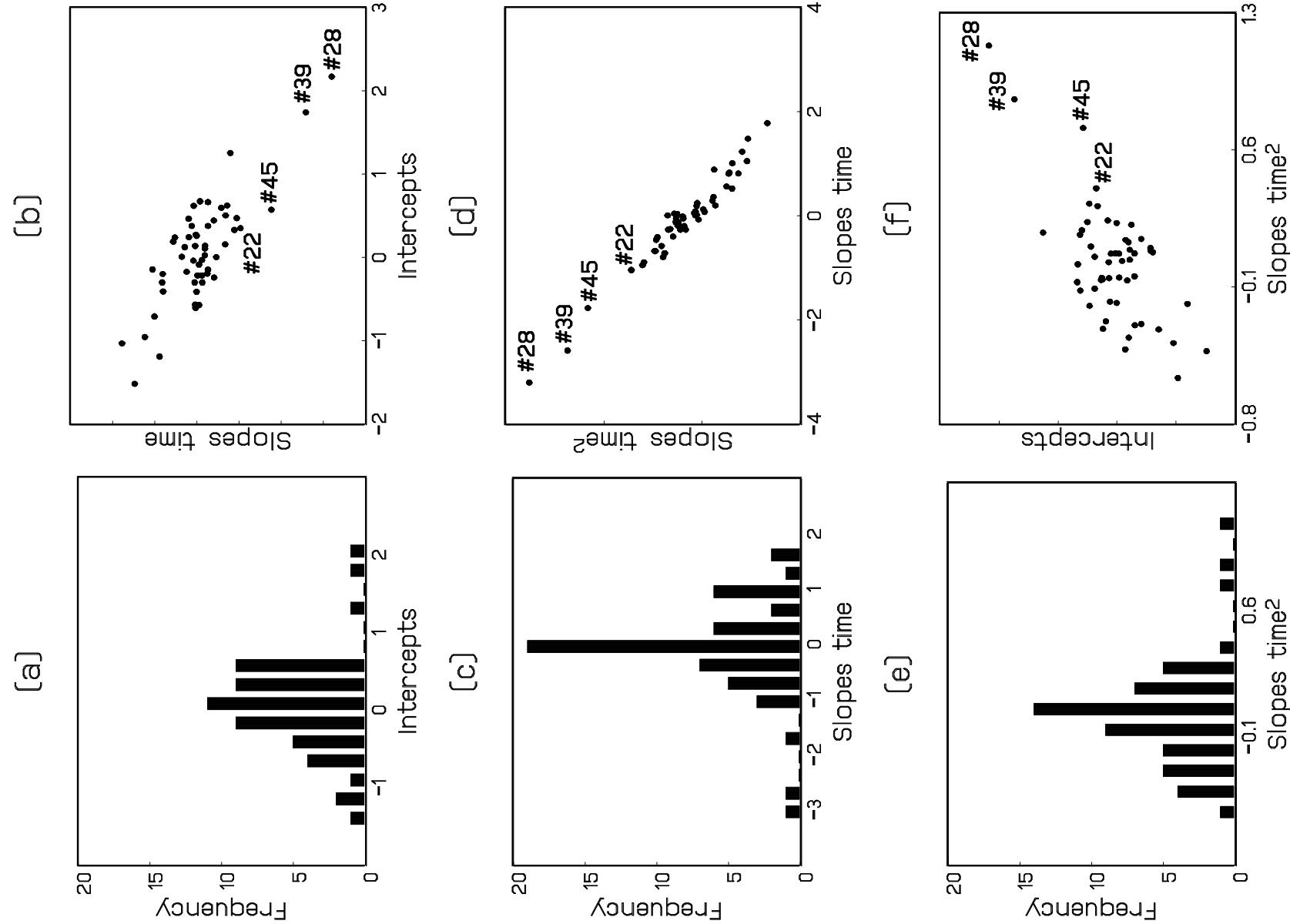
- We reconsider the reduced model:

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

- 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{b}_i are used to detect model deviations or subjects with ‘exceptional’ evolutions over time



- 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 time² 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

10.4 Shrinkage Estimators $\widehat{\boldsymbol{b}_i}$

- Consider the prediction of the evolution of the i th subject:

$$\begin{aligned}\widehat{\mathbf{Y}}_i &\equiv X_i \widehat{\boldsymbol{\beta}} + Z_i \widehat{\boldsymbol{b}_i} \\&= X_i \widehat{\boldsymbol{\beta}} + Z_i D Z'_i V_i^{-1} (\mathbf{y}_i - X_i \widehat{\boldsymbol{\beta}}) \\&= (I_{n_i} - Z_i D Z'_i V_i^{-1}) X_i \widehat{\boldsymbol{\beta}} + Z_i D Z'_i V_i^{-1} \mathbf{y}_i \\&= \Sigma_i V_i^{-1} X_i \widehat{\boldsymbol{\beta}} + (I_{n_i} - \Sigma_i V_i^{-1}) \mathbf{y}_i,\end{aligned}$$

- Hence, $\widehat{\mathbf{Y}}_i$ is a weighted mean of the population-averaged profile $X_i \widehat{\boldsymbol{\beta}}$ and the observed data \mathbf{y}_i , with weights $\widehat{\Sigma}_i V_i^{-1}$ and $I_{n_i} - \widehat{\Sigma}_i V_i^{-1}$ respectively.

- Note that $X_i\hat{\beta}$ gets much weight if the residual variability is ‘large’ in comparison to the total variability.
- This phenomenon is usually called shrinkage :

The observed data are shrunk towards the prior average profile $X_i\beta$.

- This is also reflected in the fact that for any linear combination $\lambda' b_i$ of random effects,

$$\text{var}(\lambda' \widehat{b}_i) \leq \text{var}(\lambda' b_i).$$

10.5 Example: Random-intercepts Model

- Consider the random-intercepts model, without serial correlation:

▷ $Z_i = \mathbf{1}_{n_i}$, vector of ones

▷ $D = \sigma_b^2$, scalar

▷ $\Sigma_i = \sigma^2 I_{n_i}$

- The EB estimate for the random intercept b_i then equals

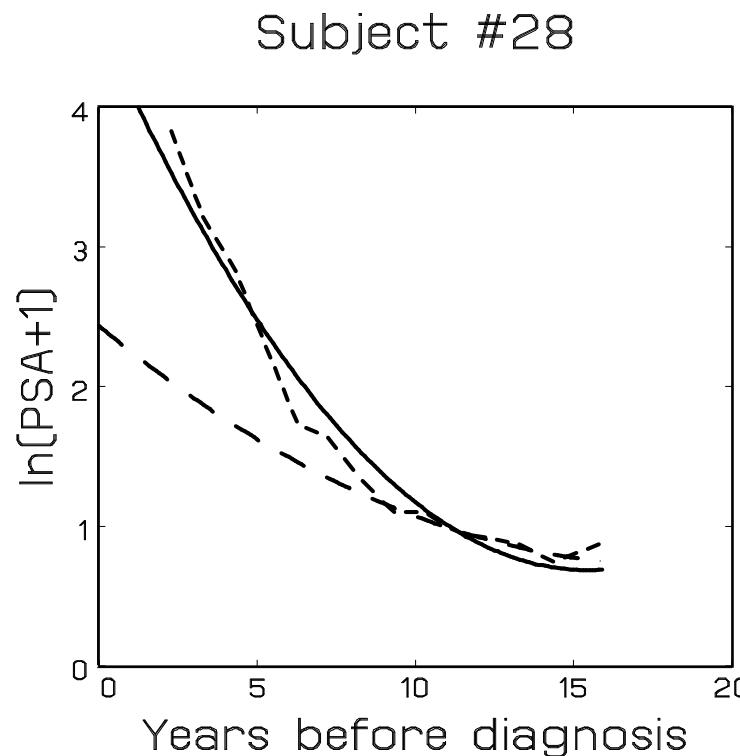
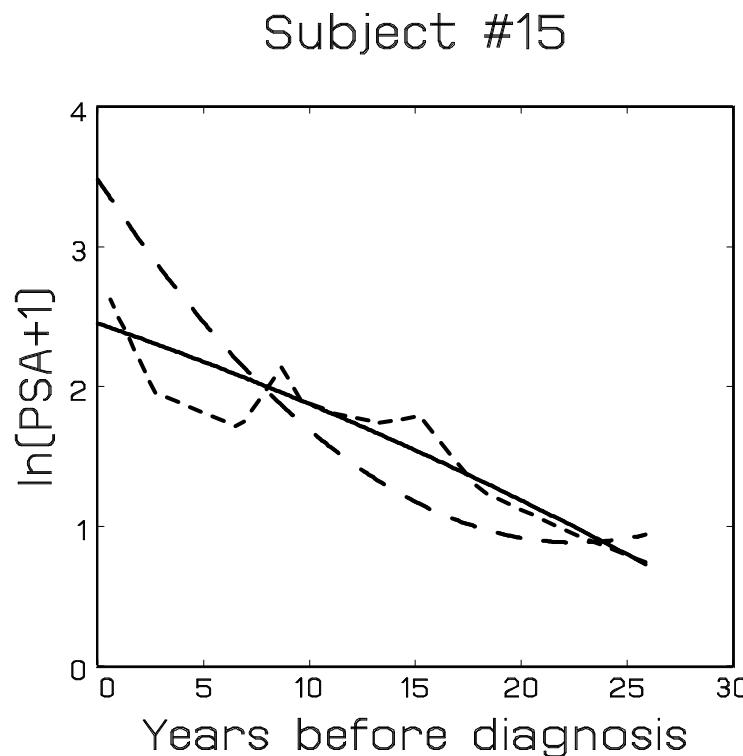
$$\begin{aligned}\widehat{b}_i &= \sigma_b^2 \mathbf{1}_{n_i}' (\sigma_b^2 \mathbf{1}_{n_i} \mathbf{1}_{n_i}' + \sigma^2 I_{n_i})^{-1} (\mathbf{y}_i - X_i \boldsymbol{\beta}) \\ &= \frac{\sigma_b^2}{\sigma^2} \mathbf{1}_{n_i}' \left(I_{n_i} - \frac{\sigma_b^2}{\sigma^2 + n_i \sigma_b^2} \mathbf{1}_{n_i} \mathbf{1}_{n_i}' \right) (\mathbf{y}_i - X_i \boldsymbol{\beta}) \\ &= \frac{n_i \sigma_b^2}{\sigma^2 + n_i \sigma_b^2} \frac{1}{n_i} \sum_{j=1}^{n_i} (y_{ij} - X_i^{[j]} \boldsymbol{\beta})\end{aligned}$$

- Remarks:

- ▷ \widehat{b}_i is weighted average of 0 (prior mean) and the average residual for subject i
- ▷ less shrinkage the larger n_i
- ▷ less shrinkage the smaller σ^2 relative to σ_b^2

10.6 Example: Prostate Data

- Comparison of predicted, average, and observed profiles for the subjects #15 and #28, obtained under the reduced model:



- Illustration of the shrinkage effect :

$$\widehat{\text{Var}}(\widehat{\boldsymbol{b}_i}) = \begin{pmatrix} 0.403 & -0.440 & 0.131 \\ -0.440 & 0.729 & -0.253 \\ 0.131 & -0.253 & 0.092 \end{pmatrix}, \quad \widehat{\boldsymbol{D}} = \begin{pmatrix} 0.443 & -0.490 & 0.148 \\ -0.490 & 0.842 & -0.300 \\ 0.148 & -0.300 & 0.114 \end{pmatrix}$$

10.7 The Normality Assumption for Random Effects

- In practice, histograms of EB estimates are often used to check the normality assumption for the random effects
- However, since

$$\widehat{\boldsymbol{b}_i} = DZ'_i W_i (\boldsymbol{y}_i - X_i \boldsymbol{\beta})$$

$$\text{var}(\widehat{\boldsymbol{b}_i}) = DZ'_i \left\{ W_i - W_i X_i \left(\sum_{i=1}^N X'_i W_i X_i \right)^{-1} X'_i W_i \right\} Z_i D$$

one should at least first standardize the EB estimates

- Further, due to the shrinkage property the EB estimates do not fully reflect the heterogeneity in the data.

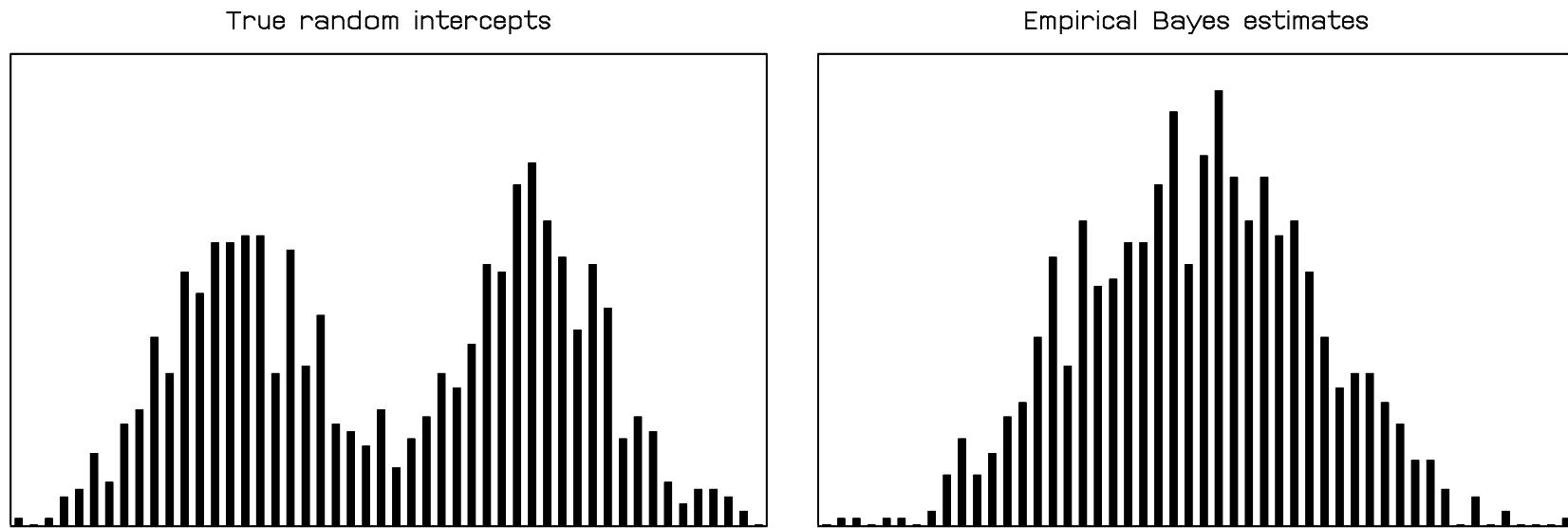
- Small simulation example:

- ▷ 1000 profiles with 5 measurements, balanced
- ▷ 1000 random intercepts sampled from

$$\frac{1}{2}N(-2, 1) + \frac{1}{2}N(2, 1)$$

- ▷ $\Sigma_i = \sigma^2 I_{n_i}$, $\sigma^2 = 30$
- ▷ Data analysed assuming normality for the intercepts

- ▷ Histogram of sampled intercepts and empirical Bayes estimates:



- ▷ Clearly, severe shrinkage forces the estimates \hat{b}_i to satisfy the normality assumption

- Conclusion:

EB estimates obtained under normality
cannot be used to check normality

- This suggests that the only possibility to check the normality assumption is to fit a more general model, with the classical linear mixed model as special case, and to compare both models using LR methods

10.8 The Heterogeneity Model

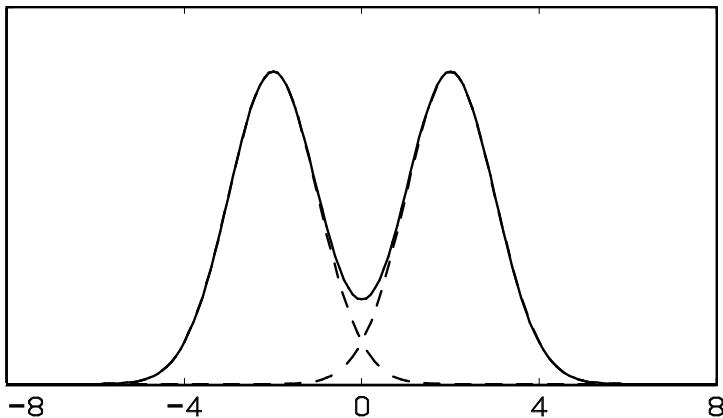
- One possible extension of the linear mixed model is to assume a finite mixture as random-effects distribution:

$$\mathbf{b}_i \sim \sum_{j=1}^g p_j N(\boldsymbol{\mu}_j, D), \quad \text{with } \sum_{j=1}^g p_j = 1 \text{ and } \sum_{j=1}^g p_j \boldsymbol{\mu}_j = \mathbf{0}$$

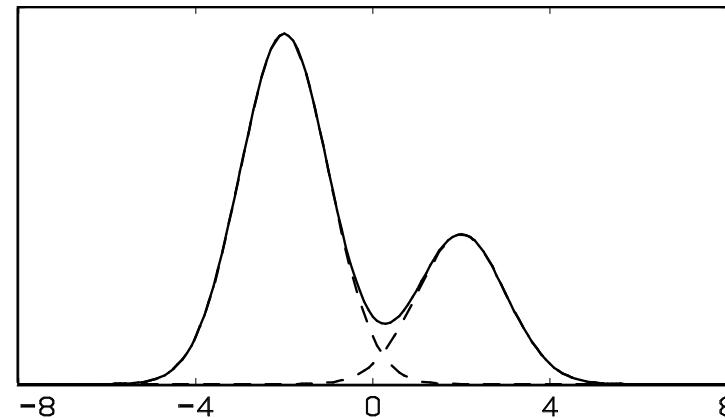
- Interpretation:
 - ▷ Population consists of g subpopulations
 - ▷ Each subpopulation contains fraction p_j of total population
 - ▷ In each subpopulation, a linear mixed model holds
- The classical model is a special case: $g = 1$

- Very flexible class of parametric models for random-effects distribution:

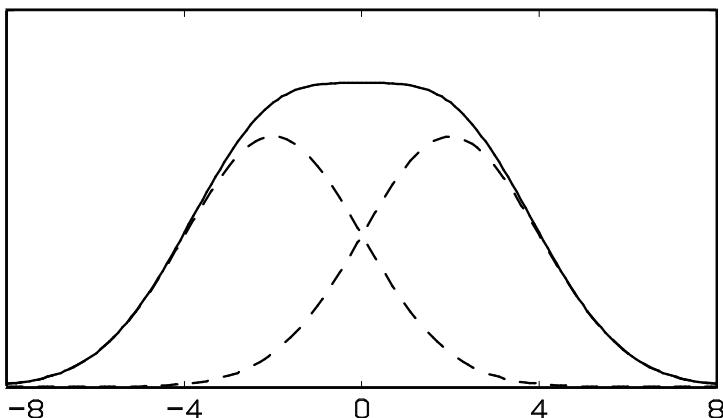
$$\mu_1 = -2, \mu_2 = 2, \sigma_b^2 = 1, p = 0.5$$



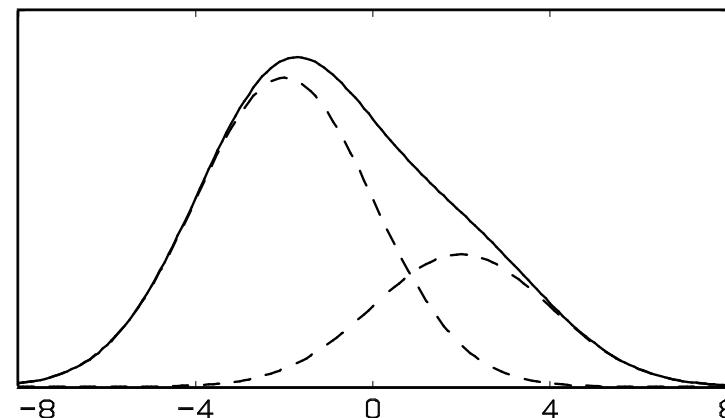
$$\mu_1 = -2, \mu_2 = 2, \sigma_b^2 = 1, p = 0.7$$



$$\mu_1 = -2, \mu_2 = 2, \sigma_b^2 = 4, p = 0.5$$

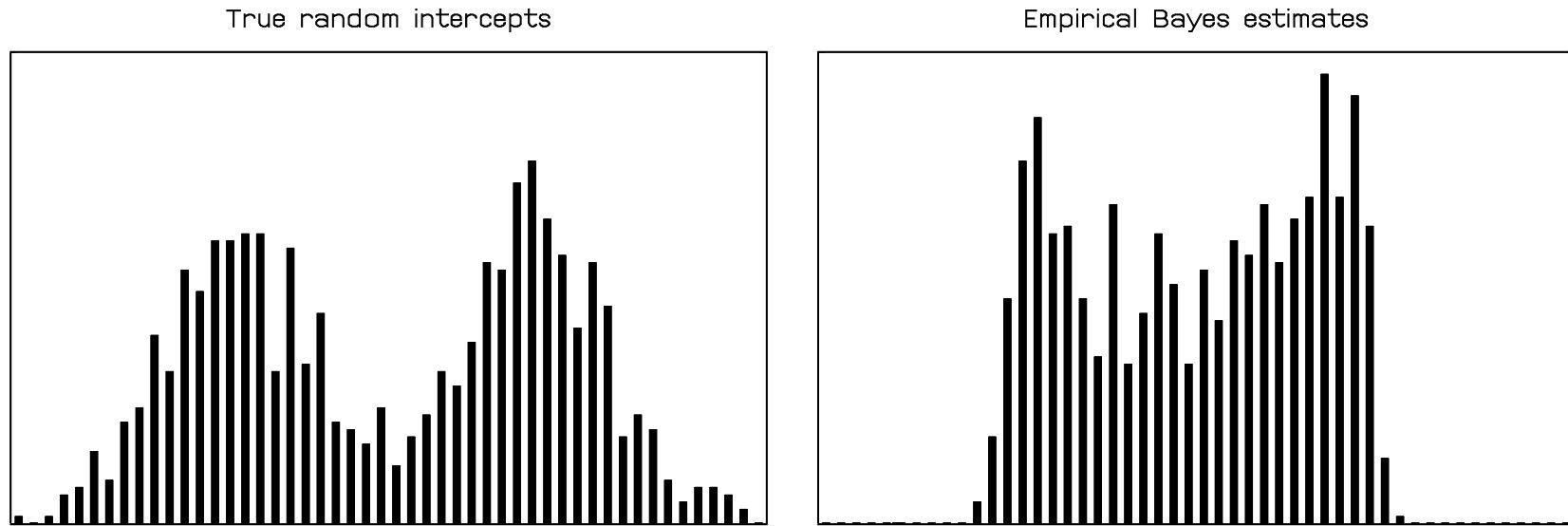


$$\mu_1 = -2, \mu_2 = 2, \sigma_b^2 = 4, p = 0.7$$



- Fitting of the model is based on the EM algorithm
- SAS macro available
- EB estimates can be calculated under the heterogeneity model
- Small simulation example:
 - ▷ 1000 profiles with 5 measurements, balanced
 - ▷ 1000 random intercepts sampled from
$$\frac{1}{2}N(-2, 1) + \frac{1}{2}N(2, 1)$$
 - ▷ $\Sigma_i = \sigma^2 I_{n_i}$, $\sigma^2 = 30$
 - ▷ Data analysed under heterogeneity model

▷ Histogram of sampled intercepts and empirical Bayes estimates:



- ▷ The correct random-effects distribution is (much) better reflected, than before under the assumption of normality

Chapter 11

General Guidelines for Model Building

- ▷ Introduction
- ▷ General strategy
- ▷ Example: The prostate data

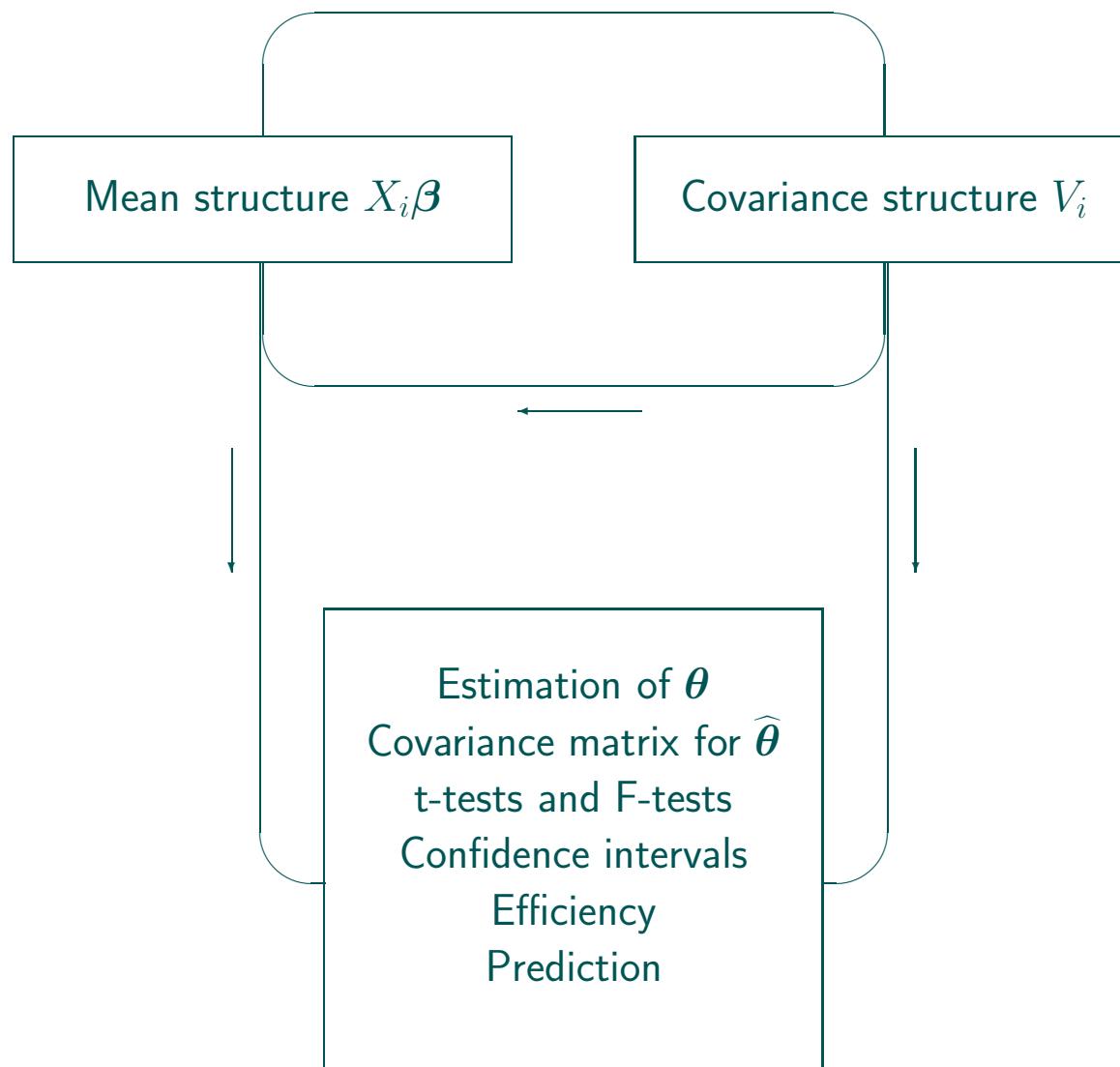
11.1 Introduction

- Marginal linear mixed model:

$$\mathbf{Y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{Z}_i D \mathbf{Z}'_i + \sigma^2 I_{n_i} + \tau^2 H_i)$$

- Fitting a linear mixed model requires specification of a mean structure, as well as covariance structure
- Mean structure:
 - ▷ Covariates
 - ▷ Time effects
 - ▷ Interactions
- Covariance structure:
 - ▷ Random effects
 - ▷ Serial correlation

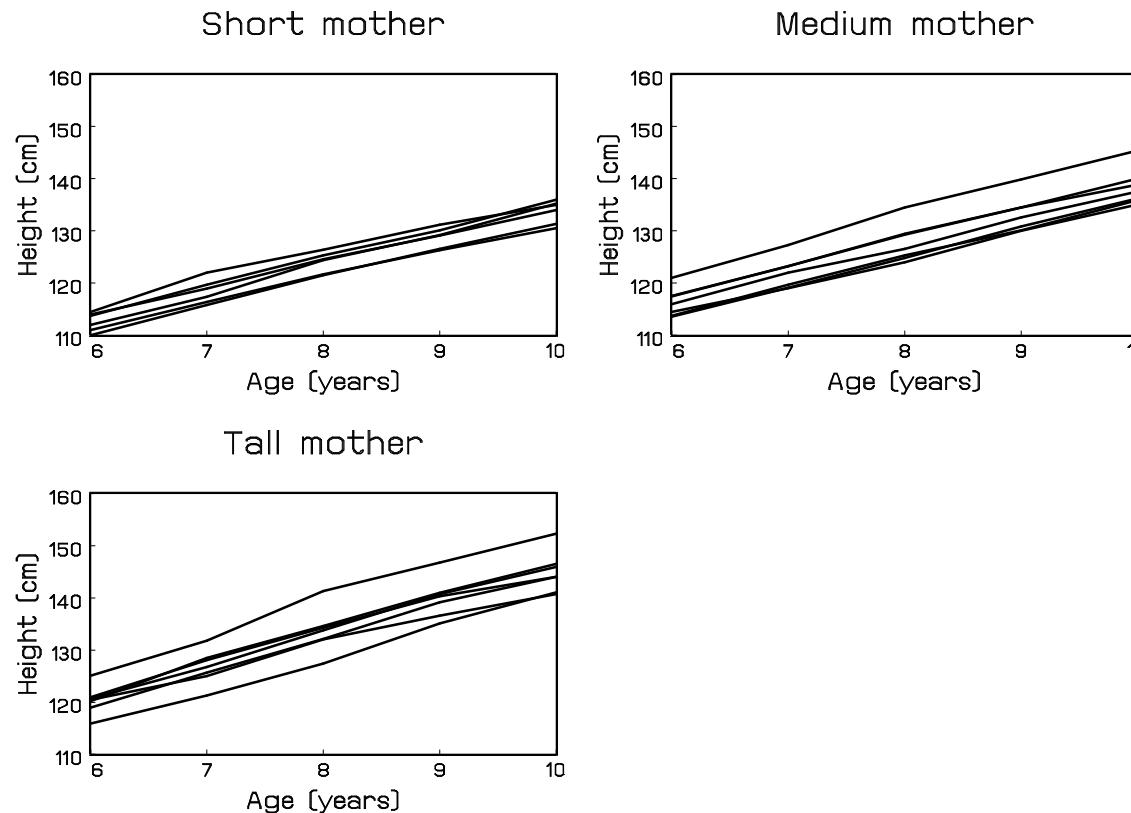
- Both components affect each other:



- When most variability is due to between-subject variability, the two-stage approach will often lead to acceptable marginal models
- In the presence of a lot within-subject variability, the two-stage approach is less straightforward
- Also, a two-stage approach may imply unrealistic marginal models

- For example, reconsider the growth curves:

- ▷ Individual profiles:



- ▷ A random-intercepts model seems reasonable

▷ However, the covariance matrix equals

$$\begin{pmatrix} 6.11 & 6.88 & 8.26 & 7.44 & 7.18 \\ 6.88 & 8.53 & 9.78 & 9.01 & 8.70 \\ 8.26 & 9.78 & 12.04 & 10.99 & 10.96 \\ 7.44 & 9.01 & 10.99 & 10.42 & 10.56 \\ 7.18 & 8.70 & 10.96 & 10.56 & 11.24 \end{pmatrix}.$$

- The aim of this chapter is to discuss some general guidelines for model building.

11.2 General Strategy

$$\mathbf{Y}_i = X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i$$

1. Preliminary mean structure $X_i\boldsymbol{\beta}$
2. Preliminary random-effects structure $Z_i\mathbf{b}_i$
3. Residual covariance structure Σ_i
4. Reduction of the random-effects structure $Z_i\mathbf{b}_i$
5. Reduction of the mean structure $X_i\boldsymbol{\beta}$

11.3 Preliminary Mean Structure

11.3.1 Strategy

- Remove all systematic trends from the data, by calculating OLS residual profiles :

$$\mathbf{r}_i = \mathbf{y}_i - X_i \hat{\boldsymbol{\beta}}_{\text{OLS}} \approx Z_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i$$

- For balanced designs with few covariates :

Saturated mean structure

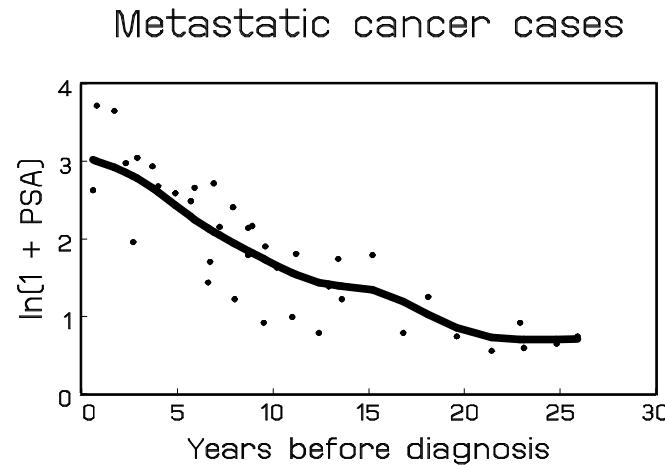
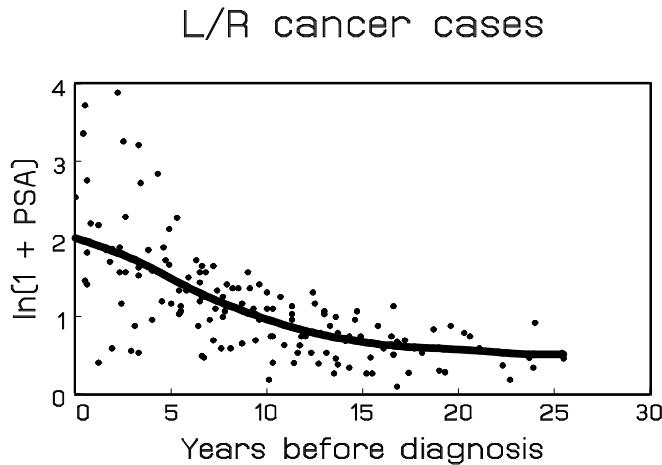
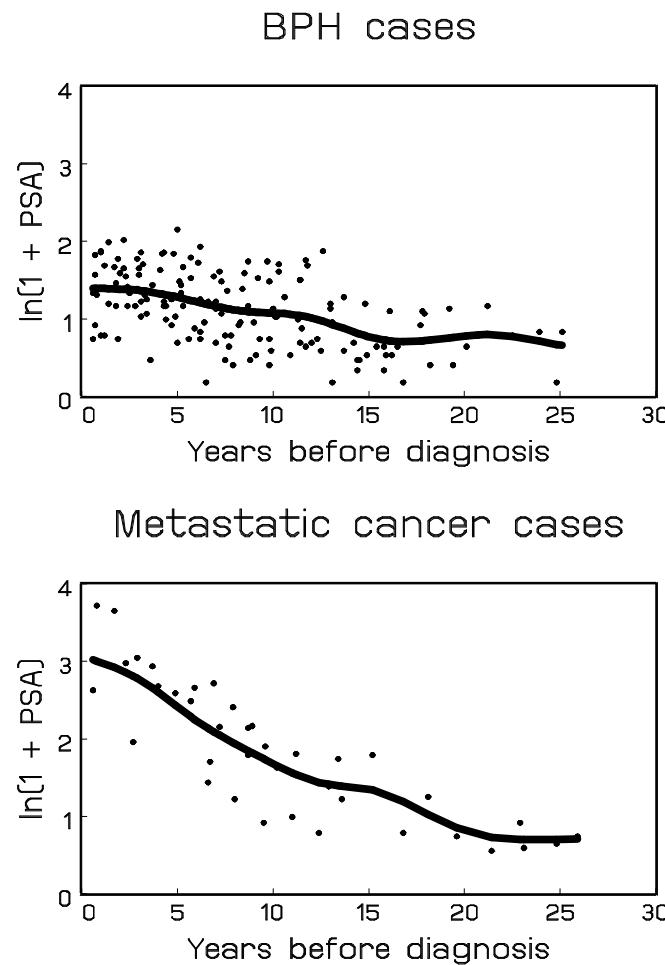
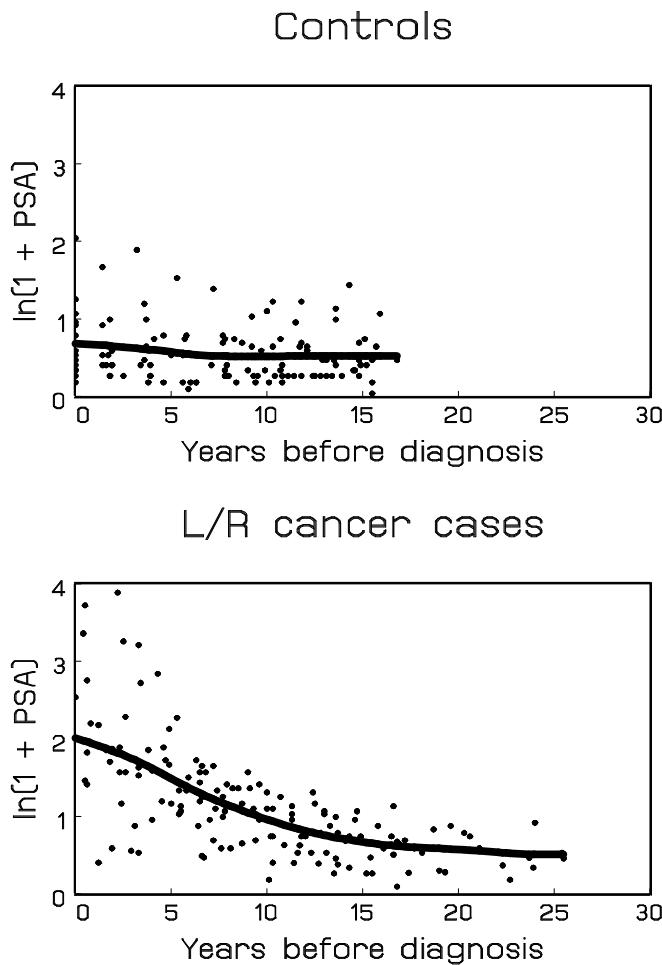
- For balanced designs with many covariates, or for highly unbalanced data sets :

The most elaborate model one is prepared
to consider for the mean structure

- Selection of preliminary mean structures will be based on exploratory tools for the mean.
- Note that the calculation of $\hat{\beta}_{OLS}$ ignores the longitudinal structure, and can be obtained in any regression module
- Provided the preliminary mean structure is ‘sufficiently right’, consistency of $\hat{\beta}_{OLS}$ follows from the theory on robust inference for the fixed effects.

11.3.2 Example: Prostate Data

- Smoothed average trend within each group:



- Quadratic function over time, within each diagnostic group
- Correction for age, via the inclusion of age , $age \times time$ and $age \times time^2$.
- Note that this yields the same model as the model originally obtained from a two-stage approach, containing 15 fixed effects

11.4 Preliminary Random-effects Structure

11.4.1 Strategy

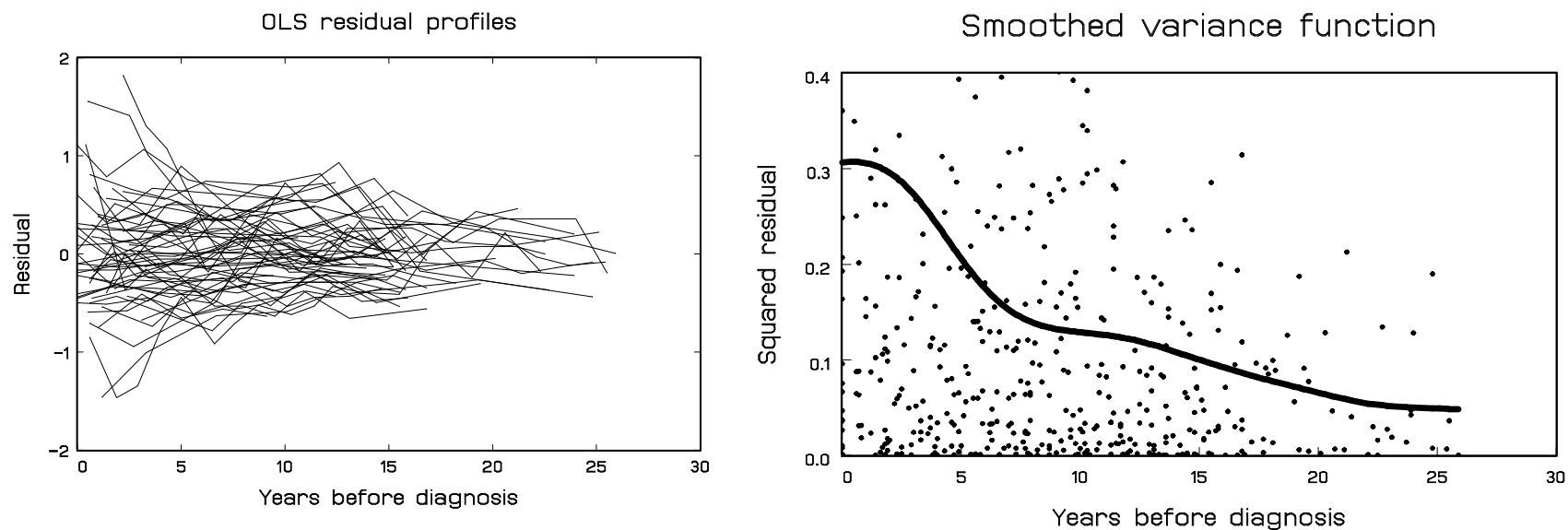
$$r_i \approx Z_i b_i + \varepsilon_i$$

- Explore the residual profiles
- Any structure left, may indicate the presence of subject-specific regression coefficients
- Try to describe the each residual profile with a (relatively) simple model.

- Do not include covariates in Z_i which are not included in X_i . Otherwise, it is not justified to assume $E(\mathbf{b}_i) = 0$.
- Use ‘well-formulated’ models: Do not include higher-order terms unless all lower-order terms are included as well.
- Compare implied variance and covariance functions with results from exploratory tools for covariance structure

11.4.2 Example: Prostate Data

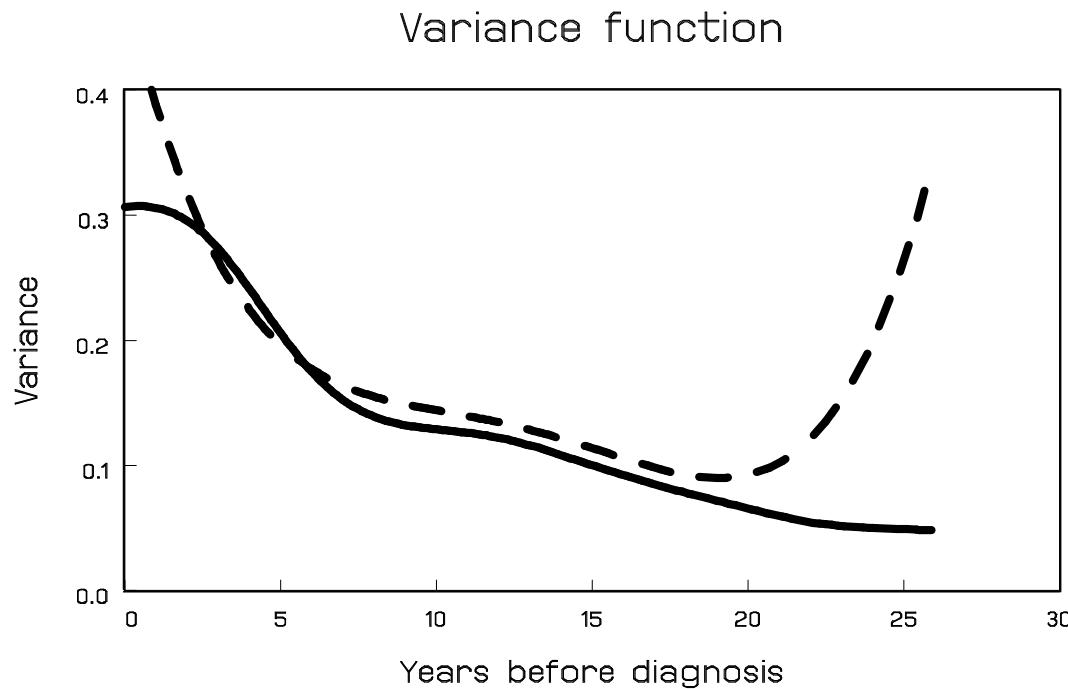
- OLS residual profiles and smoothed average of squared OLS residuals:



- We assume a quadratic function for each residual profile
- This results in a model with random intercepts, and random slopes for the linear as well as quadratic time effect.

- Variance function:
$$\begin{pmatrix} 1 & t & t^2 \end{pmatrix} D \begin{pmatrix} 1 \\ t \\ t^2 \end{pmatrix} + \sigma^2$$

- Comparison of smoothed average of squared OLS residuals and fitted variance function:



- Possible explanation for observed differences:
 - ▷ Small t : some subjects have extremely large responses close to diagnosis. This may have inflated the fitted variance
 - ▷ Large t : few observations available: only 24 out of 463 measurements taken earlier than 20 years prior to diagnosis.

11.5 Residual Covariance Structure

11.5.1 Strategy

$$\mathbf{r}_i \approx Z_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i$$

- Which covariance matrix Σ_i for $\boldsymbol{\varepsilon}_i$?
- In many applications, random effects explain most of the variability
- Therefore, in the presence of random effects other than intercepts, often $\Sigma_i = \sigma^2 I_{n_i}$ is assumed

- However, many other covariance structures can be specified as well

- A special class of parametric models for Σ_i is obtained from splitting ε_i into a measurement error component $\varepsilon_{(1)i}$ and a serial correlation component $\varepsilon_{(2)i}$:

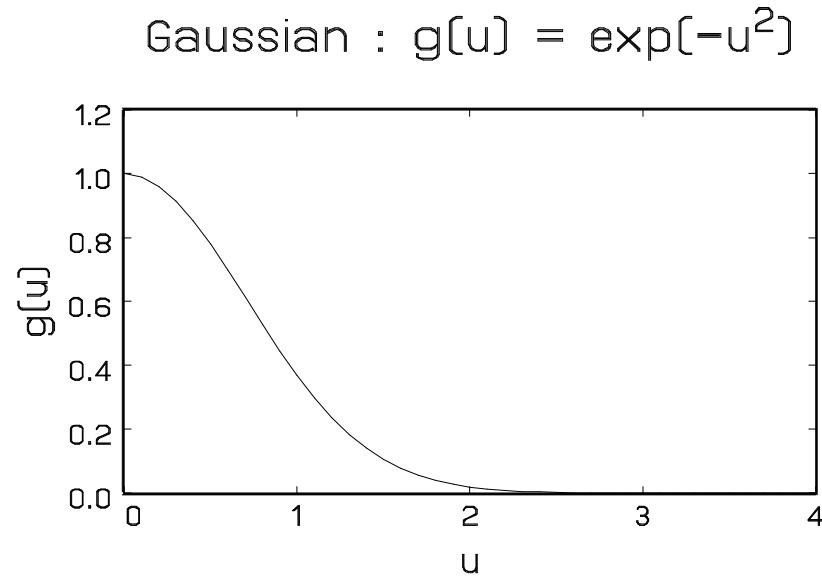
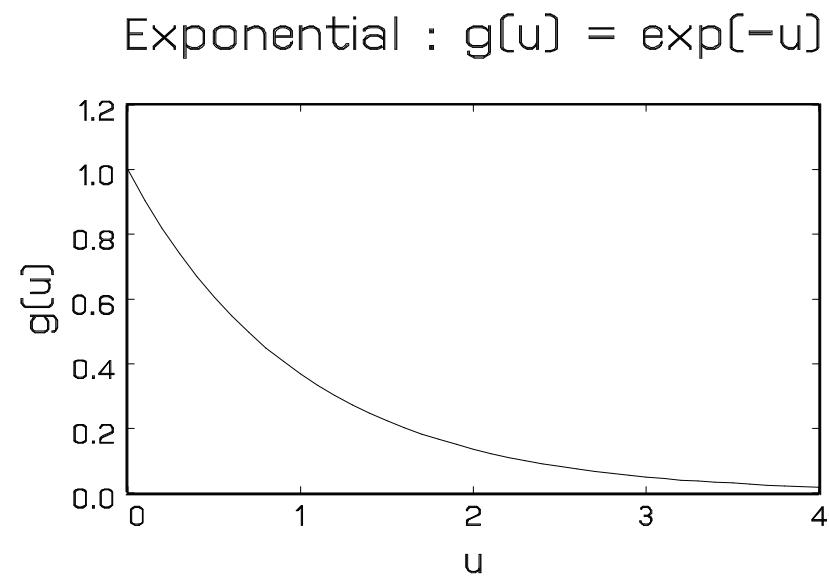
$$\mathbf{Y}_i = X_i \boldsymbol{\beta} + Z_i \mathbf{b}_i + \varepsilon_{(1)i} + \varepsilon_{(2)i}$$

$$\left. \begin{array}{l} \mathbf{b}_i \sim N(\mathbf{0}, D) \\ \varepsilon_{(1)i} \sim N(\mathbf{0}, \sigma^2 I_{n_i}) \\ \varepsilon_{(2)i} \sim N(\mathbf{0}, \tau^2 H_i) \end{array} \right\} \text{independent}$$

- Only the correlation matrix H_i then still needs to be specified
- H_i is assumed to have (j, k) element of the form $h_{ijk} = g(|t_{ij} - t_{ik}|)$ for some decreasing function $g(\cdot)$ with $g(0) = 1$

- Frequently used functions $g(\cdot)$:
 - ▷ Exponential serial correlation: $g(u) = \exp(-\phi u)$
 - ▷ Gaussian serial correlation: $g(u) = \exp(-\phi u^2)$

- Graphical representation ($\phi = 1$):



- When only random intercepts are included, the semi-variogram can be used to explore the presence and the nature of serial correlation
- When other random effects are present as well, an extension of the variogram is needed.
- Also, a variety of serial correlation functions can be fitted and compared.

11.5.2 Example: Prostate Data

- Based on the preliminary mean and random-effects structures, several serial correlation functions can be fitted.
- For example, a model with Gaussian serial correlation can be fitted in SAS using the following program:

```
proc mixed data=prostate method=reml;
  class id group timeclss;
  model lnpsa = group age group*time age*time group*time2 age*time2 / noint solution;
  random intercept time time2 / type=un subject=id g gcorr v vcorr;
  repeated timeclss / type=sp(gau)(time) local subject=id r rcorr;
  run;
```

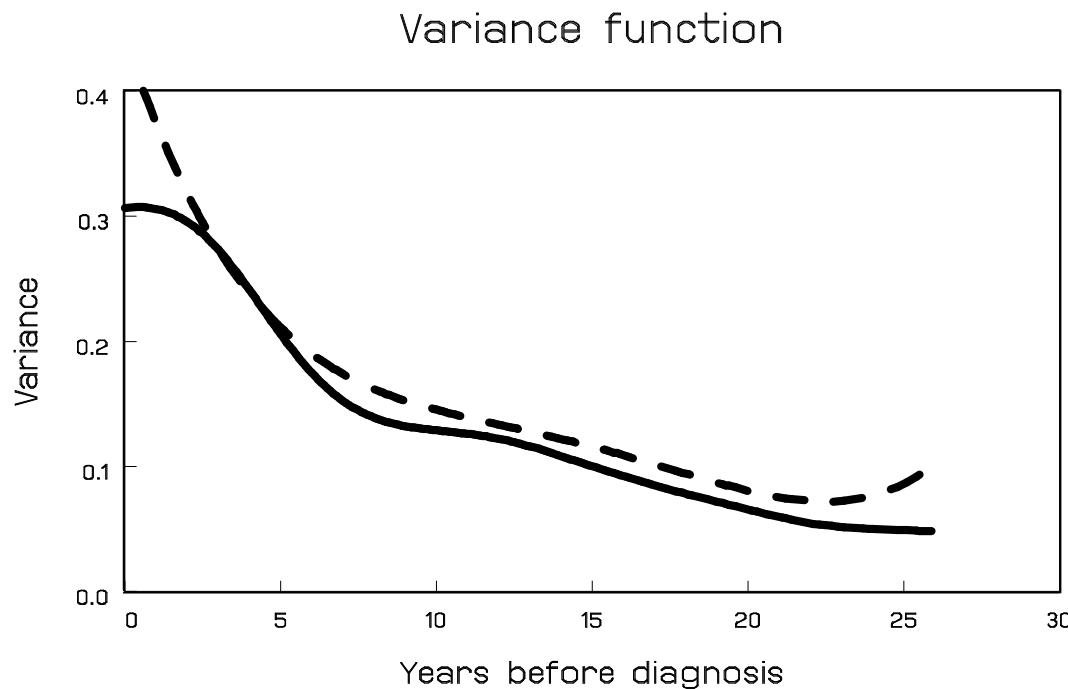
- REPEATED statement:
 - ▷ the serial correlation model is specified in the ‘type’ option
 - ▷ ‘local’ is added to include measurement error

- Summary of model fits:

Residual covariance structure	REML log-likelihood
Measurement error	-31.235
Measurement error + Gaussian	-24.787
Measurement error + exponential	-24.266

- The presence of serial correlation is clearly detected
- However, there seems to be little information in the data to distinguish between different serial correlation structures
- Practical experience suggests that including serial correlation, if present, is far more important than correctly specifying the serial correlation function.

- Variance function:
$$\begin{pmatrix} 1 & t & t^2 \end{pmatrix} D \begin{pmatrix} 1 \\ t \\ t^2 \end{pmatrix} + \sigma^2 + \tau^2$$
- Comparison of smoothed average of squared OLS residuals and fitted variance function:



- Inclusion of serial correlation leads to different estimates for the variance components in D
- Therefore, the fitted variance function differs from the one obtained before without serial correlation
- The deviation for small values of t remains, but the functions coincide better for large t .

11.6 Reduction of Preliminary Random-effects Structure

- Once an appropriate residual covariance model is obtained, one can try to reduce the number of random effects in the preliminary random-effects structure
- This is done based on inferential tools for variance components

11.7 Reduction of Preliminary Mean Structure

- Once an appropriate covariance model is obtained, one can try to reduce the number of covariates in the preliminary mean structure
- This is done based on inferential tools for fixed effects
- In case there is still some doubt about the validity of the marginal covariance structure, robust inference can be used to still obtain correct inferences.

11.8 Example: Prostate Data

- Fixed effects estimates from the final model, under Gaussian serial correlation, and without serial correlation:

Effect	Parameter	Serial corr.	No serial corr.
		Estimate (s.e.)	Estimate (s.e.)
Age effect	β_1	0.015 (0.006)	0.016 (0.006)
Intercepts:			
Control	β_2	-0.496 (0.411)	-0.564 (0.428)
BPH	β_3	0.320 (0.470)	0.275 (0.488)
L/R cancer	β_4	1.216 (0.469)	1.099 (0.486)
Met. cancer	β_5	2.353 (0.518)	2.284 (0.531)
Time effects:			
BPH	β_8	-0.376 (0.070)	-0.410 (0.068)
L/R cancer	β_9	-1.877 (0.210)	-1.870 (0.233)
Met. cancer	β_{10}	-2.274 (0.244)	-2.303 (0.262)
Time ² effects:			
Cancer	$\beta_{14} = \beta_{15}$	0.484 (0.073)	0.510 (0.088)

- Variance components estimates from the final model, under Gaussian serial correlation, and without serial correlation:

Effect	Parameter	Serial corr.	No serial corr.
		Estimate (s.e.)	Estimate (s.e.)
Covariance of b_i:			
$\text{var}(b_{1i})$	d_{11}	0.393 (0.093)	0.443 (0.093)
$\text{var}(b_{2i})$	d_{22}	0.550 (0.187)	0.842 (0.203)
$\text{var}(b_{3i})$	d_{33}	0.056 (0.028)	0.114 (0.035)
$\text{cov}(b_{1i}, b_{2i})$	$d_{12} = d_{21}$	-0.382 (0.114)	-0.490 (0.124)
$\text{cov}(b_{2i}, b_{3i})$	$d_{23} = d_{32}$	-0.170 (0.070)	-0.300 (0.082)
$\text{cov}(b_{3i}, b_{1i})$	$d_{13} = d_{31}$	0.098 (0.039)	0.148 (0.047)
Measurement error variance:			
$\text{var}(\varepsilon_{(1)ij})$	σ^2	0.023 (0.002)	0.028 (0.002)
Gaussian serial correlation:			
$\text{var}(\varepsilon_{(2)ij})$	τ^2	0.029 (0.018)	— (—)
Rate of exponential decrease	$1/\sqrt{\phi}$	0.599 (0.192)	— (—)
REML log-likelihood		-13.704	-20.165

- Many standard errors are smaller under the model which includes the Gaussian serial correlation component
- Hence, adding the serial correlation leads to more efficient inferences for most parameters in the marginal model.

11.9 Random-effects Structure versus Residual Covariance Structure

- The marginal covariance structure equals

$$V_i = Z_i D Z_i' + \Sigma_i$$

- Hence, the residual covariance Σ_i models all variation not yet been accounted for by random effects
- In practice, one therefore often observes strong competition between these two sources of stochastic variation
- This is also reflected in substantial correlations between the variance components estimates

- As an example, consider the final model for the prostate data, with Gaussian serial correlation
- Estimated correlation matrix for variance components estimates:

$$\text{Corr} \left(\widehat{d_{11}}, \widehat{d_{12}}, \widehat{d_{22}}, \widehat{d_{13}}, \widehat{d_{23}}, \widehat{d_{33}}, \widehat{\tau^2}, 1/\sqrt{\widehat{\phi}}, \widehat{\sigma}^2 \right)$$

$$= \left(\begin{array}{cccccc|ccc} 1.00 & -0.87 & 0.62 & 0.70 & -0.49 & 0.39 & -0.18 & -0.10 & -0.00 \\ -0.87 & 1.00 & -0.85 & -0.94 & 0.75 & -0.63 & 0.21 & 0.08 & -0.03 \\ 0.62 & -0.85 & 1.00 & 0.88 & -0.97 & 0.91 & -0.46 & -0.29 & 0.02 \\ 0.70 & -0.94 & 0.88 & 1.00 & -0.82 & 0.72 & -0.22 & -0.06 & 0.05 \\ -0.49 & 0.75 & -0.97 & -0.82 & 1.00 & -0.97 & 0.51 & 0.33 & -0.02 \\ 0.39 & -0.63 & 0.91 & 0.72 & -0.97 & 1.00 & -0.57 & -0.38 & 0.01 \\ \hline -0.18 & 0.21 & -0.46 & -0.22 & 0.51 & -0.57 & 1.00 & 0.81 & 0.04 \\ -0.10 & 0.08 & -0.29 & -0.06 & 0.33 & -0.38 & 0.81 & 1.00 & 0.32 \\ -0.00 & -0.03 & 0.02 & 0.05 & -0.02 & 0.01 & 0.04 & 0.32 & 1.00 \end{array} \right).$$

- Relatively large correlations between $\hat{\tau}^2$ and the estimates of some of the parameters in D
- Small correlations between $\hat{\sigma}^2$ and the other estimates, except for $1/\sqrt{\hat{\phi}}$.
- Indeed, the serial correlation component vanishes for ϕ becoming infinitely large.

Chapter 12

Power Analyses under Linear Mixed Models

- ▷ F test for fixed effects
- ▷ Calculation in SAS
- ▷ Examples

12.1 F Statistics for Fixed Effects

- Consider a general linear hypothesis

$$H_0 : L\beta = \mathbf{0}, \quad \text{versus} \quad H_A : L\beta \neq \mathbf{0}$$

- F test statistic:

$$F = \frac{\widehat{\beta}' L' \left[L \left(\sum_{i=1}^N X_i' V_i^{-1}(\widehat{\alpha}) X_i \right)^{-1} L' \right]^{-1} L \widehat{\beta}}{\text{rank}(L)}.$$

- Approximate null-distribution of F is F with numerator degrees of freedom equal to $\text{rank}(L)$

- Denominator degrees of freedom to be estimated from the data:
 - ▷ Containment method
 - ▷ Satterthwaite approximation
 - ▷ Kenward and Roger approximation
 - ▷ ...
- In general (not necessarily under H_0), F is approximately F distributed with the same numbers of degrees of freedom, but with non-centrality parameter

$$\phi = \boldsymbol{\beta}' L' \left[L \left(\sum_{i=1}^N X_i' V_i^{-1}(\boldsymbol{\alpha}) X_i \right)^{-1} L' \right]^{-1} L \boldsymbol{\beta}$$

which equals 0 under H_0 .

- This can be used to calculate powers under a variety of models, and under a variety of alternative hypotheses

- Note that ϕ is equal to $\text{rank}(L) \times F$, and with $\widehat{\beta}$ replaced by β
- The SAS procedure MIXED can therefore be used for the calculation of ϕ and the related numbers of degrees of freedom.

12.2 Calculation in SAS

- Construct a data set of the same dimension and with the same covariates and factor values as the design for which power is to be calculated
- Use as responses \mathbf{y}_i the average values $X_i\boldsymbol{\beta}$ under the alternative model
- The fixed effects estimate will then be equal to

$$\begin{aligned}\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) &= \left(\sum_{i=1}^N X_i' W_i(\boldsymbol{\alpha}) X_i \right)^{-1} \sum_{i=1}^N X_i' W_i(\boldsymbol{\alpha}) \mathbf{y}_i \\ &= \left(\sum_{i=1}^N X_i' W_i(\boldsymbol{\alpha}) X_i \right)^{-1} \sum_{i=1}^N X_i' W_i(\boldsymbol{\alpha}) X_i \boldsymbol{\beta} = \boldsymbol{\beta}\end{aligned}$$

- Hence, the F -statistic reported by SAS will equal $\phi/\text{rank}(L)$

- This calculated F value, and the associated numbers of degrees of freedom can be saved and used afterwards for calculation of the power.
- Note that this requires keeping the variance components in α fixed, equal to the assumed population values.
- Steps in calculations:
 - ▷ Use PROC MIXED to calculate ϕ , and degrees of freedom ν_1 and ν_2
 - ▷ Calculate critical value F_c :
$$P(F_{\nu_1, \nu_2, 0} > F_c) = \text{level of significance}$$
 - ▷ Calculate power:
$$\text{power} = P(F_{\nu_1, \nu_2, \phi} > F_c)$$
- The SAS functions ‘finv’ and ‘probf’ are used to calculate F_c and the power

12.3 Example 1

- Re-consider the random-intercepts model previously discussed for the rat data:

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

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

- This model is fitted in SAS as follows:

```
proc mixed data = test;
  class treat rat;
  model y = treat*t / solution ddfm=kr;
  random intercept / subject=rat;
  contrast 'Equal slopes' treat*t 1 -1 0, treat*t 1 0 -1;
run;
```

- The CONTRAST statement is added to test equality of the average slopes.
- Suppose a new experiment is to be designed, to test the above hypothesis, when the true parameter values are given by:

Effect	Parameter	True value
Intercept	β_0	68
Time effects:		
Low dose	β_1	7
High dose	β_2	7.5
Control	β_3	6.5
Covariance of b_i :		
$\text{var}(b_{1i})$	d_{11}	3.6
Residual variance:		
$\text{var}(\varepsilon_{ij})$	σ^2	1.4

- The power of a design with 10 rats per treatment group is calculated as follows:
 - ▷ Construction of data set with expected averages as response values:

```
data power;
do treat=1 to 3;
  do rat=1 to 10;
    do age=50 to 110 by 10;
      t=log(1+(age-45)/10);
      if treat=1 then y=68 + 7.5*t;
      if treat=2 then y=68 + 7.0*t;
      if treat=3 then y=68 + 6.5*t;
      output;
    end;
  end;
end;
```

- ▷ Fit model, keeping the variance components equal to their true values:

```
proc mixed data = power  noprofile;
  class treat rat;
  model y = treat*t ;
  random intercept / subject=rat(treat);
  parms (3.6) (1.4) / noiter;
  contrast 'Equal slopes' treat*t 1 -1 0,
            treat*t 1 0 -1;
  ods output contrasts=c;
run;
```

- ▷ PARMS statement to specify starting values for the variance components.
- ▷ The ‘noiter’ and ‘noprofile’ options request that no iterations be performed and that inferences are based on the specified values.
- ▷ ODS statement needed to save F , ν_1 and ν_2 .

▷ Calculation of ϕ , F_c and power:

```
data power;
set c;
alpha=0.05;
ncparm=numdf*fvalue;
fc=finv(1-alpha,numdf,dendf,0);
power=1-probf(fc,numdf,dendf,ncparm);
run;

proc print;run;
```

● Output:

Label	Num DF	Den DF	FValue	ProbF	alpha	ncparm	fc	power
Equal slopes	2	177	4.73	0.0100	0.05	9.46367	3.04701	0.78515

- Hence, there is a power of 78.5% to detect the prespecified differences at the 5% level of significance.
- Increasing the number of rats yields the following powers:

Group size	Power
10	78.5%
11	82.5%
12	85.9%
13	88.7%
14	91.0%
15	92.9%
20	97.9%

12.4 Example 2

- We continue the previous random-intercepts model and study the effect of varying the variance components values

- Results (10 rats per group):

		d_{11}		
		3.2	3.6	4.0
σ^2	1.0	89.3%	88.5%	87.9%
	1.4	79.8%	78.5%	77.4%
	1.8	71.9%	70.3%	68.9%

- Conclusions:

- ▷ The power decreases as the total variance increases
- ▷ Keeping the total variance constant, the power increases as the intraclass correlation $\rho_I = d_{11}/(d_{11} + \sigma^2)$ increases

12.5 Example 3

12.5.1 Introduction

- Experiment for the comparison of two treatments A and B
- A total of N general practitioners (GP's) involved
- Each GP treats n subjects
- Y_{ij} is the response for subject j treated by GP i
- The analysis should account for the variability between GP's

- We use the following random-intercepts model, where the random intercepts reflect random GP effects:

$$Y_{ij} = \begin{cases} \beta_1 + b_{1i} + \varepsilon_{ij} & \text{if treatment } A \\ \beta_2 + b_{1i} + \varepsilon_{ij} & \text{if treatment } B \end{cases}$$

- Assumed true parameter values:

Effect	Parameter	True value
Fixed effects:		
Average treatment A	β_1	1
Average treatment B	β_2	2
Variance components:		
$\text{var}(b_{1i})$	d_{11}	?
$\text{var}(\varepsilon_{ij})$	σ^2	?
	$d_{11} + \sigma^2$	4

- Hence, the individual variance components are unknown. Only the total variability is known to equal 4.
- Power analyses will be performed for several values for the intraclass correlation
$$\rho_I = d_{11}/(d_{11} + \sigma^2)$$

12.5.2 Case 1: Treatments Assigned to GP's

- We now consider the situation in which the treatments will be randomly assigned to GP's, and all subjects with the same GP will be treated identically.
- Powers for $2 \times 25 = 50$ GP's, each treating 10 subjects ($\alpha = 0.05$):

ρ_I	Power
0.25	86%
0.50	65%
0.75	50%

- The power decreases as the intraclass correlation increases

12.5.3 Case 2: Treatments Assigned to Subjects

- We now consider the situation in which the treatments will be randomly assigned to subjects within GP's, with the same number $n/2$ of subjects assigned to both treatments
- Powers for $2 \times 5 = 10$ subjects within 10 GP's ($\alpha = 0.05$):

ρ_I	Power
0.25	81%
0.50	94%
0.75	100%

- The power increases as the intraclass correlation increases
- Note also that Case 2 requires many less observations than Case 1

12.5.4 Conclusion

Within-'subject' correlation

increases power for inferences on within-'subject' effects,
but decreases power for inferences on between-'subject' effects

Part II

Marginal Models for Non-Gaussian Longitudinal Data

Chapter 13

The Toenail Data

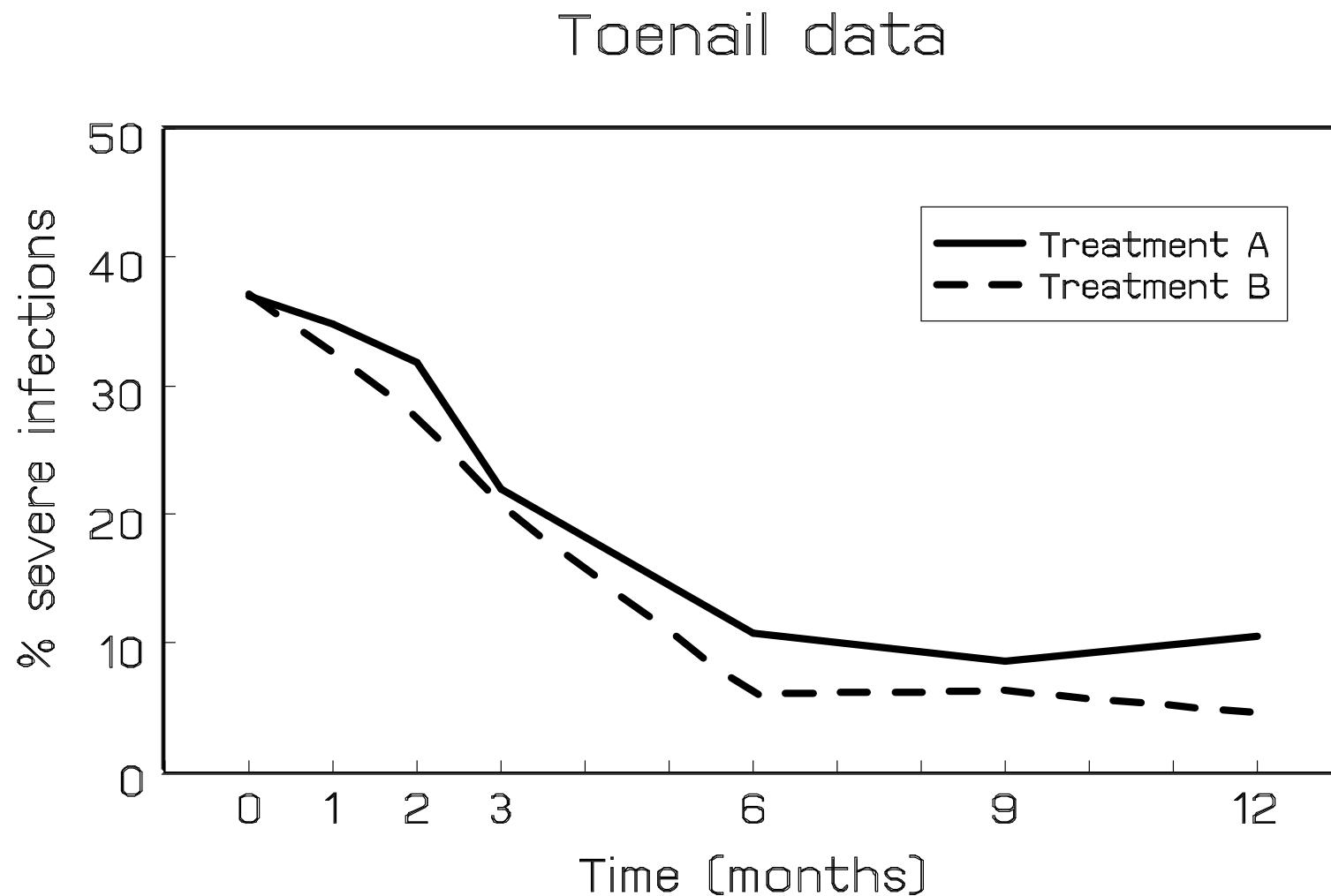
- **Toenail Dermatophyte Onychomycosis:** Common toenail infection, difficult to treat, affecting more than 2% of population.
- Classical treatments with antifungal compounds need to be administered until the whole nail has grown out healthy.
- New compounds have been developed which reduce treatment to 3 months
- Randomized, double-blind, parallel group, multicenter study for the comparison of two such new compounds (*A* and *B*) for oral treatment.

- Research question:

Severity relative to treatment of TDO ?

- 2×189 patients randomized, 36 centers
- 48 weeks of total follow up (12 months)
- 12 weeks of treatment (3 months)
- measurements at months 0, 1, 2, 3, 6, 9, 12.

- Frequencies at each visit (both treatments):



Chapter 14

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					
O	O	O	O	163	41.2
Dropouts					
O	O	O	M	51	12.91
O	O	M	M	51	12.91
O	M	M	M	63	15.95
Non-monotone missingness					
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

Chapter 15

The National Toxicology Program (NTP) Data

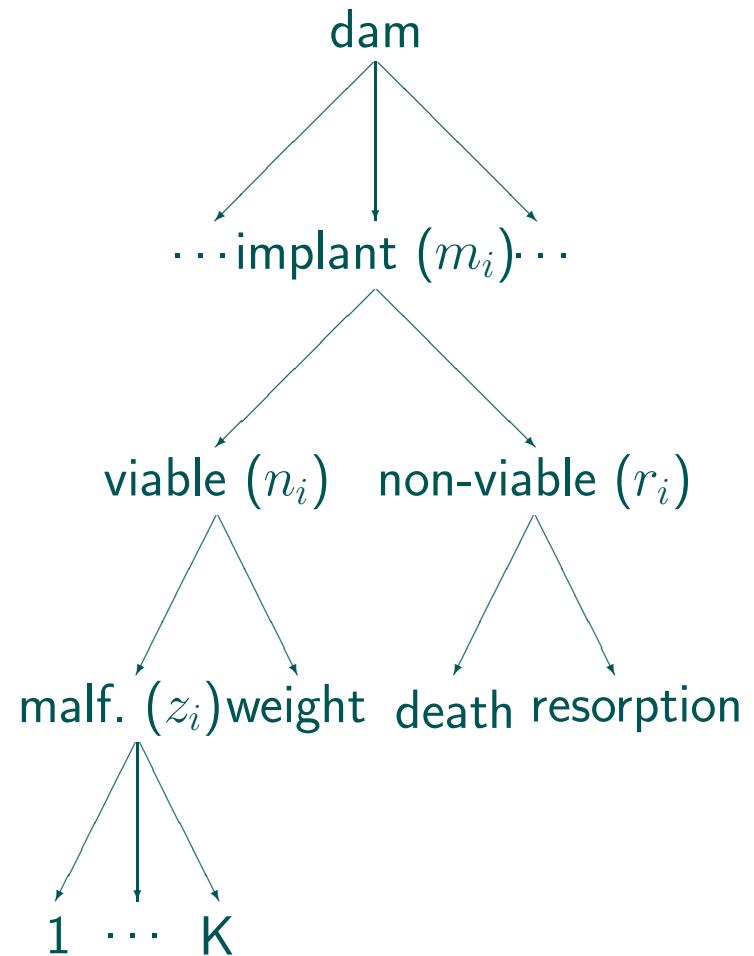
Developmental Toxicity Studies

- Research Triangle Institute
- The effect in mice of 3 chemicals:
 - ▷ DEHP: di(2-ethylhexyl)-phtalate
 - ▷ EG: ethylene glycol
 - ▷ DYME: diethylene glycol dimethyl ether

- Implanted fetuses:

- ▷ death/resorbed
- ▷ viable:
 - * weight
 - * malformations: visceral, skeletal, external

- Data structure:



Exposure	Dose	Litter						
		# Dams, ≥ 1		Live (mean)	Size	Malformations		
		Impl.	Viab.			Ext.	Visc.	Skel.
EG	0	25	25	297	11.9	0.0	0.0	0.3
	750	24	24	276	11.5	1.1	0.0	8.7
	1500	23	22	229	10.4	1.7	0.9	36.7
	3000	23	23	226	9.8	7.1	4.0	55.8
DEHP	0	30	30	330	13.2	0.0	1.5	1.2
	44	26	26	288	11.1	1.0	0.4	0.4
	91	26	26	277	10.7	5.4	7.2	4.3
	191	24	17	137	8.1	17.5	15.3	18.3
	292	25	9	50	5.6	54.0	50.0	48.0
DYME	0	21	21	282	13.4	0.0	0.0	0.0
	62.5	20	20	225	11.3	0.0	0.0	0.0
	125	24	24	290	12.1	1.0	0.0	1.0
	250	23	23	261	11.3	2.7	0.1	20.0
	500	22	22	141	6.1	66.0	19.9	79.4

Chapter 16

Generalized Linear Models

- ▷ The model
- ▷ Maximum likelihood estimation
- ▷ Examples
- ▷ McCullagh and Nelder (1989)

16.1 The Generalized Linear Model

- Suppose a sample Y_1, \dots, Y_N of independent observations is available
- All Y_i have densities $f(y_i|\theta_i, \phi)$ which belong to the exponential family:

$$f(y|\theta_i, \phi) = \exp\{\phi^{-1}[y\theta_i - \psi(\theta_i)] + c(y, \phi)\}$$

- θ_i the natural parameter
- Linear predictor: $\theta_i = \mathbf{x}_i' \boldsymbol{\beta}$
- θ is the scale parameter (overdispersion parameter)
- $\psi(\cdot)$ is a function to be discussed next

16.2 Mean and Variance

- We start from the following general property:

$$\int f(y|\theta, \phi) dy$$

$$= \int \exp \left\{ \phi^{-1}[y\theta - \psi(\theta)] + c(y, \phi) \right\} dy = 1$$

- Taking first and second-order derivatives with respect to θ yields

$$\begin{cases} \frac{\partial}{\partial \theta} \int f(y|\theta, \phi) dy = 0 \\ \frac{\partial^2}{\partial \theta^2} \int f(y|\theta, \phi) dy = 0 \end{cases}$$

$$\iff \begin{cases} \int [y - \psi'(\theta)] f(y|\theta, \phi) dy = 0 \\ \int [\phi^{-1}(y - \psi'(\theta))^2 - \psi''(\theta)] f(y|\theta, \phi) dy = 0 \end{cases}$$

$$\iff \begin{cases} \mathbb{E}(Y) = \psi'(\theta) \\ \text{Var}(Y) = \phi\psi''(\theta) \end{cases}$$

- Note that, in general, the mean μ and the variance are related:

$$\text{Var}(Y) = \phi\psi''[\psi'^{-1}(\mu)] = \phi v(\mu)$$

- The function $v(\mu)$ is called the variance function.
- The function ψ'^{-1} which expresses θ as function of μ is called the link function.
- ψ' is *the inverse link function*

16.3 Examples

16.3.1 The Normal Model

- Model:

$$Y \sim N(\mu, \sigma^2)$$

- Density function:

$$\begin{aligned} f(y|\theta, \phi) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{\sigma^2}(y - \mu)^2\right\} \\ &= \exp\left\{\frac{1}{\sigma^2}\left(y\mu - \frac{\mu^2}{2}\right) + \left(\frac{\ln(2\pi\sigma^2)}{2} - \frac{y^2}{2\sigma^2}\right)\right\} \end{aligned}$$

- Exponential family:

$$\triangleright \theta = \mu$$

$$\triangleright \phi = \sigma^2$$

$$\triangleright \psi(\theta) = \theta^2/2$$

$$\triangleright c(y, \phi) = \frac{\ln(2\pi\phi)}{2} - \frac{y^2}{2\phi}$$

- Mean and variance function:

$$\triangleright \mu = \theta$$

$$\triangleright v(\mu) = 1$$

- Note that, under this normal model, the mean and variance are not related:

$$\phi v(\mu) = \sigma^2$$

- The link function is here the identity function: $\theta = \mu$

16.3.2 The Bernoulli Model

- Model:

$$Y \sim \text{Bernoulli}(\pi)$$

- Density function:

$$\begin{aligned} f(y|\theta, \phi) &= \pi^y(1-\pi)^{1-y} \\ &= \exp\{y \ln \pi + (1-y) \ln(1-\pi)\} \\ &= \exp\left\{y \ln\left(\frac{\pi}{1-\pi}\right) + \ln(1-\pi)\right\} \end{aligned}$$

- Exponential family:

$$\triangleright \theta = \ln\left(\frac{\pi}{1-\pi}\right)$$

$$\triangleright \phi = 1$$

$$\triangleright \psi(\theta) = \ln(1 - \pi) = \ln(1 + \exp(\theta))$$

$$\triangleright c(y, \phi) = 0$$

- Mean and variance function:

$$\triangleright \mu = \frac{\exp \theta}{1 + \exp \theta} = \pi$$

$$\triangleright v(\mu) = \frac{\exp \theta}{(1 + \exp \theta)^2} = \pi(1 - \pi)$$

- Note that, under this model, the mean and variance are related:

$$\phi v(\mu) = \mu(1 - \mu)$$

- The link function here is the logit link: $\theta = \ln\left(\frac{\mu}{1-\mu}\right)$

16.3.3 The Poisson Model

- Model:

$$Y \sim \text{Poisson}(\lambda)$$

- Density function:

$$\begin{aligned} f(y|\theta, \phi) &= \frac{e^{-\lambda} \lambda^y}{y!} \\ &= \exp\{y \ln \lambda - \lambda - \ln y!\} \end{aligned}$$

- Exponential family:

$$\triangleright \theta = \ln \lambda$$

$$\triangleright \phi = 1$$

$$\triangleright \psi(\theta) = \lambda = \exp \theta$$

$$\triangleright c(y, \phi) = -\ln y!$$

- Mean and variance function:

$$\triangleright \mu = \exp \theta = \lambda$$

$$\triangleright v(\mu) = \exp \theta = \lambda$$

- Note that, under this model, the mean and variance are related:

$$\phi v(\mu) = \mu$$

- The link function is here the log link: $\theta = \ln \mu$

16.4 Generalized Linear Models (GLM)

- Suppose a sample Y_1, \dots, Y_N of independent observations is available
- All Y_i have densities $f(y_i|\theta_i, \phi)$ which belong to the exponential family
- In GLM's, it is believed that the differences between the θ_i can be explained through a linear function of known covariates:

$$\theta_i = \mathbf{x}_i' \boldsymbol{\beta}$$

- \mathbf{x}_i is a vector of p known covariates
- $\boldsymbol{\beta}$ is the corresponding vector of unknown regression parameters, to be estimated from the data.

16.5 Maximum Likelihood Estimation

- Log-likelihood:

$$\ell(\beta, \phi) = \frac{1}{\phi} \sum_i [y_i \theta_i - \psi(\theta_i)] + \sum_i c(y_i, \phi)$$

- First order derivative with respect to β :

$$\frac{\partial \ell(\beta, \phi)}{\partial \beta} = \frac{1}{\phi} \sum_i \frac{\partial \theta_i}{\partial \beta} [y_i - \psi'(\theta_i)]$$

- The score equations for β to be solved:

$$S(\beta) = \sum_i \frac{\partial \theta_i}{\partial \beta} [y_i - \psi'(\theta_i)] = 0$$

- Since $\mu_i = \psi'(\theta_i)$ and $v_i = v(\mu_i) = \psi''(\theta_i)$, we have that

$$\frac{\partial \mu_i}{\beta} = \psi''(\theta_i) \frac{\partial \theta_i}{\partial \beta} = v_i \frac{\partial \theta_i}{\partial \beta}$$

- The score equations now become

$$S(\beta) = \sum_i \frac{\partial \mu_i}{\partial \beta} v_i^{-1} (y_i - \mu_i) = 0$$

- Note that the estimation of β depends on the density only through the means μ_i and the variance functions $v_i = v(\mu_i)$.

- The score equations need to be solved numerically:
 - ▷ iterative (re-)weighted least squares
 - ▷ Newton-Raphson
 - ▷ Fisher scoring
- Inference for β is based on classical maximum likelihood theory:
 - ▷ asymptotic Wald tests
 - ▷ likelihood ratio tests
 - ▷ score tests

- In some cases, ϕ is a known constant, in other examples, estimation of ϕ may be required to estimate the standard errors of the elements in β
- Estimation can be based on $\text{Var}(Y_i) = \phi v_i$:

$$\hat{\phi} = \frac{1}{N - p} \sum_i (y_i - \hat{\mu}_i)^2 / v_i(\hat{\mu}_i)$$

- For example, under the normal model, this would yield:

$$\hat{\sigma}^2 = \frac{1}{N - p} \sum_i (y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}})^2,$$

the mean squared error used in linear regression models to estimate the residual variance.

16.6 Illustration: The Analgesic Trial

- Early dropout (did the subject drop out after the first or the second visit) ?
- Binary response
- PROC GENMOD can fit GLMs in general
- PROC LOGISTIC can fit models for binary (and ordered) responses
- SAS code for logit link:

```
proc genmod data=earlydrp;
  model earlydrp = pca0 weight psychiat physfct / dist=b;
run;

proc logistic data=earlydrp descending;
  model earlydrp = pca0 weight psychiat physfct;
run;
```

- SAS code for probit link:

```

proc genmod data=earlydrp;
  model earlydrp = pca0 weight psychiat physfct / dist=b link=probit;
run;

proc logistic data=earlydrp descending;
  model earlydrp = pca0 weight psychiat physfct / link=probit;
run;

```

- Selected output:

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits	Chi-Square	Pr > ChiSq
Intercept	1	-1.0673	0.7328	-2.5037 0.3690	2.12	0.1453
PCAO	1	0.3981	0.1343	0.1349 0.6614	8.79	0.0030
WEIGHT	1	-0.0211	0.0072	-0.0353 -0.0070	8.55	0.0034
PSYCHIAT	1	0.7169	0.2871	0.1541 1.2796	6.23	0.0125
PHYSFCT	1	0.0121	0.0050	0.0024 0.0219	5.97	0.0145
Scale	0	1.0000	0.0000	1.0000 1.0000		

NOTE: The scale parameter was held fixed.

Chapter 17

Parametric Modeling Families

- ▷ Continuous outcomes
- ▷ Longitudinal generalized linear models
- ▷ Notation

17.1 Continuous Outcomes

- Marginal Models:

$$E(Y_{ij} | \mathbf{x}_{ij}) = \mathbf{x}'_{ij} \boldsymbol{\beta}$$

- Random-Effects Models:

$$E(Y_{ij} | \mathbf{b}_i, \mathbf{x}_{ij}) = \mathbf{x}'_{ij} \boldsymbol{\beta} + \mathbf{z}'_{ij} \mathbf{b}_i$$

- Transition Models:

$$E(Y_{ij} | Y_{i,j-1}, \dots, Y_{i1}, \mathbf{x}_{ij}) = \mathbf{x}'_{ij} \boldsymbol{\beta} + \alpha Y_{i,j-1}$$

17.2 Longitudinal Generalized Linear Models

- Normal case: easy transfer between models
- Also non-normal data can be measured repeatedly (over time)
- Lack of key distribution such as the normal [\Rightarrow]
 - ▷ A lot of modeling options
 - ▷ Introduction of non-linearity
 - ▷ No easy transfer between model families

	cross-sectional	longitudinal
normal outcome	linear model	LMM
non-normal outcome	GLM	?

17.3 Notation

- Let the outcomes for subject $i = 1, \dots, N$ be denoted as $(Y_{i1}, \dots, Y_{in_i})$.
- Group into a vector \mathbf{Y}_i :
 - ▷ Binary data: each component is either 0 or 1.
 - ▷ (Binary data: each component is either 1 or 2.)
 - ▷ (Binary data: each component is either -1 or $+1$.)
 - ▷ (Categorical data: $Y_{ij} \in \{1, \dots, c\}$.)
- The corresponding covariate vector is \mathbf{x}_{ij} .
- It is convenient to use (binary 0/1 data):

$$E(Y_{ij}) = \Pr(Y_{ij} = 1) = \mu_{ij} \quad \text{and} \quad \mu_{ijk} = E(Y_{ij}Y_{ik}) = \Pr(Y_{ij} = 1, Y_{ik} = 1)$$

Chapter 18

Conditional Models

- ▷ A log-linear model
- ▷ Quadratic version of the model
- ▷ Linear version of the model
- ▷ Clustered-data versions of the model
- ▷ Transition models

18.1 A Log-linear Model

- Cox (1972)
- Joint distribution of \mathbf{Y}_i in terms of a multivariate exponential family:

$$\begin{aligned} f(\mathbf{y}_i, \boldsymbol{\theta}_i) &= \exp \left(\sum_{j=1}^n \theta_{ij} y_{ij} + \sum_{j_1 < j_2} \theta_{ij_1 j_2} y_{ij_1} y_{ij_2} + \dots + \theta_{i1\dots n} y_{i1} \dots y_{in} - A(\boldsymbol{\theta}_i) \right) \\ &= c(\boldsymbol{\theta}_i) \exp \left(\sum_{j=1}^n \theta_{ij} y_{ij} + \sum_{j_1 < j_2} \theta_{ij_1 j_2} y_{ij_1} y_{ij_2} + \dots + \theta_{i1\dots n} y_{i1} \dots y_{in} \right) \end{aligned}$$

- $A(\boldsymbol{\theta}_i)$ [equivalently, $c(\boldsymbol{\theta}_i)$] is the normalizing constant
- $\boldsymbol{\theta}_i$ is the canonical parameter, consisting of first, second, up to n th order components.

- **Interpretation of Parameters:**

- ▷ The parameters have a conditional interpretation:

$$\theta_{ij} = \ln \left(\frac{\Pr(Y_{ij} = 1 | Y_{ik} = 0; k \neq j)}{\Pr(Y_{ij} = 0 | Y_{ik} = 0; k \neq j)} \right)$$

- ▷ ⇒ the first order parameters (main effects) are interpreted as **conditional logits**.

- ▷ Similarly,

$$\theta_{ijk} = \ln \left(\frac{\Pr(Y_{ij} = 1, Y_{ik} = 1 | Y_{il} = 0; k, j \neq l) \Pr(Y_{ij} = 0, Y_{ik} = 0 | Y_{il} = 0; k, j \neq l)}{\Pr(Y_{ij} = 1, Y_{ik} = 0 | Y_{il} = 0; k, j \neq l) \Pr(Y_{ij} = 0, Y_{ik} = 1 | Y_{il} = 0; k, j \neq l)} \right)$$

- ▷ These are **conditional log odds ratios**.

- **Advantages:**

- ▷ The parameter vector is not constrained. All values of $\theta \in I\!\!R$ yield nonnegative probabilities.
- ▷ Calculation of the joint probabilities is fairly straightforward:
 - * ignore the normalizing constant
 - * evaluate the density for all possible sequences y
 - * sum all terms to yield $c(\theta)^{-1}$

- **Drawbacks:**

- ▷ Due to above conditional interpretation, the models are less useful for regression.
The dependence of $E(Y_{ij})$ on covariates involves **all** parameters, not only the main effects.
- ▷ The interpretation of the parameters depends on the length n_i of a sequence.

These drawbacks make marginal models or models that combine marginal and conditional features better suited.

18.2 Quadratic and Linear Versions

- Cox (1972) and others suggest that often the higher order interactions can be neglected. This claim is supported by empirical evidence.
- **The quadratic exponential model:**

$$\begin{aligned} f(\mathbf{y}_i, \boldsymbol{\theta}_i) &= \exp \left(\sum_{j=1}^n \theta_{ij} y_{ij} + \sum_{j_1 < j_2} \theta_{ij_1 j_2} y_{ij_1} y_{ij_2} - A(\boldsymbol{\theta}_i) \right) \\ &= c(\boldsymbol{\theta}_i) \exp \left(\sum_{j=1}^n \theta_{ij} y_{ij} + \sum_{j_1 < j_2} \theta_{ij_1 j_2} y_{ij_1} y_{ij_2} \right). \end{aligned}$$

- **The linear exponential model:**

$$f(\mathbf{y}_i, \boldsymbol{\theta}_i) = \exp \left(\sum_{j=1}^n \theta_{ij} y_{ij} - A(\boldsymbol{\theta}_i) \right)$$

then this model reflects the assumption of independence.

- The linear model equals logistic regression.

18.3 A Version for Clustered Binary Data

- NTP data: Y_{ij} is malformation indicator for fetus j in litter i
- Code Y_{ij} as -1 or 1
- d_i is dose level at which litter i is exposed
- Simplification: $\theta_{ij} = \theta_i = \beta_0 + \beta_d d_i$ and $\theta_{ij_1j_2} = \beta_a$
- Using

$$Z_i = \sum_{j=1}^{n_i} Y_{ij}$$

we obtain

$$f(z_i | \theta_i, \beta_a) = \binom{n_i}{z_i} \exp \{ \theta_i z_i + \beta_a z_i (n_i - z_i) - A(\theta_i) \}$$

18.4 Transition Models

- Molenberghs and Verbeke (2005, Section 11.5)
- Outcome Y_{ij} or error term ε_{ij} is a function of history $\mathbf{h}_{ij} = (Y_{i1}, \dots, Y_{i,j-1})$
- Order of transition model: # of previous outcomes in regression
- Stationary model: functional form of dependence independent of occurrence time
- A stationary first-order autoregressive model for *continuous* data is:

$$Y_{i1} = \mathbf{x}'_{i1}\boldsymbol{\beta} + \varepsilon_{i1}$$

$$Y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \alpha Y_{i,j-1} + \varepsilon_{ij}$$

- Assume

$$\varepsilon_{i1} \sim N(0, \sigma^2) \quad \text{and} \quad \varepsilon_{ij} \sim N(0, \sigma^2(1 - \alpha^2))$$

then

$$\text{cov}(Y_{ij}, Y_{ij'}) = \alpha^{|j' - j|} \sigma^2$$

⇒ a marginal multivariate normal model with AR(1) variance-covariance matrix.

- For non-Gaussian outcomes, first write

$$Y_{ij} = \mu_{ij}^c + \varepsilon_{ij}^c$$

and then

$$\mu_{ij}^c = E(Y_{ij} | \mathbf{h}_{ij})$$

$$\phi v^c(\mu_{ij}^c) = \text{var}(Y_{ij} | \mathbf{h}_{ij})$$

- Example of a linear predictor:

$$\eta_{ij}(\mu_{ij}^c) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \kappa(\mathbf{h}_{ij}, \boldsymbol{\beta}, \boldsymbol{\alpha})$$

- κ is a function of the history.
- This model is easy to fit since it leads to independent GLM contributions:

$$\begin{aligned} f(y_{i1}, \dots, y_{in_i}) &= f(y_{i1}) \cdot f(y_{i2}|y_{i1}) \cdot f(y_{i3}|y_{i1}, y_{i2}) \cdot f(y_{in_i}|y_{i1}, \dots, y_{i,n_i-1}) \\ &= f(y_{i1}) \cdot \prod_{j=2}^{n_i} f(y_{ij}|\mathbf{h}_{ij}) = f(y_{i1}, \dots, y_{iq}) \cdot \prod_{j=q+1}^{n_i} f(y_{ij}|\mathbf{h}_{ij}) \end{aligned}$$

- This product yields $n_i - q$ independent univariate GLM contributions.
- A separate model may need to be considered for the first q measurements.

- A logistic-regression type example:

$$\text{logit}[P(Y_{ij} = 1 | \boldsymbol{x}_{ij}, Y_{i,j-1} = y_{i,j-1}, \boldsymbol{\beta}, \alpha)] = \boldsymbol{x}'_{ij} \boldsymbol{\beta} + \alpha y_{i,j-1}.$$

- The marginal means and variances do not follow easily, except in the normal case.
- Recursive formulas are:

$$\mu_{ij} = \mu_{ij}^c(0)[1 - \mu_{i,j-1}] + \mu_{ij}^c(1)\mu_{i,j-1}$$

$$v_{ij} = [\mu_{ij}^c(1) - \mu_{ij}^c(0)]^2 v_{i,j-1} + v_{ij}^c(0)[1 - \mu_{i,j-1}] + v_{ij}^c(1)\mu_{i,j-1}$$

18.4.1 Analysis of the Toenail Data

- Formulate a transition model (Model I):

$$Y_{ij} \sim \text{Bernoulli}(\mu_{ij})$$

$$\text{logit} \left(\frac{\mu_{ij}}{1 - \mu_{ij}} \right) = \beta_0 + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij} + \alpha_1 y_{i,j-1}$$

- To account for unequal spacing (Model II):

- ▷ α_1 describes the transition effect for the later measurements
- ▷ α_{1a} is the ‘excess’ during the first quarter
- ▷ hence: autoregressive effect at months 1, 2, and 3 is $\alpha_1 + \alpha_{1a}$

- Alternatively: dependence on two prior occasions:

$$\text{logit} \left(\frac{\mu_{ij}}{1 - \mu_{ij}} \right) = \beta_0 + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij} + \alpha_1 y_{i,j-1} + \alpha_2 y_{i,j-2}$$

- Fitted models:

Effect	Par.	First order		Second order
		I	II	
Intercept	β_0	-3.14 (0.27)	-3.77 (0.34)	-3.28 (0.34)
T_i	β_1	0.00 (0.31)	-0.08 (0.32)	0.13 (0.39)
t_{ij}	β_2	-0.09 (0.04)	0.03 (0.05)	-0.05 (0.04)
$T_i \cdot t_{ij}$	β_3	-0.08 (0.06)	-0.06 (0.06)	-0.09 (0.07)
Dep. on $Y_{i,j-1}$	α_1	4.48 (0.22)	3.59 (0.29)	4.01 (0.39)
Dep. on $Y_{i,j-1}$	α_{1a}		1.56 (0.35)	
Dep. on $Y_{i,j-2}$	α_2			0.25 (0.38)

- Two separate models, depending on the level of the previous outcome:

$$\text{logit} \left(\frac{\mu_{ij}}{1 - \mu_{ij}} \right) = (\beta_{00} + \beta_{10}T_i + \beta_{20}t_{ij} + \beta_{30}T_i t_{ij}) I_{Y_{i,j-1}=0} \\ + (\beta_{01} + \beta_{11}T_i + \beta_{21}t_{ij} + \beta_{31}T_i t_{ij}) I_{Y_{i,j-1}=1}$$

- Fitted model:

Effect	$Y_{i,j-1} = 0$		$Y_{i,j-1} = 1$	
	Par.	Estimate (s.e.)	Par.	Estimate (s.e.)
Intercept	β_{00}	-3.92 (0.56)	β_{01}	1.56 (1.26)
T_i	β_{10}	0.45 (0.70)	β_{11}	-0.01 (0.37)
t_{ij}	β_{20}	-0.06 (0.09)	β_{21}	-0.20 (0.06)
$T_i \cdot t_{ij}$	β_{30}	0.07 (0.10)	β_{31}	0.04 (0.07)

18.4.2 Transition Model in SAS

- Prepare models so that the previous outcome can be used as a covariate (using the same code as used to fit a model for dropout – see Part V)

```
%dropout(data=test,id=idnum,time=time,response=onlyresp,out=test2);
```

```
data test2a;  
set test2;  
prev1=prev;  
drop prev;  
run;
```

```
%dropout(data=test2a,id=idnum,time=time,response=prev1,out=test3);
```

```
data test3a;  
set test3;  
prev2=prev;  
drop prev;  
run;
```

- The result for the first subject is

Obs	idnum	time	treatn	onyresp	prev1	prev2
1	1	0	1	1	.	.
2	1	1	1	1	1	.
3	1	2	1	1	1	1
4	1	3	1	0	1	1
5	1	6	1	0	0	1
6	1	9	1	0	0	0
7	1	12	1	0	0	0

- Code to fit a transition model:

```
proc genmod data=test3a descending;
model onyresp = treatn time treatn*time prev1 / dist=binomial;
run;
```

- When both predecessors are used, one merely adds ‘prev2’ to MODEL statement:

```
model onyresp = prev1 treatn*prev1 time*prev1 treatn*time*prev1  
          / noint dist=binomial;
```

- To fit Model II, an additional variable ‘prev1a’ needs to be created:

```
data test3b;  
set test3a;  
prev1a=prev1;  
if time>3 then prev1a=0;  
run;
```

which is then added to the logistic regression, next to ‘prev1.’

Chapter 19

Full Marginal Models

- ▷ Introduction
- ▷ Link functions
- ▷ Associations
- ▷ Bahadur model
- ▷ Multivariate Probit model
- ▷ Example: POPS data

19.1 Introduction

- Choices to make:
 - ▷ Description of mean profiles (univariate parameters) and of association (bivariate and higher order parameters)
 - ▷ Degree of modeling:
 - * joint distribution fully specified \Rightarrow likelihood procedures
 - * only a limited number of moments \Rightarrow e.g., generalized estimating equations
- Minimally, one specifies:
 - ▷ $\eta_i(\mu_i) = \{\eta_{i1}(\mu_{i1}), \dots, \eta_{in}(\mu_{in})\}$
 - ▷ $E(\mathbf{Y}_i) = \mu_i$ and $\eta_i(\mu_i) = \mathbf{X}_i\beta$
 - ▷ $\text{var}(\mathbf{Y}_i) = \phi v(\mu_i)$ where $v(\cdot)$ is a known variance function
 - ▷ $\text{corr}(\mathbf{Y}_i) = R(\alpha)$

19.2 Univariate Link Functions

- The marginal **logit link**:

$$\eta_{ij} = \ln(\mu_{ij}) - \ln(1 - \mu_{ij}) = \text{logit}(\mu_{ij}).$$

- The **probit link**:

$$\eta_{ij} = \Phi_1^{-1}(\mu_{ij}).$$

- The complementary log-log link

- ...

19.3 Pairwise Association

- **Success probability approach.** (Ekholm 1991)

Logit link for two-way probabilities

$$\eta_{ijk} = \ln(\mu_{ijk}) - \ln(1 - \mu_{ijk}) = \text{logit}(\mu_{ijk}),$$

- **Marginal correlation coefficient.** (Bahadur model)

$$\rho_{ijk} = \frac{\mu_{ijk} - \mu_{ij}\mu_{ik}}{\sqrt{\mu_{ij}(1 - \mu_{ij})\mu_{ik}(1 - \mu_{ik})}}$$

$$\eta_{ijk} = \ln(1 + \rho_{ijk}) - \ln(1 - \rho_{ijk}) \quad (\text{Fisher's } z \text{ transform})$$

- Marginal odds ratio. (Dale model)

$$\begin{aligned}\psi_{ijk} &= \frac{(\mu_{ijk})(1 - \mu_{ij} - \mu_{ik} + \mu_{ijk})}{(\mu_{ik} - \mu_{ijk})(\mu_{ij} - \mu_{ijk})} \\ &= \left(\frac{\Pr(Y_{ij} = 1, Y_{ik} = 1)\Pr(Y_{ij} = 0, Y_{ik} = 0)}{\Pr(Y_{ij} = 0, Y_{ik} = 1)\Pr(Y_{ij} = 1, Y_{ik} = 0)} \right)\end{aligned}$$

$$\eta_{ijk} = \ln(\psi_{ijk}) \quad (\text{log odds ratio})$$

- Higher order association defined similarly
- Calculations can become cumbersome

19.4 The Bahadur Model

- Univariate: $E(Y_{ij}) = P(Y_{ij} = 1) \equiv \pi_{ij}.$
- Bivariate: $E(Y_{ij}Y_{ik}) = P(Y_{ij} = 1, Y_{ik} = 1) \equiv \pi_{ijk}.$

- Correlation structure:

$$\text{Corr}(Y_{ij}, Y_{ik}) \equiv \rho_{ijk} = \frac{\pi_{ijk} - \pi_{ij}\pi_{ik}}{[\pi_{ij}(1 - \pi_{ij})\pi_{ik}(1 - \pi_{ik})]^{1/2}}.$$

- This yields expression for pairwise probabilities:

$$\pi_{ijk} = \pi_{ij}\pi_{ik} + \rho_{ijk}[\pi_{ij}(1 - \pi_{ij})\pi_{ik}(1 - \pi_{ik})]^{1/2}.$$

- Similarly for the full joint distribution $f(\mathbf{y}).$

- Let

$$\varepsilon_{ij} = \frac{Y_{ij} - \pi_{ij}}{\sqrt{\pi_{ij}(1 - \pi_{ij})}} \quad \text{and} \quad e_{ij} = \frac{y_{ij} - \pi_{ij}}{\sqrt{\pi_{ij}(1 - \pi_{ij})}},$$

and

$$\rho_{ijk} = E(\varepsilon_{ij}\varepsilon_{ik}),$$

$$\rho_{ijkl} = E(\varepsilon_{ij}\varepsilon_{ik}\varepsilon_{il}),$$

⋮

$$\rho_{i12\dots n_i} = E(\varepsilon_{i1}\varepsilon_{i2} \dots \varepsilon_{in_i}).$$

- A general expression:

$$f(\mathbf{y}_i) = f_1(\mathbf{y}_i)c(\mathbf{y}_i),$$

with

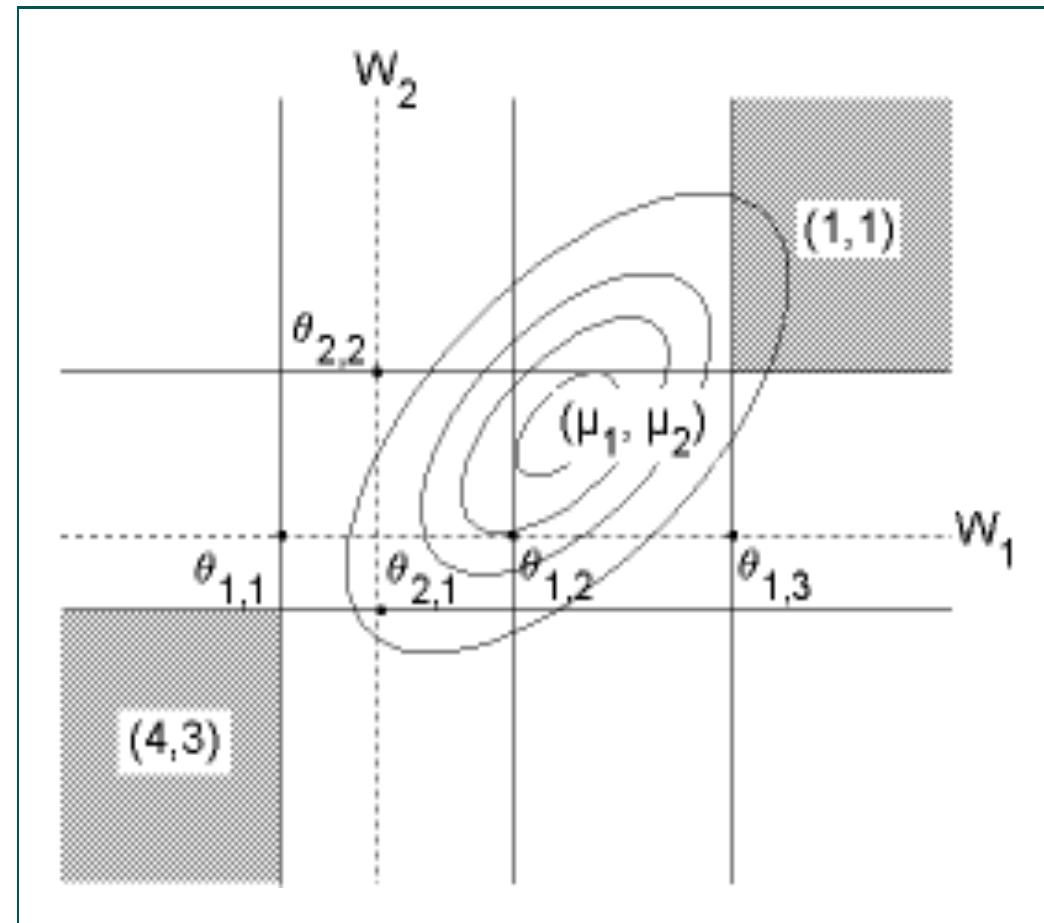
$$f_1(\mathbf{y}_i) = \prod_{j=1}^{n_i} \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{1-y_{ij}}$$

and

$$c(\mathbf{y}_i) = 1 + \sum_{j < k} \rho_{ijk} e_{ij} e_{ik} + \sum_{j < k < \ell} \rho_{ijkl} e_{ij} e_{ik} e_{il} + \dots + \rho_{i1\dots n_i} e_{i1} e_{i2} \dots e_{in_i}.$$

19.5 The Multivariate Probit Model

- E.g., 4×3 categorical outcome arises from underlying bivariate normal
- Covariate effects \equiv shift of cut off points
- Correlation = polychoric correlation: allowed to depend on covariates



19.6 The POPS Data

- Project On Preterm and Small for Gestational Age Infants
- 1530 Dutch children (1983)
- Collected data:

Perinatal information:

- ▷ Bilirubin value
- ▷ Neonatal seizures
- ▷ Congenital malformations

Ability scores at the age of 2:

- ▷ Are the child's movements natural ?
- ▷ Can the child pile three bricks ?
- ▷ Can the child put a ball in a boxed when asked to ?

19.7 Application to POPS Data

	Bahad	Probit	Dale-Norm	Dale-Logist
First Ability Score				
Intercept	3.67(0.49)	2.01(0.26)	2.03(0.27)	3.68(0.52)
Neonatal seiz.	-1.94(0.42)	-1.12(0.26)	-1.16(0.26)	-2.06(0.44)
Congenital malf.	-1.21(0.31)	-0.61(0.18)	-0.62(0.18)	-1.17(0.33)
100× Bilirubin	-0.69(0.25)	-0.32(0.14)	-0.32(0.14)	-0.64(0.27)
Second Ability Score				
Intercept	4.03(0.51)	2.19(0.27)	2.21(0.27)	4.01(0.54)
Neonatal seiz.	-2.26(0.43)	-1.27(0.26)	-1.29(0.26)	-2.28(0.44)
Congenital malf.	-1.08(0.32)	-0.56(0.19)	-0.59(0.19)	-1.11(0.34)
100× Bilirubin	-0.85(0.26)	-0.42(0.14)	-0.41(0.14)	-0.80(0.27)
Third Ability Score				
Intercept	3.32(0.50)	1.84(0.27)	1.91(0.27)	3.49(0.54)
Neonatal seiz.	-1.55(0.44)	-0.88(0.27)	-0.93(0.27)	-1.70(0.46)
Congenital malf.	-0.96(0.32)	-0.47(0.19)	-0.49(0.19)	-0.96(0.35)
100× Bilirubin	-0.44(0.26)	-0.21(0.14)	-0.24(0.14)	-0.49(0.28)

	Bahad	Probit	Dale-Norm	Dale-Logist
	Association parameters			
	ρ	ρ	ψ	ψ
(1,2): ρ or ψ	0.27(0.05)	0.73(0.05)	17.37(5.19)	17.35(5.19)
(1,2): $z(\rho)$ or $\ln \psi$	0.55(0.11)	1.85(0.23)	2.85(0.30)	2.85(0.30)
(1,3): ρ or ψ	0.39(0.05)	0.81(0.04)	30.64(9.78)	30.61(9.78)
(1,3): $z(\rho)$ or $\ln \psi$	0.83(0.12)	2.27(0.25)	3.42(0.32)	3.42(0.32)
(2,3): ρ or ψ	0.23(0.05)	0.72(0.05)	17.70(5.47)	17.65(5.47)
(2,3): $z(\rho)$ or $\ln \psi$	0.47(0.10)	1.83(0.23)	2.87(0.31)	2.87(0.31)
(1,2,3): ρ or ψ	—	—	0.91(0.69)	0.92(0.69)
(1,2,3): $z(\rho)$ or $\ln \psi$	—	—	-0.09(0.76)	-0.09(0.76)
Log-likelihood	-598.44	-570.69	-567.11	-567.09

Chapter 20

Generalized Estimating Equations

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

20.1 General Idea

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

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

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

$$S(\boldsymbol{\beta}) = \sum_i \sum_j \frac{\partial \mu_{ij}}{\partial \boldsymbol{\beta}} v_{ij}^{-1} (y_{ij} - \mu_{ij}) = \sum_{i=1}^N D'_i [V_i(\boldsymbol{\alpha})]^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}$$

where

- ▷ D_i is an $n_i \times p$ matrix with (i, j) th elements $\frac{\partial \mu_{ij}}{\partial \boldsymbol{\beta}}$
- ▷ \mathbf{y}_i and $\boldsymbol{\mu}_i$ are n_i -vectors with elements y_{ij} and μ_{ij}
- ▷ Is V_i $n_i \times n_i$ diagonal or more complex?

- $V_i = \text{Var}(\mathbf{Y}_i)$ is more complex since it involves a set of nuisance parameters α , determining the covariance structure of \mathbf{Y}_i :

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

in which

$$A_i^{1/2}(\boldsymbol{\beta}) = \begin{pmatrix} \sqrt{v_{i1}(\mu_{i1}(\boldsymbol{\beta}))} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sqrt{v_{in_i}(\mu_{in_i}(\boldsymbol{\beta}))} \end{pmatrix}$$

and $R_i(\boldsymbol{\alpha})$ is the correlation matrix of \mathbf{Y}_i , parameterized by α .

- Same form as for full likelihood procedure, **but** we restrict specification to the first moment only
- Liang and Zeger (1986)

20.2 Large Sample Properties

As $N \rightarrow \infty$

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

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 $V_i(\boldsymbol{\alpha})$ is known
- This is the naive=purely model based variance estimator
- Solution: working correlation matrix

20.3 Unknown Covariance Structure

Keep the score equations

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

BUT

- suppose $V_i(\cdot)$ is not the true variance of \mathbf{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

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

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

$$I_1 = \sum_{i=1}^N D_i' [V_i(\alpha)]^{-1} \text{Var}(Y_i) [V_i(\alpha)]^{-1} D_i.$$

- This is the **robust** \equiv **empirically corrected** \equiv **sandwich** variance estimator
 - $\triangleright I_0$ is the bread
 - $\triangleright I_1$ is the filling (ham or cheese)
- Correct guess \implies likelihood variance

- The estimators $\hat{\beta}$ are consistent even if the working correlation matrix is incorrect
- An estimate is found by replacing the unknown variance matrix $\text{Var}(\mathbf{Y}_i)$ by

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

- Even if this estimator is bad for $\text{Var}(\mathbf{Y}_i)$ it leads to a good estimate of I_1 , provided that:
 - ▷ replication in the data is sufficiently large
 - ▷ same model for μ_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 $\hat{\beta}$
- Care needed with incomplete data (see Part V)

20.4 The Working Correlation Matrix

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

- Variance function: A_i is $(n_i \times n_i)$ diagonal with elements $v(\mu_{ij})$, the known GLM variance function.
- Working correlation: $R_i(\alpha)$ possibly depends on a different set of parameters α .
- 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 β .

20.5 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	$\hat{\alpha} = \frac{1}{N} \sum_{i=1}^N \frac{1}{n_i(n_i-1)} \sum_{j \neq k} e_{ij}e_{ik}$
AR(1)	α	$\hat{\alpha} = \frac{1}{N} \sum_{i=1}^N \frac{1}{n_i-1} \sum_{j \leq n_i-1} e_{ij}e_{i,j+1}$
Unstructured	α_{jk}	$\hat{\alpha}_{jk} = \frac{1}{N} \sum_{i=1}^N e_{ij}e_{ik}$

Dispersion parameter:

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

20.6 Fitting GEE

The standard procedure, implemented in the SAS procedure GENMOD.

1. Compute initial estimates for β , using a univariate GLM (i.e., assuming independence).
2.
 - ▷ Compute Pearson residuals e_{ij} .
 - ▷ Compute estimates for α and ϕ .
 - ▷ Compute $R_i(\alpha)$ and $V_i(\beta, \alpha) = \phi A_i^{1/2}(\beta) R_i(\alpha) A_i^{1/2}(\beta)$.
3. Update estimate for β :

$$\beta^{(t+1)} = \beta^{(t)} - \left[\sum_{i=1}^N D_i' V_i^{-1} D_i \right]^{-1} \left[\sum_{i=1}^N D_i' V_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) \right].$$

4. Iterate 2.–3. until convergence.

Estimates of precision by means of I_0^{-1} and/or $I_0^{-1} I_1 I_0^{-1}$.

20.7 Special Case: Linear Mixed Models

- Estimate for β :

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

with α replaced by its ML or REML estimate

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

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

provided that $E(Y_i) = X_i \beta$

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

- Conditional on α , $\hat{\beta}$ has covariance

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

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

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

Chapter 21

A Family of GEE Methods

- ▷ Classical approach
- ▷ Prentice's two sets of GEE
- ▷ Linearization-based version
- ▷ GEE2
- ▷ Alternating logistic regressions

21.1 Prentice's GEE

$$\sum_{i=1}^N D_i' V_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) = \mathbf{0}, \quad \sum_{i=1}^N E_i' W_i^{-1} (\mathbf{Z}_i - \boldsymbol{\delta}_i) = \mathbf{0}$$

where

$$Z_{ijk} = \frac{(Y_{ij} - \mu_{ij})(Y_{ik} - \mu_{ik})}{\sqrt{\mu_{ij}(1 - \mu_{ij})\mu_{ik}(1 - \mu_{ik})}}, \quad \delta_{ijk} = E(Z_{ijk})$$

The joint asymptotic distribution of $\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ and $\sqrt{N}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$ normal with variance-covariance matrix consistently estimated by

$$N \begin{pmatrix} A & 0 \\ B & C \end{pmatrix} \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix} \begin{pmatrix} A & B' \\ 0 & C \end{pmatrix}$$

where

$$\begin{aligned}
 A &= \left(\sum_{i=1}^N D_i' V_i^{-1} D_i \right)^{-1}, & \Lambda_{11} &= \sum_{i=1}^N D_i' V_i^{-1} \text{Cov}(\mathbf{Y}_i) V_i^{-1} D_i, \\
 B &= \left(\sum_{i=1}^N E_i' W_i^{-1} E_i \right)^{-1} \left(\sum_{i=1}^N E_i' W_i^{-1} \frac{\partial \mathbf{Z}_i}{\partial \boldsymbol{\beta}} \right) \left(\sum_{i=1}^N D_i' V_i^{-1} D_i \right)^{-1}, & \Lambda_{12} &= \sum_{i=1}^N D_i' V_i^{-1} \text{Cov}(\mathbf{Y}_i, \mathbf{Z}_i) W_i^{-1} E_i, \\
 C &= \left(\sum_{i=1}^N E_i' W_i^{-1} E_i \right)^{-1}, & \Lambda_{21} &= \Lambda_{12}, \\
 && \Lambda_{22} &= \sum_{i=1}^N E_i' W_i^{-1} \text{Cov}(\mathbf{Z}_i) W_i^{-1} E_i,
 \end{aligned}$$

and

Statistic	Estimator
$\text{Var}(\mathbf{Y}_i)$	$(\mathbf{Y}_i - \boldsymbol{\mu}_i)(\mathbf{Y}_i - \boldsymbol{\mu}_i)'$
$\text{Cov}(\mathbf{Y}_i, \mathbf{Z}_i)$	$(\mathbf{Y}_i - \boldsymbol{\mu}_i)(\mathbf{Z}_i - \boldsymbol{\delta}_i)'$
$\text{Var}(\mathbf{Z}_i)$	$(\mathbf{Z}_i - \boldsymbol{\delta}_i)(\mathbf{Z}_i - \boldsymbol{\delta}_i)'$

21.2 GEE Based on Linearization

21.2.1 Model formulation

- Previous version of GEE are formulated directly in terms of binary outcomes
- This approach is based on a **linearization**:

$$\mathbf{y}_i = \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_i$$

with

$$\boldsymbol{\eta}_i = g(\boldsymbol{\mu}_i), \quad \boldsymbol{\eta}_i = \mathbf{X}_i \boldsymbol{\beta}, \quad \text{Var}(\mathbf{y}_i) = \text{Var}(\boldsymbol{\varepsilon}_i) = \Sigma_i.$$

- $\boldsymbol{\eta}_i$ is a vector of linear predictors,
- $g(\cdot)$ is the (vector) link function.

21.2.2 Estimation (Nelder and Wedderburn 1972)

- Solve iteratively:

$$\sum_{i=1}^N X_i' W_i X_i \boldsymbol{\beta} = \sum_{i=1}^N W_i \mathbf{y}_i^*,$$

where

$$W_i = F_i' \Sigma_i^{-1} F_i, \quad \mathbf{y}_i^* = \hat{\boldsymbol{\eta}}_i + (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i) F_i^{-1},$$

$$F_i = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_i}, \quad \Sigma_i = \text{Var}(\boldsymbol{\varepsilon}), \quad \boldsymbol{\mu}_i = E(\mathbf{y}_i).$$

- Remarks:

- ▷ \mathbf{y}_i^* is called ‘working variable’ or ‘pseudo data’.
- ▷ Basis for SAS macro and procedure GLIMMIX
- ▷ For linear models, $D_i = I_{n_i}$ and standard linear regression follows.

21.2.3 The Variance Structure

$$\Sigma_i = \phi A_i^{1/2}(\beta) R_i(\alpha) A_i^{1/2}(\beta)$$

- ϕ is a scale (overdispersion) parameter,
- $A_i = v(\mu_i)$, expressing the mean-variance relation (this is a function of β),
- $R_i(\alpha)$ describes the correlation structure:
 - ▷ If independence is assumed then $R_i(\alpha) = I_{n_i}$.
 - ▷ Other structures, such as compound symmetry, AR(1),... can be assumed as well.

21.3 GEE2

- **Model:**

- ▷ Marginal mean structure
- ▷ Pairwise association:
 - * Odds ratios
 - * Correlations

- **Working assumptions:** Third and fourth moments

- **Estimation:**

- ▷ Second-order estimating equations
- ▷ Likelihood (assuming 3rd and 4th moments are correctly specified)

21.4 Alternating Logistic Regression

- Diggle, Heagerty, Liang, and Zeger (2002) and Molenberghs and Verbeke (2005)
- When marginal odds ratios are used to model association, α can be estimated using ALR, which is
 - ▷ almost as efficient as GEE2
 - ▷ almost as easy (computationally) than GEE1
- μ_{ijk} as before and $\alpha_{ijk} = \ln(\psi_{ijk})$ the marginal log odds ratio:

$$\text{logit } \Pr(Y_{ij} = 1 | \mathbf{x}_{ij}) = \mathbf{x}_{ij}\boldsymbol{\beta}$$

$$\text{logit } \Pr(Y_{ij} = 1 | Y_{ik} = y_{ik}) = \alpha_{ijk}y_{ik} + \ln \left(\frac{\mu_{ij} - \mu_{ijk}}{1 - \mu_{ij} - \mu_{ik} + \mu_{ijk}} \right)$$

- α_{ijk} can be modelled in terms of predictors
- the second term is treated as an offset
- the estimating equations for β and α are solved in turn, and the ‘alternating’ between both sets is repeated until convergence.
- this is needed because the offset clearly depends on β .

21.5 Application to the Toenail Data

21.5.1 The model

- Consider the model:

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

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

21.5.2 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 parameters:

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	

- ▷ Estimates from fitting the model, ignoring the correlation structure, i.e., from fitting a classical GLM to the data, using proc GENMOD.
 - ▷ The reported log-likelihood also corresponds to this model, and therefore should not be interpreted.
 - ▷ The reported estimates are used as starting values in the iterative estimation procedure for fitting the GEE's.

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence		Z	Pr > Z
			Limits			
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		Z	Pr > Z
			Limits			
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

21.5.3 Alternating Logistic Regression

- ‘type=exch’ —→ ‘logor=exch’
- Note that α now is a genuine parameter

- Selected output:

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence		Z	Pr > Z
			Limits			
Intercept	-0.5244	0.1686	-0.8548	-0.1940	-3.11	0.0019
treatn	0.0168	0.2432	-0.4599	0.4935	0.07	0.9448
time	-0.1781	0.0296	-0.2361	-0.1200	-6.01	<.0001
treatn*time	-0.0837	0.0520	-0.1856	0.0182	-1.61	0.1076
Alpha1	3.2218	0.2908	2.6519	3.7917	11.08	<.0001

Analysis Of GEE Parameter Estimates
Model-Based Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence		Z	Pr > Z
			Limits			
Intercept	-0.5244	0.1567	-0.8315	-0.2173	-3.35	0.0008
treatn	0.0168	0.2220	-0.4182	0.4519	0.08	0.9395
time	-0.1781	0.0233	-0.2238	-0.1323	-7.63	<.0001
treatn*time	-0.0837	0.0392	-0.1606	-0.0068	-2.13	0.0329

21.5.4 Linearization Based Method

- **GLIMMIX macro:**

```
%glimmix(data=test, procopt=%str(method=ml empirical),
stmts=%str(
    class idnum timeclss;
    model onyresp = treatn time treatn*time / solution;
    repeated timeclss / subject=idnum type=cs rcorr;
),
error=binomial,
link=logit);
```

- **GLIMMIX procedure:**

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

- Both produce the same results
- The GLIMMIX macro is a MIXED core, with GLM-type surrounding statements
- The GLIMMIX procedure does not call MIXED, it has its own engine
- PROC GLIMMIX combines elements of MIXED and of GENMOD
- RANDOM _residual_ is the PROC GLIMMIX way to specify residual correlation

21.5.5 Results of Models Fitted to Toenail Data

Effect	Par.	IND	EXCH	UN
GEE1				
Int.	β_0	-0.557(0.109;0.171)	-0.584(0.134;0.173)	-0.720(0.166;0.173)
T_i	β_1	0.024(0.157;0.251)	0.012(0.187;0.261)	0.072(0.235;0.246)
t_{ij}	β_2	-0.177(0.025;0.030)	-0.177(0.021;0.031)	-0.141(0.028;0.029)
$T_i \cdot t_{ij}$	β_3	-0.078(0.039;0.055)	-0.089(0.036;0.057)	-0.114(0.047;0.052)
ALR				
Int.	β_0		-0.524(0.157;0.169)	
T_i	β_1		0.017(0.222;0.243)	
t_{ij}	β_2		-0.178(0.023;0.030)	
$T_i \cdot t_{ij}$	β_3		-0.084(0.039;0.052)	
Ass.	α		3.222(;0.291)	
Linearization based method				
Int.	β_0	-0.557(0.112;0.171)	-0.585(0.142;0.174)	-0.630(0.171;0.172)
T_i	β_1	0.024(0.160;0.251)	0.011(0.196;0.262)	0.036(0.242;0.242)
t_{ij}	β_2	-0.177(0.025;0.030)	-0.177(0.022;0.031)	-0.204(0.038;0.034)
$T_i \cdot t_{ij}$	β_3	-0.078(0.040;0.055)	-0.089(0.038;0.057)	-0.106(0.058;0.058)

estimate (model-based s.e.; empirical s.e.)

21.5.6 Discussion

- GEE1: All empirical standard errors are correct, but the efficiency is higher for the more complex working correlation structure, as seen in p -values for $T_i \cdot t_{ij}$ effect:

Structure	p -value
IND	0.1515
EXCH	0.1208
UN	0.0275

Thus, opting for reasonably adequate correlation assumptions still pays off, in spite of the fact that all are consistent and asymptotically normal

- Similar conclusions for linearization-based method

- Model-based s.e. and empirically corrected s.e. in reasonable agreement for UN
- Typically, the model-based standard errors are much too small as they are based on the assumption that all observations in the data set are independent, hereby overestimating the amount of available information, hence also overestimating the precision of the estimates.
- ALR: similar inferences but now also α part of the inferences

Part III

Generalized Linear Mixed Models for Non-Gaussian Longitudinal Data

Chapter 22

The Beta-binomial Model

- ▷ Genesis of the model
- ▷ Implied marginal distribution

22.1 Genesis of the Beta-binomial Model

- Skellam (1948), Kleinman (1973)
- Let \mathbf{Y}_i be a n_i -dimensional vector of Bernoulli-distributed outcomes, with success probability b_i .
- Assume the elements in \mathbf{Y}_i to be independent, conditionally on b_i
- Then, the conditional density of \mathbf{Y}_i , given b_i is proportional to the density of

$$Z_i = \sum_{j=1}^{n_i} Y_{ij}$$

- The density of Z_i , given b_i is binomial with n_i trials and success probability b_i .

- The beta-binomial model assumes the b_i to come from a beta distribution with parameters α and β :

$$f(b_i|\alpha, \beta) = \frac{b_i^{\alpha-1}(1-b_i)^{\beta-1}}{B(\alpha, \beta)}$$

$B(., .)$: the beta function

- α and β can depend on covariates, but this dependence is temporarily dropped from notation

22.2 Implied Marginal Model

- The marginal density of Z_i is the so-called **beta-binomial** density:

$$\begin{aligned} f_i(z_i|\alpha, \beta) &= \int \binom{n_i}{z_i} b_i^{z_i} (1 - b_i)^{n_i - z_i} f(b_i|\alpha, \beta) db_i \\ &= \binom{n_i}{z_i} \frac{B(z_i + \alpha, n_i - z_i + \beta)}{B(\alpha, \beta)} \end{aligned}$$

- Useful moments and relationships ($\pi = \mu_i/n_i$):

	$\alpha = \pi(\rho^{-1} - 1)$ $\beta = (1 - \pi)(\rho^{-1} - 1)$
Mean	$\mu_i = \mathbb{E}(Z_i) = n_i \frac{\alpha}{\alpha + \beta}$
Correlation	$\rho = \text{Corr}(Y_{ij}, Y_{ik}) = \frac{1}{\alpha + \beta + 1}$
Variance	$\text{Var}(Z_i) = n_i \pi(1 - \pi)[1 + (n_i - 1)\rho]$

- The density can now be written as:

$$f_i(z_i|\pi, \rho) = \binom{n_i}{z_i} \frac{B[z_i + \pi(\rho^{-1} - 1), n_i - z_i + (1 - \pi)(\rho^{-1} - 1)]}{B[\pi(\rho^{-1} - 1), (1 - \pi)(\rho^{-1} - 1)]}$$

- When there are covariates (e.g., sub-populations, dose groups), rewrite π and/or ρ as π_i and/or ρ_i , respectively.
- It is then easy to formulate a model through the marginal parameters π_i and ρ_i :
 - ▷ π_i can be modeled through, e.g., a logit link
 - ▷ ρ_i can be modeled through, e.g., Fisher's z transformation
- In Part IV, the NTP data will be analyzed using the beta-binomial model

Chapter 23

Generalized Linear Mixed Models (GLMM)

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

23.1 Introduction: LMM Revisited

- We re-consider the linear mixed model:

$$\mathbf{Y}_i | \mathbf{b}_i \sim N(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i, \Sigma_i), \quad \mathbf{b}_i \sim N(\mathbf{0}, D)$$

- The implied marginal model equals $\mathbf{Y}_i \sim N(\mathbf{X}_i \boldsymbol{\beta}, \mathbf{Z}_i D \mathbf{Z}'_i + \Sigma_i)$
- 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.

23.2 Generalized Linear Mixed Models (GLMM)

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

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

- θ_{ij} is now modelled as $\theta_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\mathbf{b}_i$
- As before, it is assumed that $\mathbf{b}_i \sim N(\mathbf{0}, D)$
- Let $f_{ij}(y_{ij}|\mathbf{b}_i, \boldsymbol{\beta}, \phi)$ denote the conditional density of Y_{ij} given \mathbf{b}_i , the conditional density of \mathbf{Y}_i equals

$$f_i(\mathbf{y}_i|\mathbf{b}_i, \boldsymbol{\beta}, \phi) = \prod_{j=1}^{n_i} f_{ij}(y_{ij}|\mathbf{b}_i, \boldsymbol{\beta}, \phi)$$

- The marginal distribution of \mathbf{Y}_i is given by

$$\begin{aligned} f_i(\mathbf{y}_i | \boldsymbol{\beta}, D, \phi) &= \int f_i(\mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\beta}, \phi) f(\mathbf{b}_i | D) d\mathbf{b}_i \\ &= \int \prod_{j=1}^{n_i} f_{ij}(y_{ij} | \mathbf{b}_i, \boldsymbol{\beta}, \phi) f(\mathbf{b}_i | D) d\mathbf{b}_i \end{aligned}$$

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

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

$$\begin{aligned} L(\boldsymbol{\beta}, D, \phi) &= \prod_{i=1}^N f_i(\mathbf{y}_i | \boldsymbol{\beta}, D, \phi) \\ &= \prod_{i=1}^N \int \prod_{j=1}^{n_i} f_{ij}(y_{ij} | \mathbf{b}_i, \boldsymbol{\beta}, \phi) f(\mathbf{b}_i | D) d\mathbf{b}_i \end{aligned}$$

- 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

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

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

23.3 Laplace Approximation of Integrand

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

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

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

$$I \approx (2\pi)^{q/2} |-Q''(\hat{\mathbf{b}})|^{-1/2} e^{Q(\hat{\mathbf{b}})}.$$

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

23.4 Approximation of Data

23.4.1 General Idea

- Re-write GLMM as:

$$Y_{ij} = \mu_{ij} + \varepsilon_{ij} = h(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i) + \varepsilon_{ij}$$

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

- Linear Taylor expansion for μ_{ij} :

- ▷ Penalized quasi-likelihood (PQL): Around current $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{b}}_i$
- ▷ Marginal quasi-likelihood (MQL): Around current $\widehat{\boldsymbol{\beta}}$ and $\mathbf{b}_i = \mathbf{0}$

23.4.2 Penalized quasi-likelihood (PQL)

- Linear Taylor expansion around current $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{b}_i}$:

$$\begin{aligned} Y_{ij} &\approx h(\mathbf{x}'_{ij}\widehat{\boldsymbol{\beta}} + \mathbf{z}'_{ij}\widehat{\mathbf{b}_i}) + h'(\mathbf{x}'_{ij}\widehat{\boldsymbol{\beta}} + \mathbf{z}'_{ij}\widehat{\mathbf{b}_i})\mathbf{x}'_{ij}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + h'(\mathbf{x}'_{ij}\widehat{\boldsymbol{\beta}} + \mathbf{z}'_{ij}\widehat{\mathbf{b}_i})\mathbf{z}'_{ij}(\mathbf{b}_i - \widehat{\mathbf{b}_i}) + \varepsilon_{ij} \\ &\approx \widehat{\mu}_{ij} + v(\widehat{\mu}_{ij})\mathbf{x}'_{ij}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + v(\widehat{\mu}_{ij})\mathbf{z}'_{ij}(\mathbf{b}_i - \widehat{\mathbf{b}_i}) + \varepsilon_{ij} \end{aligned}$$

- In vector notation: $\mathbf{Y}_i \approx \widehat{\boldsymbol{\mu}}_i + \widehat{V}_i X_i (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + \widehat{V}_i Z_i (\mathbf{b}_i - \widehat{\mathbf{b}}_i) + \boldsymbol{\varepsilon}_i$
- Re-ordering terms yields:

$$\mathbf{Y}_i^* \equiv \widehat{V}_i^{-1}(\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}_i) + X_i \widehat{\boldsymbol{\beta}} + Z_i \widehat{\mathbf{b}}_i \approx X_i \boldsymbol{\beta} + Z_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i^*,$$

- Model fitting by iterating between updating the **pseudo responses** \mathbf{Y}_i^* and fitting the above linear mixed model to them.

23.4.3 Marginal quasi-likelihood (MQL)

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

$$\begin{aligned} Y_{ij} &\approx h(\mathbf{x}'_{ij}\widehat{\boldsymbol{\beta}}) + h'(\mathbf{x}'_{ij}\widehat{\boldsymbol{\beta}})\mathbf{x}'_{ij}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + h'(\mathbf{x}'_{ij}\widehat{\boldsymbol{\beta}})\mathbf{z}'_{ij}\mathbf{b}_i + \varepsilon_{ij} \\ &\approx \widehat{\mu}_{ij} + v(\widehat{\mu}_{ij})\mathbf{x}'_{ij}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + v(\widehat{\mu}_{ij})\mathbf{z}'_{ij}\mathbf{b}_i + \varepsilon_{ij} \end{aligned}$$

- In vector notation: $\mathbf{Y}_i \approx \widehat{\boldsymbol{\mu}}_i + \widehat{V}_i X_i (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + \widehat{V}_i Z_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i$
- Re-ordering terms yields:

$$\mathbf{Y}_i^* \equiv \widehat{V}_i^{-1}(\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}_i) + X_i \widehat{\boldsymbol{\beta}} \approx X_i \boldsymbol{\beta} + Z_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i^*$$

- Model fitting by iterating between updating the **pseudo responses** \mathbf{Y}_i^* and fitting the above linear mixed model to them.

23.4.4 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

23.5 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

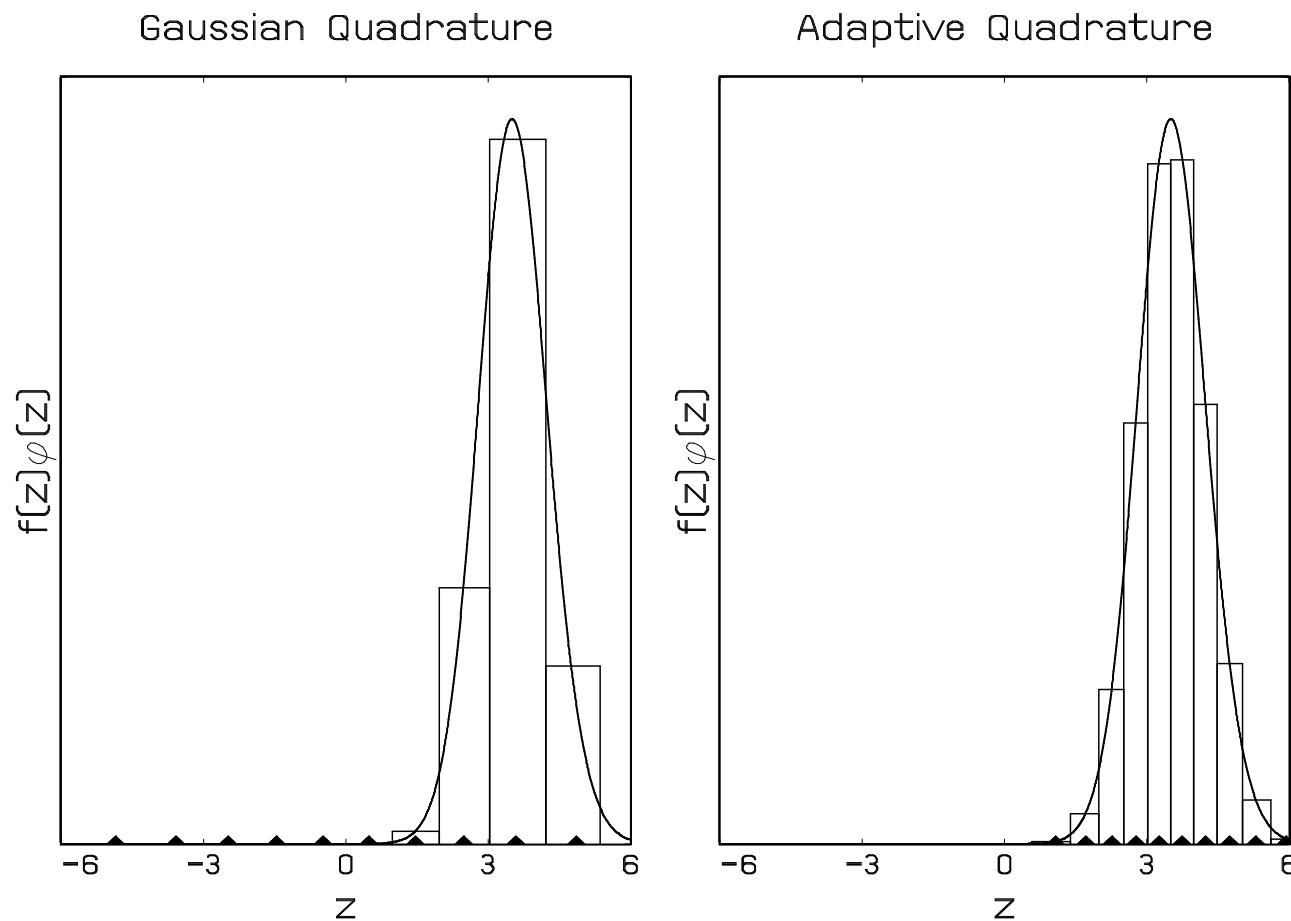
- Gaussian quadrature methods replace the integral by a weighted sum:

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

- 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 Q th 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.
- Ample detail can be found in Molenberghs and Verbeke (2005, Sections 14.3–14.5)

23.6 Example: Toenail Data

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

- Model:

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

- Notation:

- ▷ T_i : treatment indicator for subject i
- ▷ t_{ij} : time point at which j th measurement is taken for i th 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 24

Fitting GLMM's in SAS

- ▷ Proc GLIMMIX for PQL and MQL
- ▷ Proc NL MIXED for Gaussian quadrature

24.1 Procedure GLIMMIX for PQL and MQL

- 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’
- Inclusion of random slopes:

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

- Selected SAS output (PQL):

Covariance Parameter Estimates				
Cov Parm	Subject	Estimate	Standard Error	
Intercept	idnum	4.7095	0.6024	

Solutions for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	-0.7204	0.2370	292	-3.04	0.0026
treatn	-0.02594	0.3360	1612	-0.08	0.9385
time	-0.2782	0.03222	1612	-8.64	<.0001
treatn*time	-0.09583	0.05105	1612	-1.88	0.0607

24.2 Procedure NLMIXED for Gaussian Quadrature

- Re-consider logistic model with random intercepts for toenail data
- SAS program (non-adaptive, $Q = 3$):

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

- Adaptive Gaussian quadrature obtained by omitting option ‘noad’

- Automatic search for ‘optimal’ value of Q in case of no option ‘qpoints=’
- Selected SAS output (non-adaptive, $Q = 3$):

Parameter Estimates										
	Standard									
Parameter	Estimate	Error	DF	t Value	Pr > t	Alpha	Lower	Upper	Gradient	
beta0	-1.5311	0.2961	293	-5.17	<.0001	0.05	-2.1139	-0.9483	2.879E-7	
beta1	-0.4294	0.3728	293	-1.15	0.2503	0.05	-1.1631	0.3043	-2.11E-6	
beta2	-0.3107	0.03373	293	-9.21	<.0001	0.05	-0.3771	-0.2443	-0.00003	
beta3	-0.07539	0.04998	293	-1.51	0.1325	0.05	-0.1738	0.02298	-0.00003	
sigma	2.2681	0.1220	293	18.58	<.0001	0.05	2.0279	2.5083	-3.6E-6	

- Good starting values needed !

- The inclusion of random slopes can be specified as follows:

```
proc nlmixed data=test noad qpoints=3;
parms beta0=-1.6 beta1=0 beta2=-0.4 beta3=-0.5
      d11=3.9 d12=0 d22=0.1;
teta = beta0 + b1 + beta1*treatn + beta2*time
      + b2*time + beta3*timetr;
expteta = exp(teta);
p = expteta/(1+expteta);
model onyresp ~ binary(p);
random b1 b2 ~ normal([0, 0] , [d11, d12, d22])
      subject=idnum;
run;
```

24.2.1 Some Comments on the NLMIXED Procedure

- Different optimization algorithms are available to carry out the maximization of the likelihood.
- Constraints on parameters are also allowed in the optimization process.
- The conditional distribution (given the random effects) can be specified as Normal, Binomial, Poisson, or as any distribution for which you can specify the likelihood by programming statements.
- E-B estimates of the random effects can be obtained.
- Only one RANDOM statement can be specified.
- Only normal random effects are allowed.

- Does not calculate automatic initial values.
- Make sure your data set is sorted by cluster ID!
- PROC NL MIXED can perform Gaussian quadrature by using the options NOAD and NOADSCALE. The number of quadrature points can be specified with the option QPOINTS= m .
- PROC NL MIXED can maximize the marginal likelihood using the Newton-Raphson algorithm by specifying the option TECHNIQUE=NEWRAP.

24.2.2 The Main Statements

- NL MIXED statement:
 - ▷ option ‘noad’ to request no adaptive quadrature
 - ▷ by default, adaptive Gaussian quadrature is used
 - ▷ the option ‘qpoints’ specifies the number of quadrature points
 - ▷ by default, the number of quadrature points is selected adaptively by evaluating the log-likelihood function at the starting values of the parameters until two successive evaluations show sufficiently small relative change.
- PARMS statement:
 - ▷ starting values for all parameters in the model
 - ▷ by default, parameters not listed in the PARMS statement are given an initial value of 1

- MODEL statement:
 - ▷ conditional distribution of the data, given the random effects
 - ▷ valid distributions:
 - * $\text{normal}(m,v)$: Normal with mean m and variance v
 - * $\text{binary}(p)$: Bernoullie with probability p
 - * $\text{binomial}(n,p)$: Binomial with count n and probability p
 - * $\text{poisson}(m)$: Poisson with mean m
 - * $\text{general}(ll)$: General model with log-likelihood ll
 - ▷ since no factors can be defined, explicit creation of dummies is required
- RANDOM statement:
 - ▷ specification of the random effects
 - ▷ the procedure requires the data to be ordered by subject !
 - ▷ empirical Bayes estimates can be obtained by adding `out=eb`

Part IV

Marginal Versus Random-effects Models and Case Studies

Chapter 25

Marginal Versus Random-effects Models

- ▷ Interpretation of GLMM parameters
- ▷ Marginalization of GLMM
- ▷ Conclusion

25.1 Interpretation of GLMM Parameters: Toenail Data

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

Parameter	GLMM	GEE
	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:

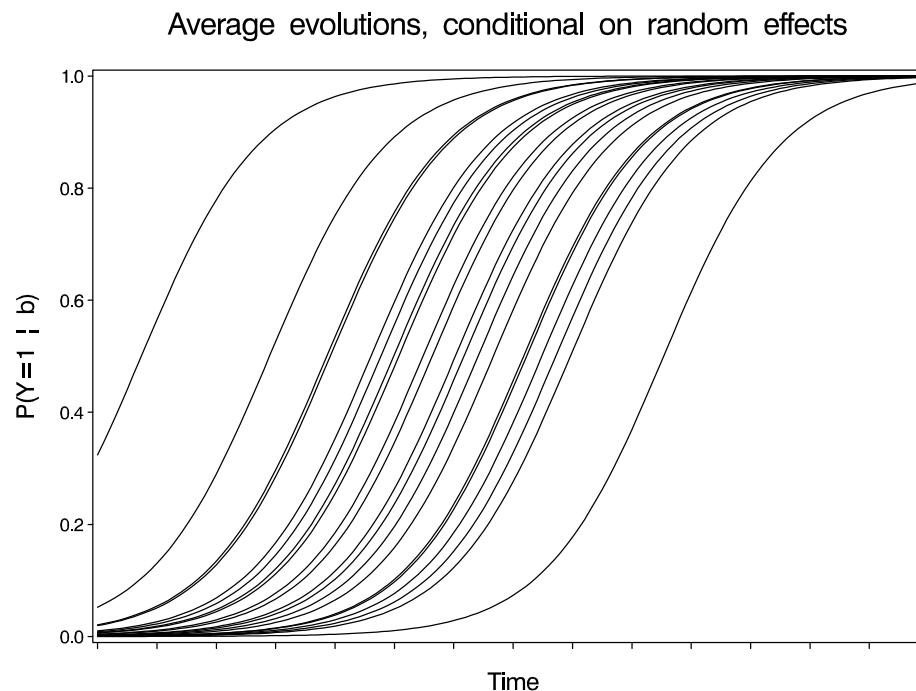
▷ Consider the following GLMM:

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

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

$$E(Y_{ij}|b_i)$$

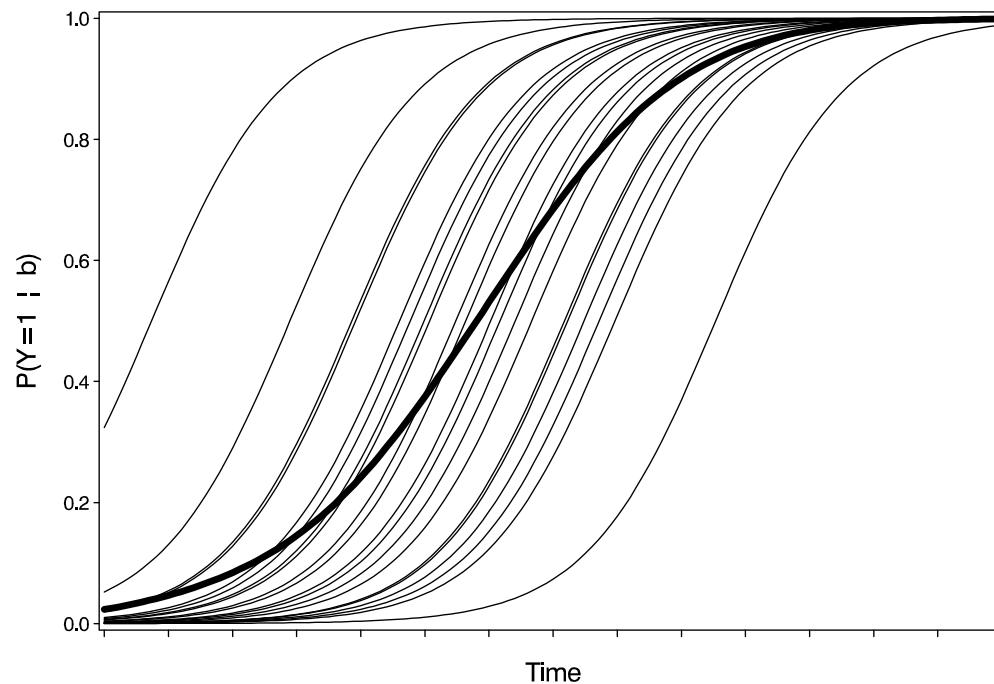
$$= \frac{\exp(\beta_0 + b_i + \beta_1 t_{ij})}{1 + \exp(\beta_0 + b_i + \beta_1 t_{ij})}$$



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

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

Average evolutions, conditional on random effects



- Hence, the parameter vector β in the GEE model needs to be interpreted completely different from the parameter vector β 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

$$\frac{\hat{\beta}^{\text{RE}}}{\hat{\beta}^{\text{M}}} = \sqrt{c^2\sigma^2 + 1} > 1, \quad \sigma^2 = \text{variance random intercepts}$$

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

- 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:

Parameter	GLMM	GEE	Ratio
	Estimate (s.e.)	Estimate (s.e.)	
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: $E(\mathbf{Y}_i|\mathbf{b}_i) = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i$
 - ▷ Specifically: $E(\mathbf{Y}_i|\mathbf{b}_i = \mathbf{0}) = \mathbf{X}_i\boldsymbol{\beta}$
 - ▷ Marginal mean: $E(\mathbf{Y}_i) = \mathbf{X}_i\boldsymbol{\beta}$

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

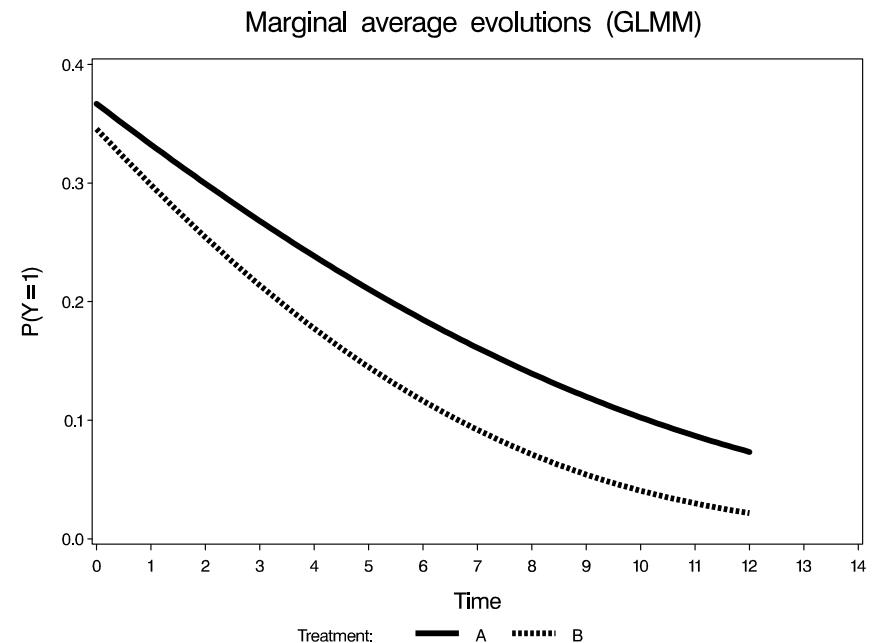
$$E[g(Y)] \neq g[E(Y)]$$

- 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.

25.2 Marginalization of GLMM: Toenail Data

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

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



- SAS code (averaging over 1000 draws):

```

data h;
do treat=0 to 1 by 1;
  do subject=1 to 1000 by 1;
    b=4.0164*rannor(-1) ;
    do t=0 to 12 by 0.1;
      if treat=0 then y=exp(-1.6308 + b -0.4043*t)
                    /(1+ exp(-1.6308 + b -0.4043*t));
      else y=exp(-1.7454 + b -0.5657*t)
                    /(1+ exp(-1.7454 + b -0.5657*t));
      output;
    end;
  end;
end;

proc sort data=h;
by t treat;
run;

proc means data=h;
var y;
by t treat;
output out=out;
run;

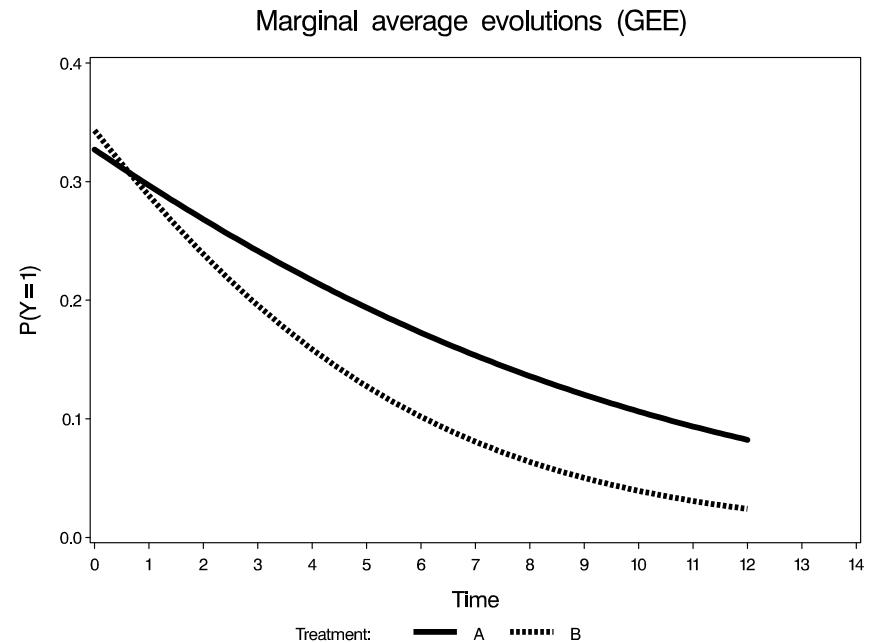
proc gplot data=out;
plot y*t=treat / haxis=axis1 vaxis=axis2 legend=legend1;
axis1 label=(h=2 'Time') value=(h=1.5)
      order=(0 to 14 by 1) minor=none;
axis2 label=(h=2 A=90 'P(Y=1)') value=(h=1.5)
      order=(0 to 0.4 by 0.1) minor=none;
legend1 label=(h=1.5 'Treatment: ')
      value=(h=1.5 'A' 'B');
title h=2.5 ' Marginal average evolutions (GLMM)';
symbol1 c=black i=join w=5 l=1 mode=include;
symbol2 c=black i=join w=5 l=2 mode=include;
where _stat_=MEAN';
run;quit;run;

```

- Average evolutions obtained from the GEE analyses:

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

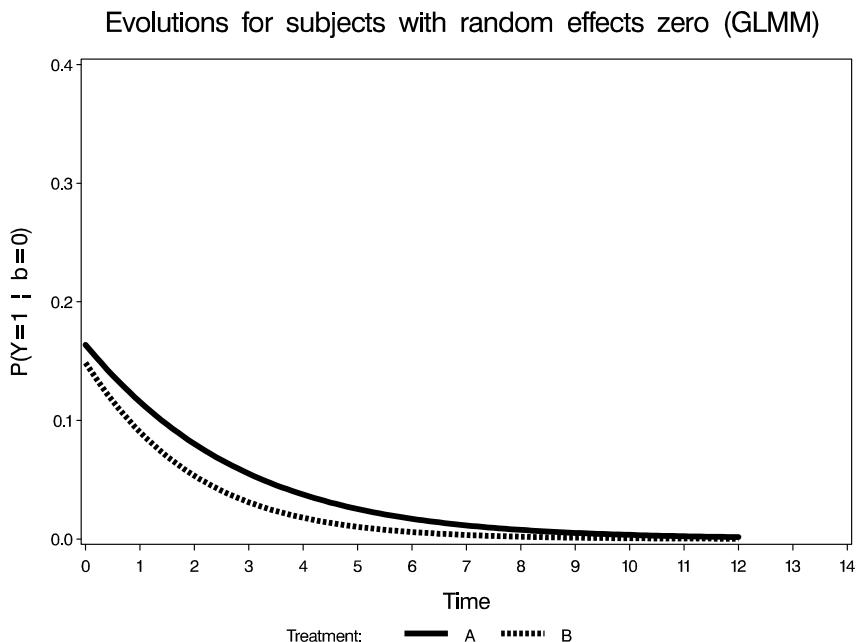
$$= \begin{cases} \frac{\exp(-0.7219 - 0.1409t_{ij})}{1 + \exp(-0.7219 - 0.1409t_{ij})} \\ \\ \frac{\exp(-0.6493 - 0.2548t_{ij})}{1 + \exp(-0.6493 - 0.2548t_{ij})} \end{cases}$$



- 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 $b_i = 0$:

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

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



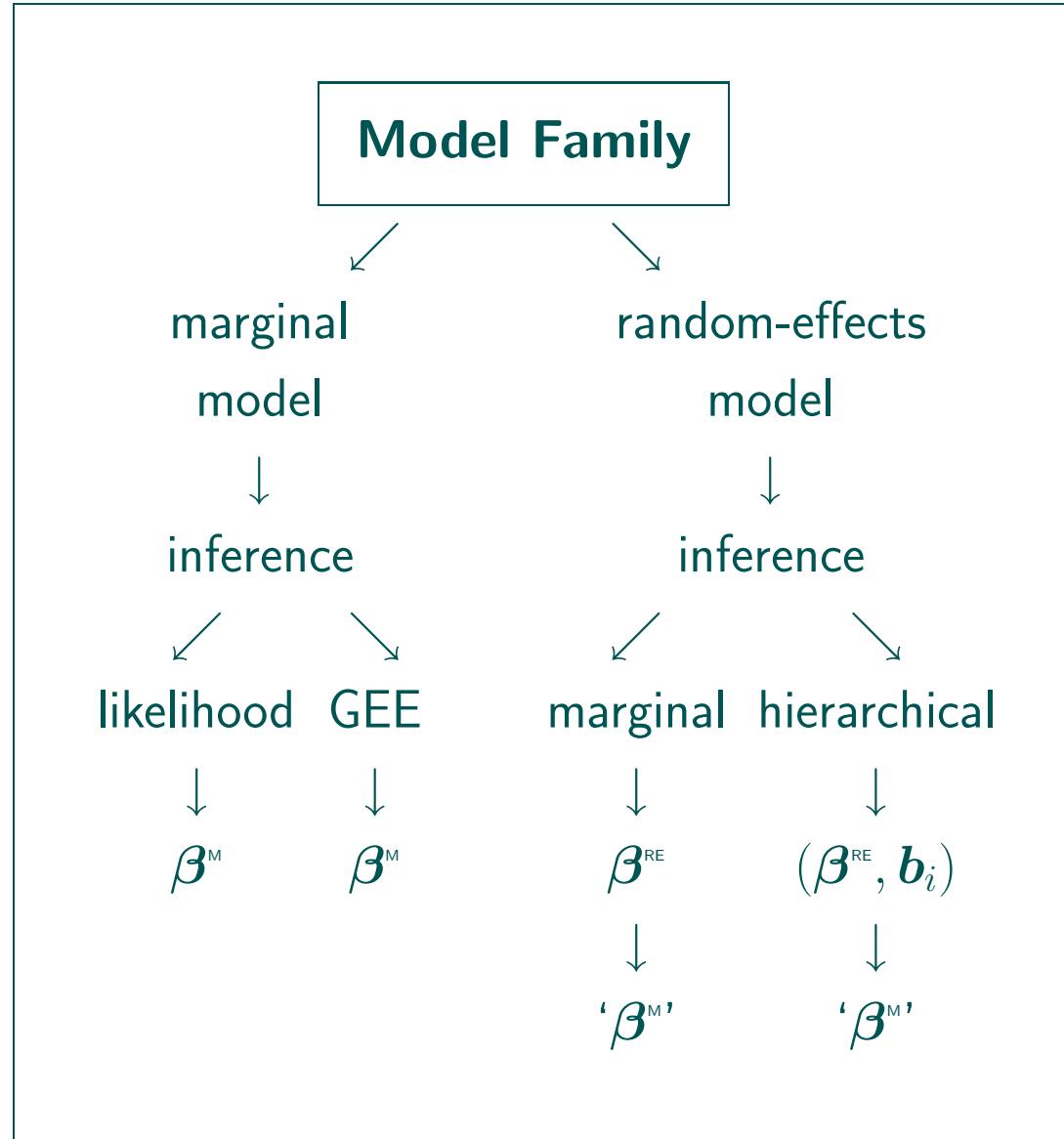
25.3 Example: Toenail Data Revisited

- 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:

$$|GEE| < |MQL| < |PQL| < |QUAD|$$



Chapter 26

Case Study: The NTP Data

- ▷ Research question
- ▷ Conditional model
- ▷ Bahadur model
- ▷ GEE1 analyses
- ▷ GEE2 analysis
- ▷ Alternating logistic regressions
- ▷ Beta-binomial model
- ▷ Generalized linear mixed model
- ▷ Discussion

26.1 Research Question

- Dose-response relationship: effect of dose on malformations

- Regression relationship:

$$\text{logit}[P(Y_{ij} = 1|d_i, \dots)] = \beta_0 + \beta_d d_i$$

- Association parameter: β_d Precise meaning is model-dependent:

- ▷ Transformed conditional odds ratio
- ▷ Transformed correlation
- ▷ Transformed marginal odds ratio

26.2 Conditional Model

- Regression relationship:

$$\text{logit}[P(Y_{ij} = 1 | d_i, Y_{ik} = 0, k \neq j)] = \beta_0 + \beta_d d_i$$

- $\delta_i = \beta_a$ is conditional log odds ratio
- Quadratic loglinear model
- Maximum likelihood estimates (model based standard errors; empirically corrected standard errors)

Outcome	Par.	DEHP	EG	DYME
External	β_0	-2.81(0.58;0.52)	-3.01(0.79;1.01)	-5.78(1.13;1.23)
	β_d	3.07(0.65;0.62)	2.25(0.68;0.85)	6.25(1.25;1.41)
	β_a	0.18(0.04;0.04)	0.25(0.05;0.06)	0.09(0.06;0.06)
Visceral	β_0	-2.39(0.50;0.52)	-5.09(1.55;1.51)	-3.32(0.98;0.89)
	β_d	2.45(0.55;0.60)	3.76(1.34;1.20)	2.88(0.93;0.83)
	β_a	0.18(0.04;0.04)	0.23(0.09;0.09)	0.29(0.05;0.05)
Skeletal	β_0	-2.79(0.58;0.77)	-0.84(0.17;0.18)	-1.62(0.35;0.48)
	β_d	2.91(0.63;0.82)	0.98(0.20;0.20)	2.45(0.51;0.82)
	β_a	0.17(0.04;0.05)	0.20(0.02;0.02)	0.25(0.03;0.03)
Collapsed	β_0	-2.04(0.35;0.42)	-0.81(0.16;0.16)	-2.90(0.43;0.51)
	β_d	2.98(0.51;0.66)	0.97(0.20;0.20)	5.08(0.74;0.96)
	β_a	0.16(0.03;0.03)	0.20(0.02;0.02)	0.19(0.03;0.03)

26.3 The Bahadur Model

- Regression relationship:

$$\text{logit}[P(Y_{ij} = 1|d_i)] = \beta_0 + \beta_d d_i$$

- β_a : Fisher's z transformed correlation
- ρ : correlation

Outcome	Parameter	DEHP	EG	DYME
External	β_0	-4.93(0.39)	-5.25(0.66)	-7.25(0.71)
	β_d	5.15(0.56)	2.63(0.76)	7.94(0.77)
	β_a	0.11(0.03)	0.12(0.03)	0.11(0.04)
	ρ	0.05(0.01)	0.06(0.01)	0.05(0.02)
Visceral	β_0	-4.42(0.33)	-7.38(1.30)	-6.89(0.81)
	β_d	4.38(0.49)	4.25(1.39)	5.49(0.87)
	β_a	0.11(0.02)	0.05(0.08)	0.08(0.04)
	ρ	0.05(0.01)	0.02(0.04)	0.04(0.02)
Skeletal	β_0	-4.67(0.39)	-2.49(0.11)	-4.27(0.61)
	β_d	4.68(0.56)	2.96(0.18)	5.79(0.80)
	β_a	0.13(0.03)	0.27(0.02)	0.22(0.05)
	ρ	0.06(0.01)	0.13(0.01)	0.11(0.02)
Collapsed	β_0	-3.83(0.27)	-2.51(0.09)	-5.31(0.40)
	β_d	5.38(0.47)	3.05(0.17)	8.18(0.69)
	β_a	0.12(0.03)	0.28(0.02)	0.12(0.03)
	ρ	0.06(0.01)	0.14(0.01)	0.06(0.01)

26.4 GEE1

- Regression relationship:

$$\text{logit}[P(Y_{ij} = 1|d_i)] = \beta_0 + \beta_d d_i$$

- ϕ : overdispersion parameter
- ρ : working correlation
- Parameter estimates (model-based standard errors; empirically corrected standard errors)
- Two sets of working assumptions:
 - ▷ Independence working assumptions
 - ▷ Exchangeable working assumptions

Outcome	Par.	Standard	Prentice	Linearized
External	β_0	-5.06(0.30;0.38)	-5.06(0.33;0.38)	-5.06(0.28;0.38)
	β_d	5.31(0.44;0.57)	5.31(0.48;0.57)	5.31(0.42;0.57)
	ϕ	0.90		0.74
Visceral	β_0	-4.47(0.28;0.36)	-4.47(0.28;0.36)	-4.47(0.28;0.36)
	β_d	4.40(0.43;0.58)	4.40(0.43;0.58)	4.40(0.43;0.58)
	ϕ	1.00		1.00
Skeletal	β_0	-4.87(0.31;0.47)	-4.87(0.31;0.47)	-4.87(0.32;0.47)
	β_d	4.89(0.46;0.65)	4.90(0.47;0.65)	4.90(0.47;0.65)
	ϕ	0.99		1.02
Collapsed	β_0	-3.98(0.22;0.30)	-3.98(0.22;0.30)	-3.98(0.22;0.30)
	β_d	5.56(0.40;0.61)	5.56(0.40;0.61)	5.56(0.41;0.61)
	ϕ	0.99		1.04

Outcome	Par.	Standard	Prentice	Linearized
External	β_0	-4.98(0.40;0.37)	-4.99(0.46;0.37)	-5.00(0.36;0.37)
	β_d	5.33(0.57;0.55)	5.32(0.65;0.55)	5.32(0.51;0.55)
	ϕ	0.88		0.65
	ρ	0.11	0.11(0.04)	0.06
Visceral	β_0	-4.50(0.37;0.37)	-4.51(0.40;0.37)	-4.50(0.36;0.37)
	β_d	4.55(0.55;0.59)	4.59(0.58;0.59)	4.55(0.54;0.59)
	ϕ	1.00		0.92
	ρ	0.08	0.11(0.05)	0.08
Skeletal	β_0	-4.83(0.44;0.45)	-4.82(0.47;0.44)	-4.82(0.46;0.45)
	β_d	4.84(0.62;0.63)	4.84(0.67;0.63)	4.84(0.65;0.63)
	ϕ	0.98		0.86
	ρ	0.12	0.14(0.06)	0.13
Collapsed	β_0	-4.05(0.32;0.31)	-4.06(0.35;0.31)	-4.04(0.33;0.31)
	β_d	5.84(0.57;0.61)	5.89(0.62;0.61)	5.82(0.58;0.61)
	ϕ	1.00		0.96
	ρ	0.11	0.15(0.05)	0.11

26.5 GEE2

- Regression relationship:

$$\text{logit}[P(Y_{ij} = 1|d_i)] = \beta_0 + \beta_d d_i$$

- β_a : Fisher's z transformed correlation
- ρ : correlation
- Working assumption: third- and fourth-order correlations are zero
- Parameter estimates (empirically corrected standard errors)

Outcome	Parameter	DEHP	EG	DYME
External	β_0	-4.98(0.37)	-5.63(0.67)	-7.45(0.73)
	β_d	5.29(0.55)	3.10(0.81)	8.15(0.83)
	β_a	0.15(0.05)	0.15(0.05)	0.13(0.05)
	ρ	0.07(0.02)	0.07(0.02)	0.06(0.02)
Visceral	β_0	-4.49(0.36)	-7.50(1.05)	-6.89(0.75)
	β_d	4.52(0.59)	4.37(1.14)	5.51(0.89)
	β_a	0.15(0.06)	0.02(0.02)	0.11(0.07)
	ρ	0.07(0.03)	0.01(0.01)	0.05(0.03)
Skeletal	β_0	-5.23(0.40)	-4.05(0.33)	
	β_d	5.35(0.60)	4.77(0.43)	
	β_a	0.18(0.02)	0.30(0.03)	
	ρ	0.09(0.01)	0.15(0.01)	
Collapsed	β_0	-5.23(0.40)	-4.07(0.71)	-5.75(0.48)
	β_d	5.35(0.60)	4.89(0.90)	8.82(0.91)
	β_a	0.18(0.02)	0.26(0.14)	0.18(0.12)
	ρ	0.09(0.01)	0.13(0.07)	0.09(0.06)

26.6 Alternating Logistic Regressions

- Regression relationship:

$$\text{logit}[P(Y_{ij} = 1|d_i)] = \beta_0 + \beta_d d_i$$

- Exchangeable association structure
- α : log odds ratio
- ψ : odds ratio
- Parameter estimates (empirically corrected standard errors)

Outcome	Parameter	DEHP	EG	DYME
External	β_0	-5.16(0.35)	-5.72(0.64)	-7.48(0.75)
	β_d	5.64(0.52)	3.28(0.72)	8.25(0.87)
	α	0.96(0.30)	1.45(0.45)	0.79(0.31)
	ψ	2.61(0.78)	4.26(1.92)	2.20(0.68)
Visceral	β_0	-4.54(0.36)	-7.61(1.06)	-7.24(0.88)
	β_d	4.72(0.57)	4.50(1.13)	6.05(1.04)
	α	1.12(0.30)	0.49(0.42)	1.76(0.59)
	ψ	3.06(0.92)	1.63(0.69)	5.81(3.43)
Skeletal	β_0	-4.87(0.49)	-3.28(0.22)	-4.92(0.34)
	β_d	4.90(0.70)	3.85(0.39)	6.73(0.65)
	α	1.05(0.40)	1.43(0.22)	1.62(0.37)
	ψ	2.86(1.14)	4.18(0.92)	5.05(1.87)
Collapsed	β_0	-4.04(0.31)	-3.19(0.22)	-5.08(0.37)
	β_d	5.93(0.63)	3.86(0.40)	7.98(0.75)
	α	1.17(0.29)	1.40(0.22)	1.26(0.31)
	ψ	3.22(0.93)	4.06(0.89)	3.53(1.09)

26.7 Beta-binomial Model

- Regression relationship:

$$\text{logit}[P(Y_{ij} = 1|d_i, \cdot)] = \beta_0 + \beta_d d_i$$

- β_a : Fisher's z transformed correlation
- ρ : correlation
- Parameter estimates (standard errors)

Outcome	Parameter	DEHP	EG	DYME
External	β_0	-4.91(0.42)	-5.32(0.71)	-7.27(0.74)
	β_d	5.20(0.59)	2.78(0.81)	8.01(0.82)
	β_a	0.21(0.09)	0.28(0.14)	0.21(0.12)
	ρ	0.10(0.04)	0.14(0.07)	0.10(0.06)
Visceral	β_0	-4.38(0.36)	-7.45(1.17)	-6.21(0.83)
	β_d	4.42(0.54)	4.33(1.26)	4.94(0.90)
	β_a	0.22(0.09)	0.04(0.09)	0.45(0.21)
	ρ	0.11(0.04)	0.02(0.04)	0.22(0.10)
Skeletal	β_0	-4.88(0.44)	-2.89(0.27)	-5.15(0.47)
	β_d	4.92(0.63)	3.42(0.40)	6.99(0.71)
	β_a	0.27(0.11)	0.54(0.09)	0.61(0.14)
	ρ	0.13(0.05)	0.26(0.04)	0.30(0.06)
Collapsed	β_0	-3.83(0.31)	-2.51(0.09)	-5.42(0.45)
	β_d	5.59(0.56)	3.05(0.17)	8.29(0.79)
	β_a	0.32(0.10)	0.28(0.02)	0.33(0.10)
	ρ	0.16(0.05)	0.14(0.01)	0.16(0.05)

26.8 Generalized Linear Mixed Model

- Regression relationship:

$$\text{logit}[P(Y_{ij} = 1|d_i, b_i)] = \beta_0 + b_i + \beta_d d_i, \quad b_i \sim N(0, \tau^2)$$

- External malformation in DEHP study
- Four ways of dealing with the integral: Laplace, adaptive Gaussian quadrature, PQL, and MQL
- Two versions of PQL and MQL: ML and REML
- Parameter estimates (standard errors)

Effect	Parameter	Laplace	QUAD
Intercept	β_0	-6.02 (0.59)	-5.97 (0.57)
Dose effect	β_d	6.50 (0.86)	6.45 (0.84)
Intercept var.	τ^2	1.42 (0.70)	1.27 (0.62)
Effect	Parameter	PQL (REML)	PQL (ML)
Intercept	β_0	-5.32 (0.40)	-5.30 (0.40)
Dose effect	β_d	5.73 (0.65)	5.71 (0.64)
Intercept var.	τ^2	0.95 (0.40)	0.89 (0.38)
Effect	Parameter	MQL (REML)	MQL (ML)
Intercept	β_0	-5.18 (0.40)	-5.17 (0.39)
Dose effect	β_d	5.70 (0.66)	5.67 (0.65)
Intercept var.	τ^2	1.20 (0.53)	1.10 (0.50)

26.9 Summary Table

- External malformation in DEHP study
- All conditional, marginal, and random-effects models considered
- Parameter estimates (standard errors)
- For non-likelihood methods, the empirically corrected standard errors are reported

Family	Model	β_0	β_d	Association	
Conditional	Quadr. loglin. (ML)	-2.81(0.58)	3.07(0.65)	LOG OR	0.18(0.04)
	Quadr. loglin. (PL)	-2.85(0.53)	3.24(0.60)	LOG OR	0.18(0.04)
Marginal	Lik. Bahadur	-4.93(0.39)	5.15(0.56)	ρ	0.05(0.01)
	St. GEE1 (exch)	-4.98(0.37)	5.33(0.55)	ρ	0.11
	St. GEE1 (ind)	-5.06(0.38)	5.31(0.57)		
	Prent. GEE1 (exch)	-4.99(0.37)	5.32(0.55)	ρ	0.11 (0.04)
	Prent. GEE1 (ind)	-5.06(0.38)	5.31(0.57)		
	Lin. based (exch)	-5.00(0.37)	5.32(0.55)	ρ	0.06
	Lin. based (ind)	-5.06(0.38)	5.31(0.57)		
	GEE2	-4.98(0.37)	5.29(0.55)	ρ	0.07(0.02)
	ALR	-.516(0.35)	5.64(0.52)	β_a	0.96(0.30)
Random-effects	Beta-binomial	-4.91(0.42)	5.20(0.59)	ρ	0.10(0.04)
	GLLM (MQL)	-5.18(0.40)	5.70(0.66)	Int. var τ^2	1.20(0.53)
	GLMM (PQL)	-5.32(0.40)	5.73(0.65)	Int. var τ^2	0.95(0.40)
	GLMM (QUAD)	-5.97(0.57)	6.45(0.84)	Int. var τ^2	1.27(0.62)

26.10 Discussion

- Relationship between regression model parameters:

$$|\text{conditional}| < |\text{marginal}| < |\text{random-effects}|$$

- Beta-binomial model behaves like a marginal model (similar to the linear mixed model)
- Marginal model parameters:
 - ▷ Mean function parameters: very similar
 - ▷ Correlation parameters:

$$|\text{Bahadur}| < |\text{GEE2}| < |\text{GEE1}| < |\text{beta-binomial}|$$

- ▷ Reason: strength of constraints:
 - * Bahadur model valid if all higher order probabilities are valid
 - * GEE2 valid if probabilities of orders 1, 2, 3, and 4 are valid
 - * GEE1 valid if probabilities of orders 1 and 2 are valid
 - * beta-binomial model is unconstrained of correlations in $[0, 1]$

- ▷ Correlation in Bahadur model really highly constrained:

For instance, the allowable range of β_a for the external outcome in the DEHP data is $(-0.0164; 0.1610)$ when β_0 and β_d are fixed at their MLE. This range excludes the MLE under a beta-binomial model. It translates to $(-0.0082; 0.0803)$ on the correlation scale.

- Additional conditional and marginal approaches can be based on **pseudo-likelihood** (Molenberghs and Verbeke 2005, Chapters 9 and 12, in particular pages 200 and 246)
- Programs: Molenberghs and Verbeke (2005, p. 219ff)

- The random effects in generalized linear mixed models

- ▷ enter linearly on the logit scale:

$$\text{logit}[P(Y_{ij} = 1|d_i, b_i)] = \beta_0 + b_i + \beta_1 d_i$$

- * mean of random intercepts is 0
 - * mean of average over litters is -3.8171
 - * mean of predicted value over litters is -3.8171

- ▷ enter non-linearly on the probability scale:

$$P(Y_{ij} = 1|d_i, b_i) = \frac{\exp(\beta_0 + b_i + \beta_1 d_i)}{1 + \exp(\beta_0 + b_i + \beta_1 d_i)}$$

- * mean of random effect is 0.0207
 - * mean of average probabilities over litters is 0.0781
 - * mean of predicted probabilities over litters is 0.0988

Chapter 27

Case Study: Binary Analysis of Analgesic Trial

- ▷ Research question
- ▷ GEE
- ▷ Alternating logistic regressions
- ▷ Further GEE analyses
- ▷ Generalized linear mixed model
- ▷ Discussion

27.1 Research Question

- Binary version of Global Satisfaction Assessment

$$\text{GSABIN} = \begin{cases} 1 & \text{if GSA} \leq 3 \text{ ('Very Good' to 'Moderate')}, \\ 0 & \text{otherwise.} \end{cases}$$

- Marginal regression relationship:

$$\text{logit}[P(Y_{ij} = 1|t_{ij}, X_i)] = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \beta_3 X_i.$$

- GLMM regression relationship:

$$\text{logit}[P(Y_{ij} = 1|t_{ij}, X_i, b_i)] = \beta_0 + b_i + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \beta_3 X_i.$$

- X_i : baseline pain control assessment (PCA0)

- Association parameters: correlation or marginal odds ratio

27.2 GEE1

- Parameter estimates (model-based standard errors; empirically corrected standard errors)
- Four sets of working assumptions:
 - ▷ Independence
 - ▷ Exchangeable
 - ▷ AR(1)
 - ▷ Unstructured

Effect	Parameter	IND	EXCH
Intercept	β_1	2.80(0.49;0.47)	2.92(0.49;0.46)
Time	β_2	-0.79(0.39;0.34)	-0.83(0.34;0.33)
Time ²	β_3	0.18(0.08;0.07)	0.18(0.07;0.07)
Basel. PCA	β_4	-0.21(0.09;0.10)	-0.23(0.10;0.10)
Correlation	ρ	—	0.22

Effect	Parameter	AR	UN
Intercept	β_1	2.94(0.49;0.47)	2.87(0.48;0.46)
Time	β_2	-0.90(0.35;0.33)	-0.78(0.33;0.32)
Time ²	β_3	0.20(0.07;0.07)	0.17(0.07;0.07)
Basel. PCA	β_4	-0.22(0.10;0.10)	-0.23(0.10;0.10)
Correlation	ρ	0.25	—
Correlation (1,2)	ρ_{12}		0.18
Correlation (1,3)	ρ_{13}		0.25
Correlation (1,4)	ρ_{14}		0.20
Correlation (2,3)	ρ_{23}		0.18
Correlation (2,4)	ρ_{24}		0.18
Correlation (3,4)	ρ_{34}		0.46

- Fitted **working** correlation matrices:

$$R_{\text{EXCH}} = \begin{pmatrix} 1 & 0.22 & 0.22 & 0.22 \\ & 1 & 0.22 & 0.22 \\ & & 1 & 0.22 \\ & & & 1 \end{pmatrix} \quad R_{\text{AR}} = \begin{pmatrix} 1 & 0.25 & 0.06 & 0.02 \\ & 1 & 0.25 & 0.06 \\ & & 1 & 0.25 \\ & & & 1 \end{pmatrix}$$

$$R_{\text{UN}} = \begin{pmatrix} 1 & 0.18 & 0.25 & 0.20 \\ & 1 & 0.18 & 0.18 \\ & & 1 & 0.46 \\ & & & 1 \end{pmatrix}$$

27.3 Alternating Logistic Regressions

- Parameter estimates (empirically corrected standard errors)
- Three sets of odds ratio structures:
 - ▷ Exchangeable
 - ▷ Unstructured \equiv full clustering (FULLCLUST)
 - ▷ User-defined design (ZREP)

Effect	Parameter	EXCH	FULLCLUST	ZREP
Intercept	β_1	2.98(0.46)	2.92(0.46)	2.92(0.46)
Time	β_2	-0.87(0.32)	-0.80(0.32)	-0.80(0.32)
Time ²	β_3	0.18(0.07)	0.17(0.06)	0.17(0.07)
Basel. PCA	β_4	-0.23(0.22)	-0.24(0.10)	-0.24(0.10)
Log OR	α	1.43(0.22)		
Log OR(1,2)	α_{12}		1.13(0.33)	
Log OR(1,3)	α_{13}		1.56(0.39)	
Log OR(1,4)	α_{14}		1.60(0.42)	
Log OR(2,3)	α_{23}		1.19(0.37)	
Log OR(2,4)	α_{24}		0.93(0.42)	
Log OR(3,4)	α_{34}		2.44(0.48)	
Log OR par.	α_0			1.26(0.23)
Log OR par.	α_1			1.17(0.47)

- In the FULLCLUST structure, there is a hint that α_{34} is different from the others, with all others being equal.
- To confirm this, a Wald test can be used for the null hypothesis:

$$H_0 : \alpha_{12} = \alpha_{13} = \alpha_{14} = \alpha_{23} = \alpha_{24}$$

- Details on the test: Molenberghs and Verbeke (2005, pp. 312–313)
- The reduced structure, fitted with ZREP, is:

$$\begin{aligned}\alpha_{12} &= \alpha_{13} = \alpha_{14} = \alpha_{23} = \alpha_{24} = \alpha_0, \\ \alpha_{34} &= \alpha_0 + \alpha_1\end{aligned}$$

- At the odds ratio level, with fitted values:

$$\begin{aligned}\widehat{\psi}_{12} &= \widehat{\psi}_{13} = \widehat{\psi}_{14} = \widehat{\psi}_{23} = \widehat{\psi}_{24} = \widehat{\psi}_0 = 3.53, \\ \widehat{\psi}_{34} &= \widehat{\psi}_0 \cdot \widehat{\psi}_1 = 11.36.\end{aligned}$$

- “Odds ratio matrices”:

$$\Psi_{\text{EXCH}} = \begin{pmatrix} 1 & 4.18 & 4.18 & 4.18 \\ & 1 & 4.18 & 4.18 \\ & & 1 & 4.18 \\ & & & 1 \end{pmatrix}$$

$$\Psi_{\text{UN}} = \begin{pmatrix} 1 & 3.10 & 4.76 & 4.95 \\ & 1 & 3.29 & 2.53 \\ & & 1 & 11.47 \\ & & & 1 \end{pmatrix}$$

$$\Psi_{\text{ZREP}} = \begin{pmatrix} 1 & 3.53 & 3.53 & 3.53 \\ & 1 & 3.53 & 3.53 \\ & & 1 & 11.36 \\ & & & 1 \end{pmatrix}$$

27.4 A Variety of GEE Methods

- Methods used:
 - ▷ Ordinary logistic regression
 - ▷ Standard GEE1
 - ▷ Prentice's GEE1
 - ▷ The linearization-based method
 - ▷ Alternating logistic regression
- Exchangeably working assumption (except for logistic regression)
- Parameter estimates (empirically corrected standard errors, unless for logistic regression)

Effect	Parameter	Log. regr.	Standard	Prentice
Intercept	β_1	2.80(0.49)	2.92(0.46)	2.94(0.46)
Time	β_2	-0.79(0.39)	-0.83(0.33)	-0.84(0.33)
Time ²	β_3	0.18(0.08)	0.18(0.07)	0.18(0.07)
Basel. PCA	β_4	-0.21(0.09)	-0.23(0.10)	-0.23(0.10)
Correlation	ρ		0.21	0.26(0.05)
Effect	Parameter	Lineariz.	ALR	
Intercept	β_1	2.94(0.46)	2.98(0.46)	
Time	β_2	-0.84(0.33)	-0.87(0.32)	
Time ²	β_3	0.18(0.07)	0.18(0.07)	
Basel. PCA	β_4	-0.23(0.10)	-0.23(0.10)	
Corr.	ρ	0.26(0.04)		
Log OR	α		1.43(0.22)	

27.5 Generalized Linear Mixed Models

- Four tools:

▷ SAS procedure GLIMMIX:

- * MQL (= MQL1)
- * PQL (= PQL1)

▷ SAS procedure NLMIXED:

- * I: non-adaptive ($Q = 10$)
- * II: non-adaptive ($Q = 10$)
≡ adaptive ($Q = 10$)
≡ adaptive ($Q = 20$)

▷ MLwiN:

- * PQL1
- * PQL2

▷ MIXOR

- Parameter estimates (standard errors)

		Integrand approximation			
Effect	Par.	SAS GLIMMIX		MLwiN	
		MQL	PQL1	PQL1	PQL2
Intercept	β_1	2.91(0.53)	3.03(0.55)	3.02(0.55)	4.07(0.70)
Time	β_2	-0.83(0.39)	-0.87(0.41)	-0.87(0.41)	-1.17(0.48)
Time ²	β_3	0.18(0.08)	0.19(0.08)	0.19(0.08)	0.25(0.10)
Basel. PCA	β_4	-0.22(0.11)	-0.22(0.11)	-0.22(0.11)	-0.31(0.15)
Rand. int s.d.	τ	1.06(0.25)	1.04(0.23)	1.01(0.12)	1.61(0.15)
Rand. int var.	τ^2	1.12(0.53)	1.08(0.48)	1.02(0.25)	2.59(0.47)

Numerical integration					
SAS NL MIXED					
Effect	Par.	I	II	MIXOR	
Intercept	β_1	4.07(0.71)	4.05(0.71)	4.05(0.55)	
Time	β_2	-1.16(0.47)	-1.16(0.47)	-1.16(0.45)	
Time ²	β_3	0.25(0.09)	0.24(0.09)	0.24(0.10)	
Basel. PCA	β_4	-0.30(0.14)	-0.30(0.14)	-0.30(0.15)	
Rand. int s.d.	τ	1.60(0.22)	1.59(0.21)	1.59(0.21)	
Rand. int var.	τ^2	2.56(0.70)	2.53(0.68)	2.53(0.67)	

27.6 Discussion

- Results are very similar, due to a relatively weak random-effects variance
- PQL1 and MQL1 perform relatively poorly
- The ratio between the RE and marginal parameters now is 1.37
- Programs: Molenberghs and Verbeke (2005, p. 219ff)

Chapter 28

Case Study: Ordinal Analysis of Analgesic Trial

- ▷ Proportional odds logistic regression
- ▷ Generalized estimating equations
- ▷ Generalized linear mixed models
- ▷ Analysis of the analgesic trial

28.1 Proportional Odds Logistic Regression

- Standard logistic regression for binary data:

$$\text{logit}[P(Y_i = 1|x_i)] = \alpha + \beta x_i$$

- An extension to ordinal data: proportional odds logistic regression

$$\text{logit}[P(Y_i \leq k|x_i)] = \alpha_k + \beta x_i, \quad (k = 1, \dots, c-1)$$

- A further extension poses problems with range-preserving restrictions:

$$\text{logit}[P(Y_i \leq k|x_i)] = \alpha_k + \beta_o x_i, \quad (k = 1, \dots, c-1)$$

and is usually not considered

- An alternative model for ordinal data is the continuation-ratio model:

$$\text{logit}[P(Y_i > k|Y_i \geq k, x_i)] = \alpha_k + \beta_k x_i, \quad (k = 1, \dots, c-1)$$

It is of use only when there is one natural directionality in the data: subjects go from the lowest category to higher categories, without ever returning. This is often not satisfied.

- Proportional-odds model for the 5-point GSA outcome in the analgesic trial:

$$\text{logit}[P(Y_{ij} \leq k | t_{ij}, X_i)] = \alpha_k + \beta_2 t_{ij} + \beta_3 t_{ij}^2 + \beta_4 X_i, \quad (k = 1, \dots, 4)$$

- SAS code:

```
proc genmod data=m.gsa2;
  title 'Analgesic, logistic regression, Ordinal';
  class patid timecls;
  model gsa = time|time pca0 / dist=multinomial link=cumlogit;
run;
```

- Note that the 'dist' and 'link' options have been adapted

- Selected output:

The GENMOD Procedure							
Analysis Of Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept1	1	-1.0048	0.3437	-1.6785	-0.3312	8.55	0.0035
Intercept2	1	0.5225	0.3407	-0.1452	1.1903	2.35	0.1251
Intercept3	1	2.3171	0.3481	1.6349	2.9994	44.31	<.0001
Intercept4	1	4.0525	0.3754	3.3166	4.7884	116.51	<.0001
TIME	1	-0.2027	0.2706	-0.7330	0.3277	0.56	0.4539
TIME*TIME	1	0.0479	0.0545	-0.0590	0.1547	0.77	0.3798
PCAO	1	-0.2141	0.0622	-0.3361	-0.0922	11.84	0.0006

- There are $5 - 1 = 4$ intercepts, as it should.

28.2 Generalized Estimating Equations

- The same regression model as in the PO logistic regression case is used:

$$\text{logit}[P(Y_{ij} \leq k | t_{ij}, X_i)] = \alpha_k + \beta_2 t_{ij} + \beta_3 t_{ij}^2 + \beta_4 X_i, \quad (k = 1, \dots, 4)$$

- This model is supplemented with working assumptions to obtain GEE
- In the SAS procedure GENMOD, only independence working assumptions are implemented for ordinal outcomes:

```
proc genmod data=m.gsa2;
  title 'Analgesic, GEE, Ordinal';
  class patid timecls;
  model gsa = time|time pca0 / dist=multinomial link=cumlogit;
  repeated subject=patid / type=ind covb corrw within=timecls modelse;
run;
```

- The output is structured in the same way as for PO logistic regression:

Analysis Of GEE Parameter Estimates Empirical Standard Error Estimates						
Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > Z
			-	-		
Intercept1	-1.0048	0.3549	-1.7004	-0.3092	-2.83	0.0046
Intercept2	0.5225	0.3568	-0.1767	1.2218	1.46	0.1430
Intercept3	2.3171	0.3669	1.5980	3.0363	6.31	<.0001
Intercept4	4.0525	0.3938	3.2807	4.8243	10.29	<.0001
TIME	-0.2027	0.2028	-0.6001	0.1948	-1.00	0.3176
TIME*TIME	0.0479	0.0399	-0.0304	0.1261	1.20	0.2304
PCAO	-0.2141	0.0911	-0.3927	-0.0356	-2.35	0.0187

28.3 Generalized Linear Mixed Models

- A generalized linear mixed model for ordinal data:

$$\text{logit}[P(Y_{ij} \leq k | X_i, Z_i)] = \alpha_k + x'_{ij}\beta + z'_{ij}\boldsymbol{b}_i, \quad (k = 1, \dots, c-1)$$

- This is the obvious counterpart for the PO logistic and GEE marginal models considered above.
- For the case of the 5-point GSA outcome in the analgesic study:

$$\text{logit}[P(Y_{ij} \leq k | t_{ij}, X_i, b_i)] = \alpha_k + b_i + \beta_2 t_{ij} + \beta_3 t_{ij}^2 + \beta_4 X_i, \quad (k = 1, \dots, 4)$$

- Code for the SAS procedure GLIMMIX:

```
proc glimmix data=m.gsa2 method=RSPL;
title 'PROC GLIMMIX analysis, ordinal, RSPL (PQL, REML)';
class patid timecls;
nloptions maxiter=50;
model gsa = time|time pca0 / dist=multinomial link=cumlogit solution;
random intercept / subject=patid type=un;
run;
```

- Also here, the 'dist' and 'link' functions have to be adapted to the ordinal setting.

- Selected output:

Covariance Parameter Estimates

Cov Parm	Subject	Standard	
		Estimate	Error
UN(1,1)	PATID	3.5348	0.4240

Solutions for Fixed Effects

Effect	GSA	Standard		DF	t Value	Pr > t
		Estimate	Error			
Intercept	1	-1.4352	0.5033	393	-2.85	0.0046
Intercept	2	0.9101	0.4999	393	1.82	0.0694
Intercept	3	3.4720	0.5084	393	6.83	<.0001
Intercept	4	5.6263	0.5358	393	10.50	<.0001
TIME		-0.4825	0.2958	737	-1.63	0.1033
TIME*TIME		0.1009	0.05972	737	1.69	0.0916
PCAO		-0.2843	0.1249	737	-2.28	0.0231

- In case the procedure NLMIXED is used, more drastic changes are needed:

```

proc nlmixed data=m.gsa2 qpoints=20;
  title 'Analgesic, PROC NLMIXED, ordinal, adaptive, q=20';
  parms int1=-1.5585 int2=1.0292 int3=3.8916 int4=6.2144
        beta1=0.5410 beta2=-0.1123 beta3=0.3173 d=2.1082;
  eta = beta1*time + beta2*time*time + beta3*pca0 + b1;
  if gsa=1 then z = 1/(1+exp(-(int1-eta)));
  else if gsa=2 then z = 1/(1+exp(-(int2-eta))) - 1/(1+exp(-(int1-eta)));
  else if gsa=3 then z = 1/(1+exp(-(int3-eta))) - 1/(1+exp(-(int2-eta)));
  else if gsa=4 then z = 1/(1+exp(-(int4-eta))) - 1/(1+exp(-(int3-eta)));
  else z = 1 - 1/(1+exp(-(int4-eta)));
  if z > 1e-8 then ll = log(z);
  else ll = -1e100;
  model gsa ~ general(ll);
  random b1 ~ normal(0,d*d) subject=patid;
  estimate 'var(d)' d*d;
run;

```

- Now, the **general likelihood** is used: a fully user-defined likelihood function.
- The probabilities are obtained as differences between cumulative probabilities:

$$P(Y_{ij} = k) = P(Y_{ij} \leq k) - P(Y_{ij} \leq k - 1), \quad (k = 1, \dots, 5)$$

with

- ▷ $P(Y_{ij} \leq 0) = 0$
- ▷ $P(Y_{ij} \leq 5) = 1$

- η is the part of the linear predictor excluding the intercept

- Selected output:

Parameter Estimates										
Parameter	Estimate	Standard		DF	t Value	Pr > t	Alpha	Lower	Upper	Gradient
		Error								
int1	-1.5585	0.5481	394	-2.84	0.0047	0.05	-2.6360	-0.4810	0.000235	
int2	1.0292	0.5442	394	1.89	0.0593	0.05	-0.04063	2.0991	-0.00004	
int3	3.8916	0.5624	394	6.92	<.0001	0.05	2.7860	4.9973	-0.00017	
int4	6.2144	0.5990	394	10.37	<.0001	0.05	5.0368	7.3920	-0.00004	
beta1	0.5410	0.3078	394	1.76	0.0796	0.05	-0.06421	1.1462	-0.00008	
beta2	-0.1123	0.06187	394	-1.82	0.0702	0.05	-0.2340	0.009311	0.000019	
beta3	0.3173	0.1386	394	2.29	0.0226	0.05	0.04475	0.5898	0.000013	
d	2.1082	0.1412	394	14.94	<.0001	0.05	1.8307	2.3858	0.000331	

Additional Estimates									
Label	Estimate	Standard		DF	t Value	Pr > t	Alpha	Lower	
		Error							
var(d)	4.4447	0.5952	394	7.47	<.0001	0.05	3.2746	5.6148	

28.4 Analysis of the Analgesic Trial

- Three approaches:
 - ▷ Logistic regression
 - ▷ GEE
 - ▷ GLMM
- For GEE: (model based standard errors; empirically corrected standard errors)
- MQL performs again rather poorly

Marginal models			
Effect	Parameter	OLR	GEE
Intercept 1	α_1	-1.00(0.34)	-1.00(0.34;0.35)
Intercept 2	α_2	0.52(0.34)	0.52(0.34;0.36)
Intercept 3	α_3	2.32(0.35)	2.32(0.34;0.37)
Intercept 4	α_4	4.05(0.38)	4.05(0.37;0.39)
Time	β_2	-0.20(0.27)	-0.20(0.27;0.20)
Time ²	β_3	0.05(0.05)	0.05(0.05;0.04)
Basel. PCA	β_4	-0.21(0.06)	-0.21(0.06;0.09)

Random-effects models				
Effect	Parameter	MQL	PQL	N.Int.
Intercept 1	α_1	-0.93(0.40)	-1.44(0.50)	-1.56(0.55)
Intercept 2	α_2	0.60(0.39)	0.51(0.50)	1.03(0.54)
Intercept 3	α_3	2.39(0.40)	3.47(0.51)	3.89(0.56)
Intercept 4	α_4	4.13(0.42)	5.63(0.54)	6.21(0.60)
Time	β_2	-0.30(0.28)	-0.48(0.30)	0.54(0.31)
Time ²	β_3	0.06(0.06)	0.10(0.06)	-0.11(0.06)
Basel. PCA	β_4	-0.21(0.09)	-0.28(0.12)	0.32(0.14)
Rand. int s.d.	τ	1.06(0.08)	1.88(0.11)	2.11(0.14)
Rand. int var.	τ^2	1.13(0.16)	3.53(0.42)	4.44(0.60)

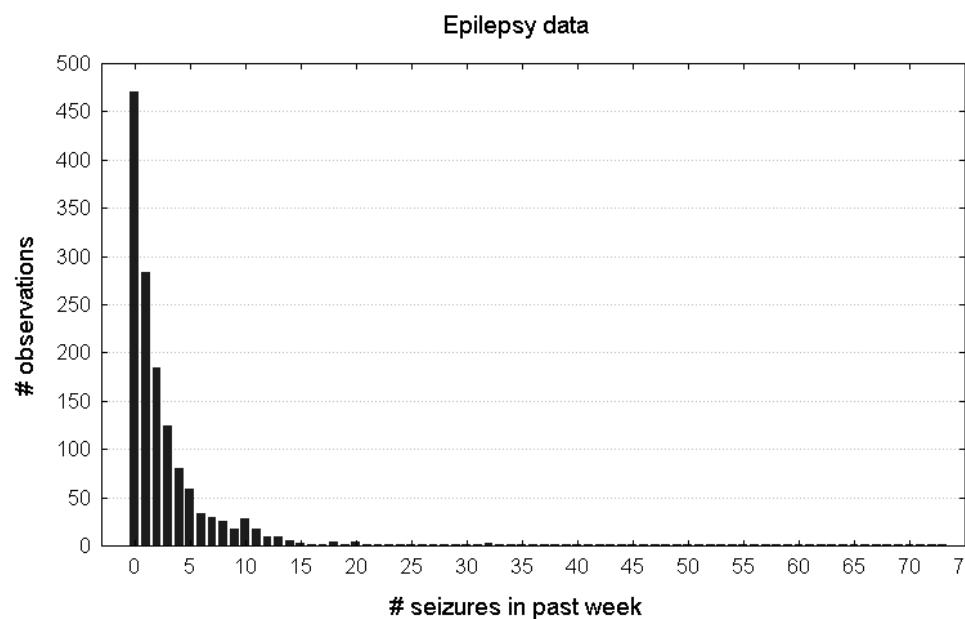
Chapter 29

Count Data: The Epilepsy Study

- ▷ The epilepsy data
- ▷ Poisson regression
- ▷ Generalized estimating equations
- ▷ Generalized linear mixed models
- ▷ Overview of analyses of the epilepsy study
- ▷ Marginalization of the GLMM

29.1 The Epilepsy Data

- Consider the epilepsy data:

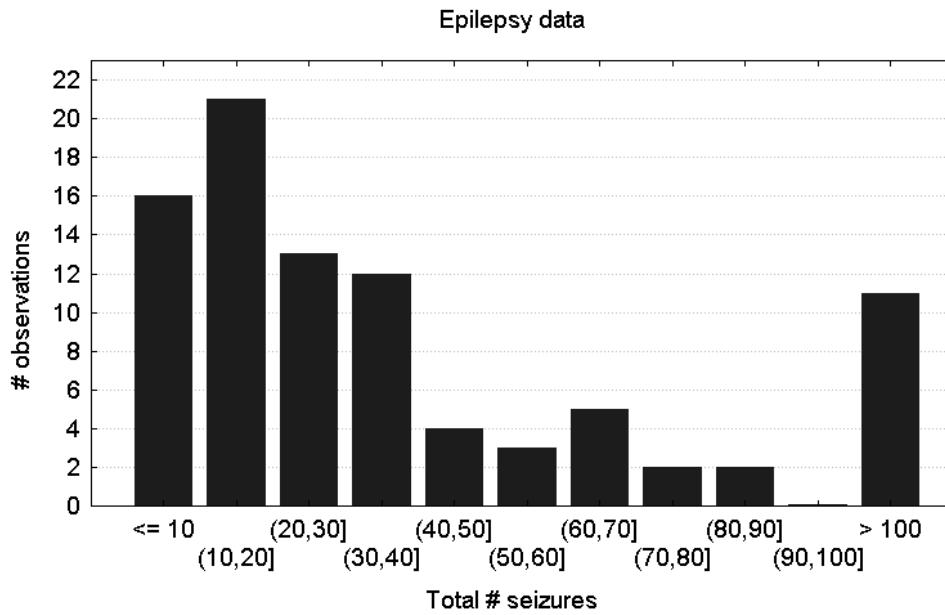


- We want to test for a treatment effect on number of seizures, correcting for the average number of seizures during the 12-week baseline phase, prior to the treatment.
- The response considered now is the total number of seizures a patient experienced, i.e., the sum of all weekly measurements.
- Let Y_i now be the total number of seizures for subject i :

$$Y_i = \sum_{j=1}^{n_i} Y_{ij}$$

where Y_{ij} was the original (longitudinally measured) weekly outcome.

- Histogram:



- As these sums are not taken over an equal number of visits for all subjects, the above histogram is not a ‘fair’ one as it does not account for differences in n_i for this.

- We will therefore use the following Poisson model:

$$Y_i \sim \text{Poisson}(\lambda_i)$$

$$\ln(\lambda_i/n_i) = \mathbf{x}_i' \boldsymbol{\beta}$$

- Note that the regression model is equivalent to

$$\lambda_i = n_i \exp(\mathbf{x}_i' \boldsymbol{\beta}) = \exp(\mathbf{x}_i' \boldsymbol{\beta} + \ln n_i)$$

- Since n_i is the number of weeks for which the number of seizures was recorded for subject i , $\exp(\mathbf{x}_i' \boldsymbol{\beta})$ is the average number of seizures per week.
- $\ln n_i$ is called an offset in the above model.
- In our application, the covariates in \mathbf{x}_i are the treatment as well as the baseline seizure rate.

- SAS statements for the calculation of outcome, offset, and for fitting the Poisson model:

```

proc sort data=test;
by id studyweek;
run;

proc means data=test sum n nmiss;
var nseizw;
by id;
output out=result
      n=n
      nmiss=nmiss
      sum=sum;
run;

data result;
set result;
offset=log(n-nmiss);
keep id offset sum;
run;

data first;
set test;
by id;
if first.id;
keep id bserate trt;
run;

data result;
merge result first;
by id;
run;

proc genmod data=result;
model sum=bserate trt
      / dist=poisson offset=offset;
run;

```

- The treatment variable **trt** is coded as 0 for placebo and 1 for treated
- Output from the GENMOD procedure:

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95%		Chi-Square
				Confidence	Limits	
Intercept	1	0.8710	0.0218	0.8283	0.9138	1596.16
bserate	1	0.0172	0.0002	0.0167	0.0177	4826.14
trt	1	-0.4987	0.0341	-0.5655	-0.4319	214.18
Scale	0	1.0000	0.0000	1.0000	1.0000	

- We obtain a highly significant reduction in the average number of seizures in the treated group, in comparison to the placebo group.

- A more general model would allow the treatment effect to depend on the baseline average number of seizures:

```
proc genmod data=result;
model sum=bserate trt bserate*trt
    / dist=poisson offset=offset;
run;
```

- Relevant part of the output:

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits	Chi-Square
Intercept	1	0.2107	0.0353	0.1415 0.2799	35.60
bserate	1	0.0450	0.0009	0.0432 0.0469	2286.94
trt	1	0.2938	0.0454	0.2047 0.3829	41.81
bserate*trt	1	-0.0295	0.0010	-0.0314 -0.0276	911.43
Scale	0	1.0000	0.0000	1.0000 1.0000	

- We get a significant interaction.
- In order to explore the nature of this interaction, we estimate the treatment effect when the baseline average number of seizures equals 6, 10.5, as well as 21 (quartiles).
- This is possible via inclusion of estimate statements:

```
proc genmod data=result;
model sum=bserate trt bserate*trt
    / dist=poisson offset=offset;
estimate 'trt, bserate=6' trt 1 bserate*trt 6;
estimate 'trt, bserate=10.5' trt 1 bserate*trt 10.5;
estimate 'trt, bserate=21' trt 1 bserate*trt 21;
run;
```

- Additional output:

Contrast Estimate Results

Label	Estimate	Standard	
		Error	Alpha
trt, bserate=6	0.1167	0.0415	0.05
trt, bserate=10.5	-0.0161	0.0388	0.05
trt, bserate=21	-0.3260	0.0340	0.05

Label	Confidence Limits	Chi-	
		Square	Pr > ChiSq
trt, bserate=6	0.0355	0.1980	7.93
trt, bserate=10.5	-0.0921	0.0600	0.17
trt, bserate=21	-0.3926	-0.2593	91.86

- On average, there are more seizures in the treatment group when there are few seizures at baseline. The opposite is true for patients with many seizures at baseline.

29.2 Generalized Estimating Equations

- Poisson regression models will be used to describe the marginal distributions, i.e., the distribution of the outcome at each time point separately:

$$Y_{ij} = \text{Poisson}(\lambda_{ij})$$

$$\log(\lambda_{ij}) = \beta_0 + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij}$$

- Notation:
 - ▷ T_i : treatment indicator for subject i
 - ▷ t_{ij} : time point at which j th measurement is taken for i th subject
- Note that, again, the randomization would allow to set β_1 equal to 0.

- More complex mean models can again be considered (e.g. including polynomial time effects, or including covariates).
- As the response is now the number of seizures during a fixed period of one week, we do not need to include an offset, as was the case in the GLM fitted previously to the epilepsy data, not in the context of repeated measurements.
- Given the long observation period, an unstructured working correlation would require estimation of many correlation parameters.
- Further, the long observation period makes the assumption of an exchangeable correlation structure quite unrealistic.
- We therefore use the AR(1) working correlation structure, which makes sense since we have equally spaced time points at which measurements have been taken.

- SAS code:

```
proc genmod data=test;
class id timeclss;
model nseizw = trt time trt*time / dist=poisson;
repeated subject=id / withinsubject=timeclss type=AR(1) corrw modelse;
run;
```

- Relevant SAS output:

	Col1	Col2	Col3	Col26	Col27
Row1	1.0000	0.5946	0.3535
Row2	0.5946	1.0000	0.5946
Row3	0.3535	0.5946	1.0000
.....
Row26	1.0000	0.5946
Row27	0.5946	1.0000

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence			Z	Pr > Z
			Limits				
Intercept	1.2259	0.1778	0.8774	1.5743	6.90	<.0001	
trt	0.1681	0.2785	-0.3777	0.7138	0.60	0.5461	
time	-0.0071	0.0229	-0.0519	0.0378	-0.31	0.7574	
trt*time	-0.0183	0.0279	-0.0730	0.0364	-0.66	0.5124	

Analysis Of GEE Parameter Estimates
Model-Based Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence			Z	Pr > Z
			Limits				
Intercept	1.2259	0.2349	0.7655	1.6862	5.22	<.0001	
trt	0.1681	0.3197	-0.4585	0.7947	0.53	0.5991	
time	-0.0071	0.0230	-0.0521	0.0380	-0.31	0.7585	
trt*time	-0.0183	0.0310	-0.0790	0.0425	-0.59	0.5553	

- The AR(1) correlation coefficient is estimated to be equal to 0.5946
- There is no difference in average evolution between both treatment groups ($p = 0.5124$).
- Note also the huge discrepancies between the results for the initial parameter estimates and the final results based on the GEE analysis.

29.3 Random-effects Model

- Conditionally on a random intercept b_i , Poisson regression models will be used to describe the marginal distributions, i.e., the distribution of the outcome at each time point separately:

$$Y_{ij} = \text{Poisson}(\lambda_{ij})$$

$$\log(\lambda_{ij}) = \beta_0 + b_i + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij}$$

- Notation:
 - ▷ T_i : treatment indicator for subject i
 - ▷ t_{ij} : time point at which j th measurement is taken for i th subject
- Similar as in our GEE analysis, we do not need to include an offset, because the response is now the number of seizures during a fixed period of one week.

- Two equivalent SAS programs:

```

proc nlmixed data=test;
parms int0=0.5 slope0=-0.1 int1=1 slope1=0.1 sigma=1;
if (trt = 0) then eta = int0 + b + slope0*time;
else if (trt = 1) then eta = int1 + b + slope1*time;
lambda = exp(eta);
model nseizw ~ poisson(lambda);
random b ~ normal(0,sigma**2) subject = id;
estimate 'difference in slope' slope1-slope0;
run;

proc nlmixed data=test;
parms int0=0.5 slope0=-0.1 int1=1 slope1=0.1 sigma=1;
eta = (1-trt)*int0 + trt*int1 + b
      + (1-trt)*slope0*time + trt*slope1*time;
lambda = exp(eta);
model nseizw ~ poisson(lambda);
random b ~ normal(0,sigma**2) subject = id;
estimate 'difference in slope' slope1-slope0;
run;

```

- As in the MIXED procedure, CONTRAST and ESTIMATE statements can be specified as well. However, under PROC NLMIXED, one is no longer restricted to linear functions of the parameters in the mean structure only.
- For example, estimation of the ratio of both slopes, as well as of the variance of the random intercepts is achieved by adding the following ESTIMATE statements:

```
estimate 'ratio of slopes' slope1/slope0;
estimate 'variance RI's sigma**2;
```

- Inference for such functions of parameters is based on the so-called ‘delta-method’:
 - ▷ Let ψ be the vector of all parameters in the marginal model.
 - ▷ Let $\widehat{\psi}$ be the MLE of ψ
 - ▷ $\widehat{\psi}$ is asymptotically normally distributed with mean ψ and covariance matrix $\text{var}(\widehat{\psi})$ (inverse Fisher information matrix).

- ▷ The ‘delta-method’ then implies that any function $\mathbf{F}(\widehat{\boldsymbol{\psi}})$ of $\widehat{\boldsymbol{\psi}}$ is asymptotically normally distributed with mean $\mathbf{F}(\boldsymbol{\psi})$ and covariance matrix equal to

$$\text{var}(\mathbf{F}(\widehat{\boldsymbol{\psi}})) = \frac{\partial \mathbf{F}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}'} \text{var}(\widehat{\boldsymbol{\psi}}) \frac{\partial \mathbf{F}'(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}$$

- ▷ Hence, a Wald-type test can be constructed, replacing the parameters in $\text{var}(\mathbf{F}(\widehat{\boldsymbol{\psi}}))$ by their estimates

- Relevant SAS output:

Parameter Estimates											
Parameter	Estimate	Error	Standard								Gradient
			DF	t Value	Pr > t	Alpha	Lower			Upper	
int0	0.8180	0.1675	88	4.88	<.0001	0.05	0.4852	1.1509	0.006008		
slope0	-0.01429	0.004404	88	-3.24	0.0017	0.05	-0.02304	-0.00554	0.022641		
int1	0.6478	0.1699	88	3.81	0.0003	0.05	0.3101	0.9855	0.010749		
slope1	-0.01200	0.004318	88	-2.78	0.0067	0.05	-0.02058	-0.00342	-0.04858		
sigma	1.0742	0.08556	88	12.55	<.0001	0.05	0.9042	1.2442	0.009566		

Additional Estimates								
Label	Standard							
	Estimate	Error	DF	t Value	Pr > t	Alpha	Lower	Upper
difference in slope	0.002287	0.006167	88	0.37	0.7116	0.05	-0.00997	0.01454
ratio of slopes	0.8399	0.3979	88	2.11	0.0376	0.05	0.04923	1.6306
variance RI	1.1539	0.1838	88	6.28	<.0001	0.05	0.7886	1.5192

- The number of quadrature points was not specified, and therefore was selected adaptively, and set equal to only one.
- In order to check whether $Q = 1$ is sufficient, we refitted the model, prespecifying $Q = 20$. This produced essentially the same output.

- Corresponding code for the GLIMMIX procedure is:

```
proc glimmix data=test method=RSPL;
class id trt;
model nseizw = trt*time / dist=poisson solution;
random intercept time / type=UNR subject=id;
estimate 'diff slopes' trt*time 1 -1;
run;
```

29.4 Overview of Epilepsy Data Analyses

- GEE analysis (empirically corrected s.e.; model based s.e.)

Effect	Parameter	Estimate (s.e.)
Common intercept	β_0	1.3140 (0.1435; 0.1601)
Slope placebo	β_1	-0.0142 (0.0234; 0.0185)
Slope treatment	β_2	-0.0192 (0.0178; 0.0174)

- Various GLMM analyses:
 - ▷ MQL
 - ▷ PQL
 - ▷ Laplace
 - ▷ Gaussian quadrature

Effect	Parameter	MQL		PQL	
		Estimate (s.e.)	Estimate (s.e.)	Estimate (s.e.)	Estimate (s.e.)
Common intercept	β_0	1.3525 (0.1492)	0.8079 (0.1261)		
Slope placebo	β_1	-0.0180 (0.0144)	-0.0242 (0.0094)		
Slope treatment	β_2	-0.0151 (0.0144)	-0.0191 (0.0094)		
Variance of intercepts	d_{11}	1.9017 (0.2986)	1.2510 (0.2155)		
Variance of slopes	d_{22}	0.0084 (0.0014)	0.0024 (0.0006)		
Correlation rand.eff.	ρ	-0.3268 (0.1039)	-0.3394 (0.1294)		
Laplace					
Effect	Parameter	Laplace		QUAD	
		Estimate (s.e.)	Estimate (s.e.)	Estimate (s.e.)	Estimate (s.e.)
Common intercept	β_0	0.7740 (0.1291)	0.7739 (0.1293)		
Slope placebo	β_1	-0.0244 (0.0096)	-0.0245 (0.0096)		
Slope treatment	β_2	-0.0193 (0.0096)	-0.0193 (0.0097)		
Variance of intercepts	d_{11}	1.2814 (0.2220)	1.2859 (0.2231)		
Variance of slopes	d_{22}	0.0024 (0.0006)	0.0024 (0.0006)		
Correlation rand.eff.	ρ	-0.3347 (0.1317)	-0.3349 (0.1318)		

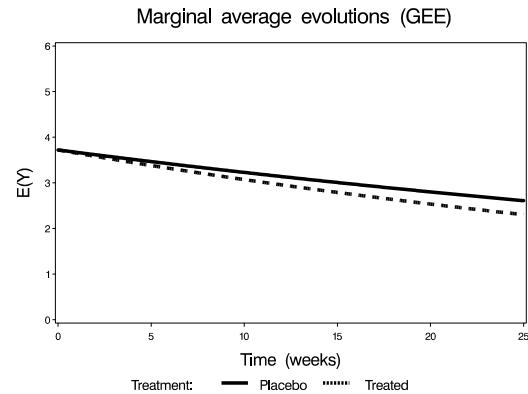
29.5 Marginalization of the Random-effects Model

- Regression coefficients in GLMM need to be interpreted conditionally on the random effects \mathbf{b}_i .
- Additional computations are needed for the population-averaged evolutions.
- The marginal expectation of Y_{ij} measured at t_{ij} in the placebo group is

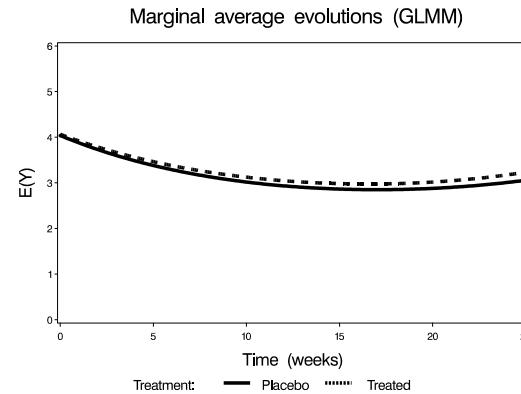
$$\begin{aligned} E[Y_{ij}] &= E[E[Y_{ij}|\mathbf{b}_i]] \\ &= E[\exp[(\beta_0 + b_{i1}) + (\beta_1 + b_{i2})t_{ij}]] \\ &\neq \exp[\beta_0 + \beta_1 t_{ij}] \end{aligned}$$

- Calculations can be done using numerical integration or numerical averaging.
- SAS code and computation: Molenberghs and Verbeke (2005, pp. 343–344)

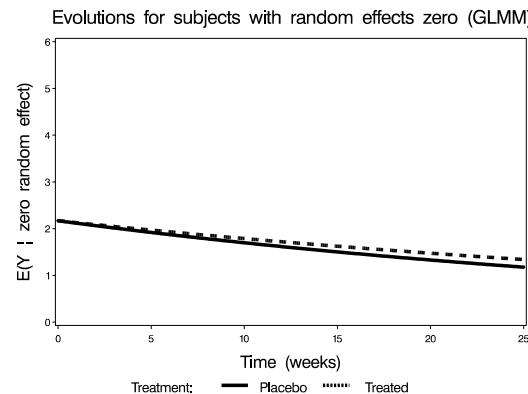
Marginal evolutions (GEE)



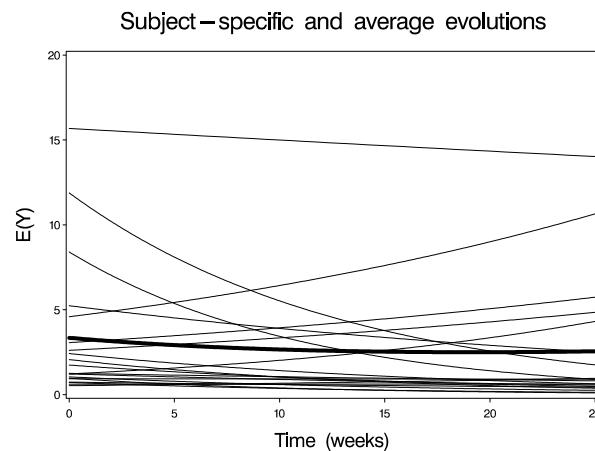
Marginal evolutions (integrated GLMM)



Evolutions average subjects ($b_i = 0$)



Sampled predicted profiles for 20 placebo patients & marginal evolution (bold)



- Curvature different in GEE and GLMM
- Ordering of treatment groups different in GEE and GLMM (although none significant)
- Watch out for the effects of missingness: many patients leave the study after week 16
- The evolution of an ‘average’ patient is completely different from the population-averaged evolution

Part V

Incomplete Data

Chapter 30

Setting The Scene

- ▷ Orthodontic growth data
- ▷ Depression trial
- ▷ Age-related macular degeneration trial
- ▷ Notation
- ▷ Taxonomies

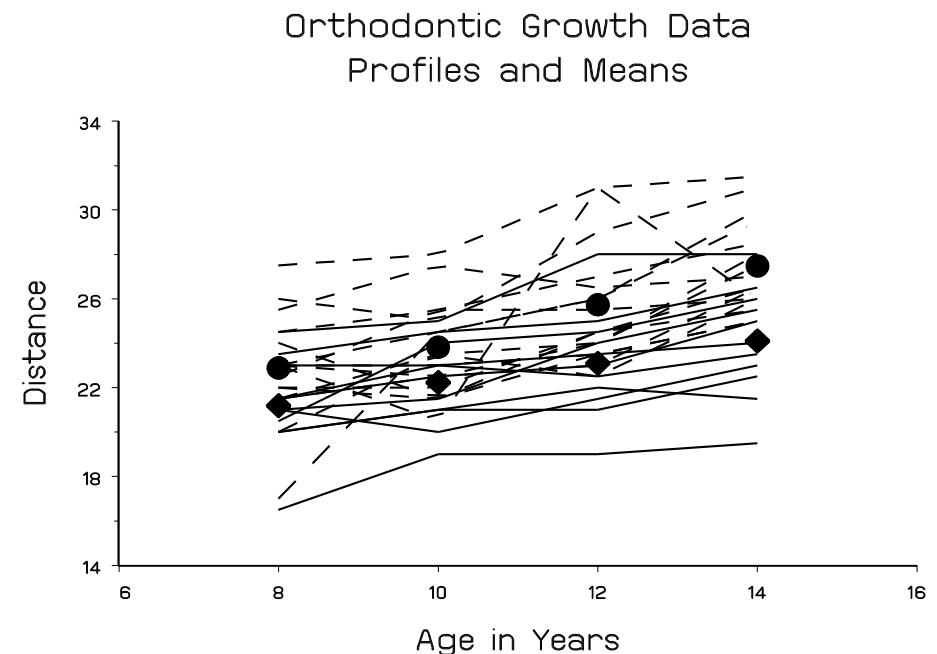
30.1 Growth Data

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

Is dental growth related to gender ?

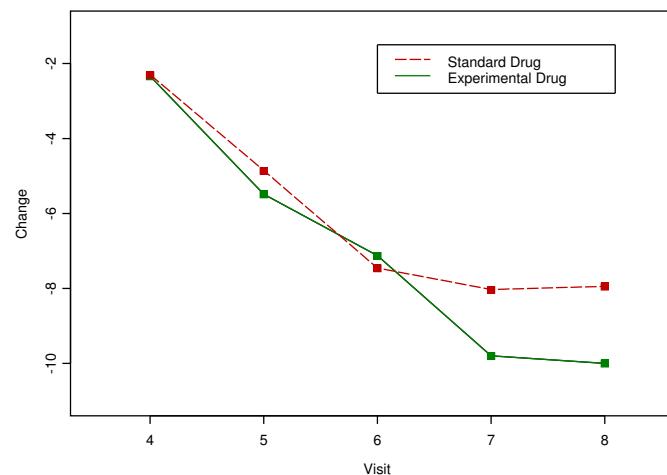
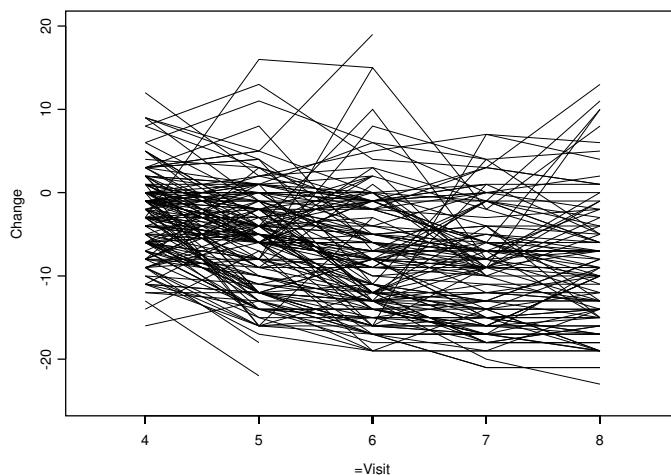
- The distance from the center of the pituitary to the maxillary fissure was recorded at ages 8, 10, 12, and 14, for 11 girls and 16 boys

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



30.2 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



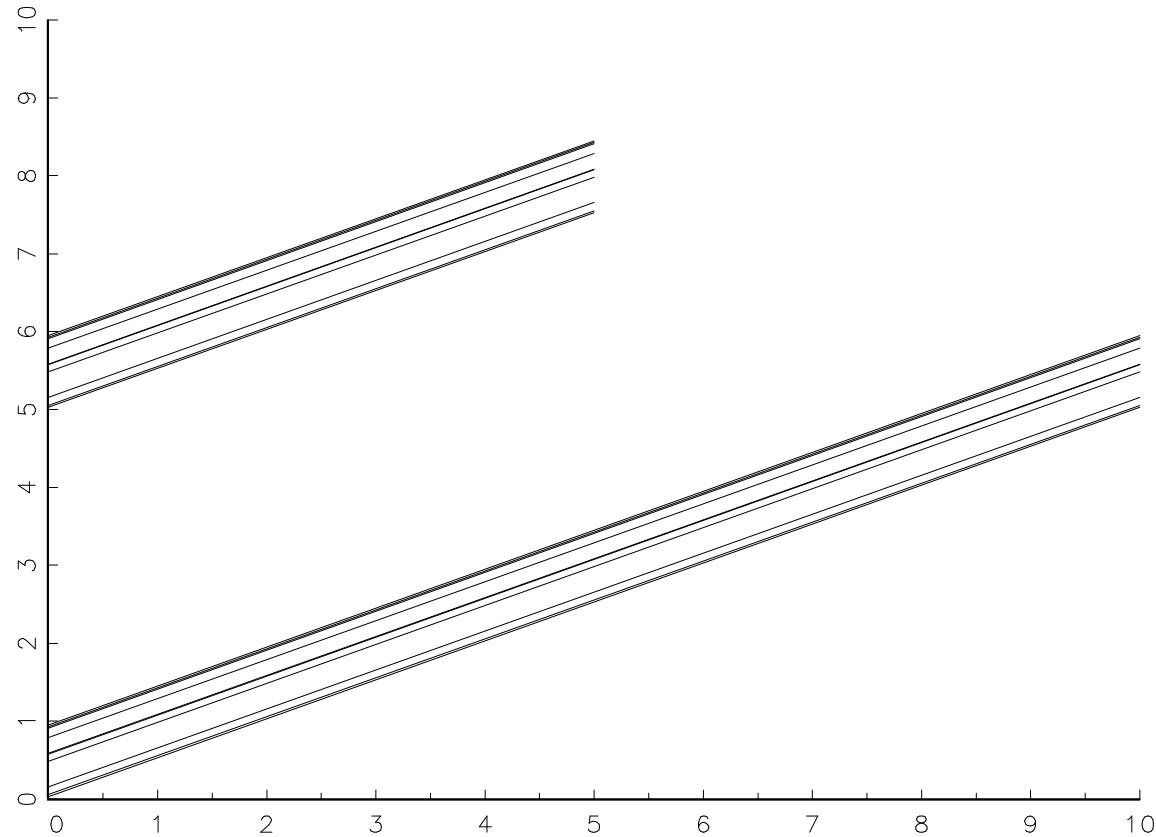
30.3 Age-related Macular Degeneration Trial

- Pharmacological Therapy for Macular Degeneration Study Group (1997)
- An ocular 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				Number	%
4 wks	12 wks	24 wks	52 wks		
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

30.4 Incomplete Longitudinal Data



30.5 Scientific Question

- In terms of **entire longitudinal profile**
- In terms of **last planned measurement**
- In terms of **last observed measurement**

30.6 Notation

- Subject i at occasion (time) $j = 1, \dots, n_i$
- **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 $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})' = (\mathbf{Y}_i^o, \mathbf{Y}_i^m)$
 - \mathbf{Y}_i^o contains Y_{ij} for which $R_{ij} = 1$,
 - \mathbf{Y}_i^m contains Y_{ij} for which $R_{ij} = 0$.
- Group R_{ij} into a vector $\mathbf{R}_i = (R_{i1}, \dots, R_{in_i})'$
- D_i : time of dropout: $D_i = 1 + \sum_{j=1}^{n_i} R_{ij}$

30.7 Framework

$$f(\mathbf{Y}_i, D_i | \boldsymbol{\theta}, \boldsymbol{\psi})$$

Selection Models: $f(\mathbf{Y}_i | \boldsymbol{\theta})$

$$f(D_i | \mathbf{Y}_i^o, \mathbf{Y}_i^m, \boldsymbol{\psi})$$

MCAR



MAR



MNAR

$$f(D_i | \boldsymbol{\psi})$$

$$f(D_i | \mathbf{Y}_i^o, \boldsymbol{\psi})$$

$$f(D_i | \mathbf{Y}_i^o, \mathbf{Y}_i^m, \boldsymbol{\psi})$$

Pattern-Mixture Models: $f(\mathbf{Y}_i | D_i, \boldsymbol{\theta}) f(D_i | \boldsymbol{\psi})$

Shared-Parameter Models: $f(\mathbf{Y}_i | \mathbf{b}_i, \boldsymbol{\theta}) f(D_i | \mathbf{b}_i, \boldsymbol{\psi})$

$$f(\mathbf{Y}_i, D_i | \boldsymbol{\theta}, \boldsymbol{\psi})$$

Selection Models: $f(\mathbf{Y}_i | \boldsymbol{\theta})$ $f(D_i | \mathbf{Y}_i^o, \mathbf{Y}_i^m, \boldsymbol{\psi})$



CC ?

direct likelihood !

joint model !?

LOCF ?

expectation-maximization (EM).

sensitivity analysis ?!

imputation ?

multiple imputation (MI).

:

(weighted) GEE !

30.8 Selection Models versus Pattern-mixture Models: A Paradox!?

- Glynn, Laird and Rubin (1986)
- Two measurements (Y_1, Y_2)
- Y_1 always observed.
- Y_2 observed ($R = 1$) or missing ($R = 0$).

- Selection model versus pattern-mixture model

$$f(y_1, y_2)g(r = 1|y_1, y_2) = f_1(y_1, y_2)p(r = 1)$$

$$f(y_1, y_2)g(r = 0|y_1, y_2) = f_0(y_1, y_2)p(r = 0)$$

or

$$f(y_1, y_2)g(y_1, y_2) = f_1(y_1, y_2)p$$

$$f(y_1, y_2)[1 - g(y_1, y_2)] = f_0(y_1, y_2)[1 - p]$$

of which the ratio yields:

$$f_0(y_1, y_2) = \frac{1 - g(y_1, y_2)}{g(y_1, y_2)} \cdot \frac{p}{1 - p} \cdot f_1(y_1, y_2)$$

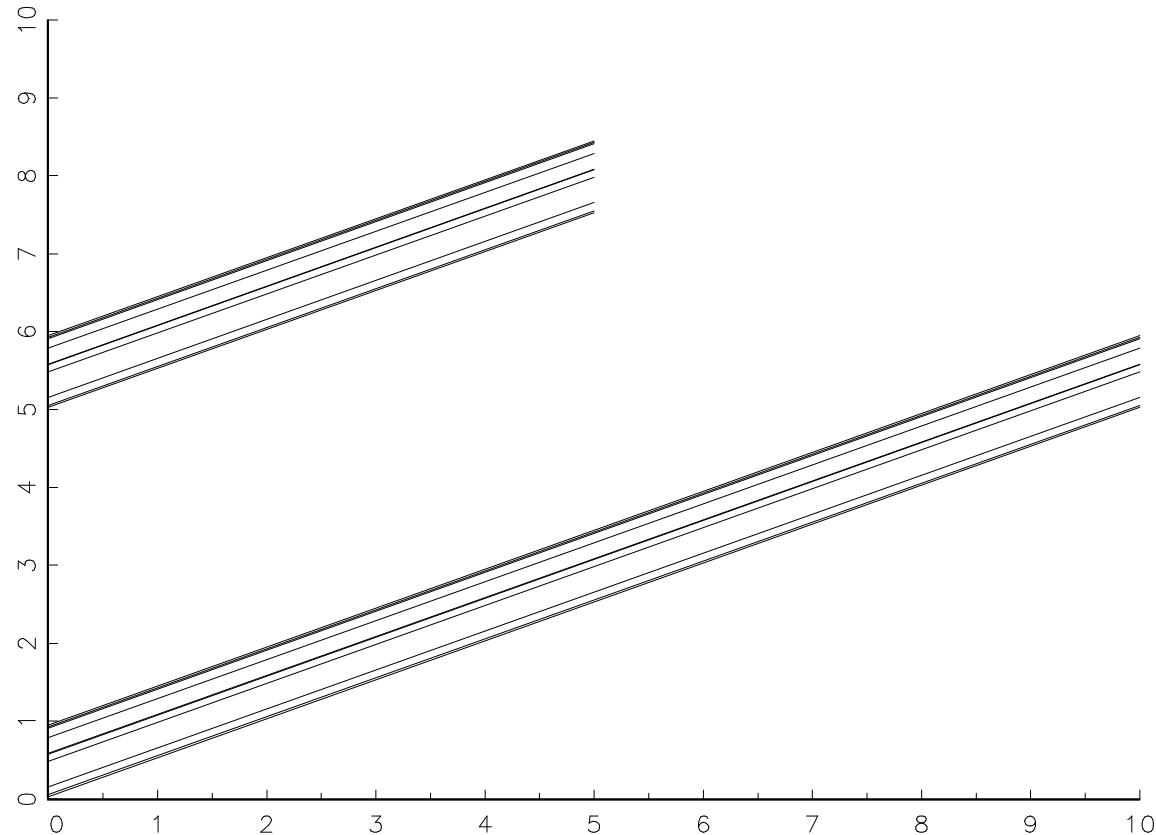
- The right hand side is identifiable \longleftrightarrow the left hand side is not...

Chapter 31

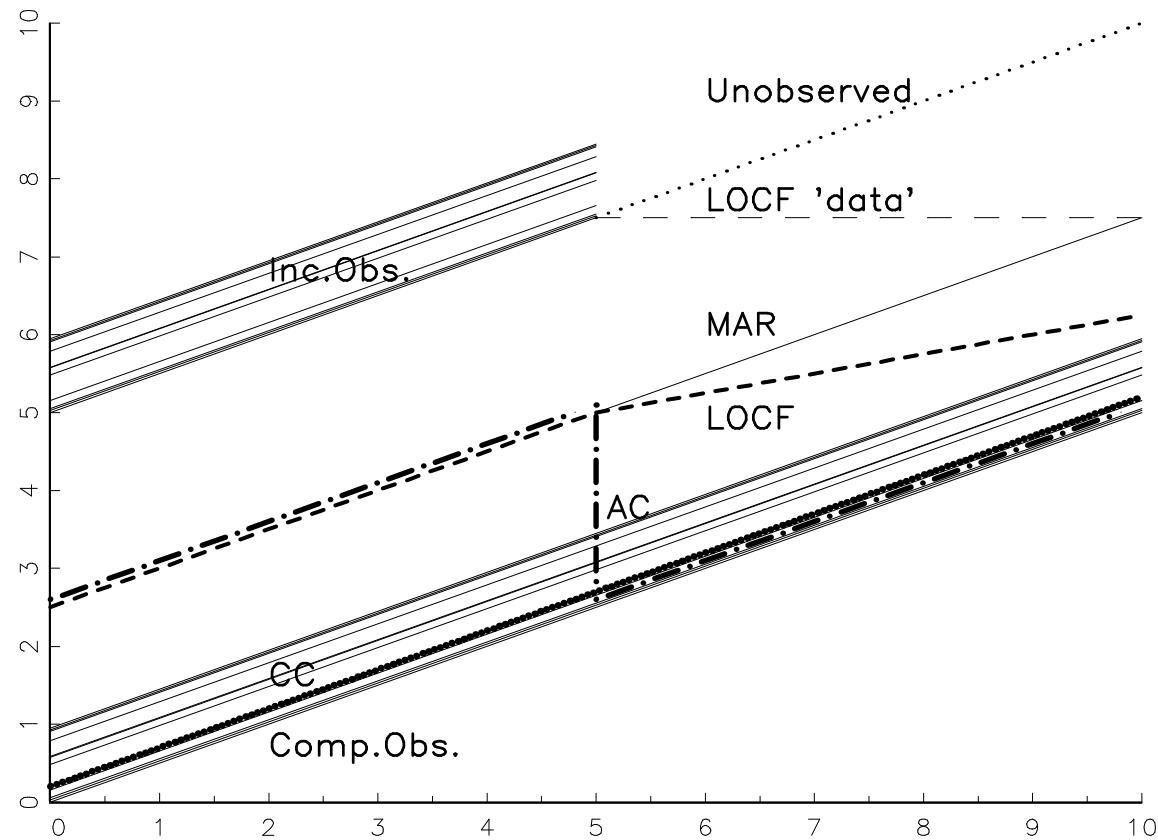
Proper Analysis of Incomplete Data

- ▷ Simple methods
- ▷ Bias for LOCF and CC
- ▷ Direct likelihood inference
- ▷ Weighted generalized estimating equations

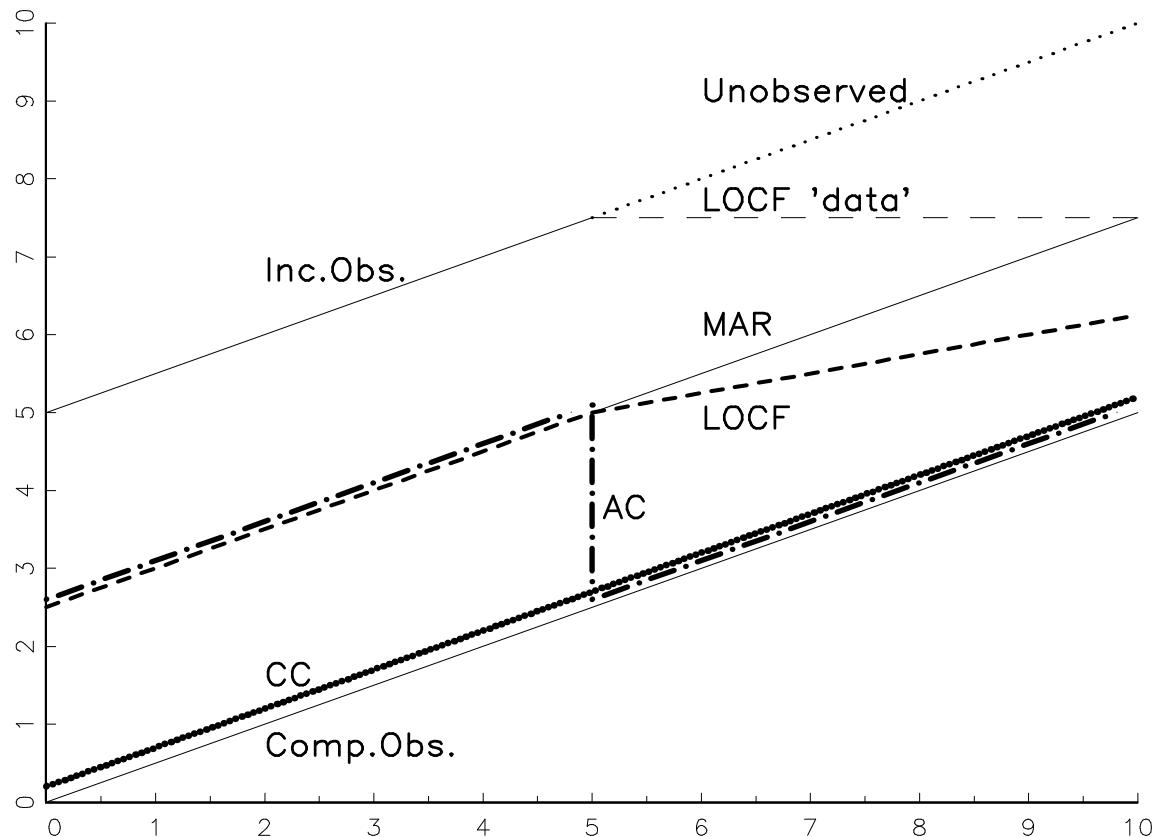
31.1 Incomplete Longitudinal Data



Data and Modeling Strategies



Modeling Strategies



31.2 Simple Methods

MCAR

Complete case analysis:

⇒ **delete** incomplete subjects

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

Last observation carried forward:

⇒ **impute** missing values

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

Quantifying the Bias

Dropouts $t_{ij} = 0$

Probability p_0

Treatment indicator $T_i = 0, 1$

$$E(Y_{ij}) = \beta_0 + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij}$$

Completers $t_{ij} = 0, 1$

Probability $1 - p_0 = p_1$

Treatment indicator $T_i = 0, 1$

$$E(Y_{ij}) = \gamma_0 + \gamma_1 T_i + \gamma_2 t_{ij} + \gamma_3 T_i t_{ij}$$

	CC	LOCF
MCAR	0	$(p_1 - p_0)\beta_2 - (1 - p_1)\beta_3$
MAR	$-\sigma[(1 - p_1)(\beta_0 + \beta_1 - \gamma_0 - \gamma_1) - (1 - p_0)(\beta_0 - \gamma_0)]$	$p_1(\gamma_0 + \gamma_1 + \gamma_2 + \gamma_3) + (1 - p_1)(\beta_0 + \beta_1)$ $-p_0(\gamma_0 + \gamma_2) - (1 - p_0)\beta_0 - \gamma_1 - \gamma_3$ $-\sigma[(1 - p_1)(\beta_0 + \beta_1 - \gamma_0 - \gamma_1) - (1 - p_0)(\beta_0 - \gamma_0)]$

31.3 Ignorability

- Let us decide to use likelihood based estimation.
- The full data likelihood contribution for subject i :

$$L^*(\boldsymbol{\theta}, \boldsymbol{\psi} | \mathbf{Y}_i, D_i) \propto f(\mathbf{Y}_i, D_i | \boldsymbol{\theta}, \boldsymbol{\psi}).$$

- Base inference on the observed data:

$$L(\boldsymbol{\theta}, \boldsymbol{\psi} | \mathbf{Y}_i, D_i) \propto f(\mathbf{Y}_i^o, D_i | \boldsymbol{\theta}, \boldsymbol{\psi})$$

with

$$\begin{aligned} f(\mathbf{Y}_i^o, D_i | \boldsymbol{\theta}, \boldsymbol{\psi}) &= \int f(\mathbf{Y}_i, D_i | \boldsymbol{\theta}, \boldsymbol{\psi}) d\mathbf{Y}_i^m \\ &= \int f(\mathbf{Y}_i^o, \mathbf{Y}_i^m | \boldsymbol{\theta}) f(D_i | \mathbf{Y}_i^o, \mathbf{Y}_i^m, \boldsymbol{\psi}) d\mathbf{Y}_i^m. \end{aligned}$$

- Under a MAR process:

$$\begin{aligned} f(\mathbf{Y}_i^o, D_i | \boldsymbol{\theta}, \boldsymbol{\psi}) &= \int f(\mathbf{Y}_i^o, \mathbf{Y}_i^m | \boldsymbol{\theta}) f(D_i | \mathbf{Y}_i^o, \boldsymbol{\psi}) d\mathbf{Y}_i^m \\ &= f(\mathbf{Y}_i^o | \boldsymbol{\theta}) f(D_i | \mathbf{Y}_i^o, \boldsymbol{\psi}), \end{aligned}$$

- The likelihood factorizes into two components.

31.3.1 Ignorability: Summary

Likelihood/Bayesian + MAR

&

Frequentist + MCAR

31.4 Direct Likelihood Maximization

$$\text{MAR} : f(\mathbf{Y}_i^o | \boldsymbol{\theta}) \cancel{f(D_i | \mathbf{Y}_i^o, \boldsymbol{\psi})}$$

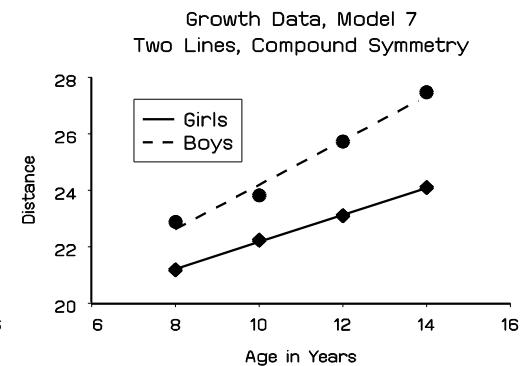
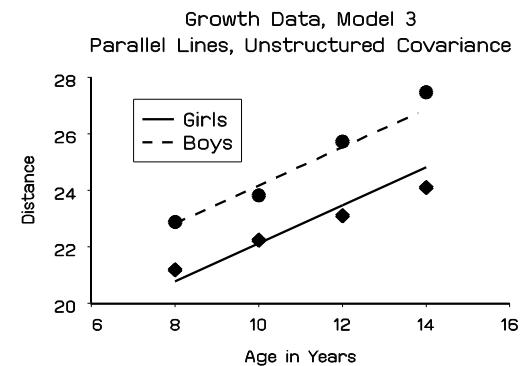
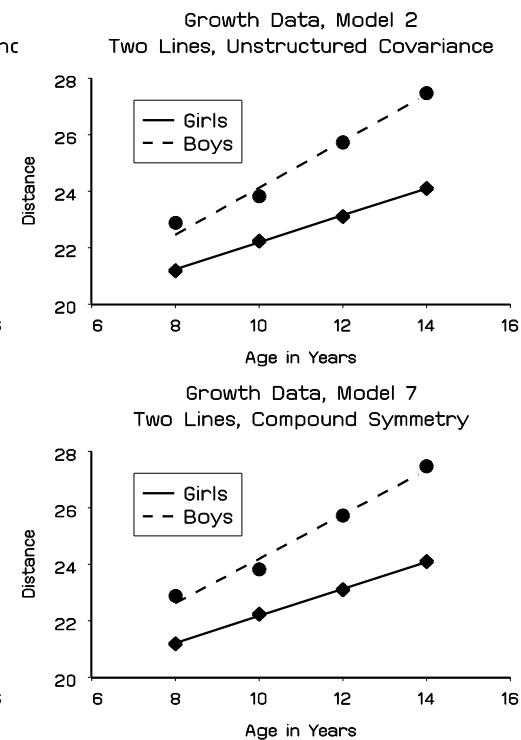
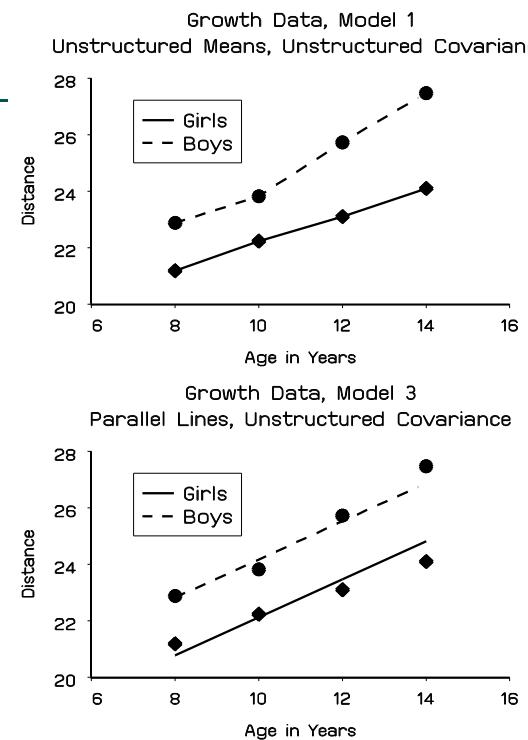
Mechanism is MAR
 $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ distinct
Interest in $\boldsymbol{\theta}$
Use observed information matrix

\Rightarrow Likelihood inference is valid

Outcome type	Modeling strategy	Software
Gaussian	Linear mixed model	SAS proc MIXED
Non-Gaussian	Generalized linear mixed model	SAS proc GLIMMIX, NLMIXED

31.4.1 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



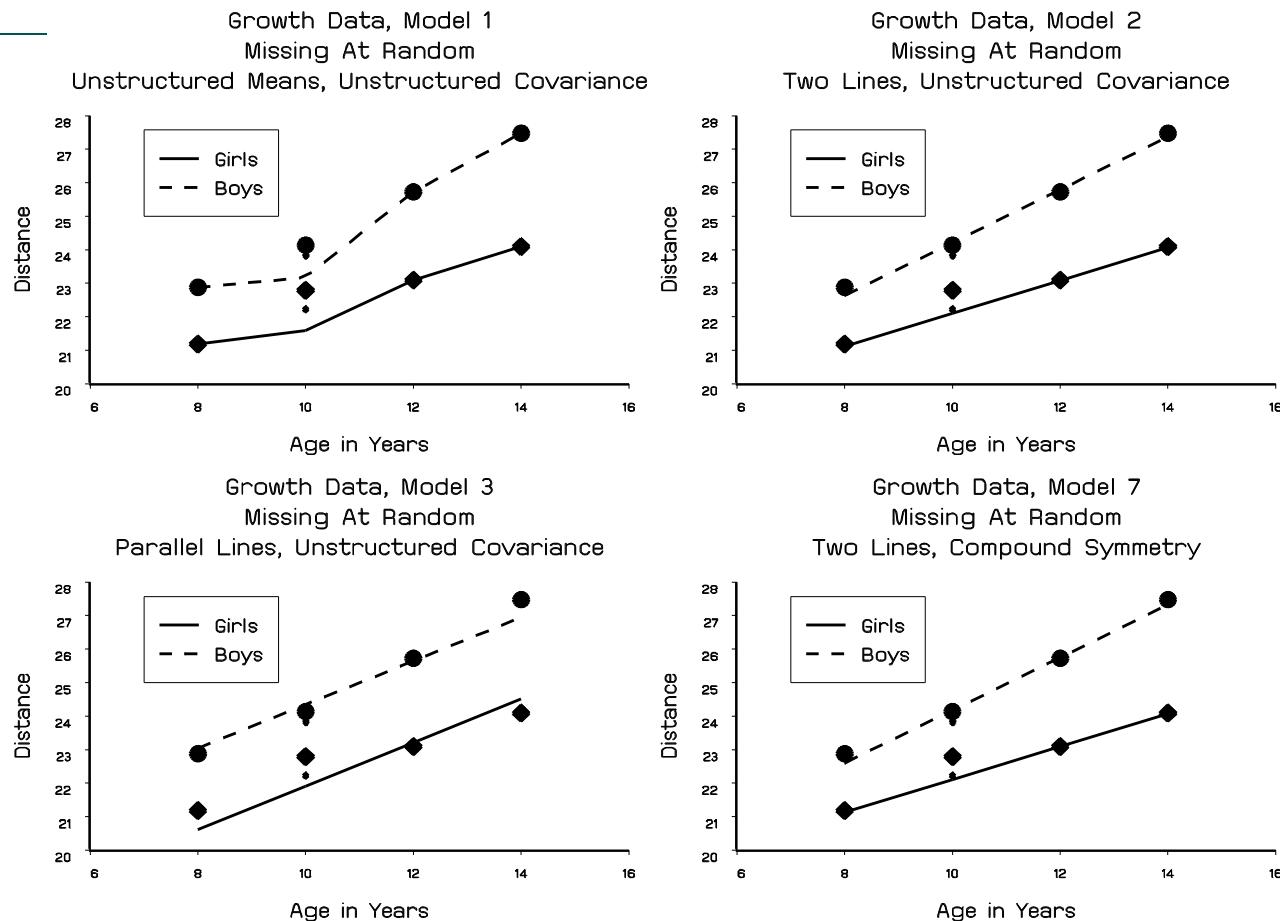
31.4.2 Trimmed Growth Data: Simple Methods

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

distorting

31.4.3 Trimmed Growth Data: Direct Likelihood

Mean	Covar	# par
7	\neq slopes	CS
7	6	

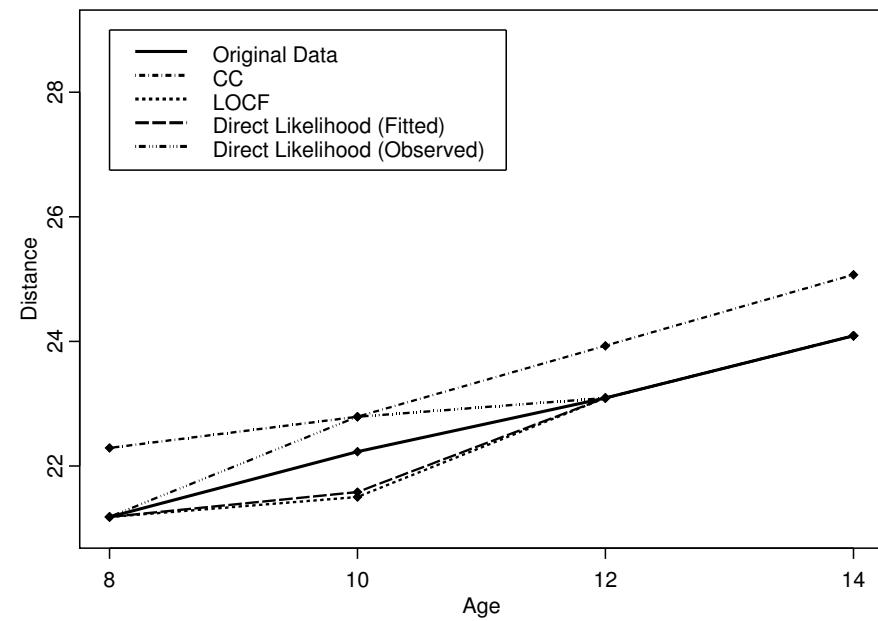
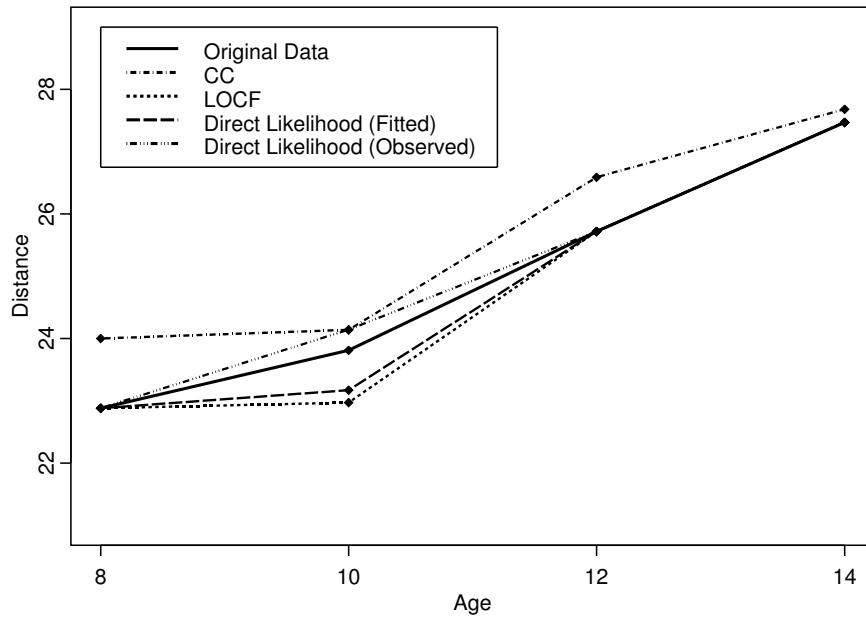


31.4.4 Growth Data: Comparison of Analyses

- Data
 - ▷ Complete cases
 - ▷ LOCF imputed data
 - ▷ All available data
- Model
 - ▷ Unstructured group by time mean
 - ▷ Unstructured covariance matrix
- Analysis methods
 - ▷ Direct likelihood
 - * ML
 - * REML
 - ▷ MANOVA
 - ▷ ANOVA per time point

Principle	Method	Boys at Age 8	Boys at Age 10
Original	<i>Direct likelihood, ML</i>	22.88 (0.56)	23.81 (0.49)
	<i>Direct likelihood, REML \equiv MANOVA</i>	22.88 (0.58)	23.81 (0.51)
	<i>ANOVA per time point</i>	22.88 (0.61)	23.81 (0.53)
Direct Lik.	Direct likelihood, ML	22.88 (0.56)	23.17 (0.68)
	Direct likelihood, REML	22.88 (0.58)	23.17 (0.71)
	MANOVA	24.00 (0.48)	24.14 (0.66)
	ANOVA per time point	22.88 (0.61)	24.14 (0.74)
CC	Direct likelihood, ML	24.00 (0.45)	24.14 (0.62)
	Direct likelihood, REML \equiv MANOVA	24.00 (0.48)	24.14 (0.66)
	ANOVA per time point	24.00 (0.51)	24.14 (0.74)
LOCF	Direct likelihood, ML	22.88 (0.56)	22.97 (0.65)
	Direct likelihood, REML \equiv MANOVA	22.88 (0.58)	22.97 (0.68)
	ANOVA per time point	22.88 (0.61)	22.97 (0.72)

31.4.5 Growth Data: Graphical Comparison of Analyses



31.4.6 Behind the Scenes

- R completers $\leftrightarrow N - R$ “incompleters”

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ & \sigma_{22} \end{pmatrix} \right)$$

- Conditional density

$$Y_{i2}|y_{i1} \sim N(\beta_0 + \beta_1 y_{i1}, \sigma_{22.1})$$

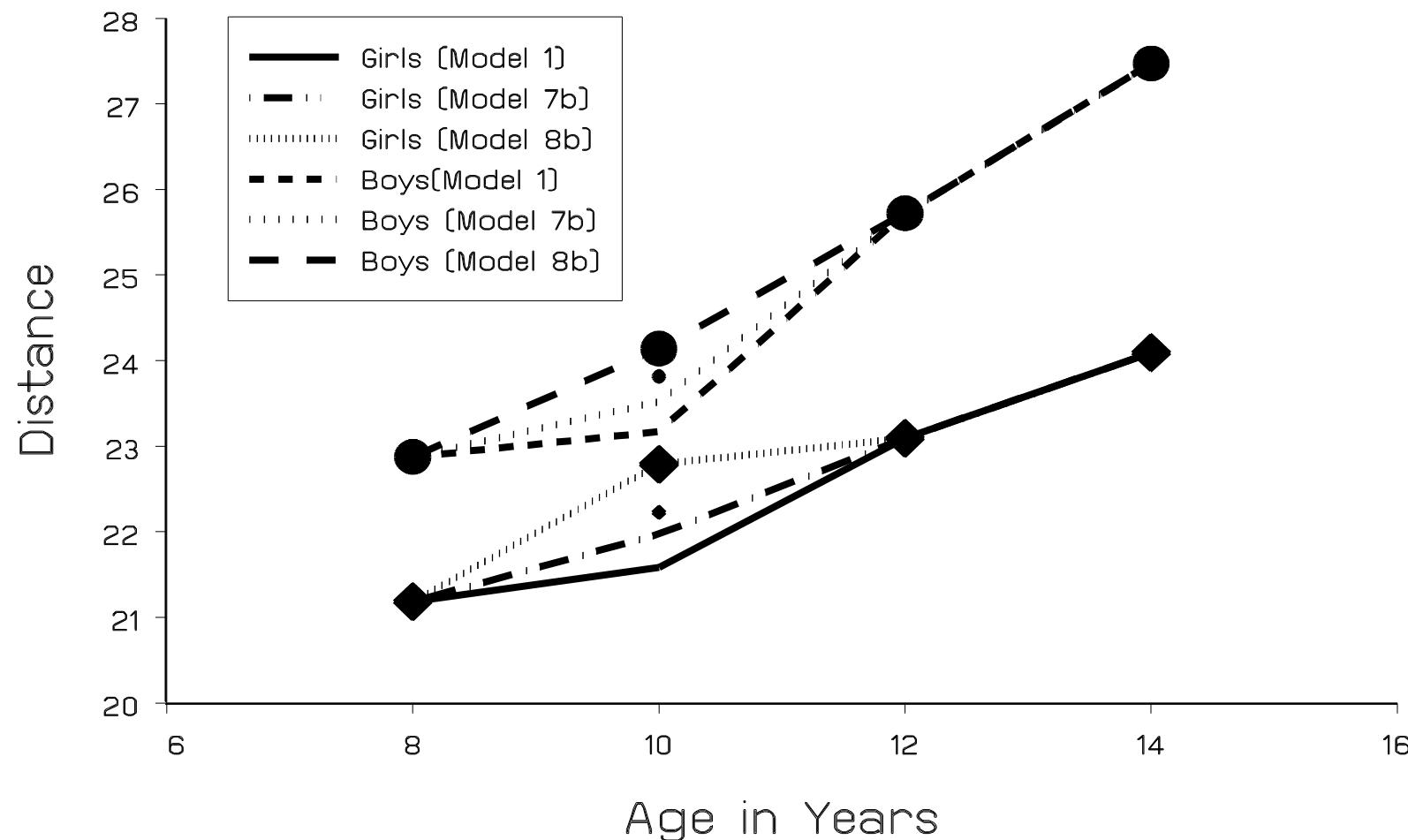
μ_1	freq. & lik.	$\widehat{\mu}_1 = \frac{1}{N} \sum_{i=1}^N y_{i1}$
μ_2	frequentist	$\widetilde{\mu}_2 = \frac{1}{R} \sum_{i=1}^R y_{i2}$
μ_2	likelihood	$\widehat{\mu}_2 = \frac{1}{N} \left\{ \sum_{i=1}^R y_{i2} + \sum_{i=R+1}^N [\bar{y}_2 + \widehat{\beta}_1(y_{i1} - \bar{y}_1)] \right\}$

31.4.7 Growth Data: Further Comparison of Analyses

Principle	Method	Boys at Age 8	Boys at Age 10
Original	<i>Direct likelihood, ML</i>	22.88 (0.56)	23.81 (0.49)
Direct Lik.	Direct likelihood, ML	22.88 (0.56)	23.17 (0.68)
CC	Direct likelihood, ML	24.00 (0.45)	24.14 (0.62)
LOCF	Direct likelihood, ML	22.88 (0.56)	22.97 (0.65)

Data	Mean	Covariance	Boys at Age 8	Boys at Age 10
Complete	Unstructured	Unstructured	22.88	23.81
	Unstructured	CS	22.88	23.81
	Unstructured	Independence	22.88	23.81
Incomplete	Unstructured	Unstructured	22.88	23.17
	Unstructured	CS	22.88	23.52
	Unstructured	Independence	22.88	24.14

Growth Data, Models 1, 7b, and 8b Missing At Random



31.4.8 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!

31.4.9 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

31.5 Analysis of the Depression Trial

- Complete case analysis:

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

⇒ performs analysis on CC data set

- LOCF analysis:

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

⇒ performs analysis on LOCF data

- Direct-likelihood analysis: ⇒ 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 32

Weighted Generalized Estimating Equations

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

32.1 General Principle

MAR and non-ignorable !

- Standard GEE inference correct only under MCAR

Robins, Rotnitzky & Zhao (JASA, 1995)

- Under MAR: *weighted* GEE

Fitzmaurice, Molenberghs & Lipsitz (JRSSB, 1995)

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

- Weigh a contribution by inverse dropout probability

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

- Adjust estimating equations

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

32.2 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 occasion, 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

32.3 Analysis of the Analgesic Trial

- A logistic regression for the dropout indicator:

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

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	Par.	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

- Working correlation matrices:

$$R_{\text{UN, GEE}} = \begin{pmatrix} 1 & 0.173 & 0.246 & 0.201 \\ & 1 & 0.177 & 0.113 \\ & & 1 & 0.456 \\ & & & 1 \end{pmatrix}$$

$$R_{\text{UN, WGEE}} = \begin{pmatrix} 1 & 0.215 & 0.253 & 0.167 \\ & 1 & 0.196 & 0.113 \\ & & 1 & 0.409 \\ & & & 1 \end{pmatrix}$$

32.4 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;
```

32.5 Analysis of the ARMD Trial

- Model for the weights:

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

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 $k = 1, \dots, 4$ (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 ($j = 1, \dots, 4$) refers to the four follow-up measurements
- ▷ Classical GEE and linearization-based GEE
- ▷ Comparison between CC, LOCF, and GEE analyses

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

- Results:

Effect	Par.	CC	LOCF	Observed data	
				Unweighted	WGEE
Standard GEE					
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

Effect	Par.	CC	LOCF	Observed data	
				Unweighted	WGEE
Linearization-based GEE					
Int.4	β_{11}	-1.01(0.24;0.24)	-0.87(0.21;0.21)	-0.87(0.21;0.21)	-0.98(0.18;0.44)
Int.12	β_{21}	-0.89(0.24;0.24)	-0.97(0.21;0.21)	-1.01(0.22;0.21)	-1.78(0.26;0.42)
Int.24	β_{31}	-1.13(0.25;0.25)	-1.05(0.21;0.21)	-1.07(0.23;0.22)	-1.19(0.25;0.38)
Int.52	β_{41}	-1.64(0.29;0.29)	-1.51(0.24;0.24)	-1.71(0.29;0.29)	-1.81(0.39;0.48)
Tr.4	β_{12}	0.40(0.32;0.32)	0.22(0.28;0.28)	0.22(0.29;0.29)	0.80(0.26;0.67)
Tr.12	β_{22}	0.49(0.31;0.31)	0.55(0.28;0.28)	0.61(0.28;0.29)	1.85(0.32;0.64)
Tr.24	β_{32}	0.48(0.33;0.33)	0.42(0.29;0.29)	0.44(0.30;0.30)	0.98(0.33;0.60)
Tr.52	β_{42}	0.40(0.38;0.38)	0.34(0.32;0.32)	0.44(0.37;0.37)	0.97(0.49;0.65)
	σ^2	0.62	0.57	0.62	1.29
	τ^2	0.39	0.44	0.39	1.85
Corr.	ρ	0.39	0.44	0.39	0.59

32.6 Analysis of the Depression Trial

- Response: create binary indicator y_{bin} 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:

Effect	WGEE			GEE		
	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 33

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

33.1 General Principles

- Valid under MAR
- An alternative to direct likelihood and WGEE
- Three steps:
 1. The missing values are filled in M times $\Rightarrow M$ 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)

33.1.1 Informal Justification

- We need to estimate θ from the data (e.g., from the complete cases)
- Plug in the estimated $\hat{\theta}$ and use

$$f(\mathbf{y}_i^m | \mathbf{y}_i^o, \hat{\theta})$$

to impute the missing data.

- We need to acknowledge that $\hat{\theta}$ is a random variable; its uncertainty needs to be included in the imputation process
- Given this distribution we:
 - ▷ draw a random θ^* from the distribution of $\hat{\theta}$
 - ▷ put this θ^* in to draw a random \mathbf{Y}_i^m from
$$f(\mathbf{y}_i^m | \mathbf{y}_i^o, \theta^*).$$

33.1.2 The Algorithm

1. Draw $\boldsymbol{\theta}^*$ from its posterior distribution
2. Draw \mathbf{Y}_i^{m*} from $f(\mathbf{y}_i^m | \mathbf{y}_i^o, \boldsymbol{\theta}^*)$.
3. To estimate $\boldsymbol{\beta}$, then calculate the estimate of the parameter of interest, and its estimated variance, using the completed data, $(\mathbf{Y}^o, \mathbf{Y}^{m*})$:

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\mathbf{Y}) = \hat{\boldsymbol{\beta}}(\mathbf{Y}^o, \mathbf{Y}^{m*})$$

The *within* imputation variance is

$$U = \widehat{\text{Var}}(\hat{\boldsymbol{\beta}})$$

4. Repeat steps 1, 2 and 3 a number of M times

$$\Rightarrow \hat{\boldsymbol{\beta}}^m \quad \& \quad U^m \quad (m = 1, \dots, M)$$

33.1.3 Pooling Information

- With M imputations, the estimate of β is

$$\hat{\beta}^* = \frac{\sum_{m=1}^M \hat{\beta}^m}{M}.$$

- Further, one can make normally based inferences for β with

$$(\beta - \hat{\beta}^*) \sim N(\mathbf{0}, V),$$

where

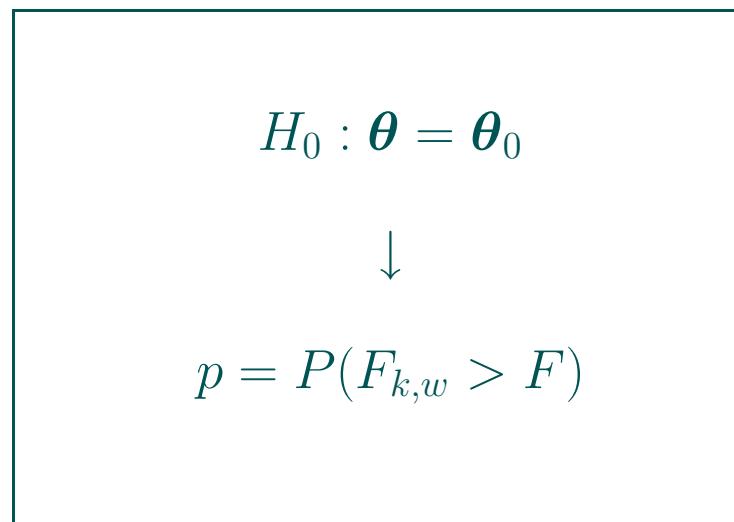
total: $V = W + \left(\frac{M+1}{M} \right) B$

within: $W = \frac{\sum_{m=1}^M \mathbf{U}^m}{M}$

between: $B = \frac{\sum_{m=1}^M (\hat{\beta}^m - \hat{\beta}^*)(\hat{\beta}^m - \hat{\beta}^*)'}{M-1}$

33.1.4 Hypothesis Testing

- Two “sample sizes”:
 - ▷ N : The sample size of the data set
 - ▷ M : The number of imputations
- Both play a role in the asymptotic distribution (Li, Raghunathan, and Rubin 1991)



where

k : length of the parameter vector $\boldsymbol{\theta}$

$$F_{k,w} \sim F$$

$$F = \frac{(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0)'W^{-1}(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0)}{k(1+r)}$$

$$w = 4 + (\tau - 4) \left[1 + \frac{(1 - 2\tau^{-1})}{r} \right]^2$$

$$r = \frac{1}{k} \left(1 + \frac{1}{M} \right) \text{tr}(BW^{-1})$$

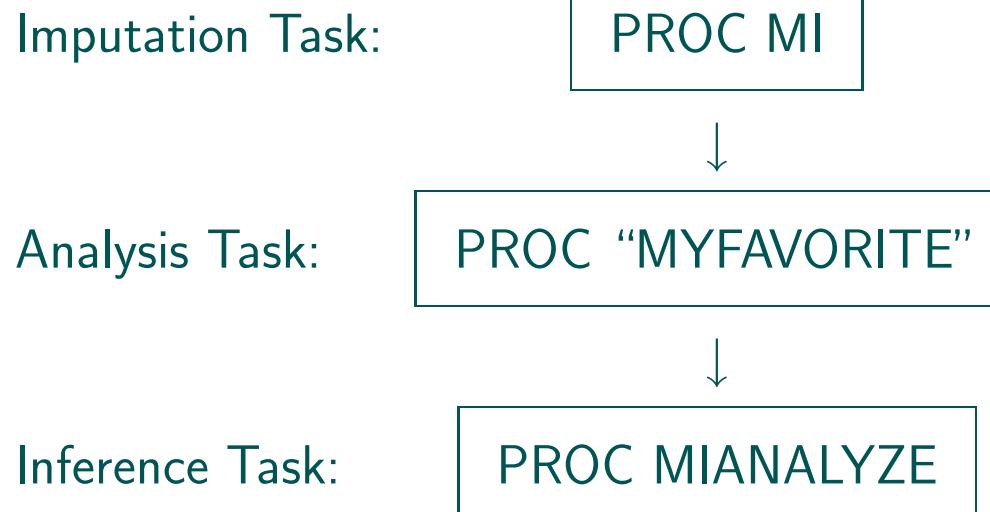
$$\tau = k(M-1)$$

- Limiting behavior:

$$F \xrightarrow{M \rightarrow \infty} F_{k,\infty} = \chi^2/k$$

33.2 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:



33.2.1 MI Analysis of the Orthodontic Growth Data

- The same Model 1 as before
- Focus on boys at ages 8 and 10
- Results

	Boys at Age 8	Boys at Age 10
Original Data	22.88 (0.58)	23.81 (0.51)
Multiple Imputation	22.88 (0.66)	22.69 (0.81)

- Between-imputation variability for age 10 measurement
- Confidence interval for Boys at age 10: [21.08,24.29]

33.3 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:

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

- $T_i = 0$ for placebo and $T_i = 1$ for interferon- α
- t_j ($j = 1, \dots, 4$) 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)

33.4 SAS Code for MI

1. Preparatory data analysis so that there is one line per subject
2. The imputation task:

```
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 analysis task (GEE):

```
proc genmod 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 ParameterEstimates=gmparms parminfo=gmpinfo CovB=gmcovb;
run;
```

5. 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=nlparsma CovMatAddEst=nlcovba;
run;
```

6. 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;
```

7. The inference task (GLMM):

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

Chapter 34

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

Part VI

Topics in Methods and Sensitivity Analysis for Incomplete Data

Chapter 35

An MNAR Selection Model and Local Influence

- ▷ The Diggle and Kenward selection model
- ▷ Mastitis in dairy cattle
- ▷ An informal sensitivity analysis
- ▷ Local influence to conduct sensitivity analysis

35.1 A Full Selection Model

MNAR

$$: \int f(\mathbf{Y}_i | \boldsymbol{\theta}) f(D_i | \mathbf{Y}_i, \boldsymbol{\psi}) d\mathbf{Y}_i^m$$

$$f(\mathbf{Y}_i | \boldsymbol{\theta})$$

Linear mixed model

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i$$

$$f(D_i | \mathbf{Y}_i, \boldsymbol{\psi})$$

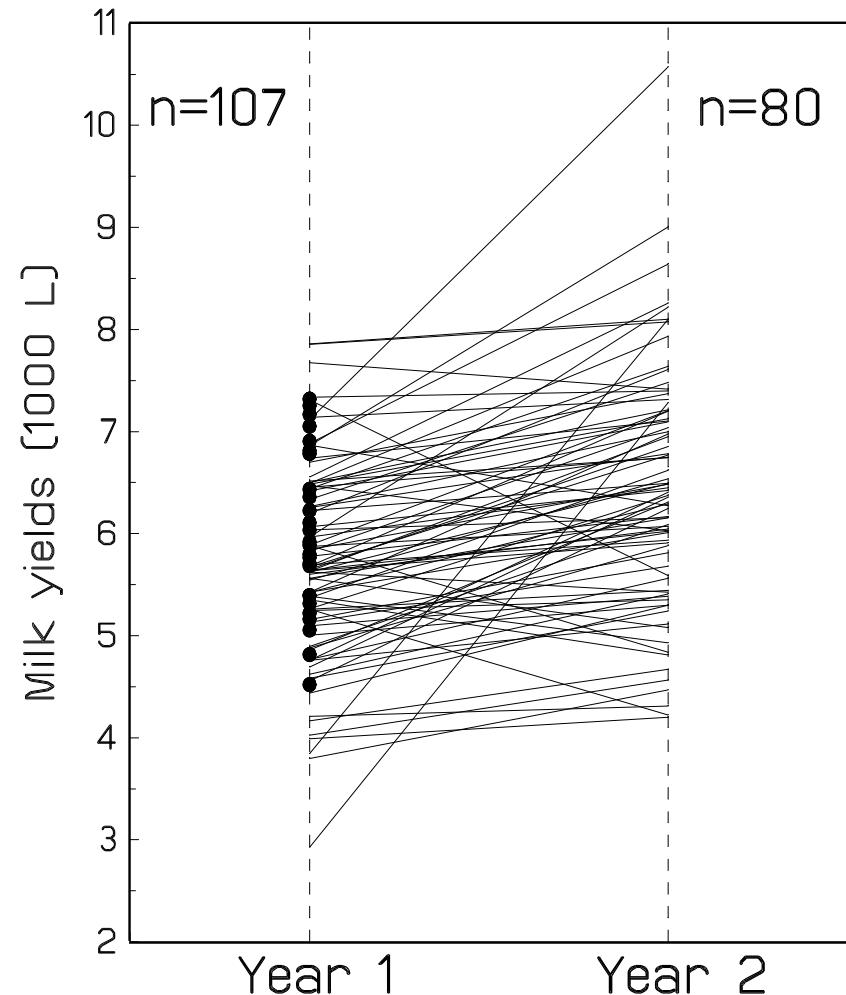
Logistic regressions for dropout

$$\begin{aligned} \text{logit} [P(D_i = j \mid D_i \geq j, Y_{i,j-1}, Y_{ij})] \\ = \psi_0 + \psi_1 Y_{i,j-1} + \psi_2 Y_{ij} \end{aligned}$$

Diggle and Kenward (JRSSC 1994)

35.2 Mastitis in Dairy Cattle

- Infectious disease of the udder
- Leads to a reduction in milk yield
- High yielding cows more susceptible?
- **But** this cannot be measured directly because of the effect of the disease: *evidence is missing* since infected cause have no reported milk yield



- Model for milk yield:

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} \sim N \left[\begin{pmatrix} \mu \\ \mu + \Delta \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} \right]$$

- Model for mastitis:

$$\begin{aligned} \text{logit}[P(R_i = 1 | Y_{i1}, Y_{i2})] &= \psi_0 + \psi_1 Y_{i1} + \psi_2 Y_{i2} \\ &= 0.37 + 2.25Y_{i1} - 2.54Y_{i2} \\ &= 0.37 - 0.29Y_{i1} - 2.54(Y_{i2} - Y_{i1}) \end{aligned}$$

- LR test for $H_0 : \psi_2 = 0$: $G^2 = 5.11$

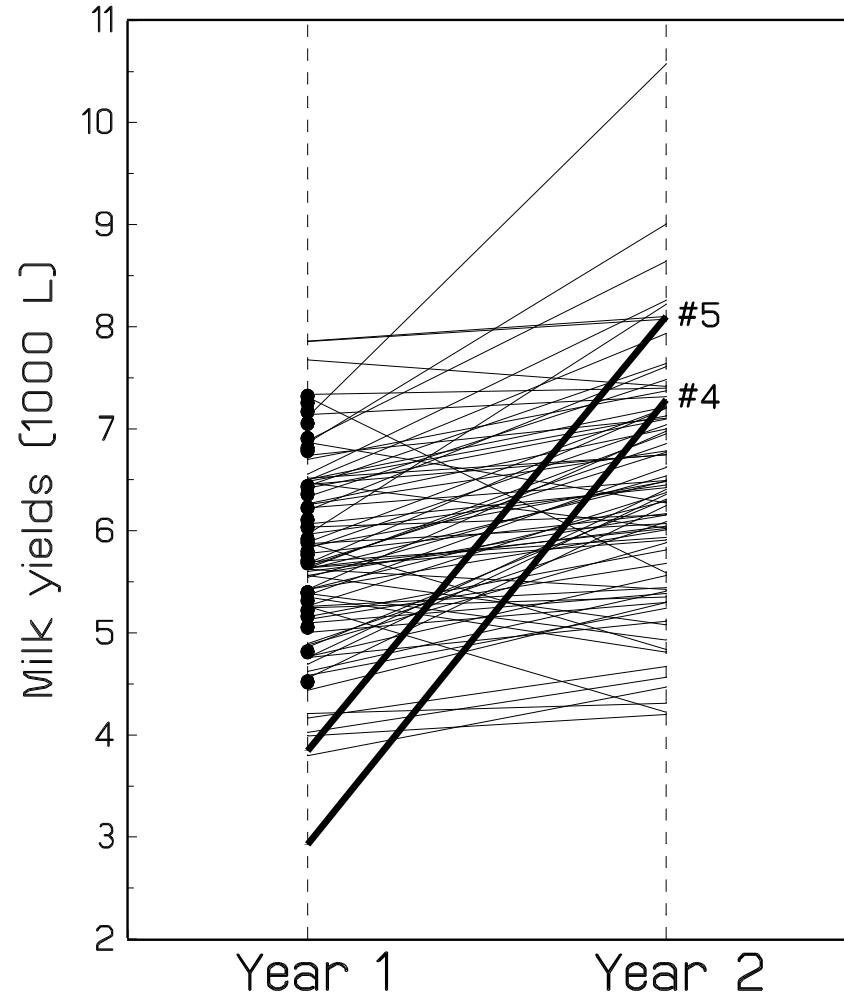
35.3 Criticism → Sensitivity Analysis

“..., estimating the ‘unestimable’ can be accomplished only by making modelling **assumptions**,.... The consequences of model **misspecification** will (...) be more severe in the non-random case.” (Laird 1994)

- Change distributional assumptions (Kenward 1998)
- Local and global influence methods
- Pattern-mixture models
- Several plausible models or ranges of inferences
- Semi-parametric framework (Scharfstein *et al* 1999)

35.4 Kenward's Sensitivity Analysis

- Deletion of #4 and #5 $\Rightarrow G^2$ for ψ_2 :
 $5.11 \rightarrow 0.08$
- Cows #4 and #5 have unusually large increments
- Kenward conjectures: #4 and #5 ill during the first year
- Kenward (SiM 1998)



35.5 Local Influence

- Verbeke, Thijs, Lesaffre, Kenward (Bcs 2001)

- Perturbed MAR dropout model:

$$\text{logit} [P(D_i = 1 | Y_{i1}, Y_{i2})]$$

$$= \psi_0 + \psi_1 Y_{i1} + \boxed{\omega_i Y_{i2}}$$

- Likelihood displacement:

$$LD(\boldsymbol{\omega}) = 2 [L_{\omega=0}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\psi}}) - L_{\omega=0}(\widehat{\boldsymbol{\theta}}_\omega, \widehat{\boldsymbol{\psi}}_\omega)] \geq 0$$

35.5 Local Influence

- Verbeke, Thijs, Lesaffre, Kenward (Bcs 2001)

- Perturbed MAR dropout model:

$$\text{logit} [P(D_i = 1 | Y_{i1}, Y_{i2})]$$

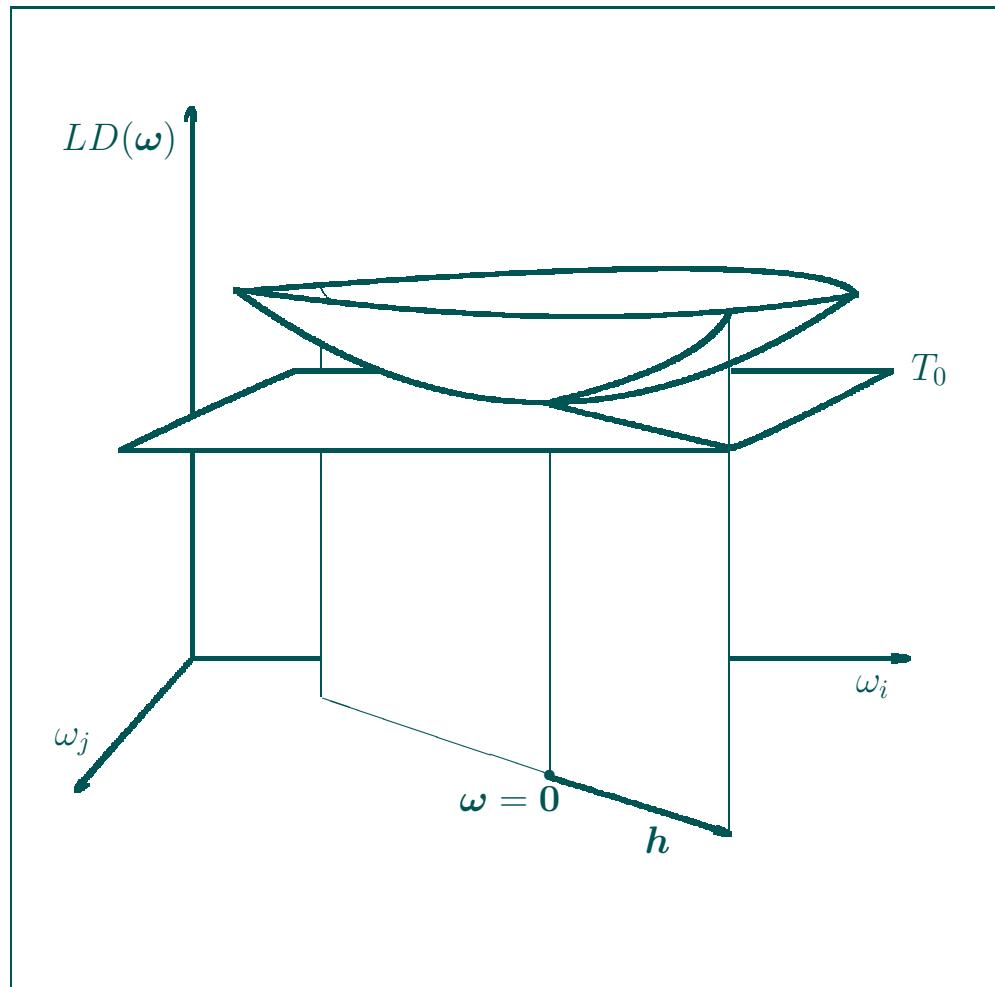
$$= \psi_0 + \psi_1 Y_{i1} + \boxed{\omega_i Y_{i2}}$$

$$\text{or } \psi_0 + \psi_1 Y_{i1} + \boxed{\omega_i (Y_{i2} - Y_{i1})}$$

- Likelihood displacement:

$$LD(\boldsymbol{\omega}) = 2 [L_{\omega=0}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\psi}}) - L_{\omega=0}(\widehat{\boldsymbol{\theta}}_\omega, \widehat{\boldsymbol{\psi}}_\omega)] \geq 0$$

35.5.1 Likelihood Displacement



Local influence direction \boldsymbol{h}



normal curvature $C_{\boldsymbol{h}}$

- Local influence for θ and ψ :

$$C_{\boldsymbol{h}} = C_{\boldsymbol{h}}(\theta) + C_{\boldsymbol{h}}(\psi)$$

35.5.2 Computational Approaches

Measuring local influence:

- Expression for $C_{\mathbf{h}}$:

$$C_{\mathbf{h}} = 2 \mid \mathbf{h}' \Delta' \ddot{\mathcal{L}}^{-1} \Delta \mathbf{h} \mid$$

- Choices for \mathbf{h}

- ▷ Direction of the i th subject $\Rightarrow C_i$
- ▷ Direction \mathbf{h}_{\max} of maximal curvature C_{\max}

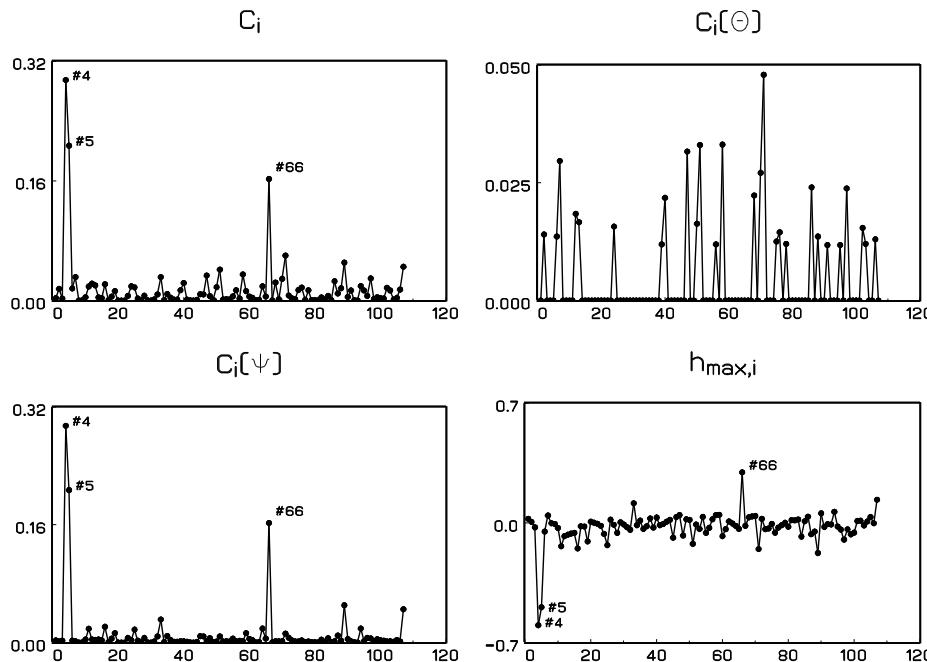
Fit for continuous outcomes:

- Fit MAR model:

- ▷ linear mixed model for outcomes
- ▷ logistic regression for dropout

- evaluate closed-form expressions for local influence

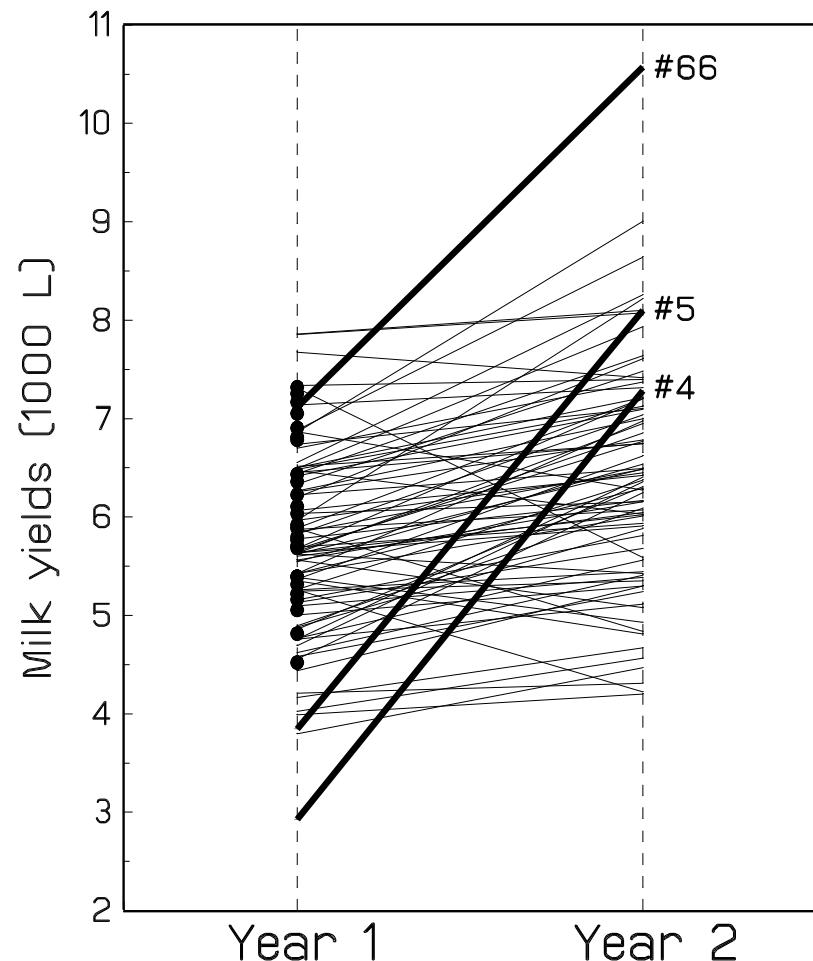
35.6 Application to Mastitis Data



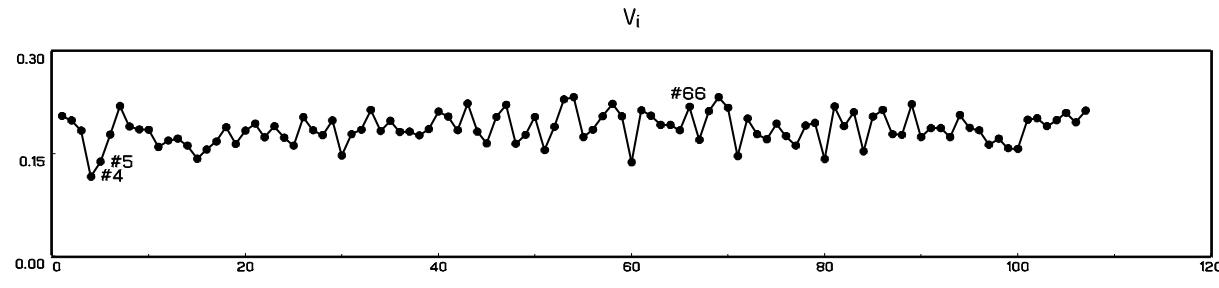
- Removing #4, #5 and #66

$$\Rightarrow G^2 = 0.005$$

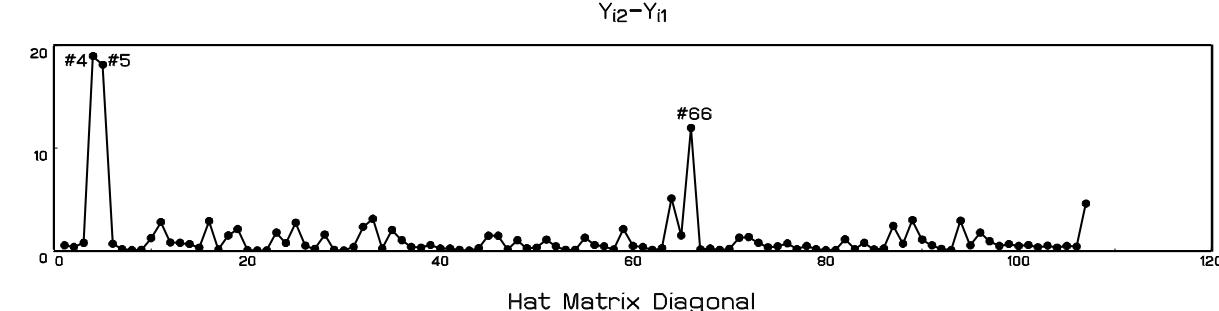
- h_{\max} : different signs for (#4,#5) and #66



35.6.1 Interpretable Components of $C_i(\psi)$



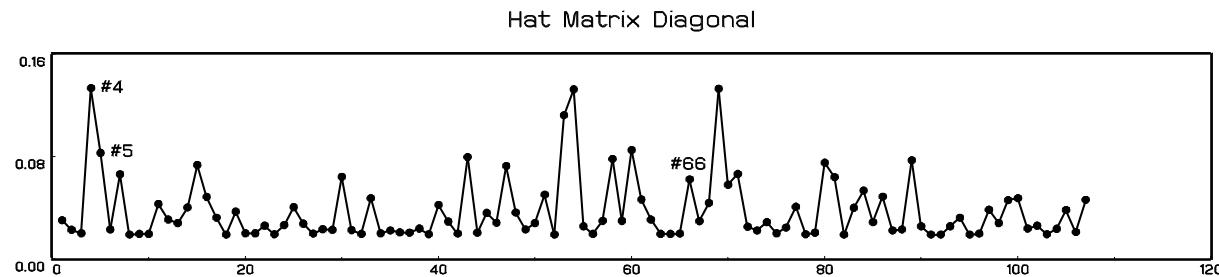
$$P(R_i = 1) [1 - P(R_i = 1)]$$



$$Y_{i2} - Y_{i1}$$

or

$$E(Y_{i2} \mid Y_{i1}) - Y_{i1}$$



$$V_i (1 \ Y_{i1}) \left\{ \sum_j V_j \begin{pmatrix} 1 \\ Y_{j1} \end{pmatrix} (1 \ Y_{j1}) \right\}^{-1} \times \begin{pmatrix} 1 \\ Y_{i1} \end{pmatrix}$$

35.7 Global Influence Analysis

- MAR versus MNAR model
- For a variety of subsets:
 - ▷ All data
 - ▷ Removal of:
 - * (53,54,66,69): from local influence on Y_{i2}
 - * (4,5): from Kenward's informal analysis
 - * (66): additional one identified from local influence on $Y_{i2} - Y_{i1}$
 - * (4,5,66): from local influence on $Y_{i2} - Y_{i1}$

Effect	Parameter	all	MAR			
			(53,54,66,69)	(4,5)	(66)	(4,5,66)
<i>Measurement model:</i>						
Intercept	μ	5.77(0.09)	5.69(0.09)	5.81(0.08)	5.75(0.09)	5.80(0.09)
Time effect	Δ	0.72(0.11)	0.70(0.11)	0.64(0.09)	0.68(0.10)	0.60(0.08)
First variance	σ_1^2	0.87(0.12)	0.76(0.11)	0.77(0.11)	0.86(0.12)	0.76(0.11)
Second variance	σ_2^2	1.30(0.20)	1.08(0.17)	1.30(0.20)	1.10(0.17)	1.09(0.17)
Correlation	ρ	0.58(0.07)	0.45(0.08)	0.72(0.05)	0.57(0.07)	0.73(0.05)
<i>Dropout model:</i>						
Intercept	ψ_0	-2.65(1.45)	-3.69(1.63)	-2.34(1.51)	-2.77(1.47)	-2.48(1.54)
First measurement	ψ_1	0.27(0.25)	0.46(0.28)	0.22(0.25)	0.29(0.24)	0.24(0.26)
Second measurement	$\omega = \psi_2$	0	0	0	0	0
-2 loglikelihood		280.02	246.64	237.94	264.73	220.23
 MNAR						
Effect	Parameter	all	(53,54,66,69)	(4,5)	(66)	(4,5,66)
<i>Measurement model:</i>						
Intercept	μ	5.77(0.09)	5.69(0.09)	5.81(0.08)	5.75(0.09)	5.80(0.09)
Time effect	Δ	0.33(0.14)	0.35(0.14)	0.40(0.18)	0.34(0.14)	0.63(0.29)
First variance	σ_1^2	0.87(0.12)	0.76(0.11)	0.77(0.11)	0.86(0.12)	0.76(0.11)
Second variance	σ_2^2	1.61(0.29)	1.29(0.25)	1.39(0.25)	1.34(0.25)	1.10(0.20)
Correlation	ρ	0.48(0.09)	0.42(0.10)	0.67(0.06)	0.48(0.09)	0.73(0.05)
<i>Dropout model:</i>						
Intercept	ψ_0	0.37(2.33)	-0.37(2.65)	-0.77(2.04)	0.45(2.35)	-2.77(3.52)
First measurement	ψ_1	2.25(0.77)	2.11(0.76)	1.61(1.13)	2.06(0.76)	0.07(1.82)
Second measurement	$\omega = \psi_2$	-2.54(0.83)	-2.22(0.86)	-1.66(1.29)	-2.33(0.86)	0.20(2.09)
-2loglikelihood		274.91	243.21	237.86	261.15	220.23
G^2 for MNAR		5.11	3.43	0.08	3.57	0.005

Chapter 36

Mechanism for Growth Data

By the way, how did Little and Rubin
delete data from the growth data set ?

36.1 Modeling Missingness

- Candidate model for missingness:

$$\text{logit}[P(R_i = 0 | \mathbf{y}_i)] = \psi_0 + \psi_1 y_{ij}, \quad \text{with } j = 1, 2, 3, \text{ or } 4$$

- When $j = 2$, then MNAR, else MAR.
- Results:

Mechanism	Effects	Deviance	p-value
MAR	Y_{i1}	19.51	<0.0001
MAR	Y_{i3}	7.43	0.0064
MAR	Y_{i4}	2.51	0.1131
MNAR	Y_{i2}	2.55	0.1105

- Including covariates:

$$\text{Boys} : \text{logit}[P(R_i = 0|y_{i1}, x_i = 0)] = \infty(22 - y_{i1})$$

$$\text{Girls} : \text{logit}[P(R_i = 0|y_{i1}, x_i = 1)] = \infty(20.75 - y_{i1})$$

- These models are interpreted as follows:

$$\text{Boys} : P(R_i = 0|y_{i1}, x_i = 0) = \begin{cases} 1 & \text{if } y_{i1} < 22, \\ 0.5 & \text{if } y_{i1} = 22, \\ 0 & \text{if } y_{i1} > 22. \end{cases}$$

$$\text{Girls} : P(R_i = 0|y_{i1}, x_i = 1) = \begin{cases} 1 & \text{if } y_{i1} < 20.75, \\ 0 & \text{if } y_{i1} > 20.75. \end{cases}$$

Chapter 37

Interval of Ignorance

- ▷ The Fluvoxamine Study
- ▷ The Slovenian Public Opinion Survey
- ▷ MAR and MNAR analyses
- ▷ Informal sensitivity analysis
- ▷ Interval of ignorance & interval of uncertainty

37.1 Fluvoxamine Trial: Side Effects

89	13
57	65
26	49
2	0
14	

- Post-marketing study of fluvoxamine in psychiatric patients
- Absence *versus* presence of side effects
- Two measurement occasions
- 315 subjects:
 - ▷ 224 completers, 75 drop out after first, 2 non-monotone, 14 without follow up
- Questions:
 - ▷ Do side effects evolve over time ?
 - ▷ Are both measurements dependent ?

37.2 The Slovenian Plebiscite

- Rubin, Stern, and Vehovar (1995)
- Slovenian Public Opinion (SPO) Survey
- Four weeks prior to decisive plebiscite
- Three questions:
 1. Are you in favor of Slovenian independence ?
 2. Are you in favor of Slovenia's secession from Yugoslavia ?
 3. Will you attend the plebiscite ?
- Political decision: ABSENCE \equiv NO
- Primary Estimand: θ : Proportion in favor of independence

- Slovenian Public Opinion Survey Data:

		Independence				
		Secession	Attendance	Yes	No	*
Secession	Yes	Yes	1191	8	21	
		No		8	0	4
	*	*	107	3	9	
Attendance	No	Yes	158	68	29	
		No		7	14	3
	*	*	18	43	31	
Yes	*	Yes	90	2	109	
		No		1	2	25
	*	*	19	8	96	

37.3 Slovenian Public Opinion: 1st Analysis

- **Pessimistic:** All who *can* say NO *will* say NO

$$\hat{\theta} = \frac{1439}{2074} = 0.694$$

- **Optimistic:** All who *can* say YES *will* say YES

$$\hat{\theta} = \frac{1439 + 159 + 144 + 136}{2074} = \frac{1878}{2076} = 0.904$$

- **Resulting Interval:**

$$\theta \in [0.694; 0.904]$$

- **Resulting Interval:**

$$\theta \in [0.694; 0.904]$$

- **Complete cases:** All who answered on 3 questions

$$\hat{\theta} = \frac{1191 + 158}{1454} = 0.928 ?$$

- **Available cases:** All who answered on both questions

$$\hat{\theta} = \frac{1191 + 158 + 90}{1549} = 0.929 ?$$

37.4 Slovenian Public Opinion: 2nd Analysis

- **Missing at Random:**

Non-response is allowed to depend on observed, but not on unobserved outcomes:

- ▷ Based on two questions:

$$\hat{\theta} = 0.892$$

- ▷ Based on three questions:

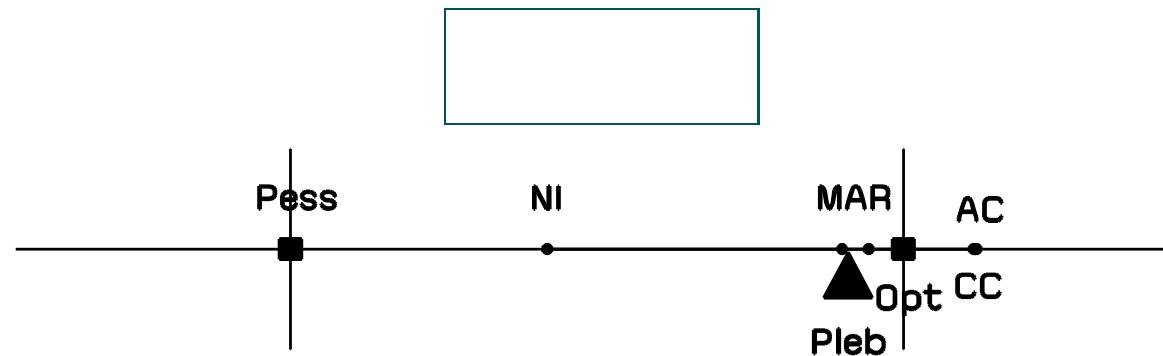
$$\hat{\theta} = 0.883$$

- **Missing Not at Random (NI):**

Non-response is allowed to depend on unobserved measurements:

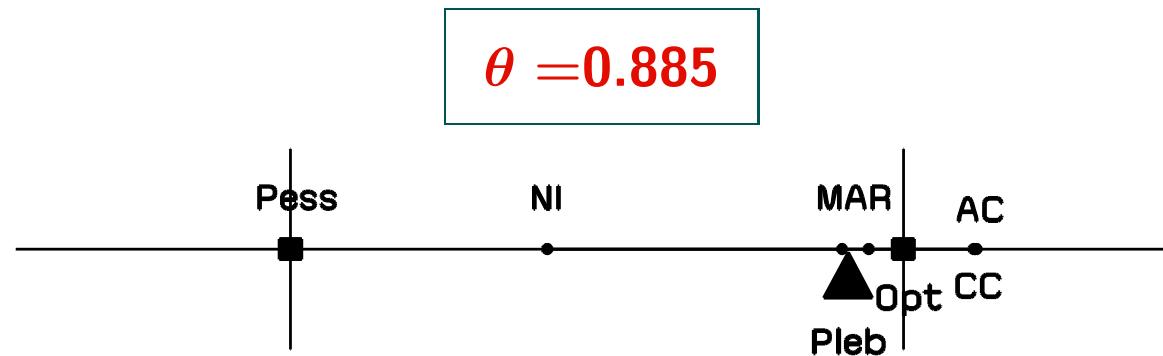
$$\hat{\theta} = 0.782$$

37.5 Slovenian Public Opinion Survey



Estimator	$\hat{\theta}$
Pessimistic bound	0.694
Optimistic bound	0.904
Complete cases	0.928 ?
Available cases	0.929 ?
MAR (2 questions)	0.892
MAR (3 questions)	0.883
MNAR	0.782

37.6 Slovenian Plebiscite: The Truth ?



Estimator	$\hat{\theta}$
Pessimistic bound	0.694
Optimistic bound	0.904
Complete cases	0.928 ?
Available cases	0.929 ?
MAR (2 questions)	0.892
MAR (3 questions)	0.883
MNAR	0.782

37.7 Did “the” MNAR model behave badly ?

Consider a family of MNAR models

- Baker, Rosenberger, and DerSimonian (1992)
- Counts $Y_{r_1 r_2 j k}$
- $j, k = 1, 2$ indicates YES/NO
- $r_1, r_2 = 0, 1$ indicates MISSING/OBSERVED

37.7.1 Model Formulation

$$E(Y_{11jk}) = m_{jk},$$

$$E(Y_{10jk}) = m_{jk}\beta_{jk},$$

$$E(Y_{01jk}) = m_{jk}\alpha_{jk},$$

$$E(Y_{00jk}) = m_{jk}\alpha_{jk}\beta_{jk}\gamma_{jk},$$

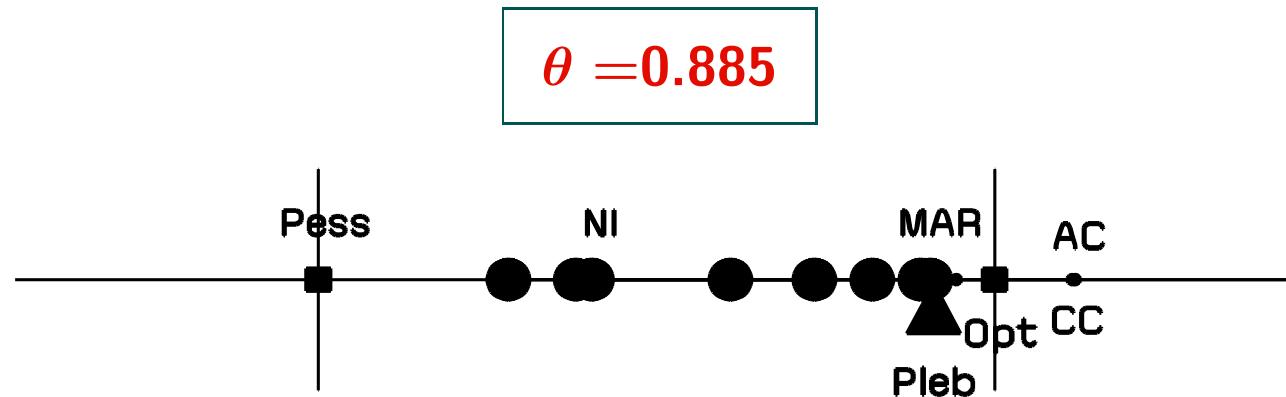
Interpretation:

- α_{jk} : models non-response on independence question
- β_{jk} : models non-response on attendance question
- γ_{jk} : interaction between both non-response indicators (cannot depend on j or k)

37.7.2 Identifiable Models

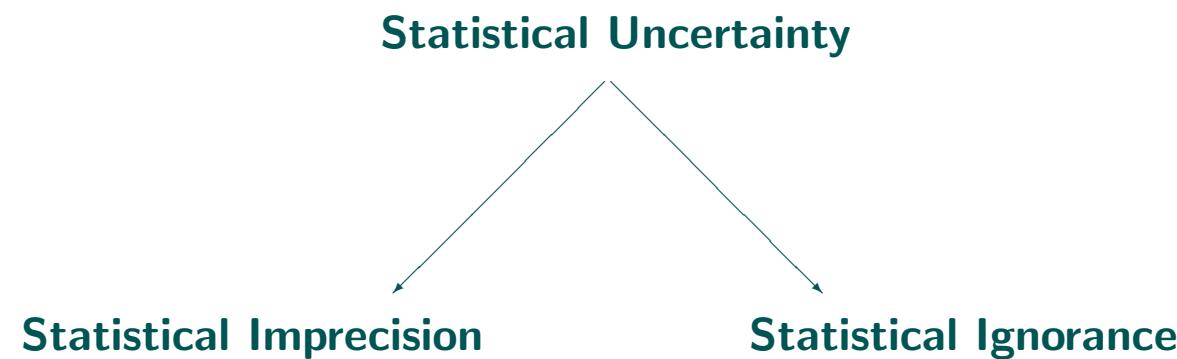
Model	Structure	d.f.	loglik	θ	C.I.
BRD1	(α, β)	6	-2495.29	0.892	[0.878;0.906]
BRD2	(α, β_j)	7	-2467.43	0.884	[0.869;0.900]
BRD3	(α_k, β)	7	-2463.10	0.881	[0.866;0.897]
BRD4	(α, β_k)	7	-2467.43	0.765	[0.674;0.856]
BRD5	(α_j, β)	7	-2463.10	0.844	[0.806;0.882]
BRD6	(α_j, β_j)	8	-2431.06	0.819	[0.788;0.849]
BRD7	(α_k, β_k)	8	-2431.06	0.764	[0.697;0.832]
BRD8	(α_j, β_k)	8	-2431.06	0.741	[0.657;0.826]
BRD9	(α_k, β_j)	8	-2431.06	0.867	[0.851;0.884]

37.7.3 An “Interval” of MNAR Estimates



Estimator	$\hat{\theta}$
[Pessimistic; optimistic]	[0.694;0.904]
Complete cases	0.928
Available cases	0.929
MAR (2 questions)	0.892
MAR (3 questions)	0.883
MNAR	0.782
MNAR “interval”	[0.741;0.892]

37.8 A More Formal Look



Statistical Imprecision: *Due to finite sampling*

- Fundamental concept of mathematical statistics
- Consistency, efficiency, precision, testing,...
- Disappears as sample size increases

Statistical Ignorance: *Due to incomplete observations*

- Received less attention
- Can invalidate conclusions
- Does not disappear with increasing sample size

Kenward, Goetghebeur, and Molenberghs (StatMod 2001)

37.8.1 Monotone Patterns

$$R = 1$$

$Y_{1,11}$	$Y_{1,12}$
$Y_{1,21}$	$Y_{1,22}$

$$R = 0$$

$Y_{0,1}$
$Y_{0,2}$



$$R = 1$$

$Y_{1,11}$	$Y_{1,12}$
$Y_{1,21}$	$Y_{1,22}$

$$R = 0$$

$Y_{0,11}$	$Y_{0,12}$
$Y_{0,21}$	$Y_{0,22}$

37.8.2 Models for Monotone Patterns

$R = 1$	$R = 0$
$\begin{array}{ c c } \hline Y_{1,11} & Y_{1,12} \\ \hline Y_{1,21} & Y_{1,22} \\ \hline \end{array}$	$\begin{array}{ c } \hline Y_{0,1} \\ \hline Y_{0,2} \\ \hline \end{array}$
↑	↑

$R = 1$	$R = 0$
$\begin{array}{ c c } \hline Y_{1,11} & Y_{1,12} \\ \hline Y_{1,21} & Y_{1,22} \\ \hline \end{array}$	$\begin{array}{ c c } \hline Y_{0,11} & Y_{0,12} \\ \hline Y_{0,21} & Y_{0,22} \\ \hline \end{array}$
↑	↑

$$\mu_{r,ij} = p_{ij} q_{r|ij}, \quad (i,j=1,2; r=0,1)$$

Model	$q_{r ij}$	# Par.	Observed d.f.	Complete d.f.
1. MCAR	q_r	4	Non-saturated	Non-saturated
2. MAR	$q_{r i}$	5	Saturated	Non-saturated
3. MNAR(0)	$q_{r j}$	5	Saturated	Non-saturated
4. MNAR(1)	$\text{logit}(q_{r ij}) = \alpha + \beta_i + \gamma_j$	6	Overspecified	Non-saturated
5. MNAR(2)	$q_{r ij}$	7	Overspecified	Saturated

37.8.3 Sensitivity Parameter Method

Sensitivity Parameter: A minimal set η

Estimable Parameter: μ , estimable, given η

Procedure:

- ▷ Given η , calculate parameter and C.I. for μ
- ▷ Set of parameter estimates: **region of ignorance**
- ▷ Set of interval estimates: **region of uncertainty**
- ▷ Single parameter case: ‘region’ becomes ‘interval’

37.9 Side Effects: Monotone Patterns

Parameter		Model 1/2	Model 3	Model 4	Model 5
1st Margin	II	0.43	0.43	0.43	0.43
	IU	[0.37;0.48]	[0.37;0.48]	[0.37;0.48]	[0.37;0.48]
2nd Margin	II	0.64	0.59	[0.49;0.74]	[0.49;0.74]
	IU	[0.58;0.70]	[0.53;0.65]	[0.43;0.79]	[0.43;0.79]
Log O.R.	II	2.06	2.06	[1.52;2.08]	[0.41;2.84]
	IU	[1.37;2.74]	[1.39;2.72]	[1.03;2.76]	[0.0013;2.84]
O.R.	II	7.81	7.81	[4.57;7.98]	[1.50;17.04]
	IU	[3.95;15.44]	[4.00;15.24]	[2.79;15.74]	[1.0013;32.89]

37.10 Side Effects: Non-Monotone Patterns

Model	d.f.	G^2	P	Marg. Prob.		Odds Ratio	
				First	Second	Orig.	Log
BRD1	6	4.5	0.104	0.43[0.37;0.49]	0.64[0.58;0.71]	7.80[3.94;15.42]	2.06[1.37;2.74]
BRD2	7	1.7	0.192	0.43[0.37;0.48]	0.64[0.58;0.70]	7.81[3.95;15.44]	2.06[1.37;2.74]
BRD3	7	2.8	0.097	0.44[0.38;0.49]	0.66[0.60;0.72]	7.81[3.95;15.44]	2.06[1.37;2.74]
BRD4	7	1.7	0.192	0.43[0.37;0.48]	0.58[0.49;0.68]	7.81[3.95;15.44]	2.06[1.37;2.74]
BRD7	8	0.0	-	0.44[0.38;0.49]	0.61[0.53;0.69]	7.81[3.95;15.44]	2.06[1.37;2.74]
BRD9	8	0.0	-	0.43[0.38;0.49]	0.66[0.60;0.72]	7.63[3.86;15.10]	2.03[1.35;2.71]
Model 10:II	9	0.0	-	[0.425;0.429]	[0.47;0.75]	[4.40;7.96]	[1.48;2.07]
Model 10:IU	9	0.0	-	[0.37;0.49]	[0.41;0.80]	[2.69;15.69]	[0.99;2.75]

37.11 Slovenian Public Opinion: 3rd Analysis

Model	Structure	d.f.	loglik	θ	C.I.
BRD1	(α, β)	6	-2495.29	0.892	[0.878;0.906]
BRD2	(α, β_j)	7	-2467.43	0.884	[0.869;0.900]
BRD3	(α_k, β)	7	-2463.10	0.881	[0.866;0.897]
BRD4	(α, β_k)	7	-2467.43	0.765	[0.674;0.856]
BRD5	(α_j, β)	7	-2463.10	0.844	[0.806;0.882]
BRD6	(α_j, β_j)	8	-2431.06	0.819	[0.788;0.849]
BRD7	(α_k, β_k)	8	-2431.06	0.764	[0.697;0.832]
BRD8	(α_j, β_k)	8	-2431.06	0.741	[0.657;0.826]
BRD9	(α_k, β_j)	8	-2431.06	0.867	[0.851;0.884]
Model 10	(α_k, β_{jk})	9	-2431.06	[0.762;0.893]	[0.744;0.907]
Model 11	(α_{jk}, β_j)	9	-2431.06	[0.766;0.883]	[0.715;0.920]
Model 12	$(\alpha_{jk}, \beta_{jk})$	10	-2431.06	[0.694;0.904]	

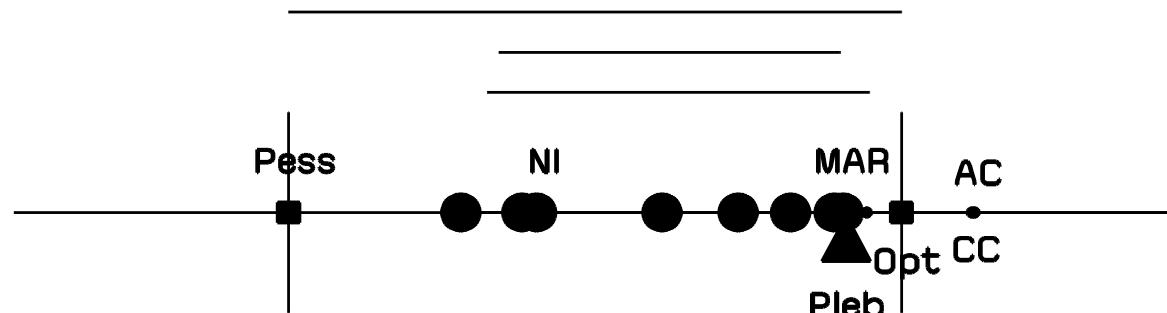
37.12 Every MNAR Model Has Got a MAR Bodyguard

- Fit an MNAR model to a set of incomplete data
- Change the conditional distribution of the unobserved outcomes, given the observed ones, to comply with MAR
- The resulting new model will have exactly the same fit as the original MNAR model
- The missing data mechanism has changed
- This implies that definitively testing for MAR *versus* MNAR is not possible

37.13 Slovenian Public Opinion: 4rd Analysis

Model	Structure	d.f.	loglik	$\hat{\theta}$	C.I.	$\hat{\theta}_{\text{MAR}}$
BRD1	(α, β)	6	-2495.29	0.892	[0.878;0.906]	0.8920
BRD2	(α, β_j)	7	-2467.43	0.884	[0.869;0.900]	0.8915
BRD3	(α_k, β)	7	-2463.10	0.881	[0.866;0.897]	0.8915
BRD4	(α, β_k)	7	-2467.43	0.765	[0.674;0.856]	0.8915
BRD5	(α_j, β)	7	-2463.10	0.844	[0.806;0.882]	0.8915
BRD6	(α_j, β_j)	8	-2431.06	0.819	[0.788;0.849]	0.8919
BRD7	(α_k, β_k)	8	-2431.06	0.764	[0.697;0.832]	0.8919
BRD8	(α_j, β_k)	8	-2431.06	0.741	[0.657;0.826]	0.8919
BRD9	(α_k, β_j)	8	-2431.06	0.867	[0.851;0.884]	0.8919
Model 10	(α_k, β_{jk})	9	-2431.06	[0.762;0.893]	[0.744;0.907]	0.8919
Model 11	(α_{jk}, β_j)	9	-2431.06	[0.766;0.883]	[0.715;0.920]	0.8919
Model 12	$(\alpha_{jk}, \beta_{jk})$	10	-2431.06	[0.694;0.904]		0.8919

$$\theta = 0.885$$



Estimator	$\hat{\theta}$
[Pessimistic; optimistic]	[0.694;0.904]
MAR (3 questions)	0.883
MNAR	0.782
MNAR “interval”	[0.753;0.891]
Model 10	[0.762;0.893]
Model 11	[0.766;0.883]
Model 12	[0.694;0.904]

Chapter 38

Pattern-mixture Models

- ▷ A selection model for the vorozole study
- ▷ Initial pattern-mixture models for the vorozole study
- ▷ Principles of pattern-mixture models
- ▷ Connection between selection models and pattern-mixture models

38.1 The Vorozole Study

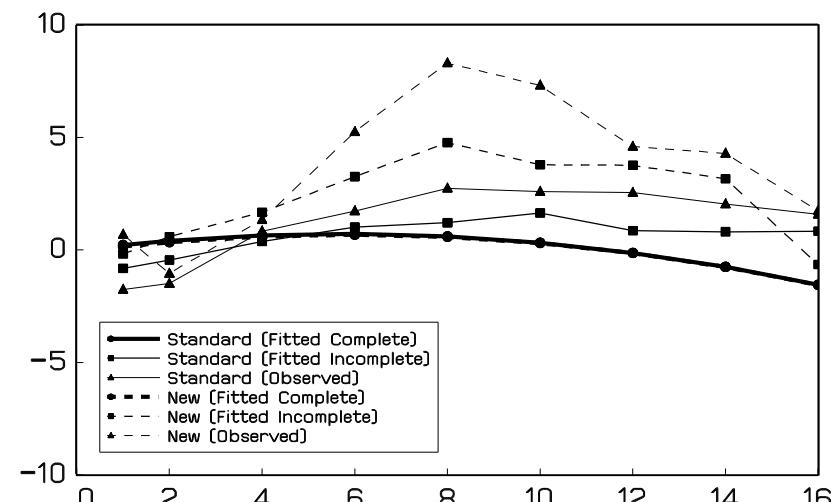
- open-label study in 67 North American centers
- postmenopausal women with metastatic breast cancer
- 452 patients, followed until disease progression/death
- two groups: **vorozole** 2.5 mg × 1 ↔ **megestrol acetate** 40 mg × 4
- several outcomes: response rate, survival, safety,...
- focus: **quality of life:** *total Function Living Index: Cancer (FLIC)*
a higher score is more desirable

38.2 A Selection Model for the Vorozole Study

Effect	Parameter	Estimate (s.e.)
<i>Fixed-Effect Parameters:</i>		
time	β_0	7.78 (1.05)
time*baseline	β_1	-0.065 (0.009)
time*treatment	β_2	0.086 (0.157)
time ²	β_3	-0.30 (0.06)
time ² *baseline	β_4	0.0024 (0.0005)
<i>Variance Parameters:</i>		
Random intercept	d	105.42
Serial variance	τ^2	77.96
Serial association	λ	7.22
Measurement error	σ^2	77.83

Treatment effect: $p = 0.5822$

Fitted Mean Profiles

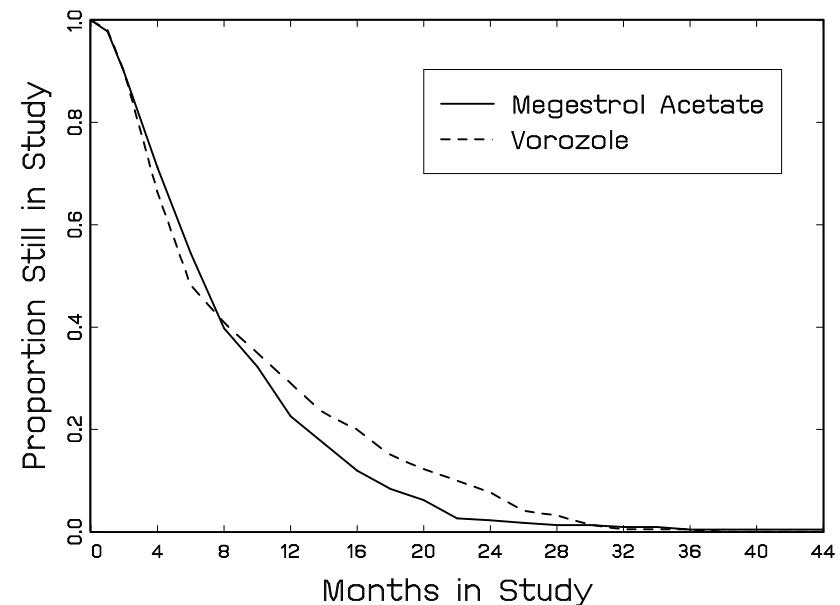


38.2.1 The Dropout Model

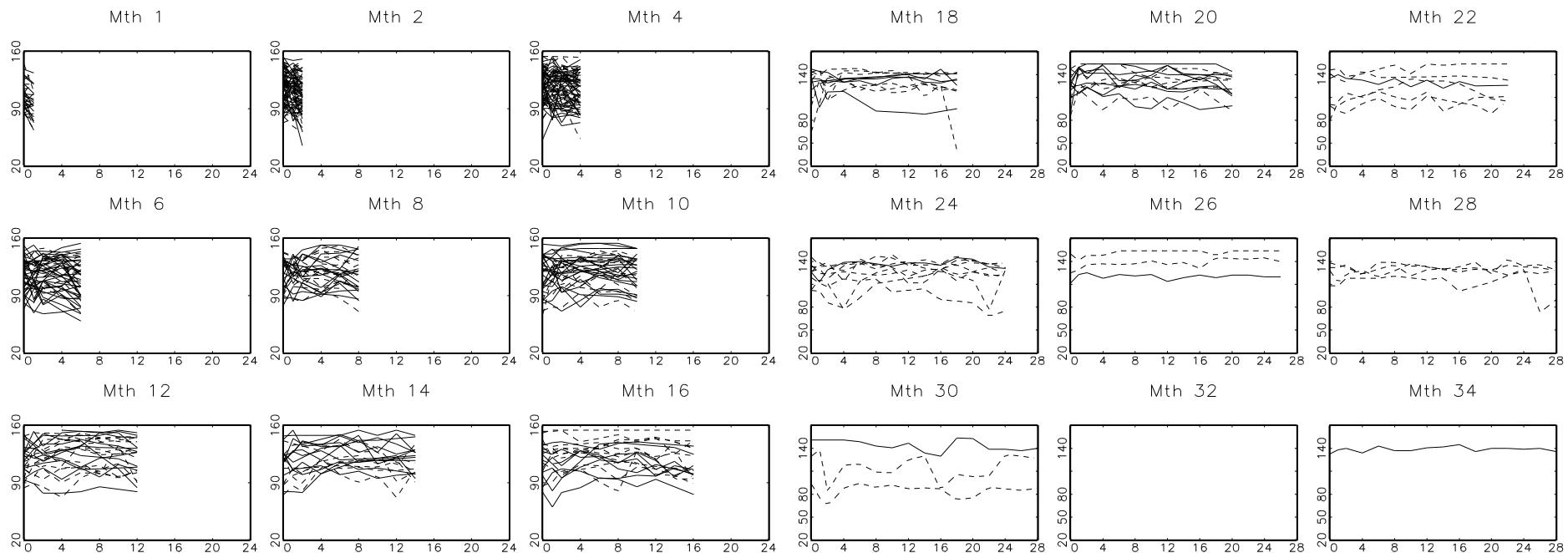
	MAR	MNAR
First	$0.080(0.341) - 0.014(0.003)\text{base}_i - 0.033(0.004)y_{i,j-1}$	$0.53 - 0.015\text{base}_i - 0.076y_{i,j-1} + 0.057y_{ij}$
Extended	$0.033(0.401) - 0.013(0.003)\text{base}_i - 0.023(0.005)\frac{y_{i,j-2}+y_{i,j-1}}{2}$ $- 0.047(0.010)\frac{y_{i,j-1}-y_{i,j-2}}{2}$	$1.38 - 0.021\text{base}_i - 0.0027y_{i,j-2}$ $- 0.064y_{i,j-1} + 0.035y_{ij}$

Dropout increases with:

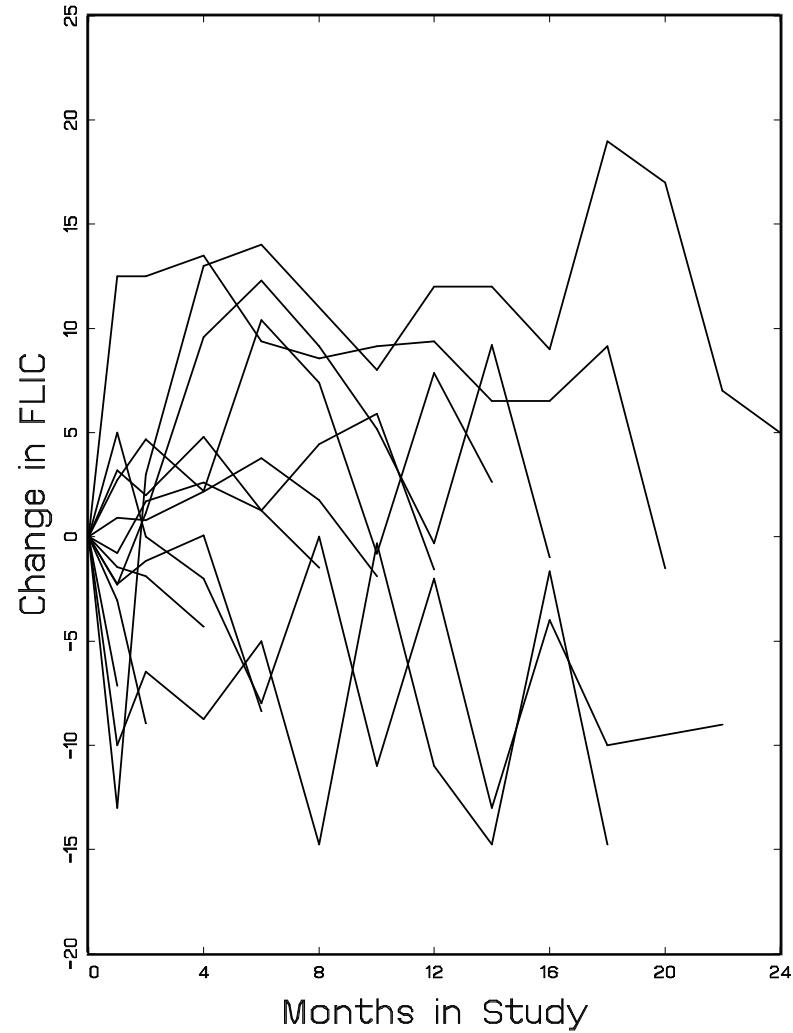
- low score
- negative trend
- lower baseline value



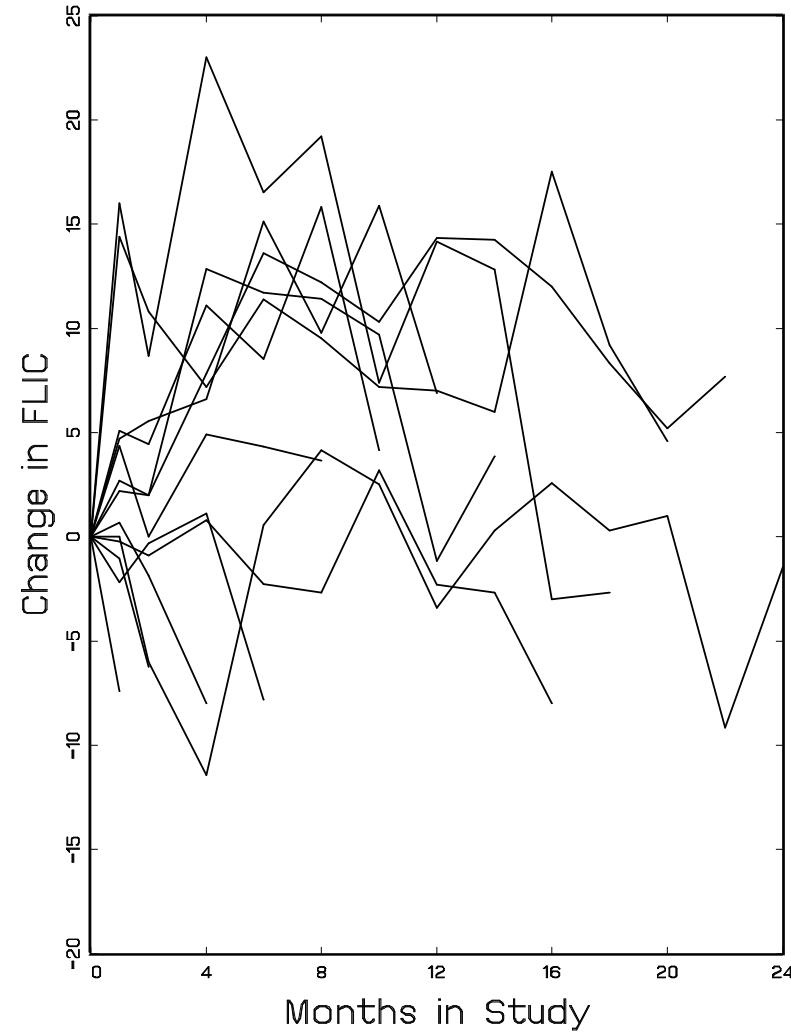
38.3 Pattern-mixture Analysis of the Vorozole Study: Profiles



Megestrol Acetate

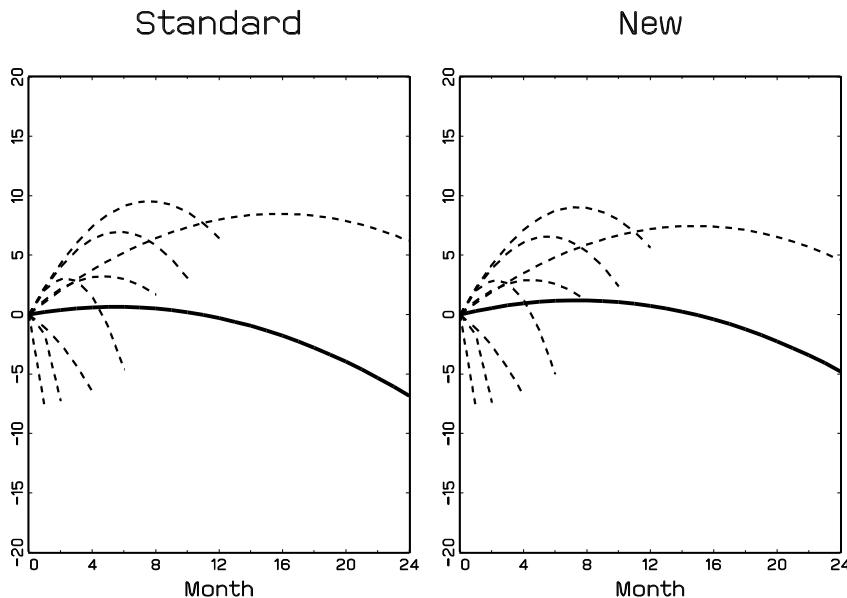


Vorozole

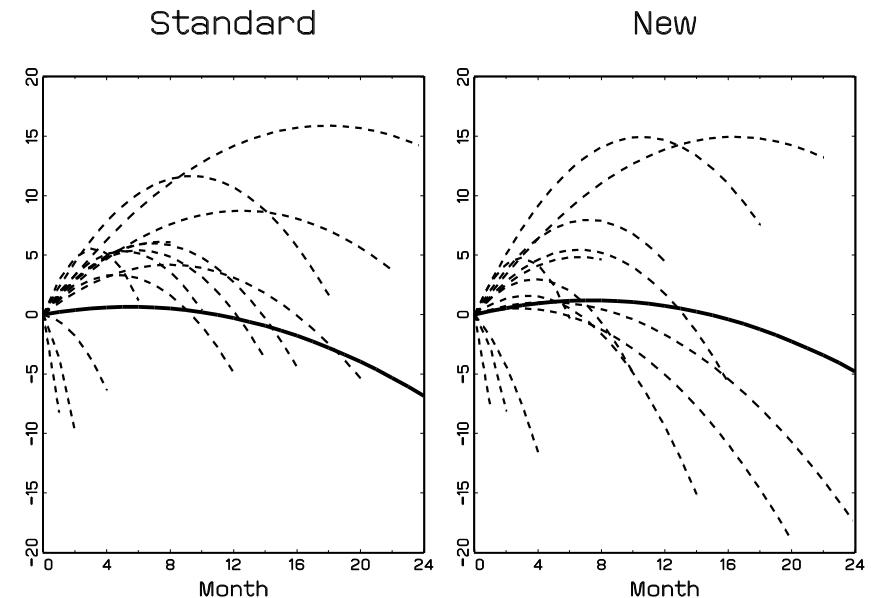


38.3.1 Two Pattern-Mixture Models

Includes: time*treatment



Includes: time*treatment*pattern



Assessment of Treatment Effect

Selection model: $p = 0.5822$ (1 df; output)

PMM1: $p = 0.6868$ (1 df; output)

PMM2: $p = 0.2403$ (13 df; output)
 $p = 0.3206$ (1 df; delta method)

38.3.2 Estimating Marginal Effects From PMM

- Pattern-membership probabilities:

$$\pi_1, \dots, \pi_t, \dots, \pi_T.$$

- The marginal effects:

$$\beta_\ell = \sum_{t=1}^n \beta_{\ell t} \pi_t, \quad \ell = 1, \dots, g$$

- Their variance:

$$\text{Var}(\beta_1, \dots, \beta_g) = A V A'$$

where

$$V = \begin{pmatrix} \text{Var}(\beta_{\ell t}) & 0 \\ 0 & \text{Var}(\pi_t) \end{pmatrix}$$

and

$$A = \frac{\partial(\beta_1, \dots, \beta_g)}{\partial(\beta_{11}, \dots, \beta_{ng}, \pi_1, \dots, \pi_n)}$$

38.3.3 Considerations

- Models fitted over the observation period within a certain pattern
- How do we extrapolate beyond dropout time ?
- Making the model simple enough ?
- Formal identifying restrictions ?
- ... ?

38.4 PMM: Three Strategies

(1a) **Simple model per pattern:**

$$Y_i = X_i\beta(d_i) + Z_ib_i + \varepsilon_i$$

$$b_i \sim N(0, D(d_i))$$

$$\varepsilon_i \sim N(0, \Sigma_i(d_i))$$

(1b) **Pattern as covariate:**

$$Y_i = X_i\beta + Z_ib_i + d_i\theta + \varepsilon_i$$

(2) **Identifying restrictions:**

CCMV: Complete Case Missing Values

ACMV: Available Case Missing Values

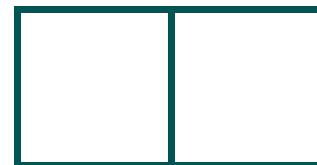
NCMV: Neighbouring Case Missing Values

38.4.1 Identifying Restrictions

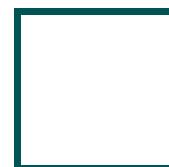
Pattern 3



Pattern 2



Pattern 1

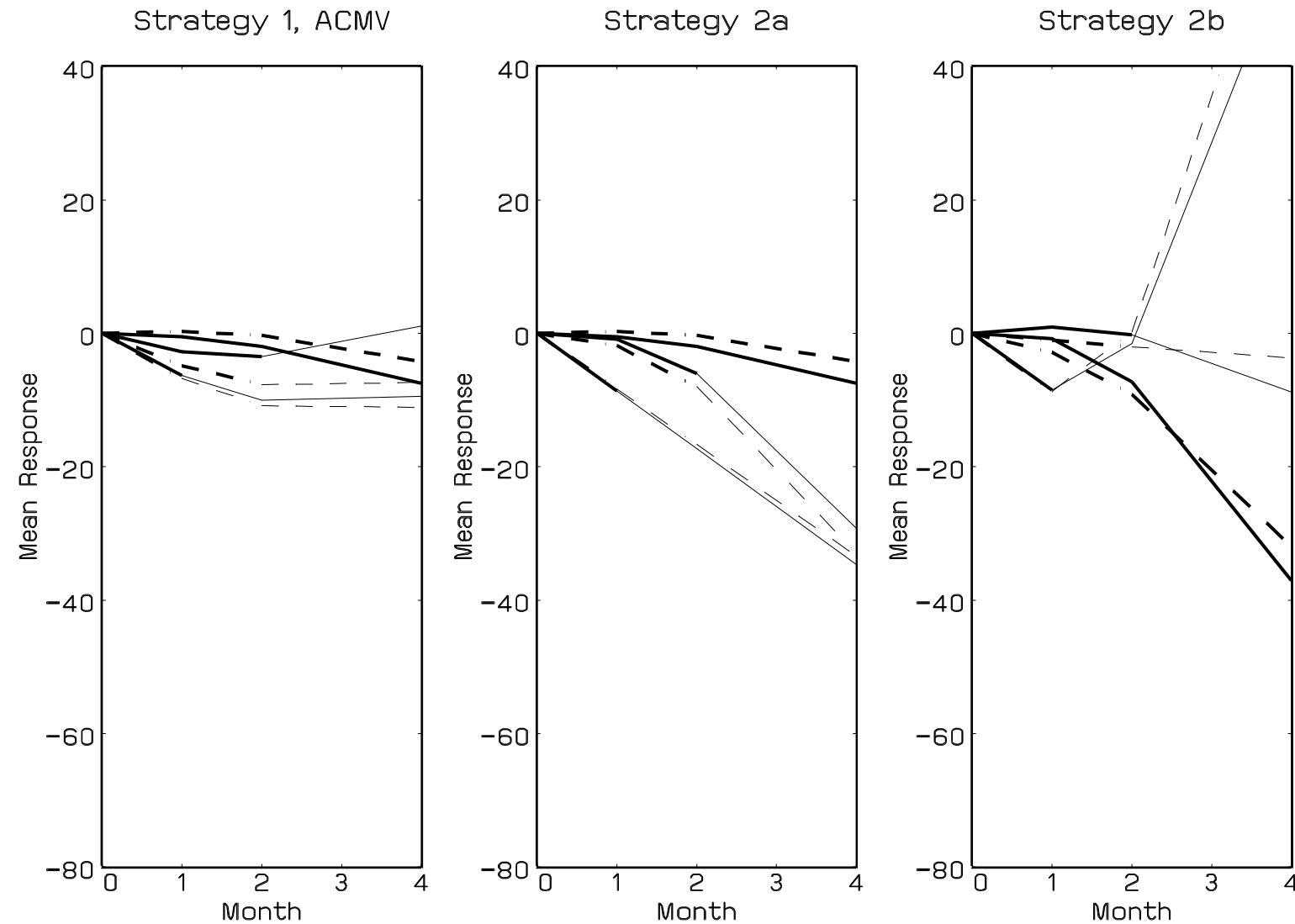


Effect	Initial	CCMV	NCMV	ACMV
<i>Pattern 1:</i>				
Time	3.40(13.94)	13.21(15.91)	7.56(16.45)	4.43(18.78)
Time*base	-0.11(0.13)	-0.16(0.16)	-0.14(0.16)	-0.11(0.17)
Time*treat	0.33(3.91)	-2.09(2.19)	-1.20(1.93)	-0.41(2.52)
Time ²		-0.84(4.21)	-2.12(4.24)	-0.70(4.22)
Time ² *base		0.01(0.04)	0.03(0.04)	0.02(0.04)
σ_{11}	131.09(31.34)	151.91(42.34)	134.54(32.85)	137.33(34.18)
σ_{12}		59.84(40.46)	119.76(40.38)	97.86(38.65)
σ_{22}		201.54(65.38)	257.07(86.05)	201.87(80.02)
σ_{13}		55.12(58.03)	49.88(44.16)	61.87(43.22)
σ_{23}		84.99(48.54)	99.97(57.47)	110.42(87.95)
σ_{33}		245.06(75.56)	241.99(79.79)	286.16(117.90)
<i>Pattern 2:</i>				
Time	53.85(14.12)	29.78(10.43)	33.74(11.11)	28.69(11.37)
Time*base	-0.46(0.12)	-0.29(0.09)	-0.33(0.10)	-0.29(0.10)
Time*treat	-0.95(1.86)	-1.68(1.21)	-1.56(2.47)	-2.12(1.36)
Time ²	-18.91(6.36)	-4.45(2.87)	-7.00(3.80)	-4.22(4.20)
Time ² *base	0.15(0.05)	0.04(0.02)	0.07(0.03)	0.05(0.04)
σ_{11}	170.77(26.14)	175.59(27.53)	176.49(27.65)	177.86(28.19)
σ_{12}	151.84(29.19)	147.14(29.39)	149.05(29.77)	146.98(29.63)
σ_{22}	292.32(44.61)	297.38(46.04)	299.40(47.22)	297.39(46.04)
σ_{13}		57.22(37.96)	89.10(34.07)	99.18(35.07)
σ_{23}		71.58(36.73)	107.62(47.59)	166.64(66.45)
σ_{33}		212.68(101.31)	264.57(76.73)	300.78(77.97)
<i>Pattern 3:</i>				
Time	29.91(9.08)	29.91(9.08)	29.91(9.08)	29.91(9.08)
Time*base	-0.26(0.08)	-0.26(0.08)	-0.26(0.08)	-0.26(0.08)
Time*treat	0.82(0.95)	0.82(0.95)	0.82(0.95)	0.82(0.95)
Time ²	-6.42(2.23)	-6.42(2.23)	-6.42(2.23)	-6.42(2.23)
Time ² *base	0.05(0.02)	0.05(0.02)	0.05(0.02)	0.05(0.02)
σ_{11}	206.73(35.86)	206.73(35.86)	206.73(35.86)	206.73(35.86)
σ_{12}	96.97(26.57)	96.97(26.57)	96.97(26.57)	96.97(26.57)
σ_{22}	174.12(31.10)	174.12(31.10)	174.12(31.10)	174.12(31.10)
σ_{13}	87.38(30.66)	87.38(30.66)	87.38(30.66)	87.38(30.66)
σ_{23}	91.66(28.86)	91.66(28.86)	91.66(28.86)	91.66(28.86)
σ_{33}	262.16(44.70)	262.16(44.70)	262.16(44.70)	262.16(44.70)

38.4.2 Pattern As Covariate

Effect	Pattern	Estimate (s.e.)
Time	1	7.29(15.69)
Time	2	37.05(7.67)
Time	3	39.40(9.97)
Time*treat	1	5.25(6.41)
Time*treat	2	3.48(5.46)
Time*treat	3	3.44(6.04)
Time*base	1	-0.21(0.15)
Time*base	2	-0.34(0.06)
Time*base	3	-0.36(0.08)
Time*treat*base		-0.06(0.04)
Time ²	1	
Time ²	2	-9.18(2.47)
Time ²	3	-7.70(2.29)
Time ² *treat		1.10(0.74)
Time ² *base		0.07(0.02)
σ_{11}		173.63(18.01)
σ_{12}		117.88(17.80)
σ_{22}		233.86(26.61)
σ_{13}		89.59(24.56)
σ_{23}		116.12(34.27)
σ_{33}		273.98(48.15)

38.4.3 Plot for Three Different Strategies

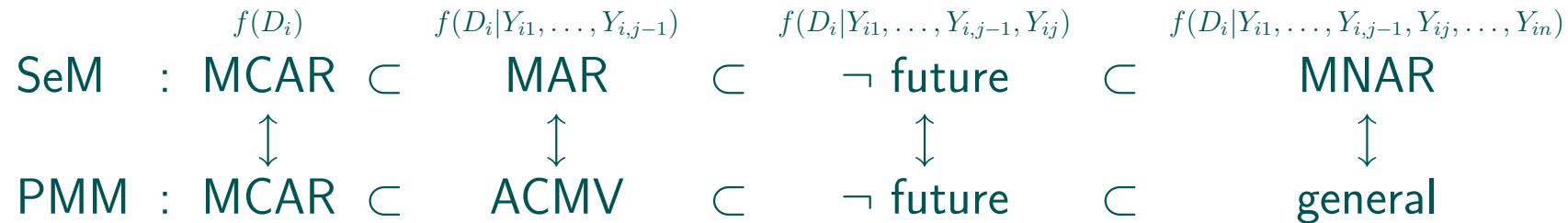


38.5 Connection SEM–PMM

Molenberghs, Michiels, Kenward, and Diggle (Stat Neerl 1998)

Kenward, Molenberghs, and Thijs (Bka 2002)

Selection Models: $f(D_i|\mathbf{Y}_i, \psi)$ \longleftrightarrow $f(\mathbf{Y}_i|D_i, \theta)$: Pattern-Mixture Models



Chapter 39

Concluding Remarks

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