

Mixed and Multilevel Models

Geert Verbeke

Biostatistical Centre, K.U.Leuven

`geert.verbeke@med.kuleuven.be`

`http://perswww.kuleuven.be/geert_verbeke`

Master in Statistics

Contents

I	Introduction	1
1	Introductory material	2
2	Course motivation	11
II	Linear Mixed Models	18
3	The Captopril data	19
4	The lizard data	28
5	The paired t -test revisited	50

6	The growth curves data	61
7	The linear mixed model	75
8	The rat data	97
9	The BLSA prostate data	107
10	The Leuven diabetes project	118
11	Estimation of Random Effects	141
12	Concluding remarks	159
III	Generalized Linear Mixed Models	178
13	The toenail data	179
14	The Leuven diabetes project	200
15	The Epilepsy data	212
16	The hierarchical versus marginal model	227

IV	Non-linear Mixed Models	248
17	Introduction	249
18	The orange trees	253
19	The Theophylline data	260
20	Remarks	269

Part I

Introduction

Chapter 1

Introductory material

- ▷ Related references
- ▷ Course material
- ▷ Software
- ▷ Course evaluation

1.1 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.
- Fitzmaurice, G., Davidian, M., Verbeke, G. and Molenberghs G. (2008). *Longitudinal Data Analysis*. Handbooks of Modern Statistical Methods series, #1. London: Chapman & Hall/CRC
- 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 Gajęcki, A.T. (2007). *Linear Mixed Models: A Practical Guide Using Statistical Software*. Boca Raton: Chapman & Hall/CRC.

1.2 Course material

- Copies of the course notes: **Toledo**
- Data sets analysed in the course: **Toledo**
- Books:
 - ▷ Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer Series in Statistics. New-York: Springer.
 - ▷ Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer-Verlag.

1.3 Software

- Many software packages nowadays allow fitting of mixed or multilevel models
- In this course, SAS will be used:
 - ▷ PROC MIXED
 - ▷ PROC GLIMMIX
 - ▷ PROC NLMIXED
- SAS is the most flexible in terms of models that can be fitted
- SAS is most up to date with the statistical literature

1.4 Course evaluation

- Take-home assignment
- Data analysis and reporting in teams
- Report submitted before final examination
- Oral defense of the report

Chapter 2

Course motivation

- ▷ Hierarchical data
- ▷ Correlated data
- ▷ Overview of model families

2.1 Hierarchical data

Hierarchical data are obtained when the sample is taken at multiple, hierarchically ordered, levels.

- Examples:
 - ▷ Measurements taken on patients, at multiple visits after their treatment
 - ▷ Growth curves of children, animals, plants, ...
 - ▷ Survey in which all members from each of a sample of families are questioned
 - ▷ Survey in which 10 habitants from each of a sample of cities are questioned
 - ▷ Exam results from students from a sample of schools
 - ▷ ...

- These are examples of **two-level** data structures, but extensions to multiple levels are possible:

10 cities

→ In each: 5 schools

→ In each: 2 classes

→ In each: 5 students

→ Each student given the test twice

- Terminologies:

▷ Repeated measures

▷ Longitudinal data

▷ Multilevel data

▷ ...

2.2 Correlated data

2.2.1 Example: Longitudinal body weight example

- Consider a body weight experiment in which body weight is measured on a daily basis, for a sample of participants
- It is natural to assume body weights from different subjects to be independent from each other
- Body weights measured on the same subject are expected to be correlated

**Should this correlation be accounted for in analysis ?
If yes, how ?**

2.2.2 Example: Comparing BMI between males and females

- Suppose interest is in comparing the average BMI between males and females, based on 100 observations from each population
- Natural analysis: **Two-sample, unpaired t -test**
- Suppose the 100 males and 100 females are married couples
- The BMI of spouses is likely to be correlated
- Natural analysis: **Paired t -test**

2.2.3 Conclusion

- Hierarchical data structures often yield data which cannot be assumed independent
- From a statistical perspective, the key issue in modelling hierarchical data is how to account for the association between observations
- Alternative terminology:
 - ▷ Repeated measures
 - ▷ Longitudinal data
 - ▷ Multilevel data
 - ▷ **Correlated data**
 - ▷ ...

2.3 Overview of model families

- Since hierarchical data are correlated, all traditional models in statistics need a counterpart for correlated data
- Many different models have been proposed in the statistical literature
- We focus on mixed models which explicitly model the various levels in the data structure

Cross-sectional data	→	Hierarchical data
Linear regression models	→	Linear mixed models
Generalized linear models	→	Generalized linear mixed models
Non-linear regression models	→	Non-linear mixed models

Part II

Linear Mixed Models

Chapter 3

The Captopril data

- ▷ Example
- ▷ Paired t -test
- ▷ Paired versus unpaired t -test
- ▷ Conclusion

3.1 Example

- 15 patients with hypertension
- The response of interest is the supine blood pressure, before and after treatment with CAPTOPRIL
- Research question:

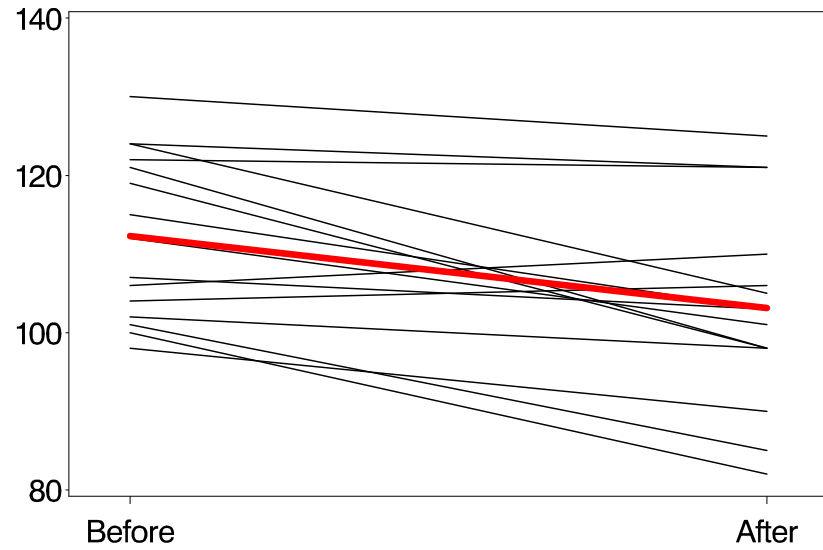
How does treatment affect BP ?

- Dataset 'Captopril'

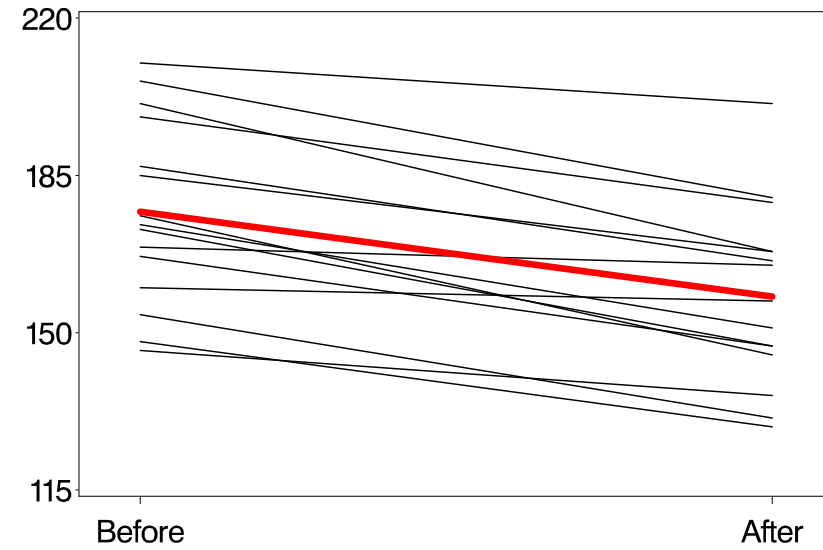
Patiënt	Before		After	
	SBP	DBP	SBP	DBP
1	210	130	201	125
2	169	122	165	121
3	187	124	166	121
4	160	104	157	106
5	167	112	147	101
6	176	101	145	85
7	185	121	168	98
8	206	124	180	105
9	173	115	147	103
10	146	102	136	98
11	174	98	151	90
12	201	119	168	98
13	198	106	179	110
14	148	107	129	103
15	154	100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

Average (mm Hg)	
Diastolic before:	112.3
Diastolic after:	103.1
Systolic before:	176.9
Systolic after:	158.0

Diastolic BP (mm Hg)



Systolic BP (mm Hg)



3.2 Paired t -test

- Let's focus on the analysis of the diastolic BP:

	Average (mm Hg)
Diastolic before:	112.3
Diastolic after:	103.1

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

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

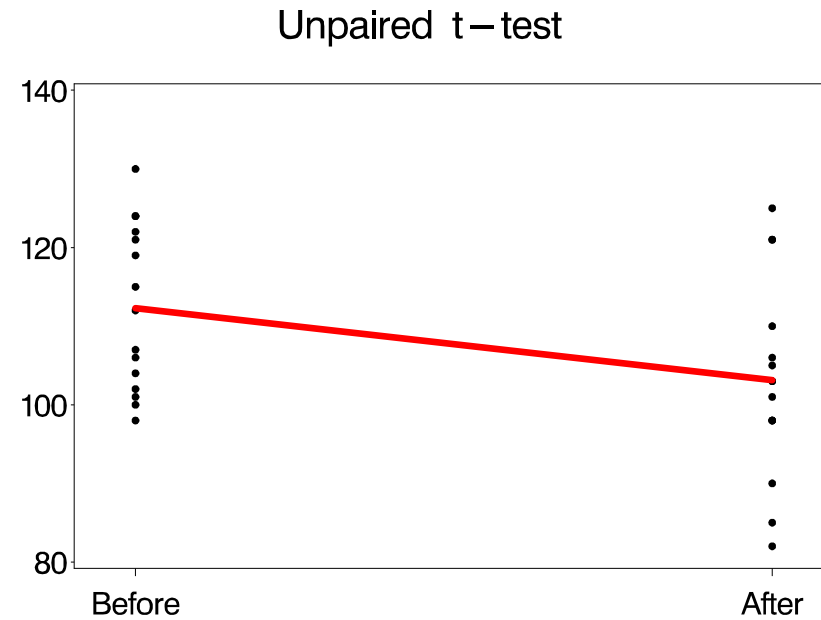
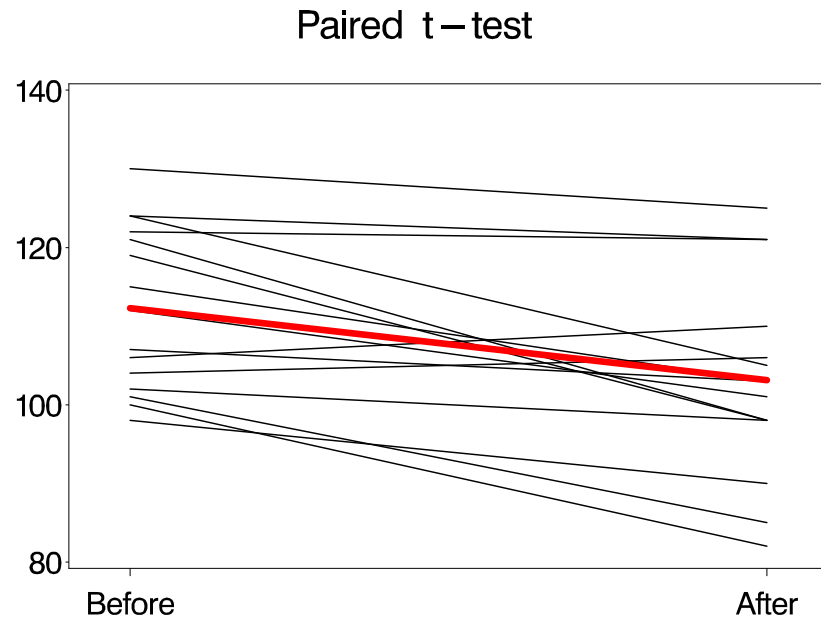
- A positive Δ_i corresponds to a decrease of the BP, while a negative Δ_i is equivalent to an increase.
- Testing for treatment effect is now equivalent to testing whether the average difference μ_{Δ} equals zero.
- Statistica output:

Variable	Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	p
DIA_VOOR	112,3333	10,47219				
DIA_NA	103,0667	12,55540	15	9,266667	8,614495	0,000951

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

3.3 Paired versus unpaired t -test

- What if the Captopril data were analysed using an unpaired t -test ?



- Results from unpaired and paired t -tests, respectively:

▷ Unpaired:

Group 1 vs. Group 2	Mean Group 1	Mean Group 2	p	t separ. var.est.	p 2-sided	Std.Dev. Group 1	Std.Dev. Group 2	F-ratio Variances	p Variances
DIA_VOOR vs. DIA_NA	112,3333	103,0667	0,036607	2,195158	0,036887	10,47219	12,55540	1,437429	0,506015

▷ Paired:

Variable	Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	p
DIA_VOOR	112,3333	10,47219				
DIA_NA	103,0667	12,55540	15	9,266667	8,614495	0,000951

- Although both tests lead to a significant result, there is a serious difference in p -values, showing that ignoring the paired nature of the data can lead to wrong conclusions.

3.4 Conclusion

15 × 2 measurements ≠ 30 × 1 measurement

- The correlation cannot be ignored in the analyses
- In the paired t -test, the correlation problem is circumvented by taking within-subject differences $\Delta_i = Y_{i1} - Y_{i2}$, $i = 1, \dots, 15$
- How to extend this to:
 - ▷ multiple measurements per subject ?
 - ▷ include covariate information ?
 - ▷ multiple levels in the data structure ?

Chapter 4

The lizard data

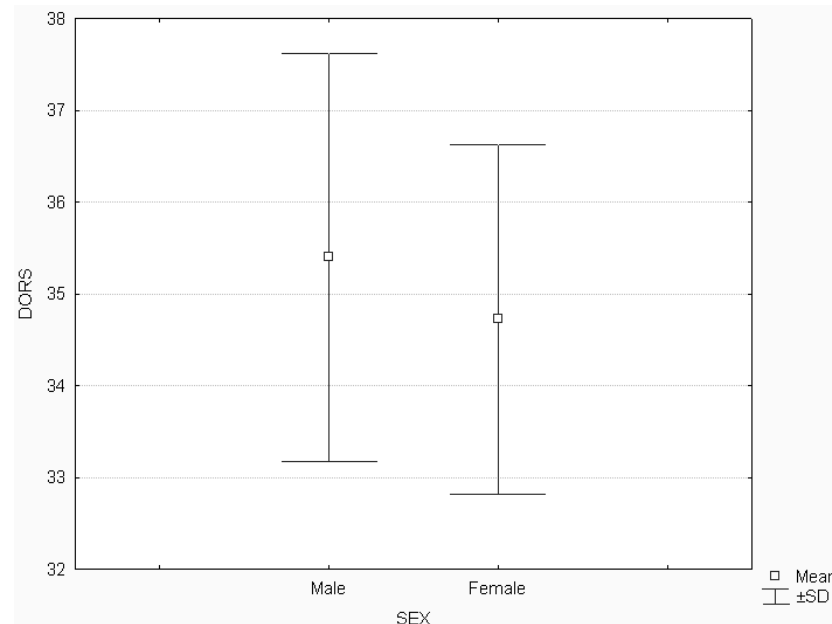
- ▷ Example
- ▷ Two-way ANOVA
- ▷ Mixed models
- ▷ Fitting mixed models in SAS
- ▷ The hierarchical versus marginal model

4.1 Example

- Data on 102 lizards
- Response of interest: Number of dorsal shells
- Research question:

Is number of dorsal shells gender-related ?

- Graphically:

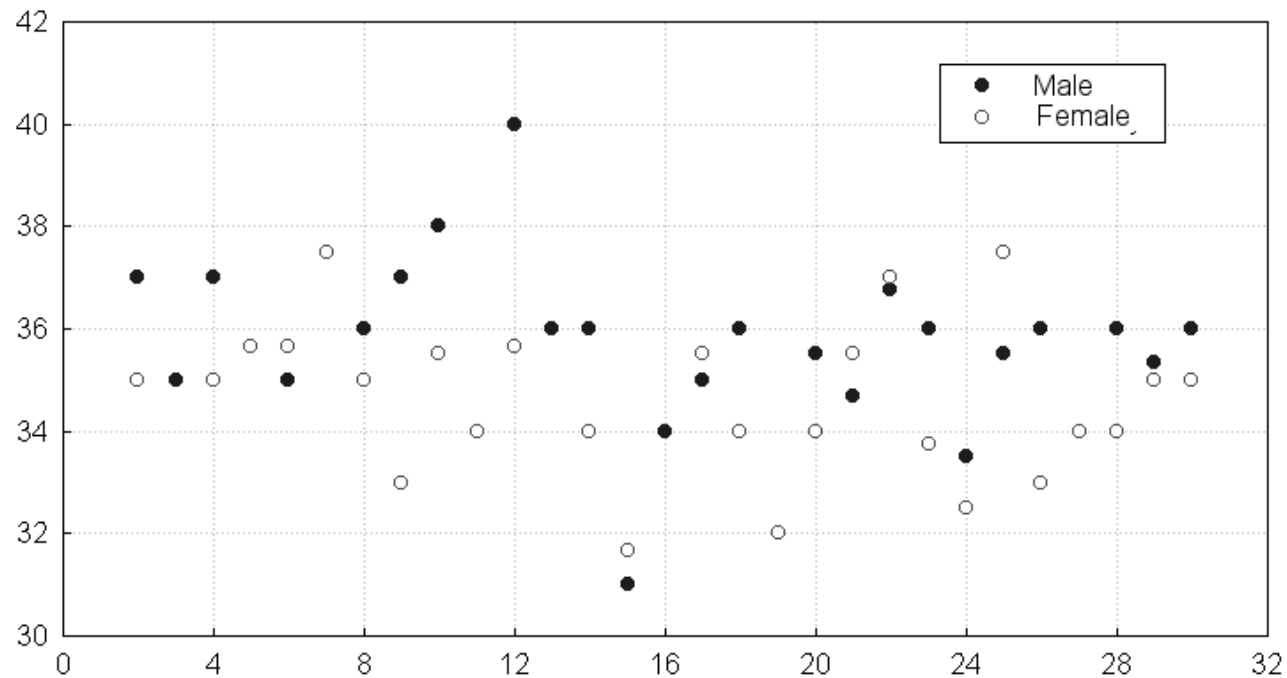


- Two-sample *t*-test:

T-tests; Grouping: SEX (schildformengels.STA)					
Group 1: Male					
Group 2: Female					
Variable	Mean Male	Mean Female	t-value	df	p
DORS	35,39583	34,72222	1,648480	100	0,102393

- Hence, the small observed difference is not significant ($p = 0.1024$).
- A typical aspect of the data is that some animals have the same mother.
- We have 102 lizards from 29 mothers
- Mother effects might be present
- Hence a comparison between male and female animals should be based on within-mother comparisons.

- Graphically:



- Observations:

- ▷ Much between-mother variability
- ▷ Often, males (considerably) higher than females
- ▷ In cases where females higher than males, small differences

- Hence the non-significant t -test result may be due to the between-mother variability
- This is an example of **clustered data**: Observations are clustered within mothers
- It is to be expected that measurements within mothers are more alike than measurements from different mothers.
- We expect correlated observations within mothers and independent observations between mothers.
- How to correct for differences between mothers ?

4.2 Two-way ANOVA

- An obvious first choice to test for a 'sex' effect, correcting for 'mother' effects, is 2-way ANOVA with factors 'sex' and 'mother'.
- The mother effect then represents the variability between mothers.
- Let Y_{ijk} be the k th outcome in the j th gender group for the i th mother.
- Our two-way ANOVA model then equals:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk},$$

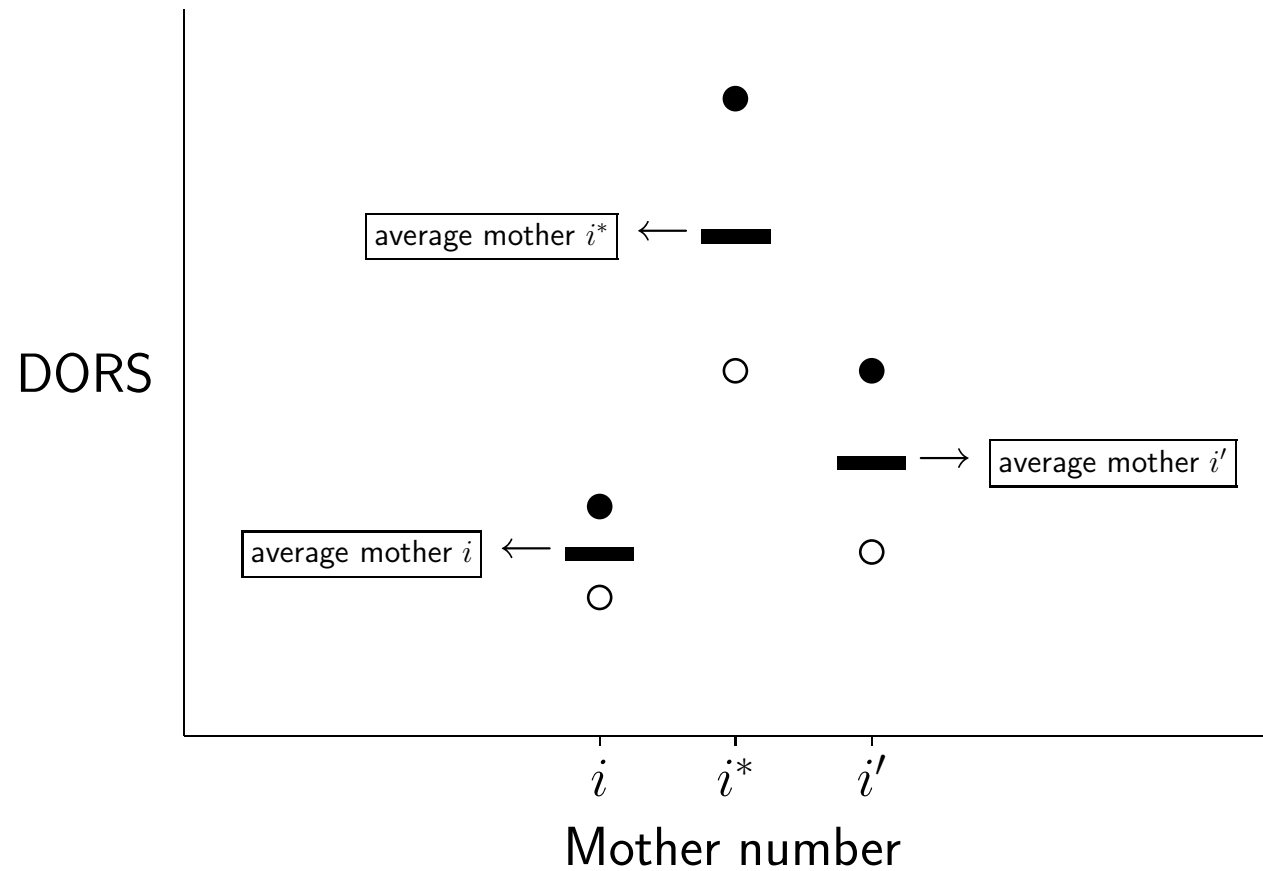
- Parameter interpretation:
 - ▷ Overall mean μ
 - ▷ Gender effect β_j
 - ▷ Mother effect α_i
- The parameter of interest is $\beta_2 - \beta_1$, the average difference between males and females
- Since the model is overparameterized, restrictions are needed, e.g.,
$$\sum_i \alpha_i = \sum_j \beta_j = 0$$
- Residual distribution: $\varepsilon_{ijk} \sim N(0, \sigma_{res}^2)$

- In order to better reflect the multilevel nature of the data, we will use an alternative parameterization **of the same model**, with one index for each level in the data structure.
- Let Y_{ij} be the j th measurement on the i th mother, and let x_{ij} be 0 for males and 1 for females.
- The model then equals:

$$Y_{ij} = \mu + \alpha_i + \beta x_{ij} + \varepsilon_{ij}$$

- The parameter of interest is β , the average difference between males and females
- We still need restrictions on the parameters α_i , e.g., $\sum_i \alpha_i = 0$
- Residual distribution: $\varepsilon_{ij} \sim N(0, \sigma_{res}^2)$

- Graphically:



- SAS program:

```
proc glm data = lizard;  
class mothc;  
model dors = sex mothc;  
run;
```

- Relevant SAS output:

Class Level Information

Class	Levels	Values
MOTHC	30	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

Dependent Variable: DORS

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	29	268.4685062	9.2575347	3.98	<.0001
Error	72	167.3746310	2.3246477		
Corrected Total	101	435.8431373			

R-Square	Coeff Var	Root MSE	DORS Mean
0.615975	4.351352	1.524680	35.03922

Source	DF	Type III SS	Mean Square
SEX	1	16.7253690	16.7253690
MOTH	28	256.9378690	9.1763525

Source	F Value	Pr > F
SEX	7.19	0.0091
MOTH	3.95	<.0001

- Note the highly significant mother effect.
- We now also obtain a significant gender effect.
- Many degrees of freedom are spent to the estimation of the mother effect, which is not even of interest.

4.3 Mixed models

- Note the different nature of the two factors:
 - ▷ SEX: defines 2 groups of interest
 - ▷ MOTHER: defines 29 groups not of real interest. A new sample would imply other mothers.
- In practice, one therefore considers the factor 'mother' as a **random factor**.
- The factor 'sex' is a **fixed effect**.
- Thus the model is a **mixed model**.
- In general, models can contain multiple fixed and/or random factors.

- The model is still of the form:

$$Y_{ij} = \mu + \alpha_i + \beta x_{ij} + \varepsilon_{ij}$$

- But the fact that mothers can be assumed to be randomly selected from a population of mothers is reflected in the additional assumption

$$\alpha_i \sim N(0, \sigma_{moth}^2)$$

- Note that we still have that the α_i have mean zero. Before, we had the restriction $\sum_i \alpha_i = 0$
- The normality assumption for the α_i is natural and mathematically convenient, **but not necessarily realistic.**
- Finally, all α_i and ε_{ij} are assumed independent.

4.4 Fitting mixed models in SAS

- Mixed model with 'sex' as fixed and 'mother' as random effect:

```
proc mixed data = lizard;  
class mothc;  
model dors = sex / solution;  
random mothc;  
run;
```

- Fixed effects are specified in the MODEL statement.
- Random effects are specified in the RANDOM statement.

- Relevant SAS-output:

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	436.17789431	
1	3	407.96849207	0.00072385
2	1	407.88032382	0.00001530
3	1	407.87858406	0.00000001

Convergence criteria met.

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	35.4904	0.3422	28	103.71	<.0001
SEX	-0.8289	0.3220	72	-2.57	0.0121

Covariance Parameter Estimates

Cov Parm	Estimate
MOTHC	1.7799
Residual	2.2501

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
SEX	1	72	6.63	0.0121

- Estimation method is iterative
- Note the significant difference between male and female animals ($p = 0.0121$)
- With the t -test, ignoring the mother effect, this was $p = 0.1024$.
- The average difference between males and females is estimated as $\hat{\beta} = 0.8289$
- Covariance parameter estimates:
 - ▷ σ_{moth}^2 represents the variability between mothers: $\hat{\sigma}_{moth}^2 = 1.78$
 - ▷ σ_{res}^2 represents the variability within mothers: $\hat{\sigma}_{res}^2 = 2.25$

4.5 The hierarchical versus marginal model

- Our mixed model was given by

$$Y_{ij} = \mu + \alpha_i + \beta x_{ij} + \varepsilon_{ij},$$

$$\alpha_i \sim N(0, \sigma_{moth}^2), \quad \varepsilon_{ij} \sim N(0, \sigma_{res}^2), \quad \text{independent}$$

- The above model can be rewritten as

$$Y_{ij} | \alpha_i \sim N(\mu + \alpha_i + \beta x_{ij}, \sigma_{res}^2), \quad \text{independent}$$

$$\alpha_i \sim N(0, \sigma_{moth}^2), \quad \text{independent}$$

- Each equation then corresponds to one level in the multilevel data structure
- The model is therefore called **the hierarchical model**

- The hierarchical model implies a specific **marginal model**, i.e., the model which describes the marginal distribution of the outcomes:

- ▷ Normal distribution

- ▷ Mean:

$$E(Y_{ij}) = \mu + \beta x_{ij}$$

- ▷ Variance:

$$\begin{aligned}\text{Var}(Y_{ij}) &= \text{Var}(\mu + \alpha_i + \beta x_{ij} + \varepsilon_{ij}) = \text{Var}(\alpha_i + \varepsilon_{ij}) \\ &= \text{Var}(\alpha_i) + \text{Var}(\varepsilon_{ij}) = \sigma_{moth}^2 + \sigma_{res}^2\end{aligned}$$

- ▷ Covariance between observations from different mothers i and i^* :

$$\begin{aligned}\text{Cov}(Y_{ij}, Y_{i^*k}) &= \text{Cov}(\mu + \alpha_i + \beta x_{ij} + \varepsilon_{ij}, \mu + \alpha_{i^*} + \beta x_{i^*k} + \varepsilon_{i^*k}) \\ &= \text{Cov}(\alpha_i, \alpha_{i^*}) + \text{Cov}(\alpha_i, \varepsilon_{i^*k}) + \text{Cov}(\varepsilon_{ij}, \alpha_{i^*}) + \text{Cov}(\varepsilon_{ij}, \varepsilon_{i^*k}) \\ &= 0\end{aligned}$$

▷ Covariance between observations j and k from the same mother i ($j \neq k$):

$$\begin{aligned}\text{Cov}(Y_{ij}, Y_{ik}) &= \text{Cov}(\mu + \alpha_i + \beta x_{ij} + \varepsilon_{ij}, \mu + \alpha_i + \beta x_{ik} + \varepsilon_{ik}) \\ &= \text{Cov}(\alpha_i, \alpha_i) + \text{Cov}(\alpha_i, \varepsilon_{ik}) + \text{Cov}(\varepsilon_{ij}, \alpha_i) + \text{Cov}(\varepsilon_{ij}, \varepsilon_{ik}) \\ &= \text{Var}(\alpha_i) = \sigma_{moth}^2\end{aligned}$$

- The total variability, correcting for gender differences is decomposed as within-cluster variability and between-cluster variability:

$$\begin{aligned}\sigma^2 &= \sigma_{moth}^2 + \sigma_{res}^2 \\ 4.03 &= 1.78 + 2.25\end{aligned}$$

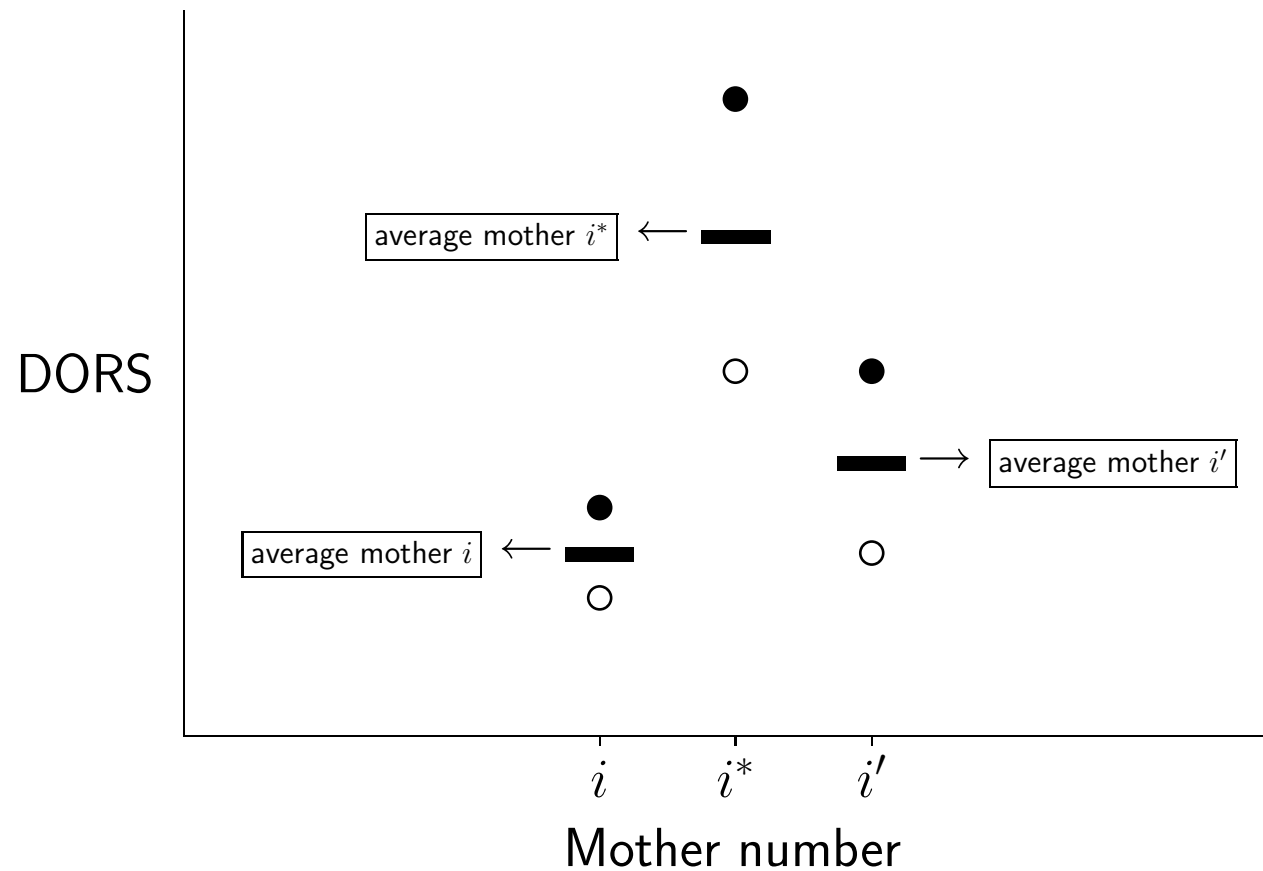
- The ‘mother’ factor explains $1.78/4.03 = 44\%$ of the total variability, after correction for gender
- Observations from different mothers are assumed independent

- Observations from the same mother are correlated with correlation coefficient

$$\rho_I = \text{Corr}(Y_{ij}, Y_{ik}) = \frac{\sigma_{moth}^2}{\sigma_{moth}^2 + \sigma_{res}^2} = \frac{1.78}{1.78 + 2.25} = 0.44$$

- The correlation ρ_I is called **intracluster correlation**
- Note how the mixed model accounts for the correlation in the data through the random effects α_i .
- The correlation will be high in cases with much between-cluster variability, relative to the within-cluster variability
- The correlation will be low in cases with little between-cluster variability, relative to the within-cluster variability

- Graphically:



- Much between-cluster variability implies that observations from the same cluster are 'more alike' than observations from different clusters

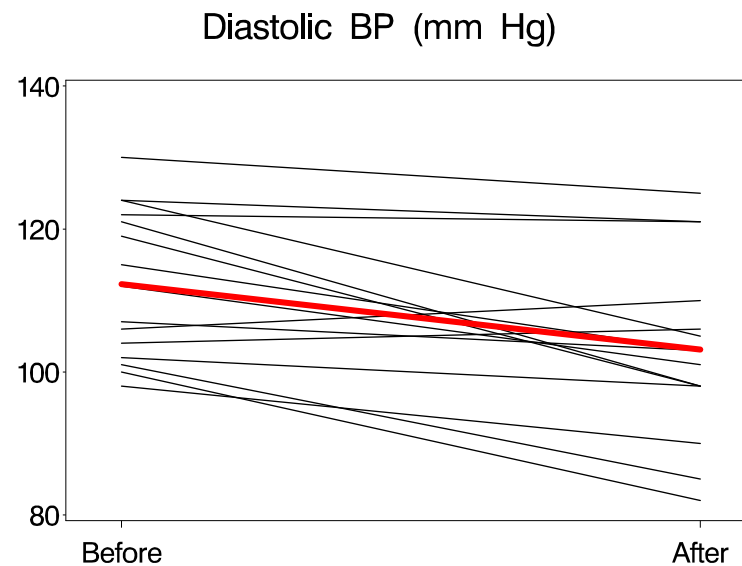
Chapter 5

The paired t -test revisited

- ▷ Example: The Captopril data
- ▷ Analysis in SAS
- ▷ The hierarchical versus marginal model
- ▷ Conclusion

5.1 Example: The Captopril data

- A paired t -test analysis of the Captopril data yields:



Variable	Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	p
DIA_VOOR	112,3333	10,47219				
DIA_NA	103,0667	12,55540	15	9,266667	8,614495	0,000951

- An alternative analysis could be based on a mixed model
- Let Y_{ij} be the observation for the i th subject, taken at time point $t_j = 0, 1$:

$$t_j = \begin{cases} 0 & \text{if before treatment} \\ 1 & \text{if after treatment} \end{cases}$$

- The mixed model is then of the form:

$$Y_{ij} = \mu + \alpha_i + \beta t_j + \varepsilon_{ij},$$

$$\alpha_i \sim N(0, \sigma_{subj}^2), \quad \varepsilon_{ij} \sim N(0, \sigma_{res}^2), \quad \text{independent}$$

- The α_i are subject-specific effects, reflecting that some patients naturally have higher BP's than others, irrespective of the treatment

- Assuming that subjects are randomly sampled from a population of patients, it is natural to assume the α_i to be random.
- The α_i reflect the variability between patients

5.2 Analysis in SAS

- SAS program:

```
proc mixed data=capto;  
class subject;  
model y = time / solution;  
random subject;  
run;
```

- Relevant SAS-output:

```
Class Level Information  
  
Class      Levels  Values  
  
subject      15    1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
```

```
Covariance Parameter  
Estimates  
  
Cov Parm      Estimate  
  
subject      96.5476  
Residual     37.1048
```

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	112.33	2.9850	14	37.63	<.0001
time	-9.2667	2.2243	14	-4.17	0.0010

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
time	1	14	17.36	0.0010

- The average difference in BP is estimated as $\hat{\beta} = 9.27$
- We obtain the same result as with the paired t -test:

$$F = t^2 = \left[\frac{9.27}{\frac{8.61}{\sqrt{15}}} \right]^2 = 17.36, \quad 14 \text{ degrees of freedom}$$

- Covariance parameter estimates:
 - ▷ σ_{subj}^2 represents the variability between patients: $\hat{\sigma}_{subj}^2 = 96.55$
 - ▷ σ_{res}^2 represents the variability within patients: $\hat{\sigma}_{res}^2 = 37.10$

5.3 The hierarchical versus marginal model

- The mixed model can again be viewed as a hierarchical model:

$$Y_{ij}|\alpha_i \sim N(\mu + \alpha_i + \beta t_j, \sigma_{res}^2), \quad \text{independent}$$

$$\alpha_i \sim N(0, \sigma_{subj}^2), \quad \text{independent}$$

- The implied marginal model is again a normal one:

- ▷ Expectation $E(Y_{ij}) = \mu + \beta t_j$

- ▷ Variance $\sigma^2 = \text{Var}(Y_{ij}) = \sigma_{subj}^2 + \sigma_{res}^2$

- ▷ Observations from different patients independent

- ▷ Observations from the same patient correlated:

$$\rho_I = \text{Corr}(Y_{i1}, Y_{i2}) = \frac{\sigma_{subj}^2}{\sigma_{subj}^2 + \sigma_{res}^2}$$

- In our example, the total variability, not explained by the systematic treatment effect, equals:

$$\sigma^2 = \sigma_{subj}^2 + \sigma_{res}^2 = 96.55 + 37.10 = 133.65$$

- The between-subject variability accounts for $96.55/133.65 = 72.24\%$ of all variability
- The within-subject correlation is given by

$$\rho_I = \frac{\sigma_{subj}^2}{\sigma_{subj}^2 + \sigma_{res}^2} = \frac{96.55}{96.55 + 37.10} = 0.7224$$

- The above intraclass correlation does not equal the Pearson correlation between the BP before and after treatment, which equals $\rho = 0.7343$.
- The reason for this difference is that the Pearson correlation does not assume the variances of the BP before and after treatment to be equal.

- The mixed model used assumes constant variance:

$$\sigma^2 = \text{Var}(Y_{i1}) = \text{Var}(Y_{i2}) = \sigma_{subj}^2 + \sigma_{res}^2 = 133.65, \quad \sigma = 11.56$$

- Summary statistics for both measurements:

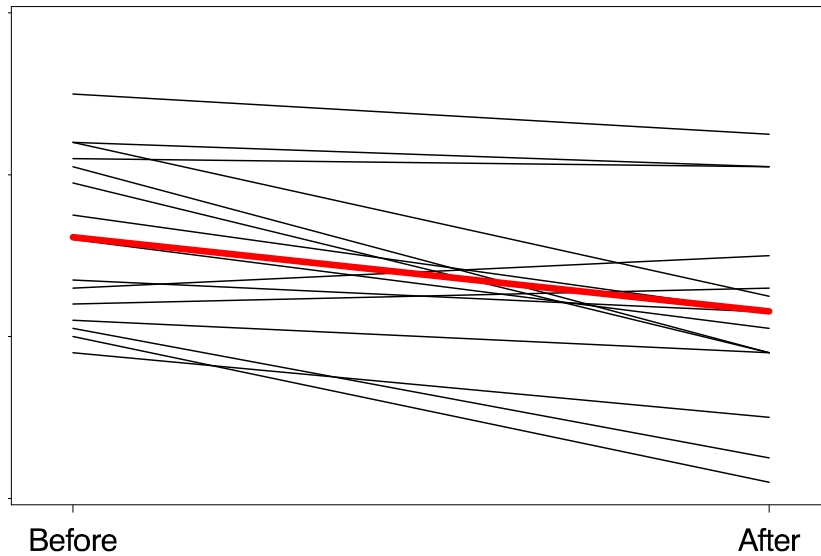
Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Before	15	112.33333	10.47219	1685	98.00000	130.00000
After	15	103.06667	12.55540	1546	82.00000	125.00000

- Note again that the correlation arises from the large amount of between-subject variability, relative to the within-subject variability:

$$\rho_I = \text{Corr}(Y_{i1}, Y_{i2}) = \frac{\sigma_{subj}^2}{\sigma_{subj}^2 + \sigma_{res}^2}$$

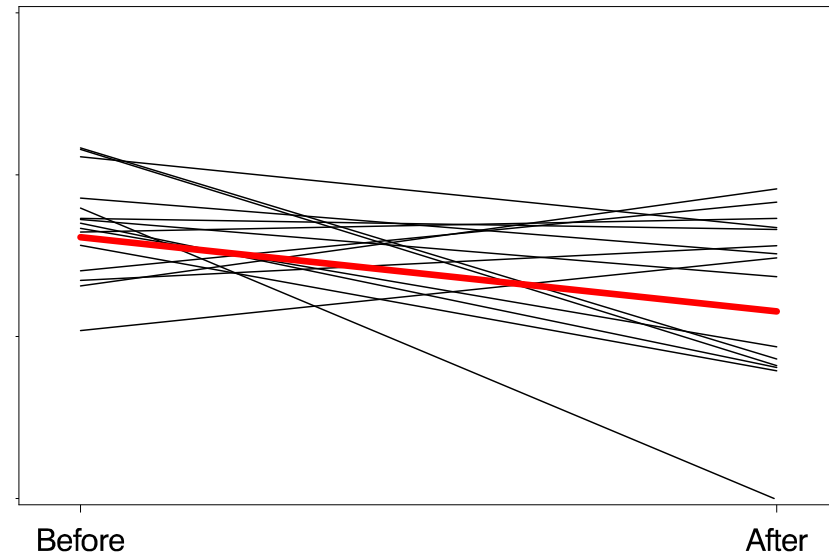
- Graphically:

Much between – cluster variability



Strong correlation

Little between – cluster variability



Weak correlation

5.4 Conclusion

- The simplest example of clustered data are paired observations, typically analyzed using a paired t -test.
- Traditionally, the within-pair correlation is circumvented by taking within-pair differences $\Delta_i = Y_{i1} - Y_{i2}$ which are then analysed using a one-sample t -test
- Hence, mixed models can be viewed as an extension of the paired t -test to:
 - ▷ more than 2 observations per cluster
 - ▷ unbalanced data: unequal number of measurements per cluster
 - ▷ models with covariates, e.g., 'sex', or others
 - ▷ models with multiple random effects (see later)

Chapter 6

The growth curves data

- ▷ Example
- ▷ The model
- ▷ Analysis in SAS
- ▷ The hierarchical versus marginal model
- ▷ ESTIMATE and CONTRAST statements

6.1 Example

- Taken from Goldstein (1979).

- Research question:

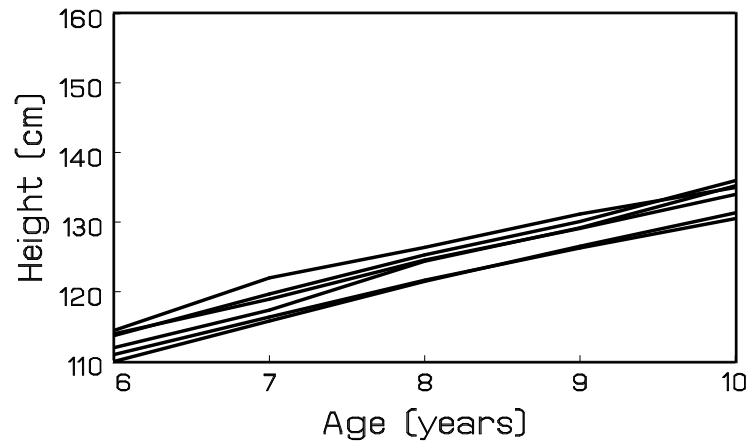
Is growth related to height of mother ?

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

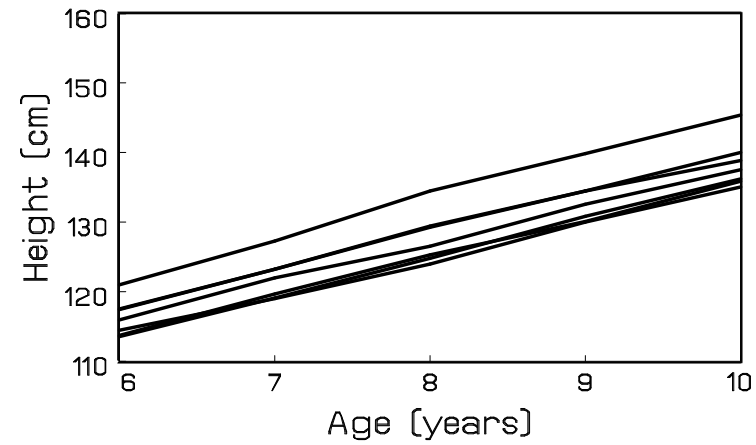
	Mothers height	Children numbers
Small mothers	< 155 cm	1 → 6
Medium mothers	[155cm; 164cm]	7 → 13
Tall mothers	> 164 cm	14 → 20

- Individual profiles:

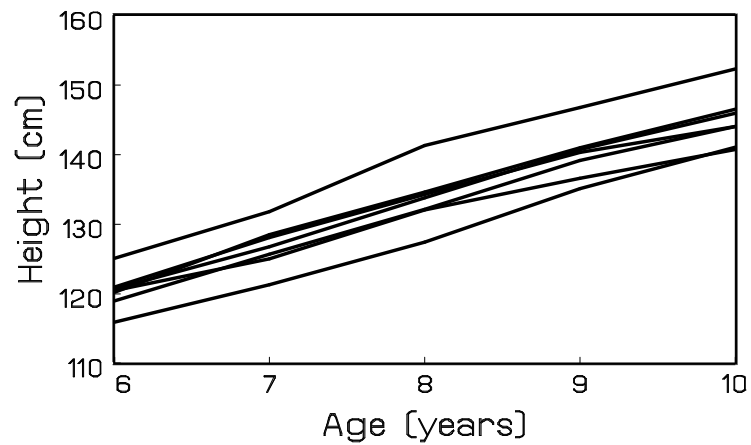
Short mother



Medium mother



Tall mother



- 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

6.2 The model

- We will assume a linear relation between Age and Height, possibly different for the different groups.
- With cross-sectional data, the appropriate model would be an ANCOVA model:
 - ▷ Covariate Age
 - ▷ Factor Group
 - ▷ Interaction Age*Group
- With longitudinal data, the observations are clustered within children, implying within-child correlation
- Correction for the variability between children is done through a random child effect.

- As before, let Y_{ij} be the j th measurement of height for the i th cluster (child), taken at time t_j (age). Our model is then of the form:

$$Y_{ij} = \begin{cases} \beta_1 + b_i + \beta_2 t_j + \varepsilon_{ij}, & \text{if short mother} \\ \beta_3 + b_i + \beta_4 t_j + \varepsilon_{ij}, & \text{if medium mother} \\ \beta_5 + b_i + \beta_6 t_j + \varepsilon_{ij}, & \text{if tall mother} \end{cases}$$

- As before, it is assumed that random effects b_i are normal with mean zero and variance σ_{child}^2 .
- The errors ε_{ij} are normal with mean zero and variance σ_{res}^2 .

6.3 Analysis in SAS

- SAS program:

```
proc mixed data = growth;  
class group child;  
model height = age group age*group / solution;  
random child;  
run;
```

- Relevant SAS output:

Solution for Fixed Effects

Covariance Parameter Estimates		Effect	GROUP	Estimate	Standard Error	DF	t Value	Pr > t
		Intercept		83.1229	1.4162	17	58.69	<.0001
Cov Parm	Estimate	AGE		6.2486	0.1049	77	59.59	<.0001
		GROUP	1	-1.8229	2.0846	77	-0.87	0.3846
CHILD	8.9603	GROUP	2	-0.1486	2.0028	77	-0.07	0.9411
Residual	0.7696	GROUP	3	0
		AGE*GROUP	1	-0.9786	0.1543	77	-6.34	<.0001
		AGE*GROUP	2	-0.6814	0.1483	77	-4.60	<.0001
		AGE*GROUP	3	0

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
AGE	1	77	8385.15	<.0001
GROUP	2	77	0.46	0.6330
AGE*GROUP	2	77	21.66	<.0001

- The hypothesis of interest is $H_0 : \beta_2 = \beta_4 = \beta_6$, which corresponds to testing the interaction Age*Group
- We find a highly significant difference between the slopes from the three groups ($p < 0.0001$)
- Covariance parameter estimates:
 - ▷ σ_{child}^2 represents the variability between children: $\hat{\sigma}_{child}^2 = 8.96$
 - ▷ σ_{res}^2 represents the variability within children: $\hat{\sigma}_{res}^2 = 0.77$

6.4 The hierarchical versus marginal model

- The mixed model can again be viewed as a hierarchical model:

$$Y_{ij}|b_i \sim \begin{cases} N(\beta_1 + b_i + \beta_2 t_j, \sigma_{res}^2), & \text{if short mother} \\ N(\beta_3 + b_i + \beta_4 t_j, \sigma_{res}^2), & \text{if medium mother} \\ N(\beta_5 + b_i + \beta_6 t_j, \sigma_{res}^2), & \text{if tall mother} \end{cases}$$

- The implied marginal model is again a normal one:

▷ Expectation

$$E(Y_{ij}) = \begin{cases} \beta_1 + \beta_2 t_j, & \text{if short mother} \\ \beta_3 + \beta_4 t_j, & \text{if medium mother} \\ \beta_5 + \beta_6 t_j, & \text{if tall mother} \end{cases}$$

- ▷ Variance $\sigma^2 = \text{Var}(Y_{ij}) = \sigma_{child}^2 + \sigma_{res}^2$
- ▷ Observations from different children independent
- ▷ Observations from the same child correlated:

$$\rho_I = \text{Corr}(Y_{i1}, Y_{i2}) = \frac{\sigma_{child}^2}{\sigma_{child}^2 + \sigma_{res}^2}$$

- In our example, the total variability, not explained by the systematic trends, equals:

$$\sigma^2 = \sigma_{child}^2 + \sigma_{res}^2 = 8.96 + 0.77 = 9.73$$

- The between-child variability accounts for $8.96/9.73 = 92\%$ of all variability
- The within-child correlation is given by

$$\rho_I = \frac{\sigma_{child}^2}{\sigma_{child}^2 + \sigma_{res}^2} = \frac{8.96}{8.96 + 0.77} = 0.9209$$

6.5 ESTIMATE and CONTRAST statements

- As in many other SAS procedures, ESTIMATE and CONTRAST statements can be used to obtain inferences about specific contrasts of the fixed effects.
- Slopes for each group separately, as well as pairwise comparisons are obtained using the following program:

```
proc mixed data=growth;
class child group;
model height = group age*group / noint solution;
random child;
contrast 'small-medium' group*age 1 -1 0;
contrast 'small-tall' group*age 1 0 -1;
contrast 'medium-tall' group*age 0 1 -1;
estimate 'small' group*age 1 0 0 / cl;
estimate 'medium' group*age 0 1 0 / cl;
estimate 'tall' group*age 0 0 1 / cl;
run;
```

- Note the different parameterization for the fixed effects, when compared to the original program:

```
proc mixed data = growth;
class group child;
model height = age group age*group / solution;
random child;
run;
```

- Relevant SAS output:

Solution for Fixed Effects

Effect	GROUP	Estimate	Standard Error	DF	t Value	Pr > t
GROUP	1	81.3000	1.5297	77	53.15	<.0001
GROUP	2	82.9743	1.4162	77	58.59	<.0001
GROUP	3	83.1229	1.4162	77	58.69	<.0001
AGE*GROUP	1	5.2700	0.1133	77	46.53	<.0001
AGE*GROUP	2	5.5671	0.1049	77	53.10	<.0001
AGE*GROUP	3	6.2486	0.1049	77	59.59	<.0001

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
GROUP	3	77	3234.13	<.0001
AGE*GROUP	3	77	2845.30	<.0001

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
small	5.2700	0.1133	77	46.53	<.0001	0.05	5.0445	5.4955
medium	5.5671	0.1049	77	53.10	<.0001	0.05	5.3584	5.7759
tall	6.2486	0.1049	77	59.59	<.0001	0.05	6.0398	6.4574

Contrasts

Label	Num DF	Den DF	F Value	Pr > F
small-medium	1	77	3.71	0.0579
small-tall	1	77	40.20	<.0001
medium-tall	1	77	21.12	<.0001

- The new parameterization implies completely different tests.
- For example, the tests reported for the Age*Group effect, under both parameterizations correspond to the hypotheses:

model height = age group age*group;	model height = group age*group / noint;
↓	↓
$H_0 : \beta_2 = \beta_4 = \beta_6$	$H_0 : \beta_2 = \beta_4 = \beta_6 = 0$

- The difference between the slopes is mainly explained from the difference between the third group on one hand, and the other two groups on the other hand.

Chapter 7

The linear mixed model

- ▷ Random intercepts model
- ▷ Remarks
- ▷ The linear mixed model
- ▷ Analysis in SAS
- ▷ The hierarchical versus marginal model
- ▷ Conclusion and terminology

7.1 Random intercepts model

- The model, used to describe the growth curves, was:

$$Y_{ij} = \begin{cases} (\beta_1 + b_i) + \beta_2 t_{ij} + \varepsilon_{ij}, & \text{if short mother} \\ (\beta_3 + b_i) + \beta_4 t_{ij} + \varepsilon_{ij}, & \text{if medium mother} \\ (\beta_5 + b_i) + \beta_6 t_{ij} + \varepsilon_{ij}, & \text{if tall mother} \end{cases}$$

- This can be interpreted as a ANCOVA model, but with child-specific intercepts b_i
- Such a b_i represents the deviation of the intercept of a specific child from the average intercept in the group to which that child belongs, i.e., deviation from β_1 , β_2 , or β_3 .

- An alternative way to fit a random intercepts model in PROC MIXED is:

```
proc mixed data = growth;  
class group child;  
model height = age group age*group / solution;  
random intercept / subject=child;  
run;
```

- The results are identical to those discussed earlier.
- From now on, the mixed model can also be interpreted as a subject-specific regression model, i.e., a regression model with subject-specific regression parameters.

7.2 Remarks

- The growth-curve dataset is an example of a longitudinal dataset
- In longitudinal data, there is a natural ordering of the measurements within clusters
- The ordering is of primary interest
- Our random-intercepts model implies very strong assumptions:
 - ▷ Parallel profiles within all 3 groups
 - ▷ Constant variance $\sigma^2 = \sigma_{child}^2 + \sigma_{res}^2$
 - ▷ Constant correlation within children: $\sigma_{child}^2 / (\sigma_{child}^2 + \sigma_{res}^2)$

- Hence, the marginal model implicitly assumes that the variance remains constant over time and that the correlation is the same between **any two** measurements from the same subject
- In the case of longitudinal data, this is often not realistic
- For example, the covariance and correlation matrix of the residuals from the ANCOVA model equal:

$$\begin{pmatrix} 8.7041 & 9.6119 & 11.4005 & 10.2351 & 8.5174 \\ 9.6119 & 11.3896 & 13.1437 & 11.9719 & 10.2474 \\ 11.4005 & 13.1437 & 15.8781 & 14.3981 & 12.6611 \\ 10.2351 & 11.9719 & 14.3981 & 13.4490 & 12.0644 \\ 8.5174 & 10.2474 & 12.6611 & 12.0644 & 12.0655 \end{pmatrix} \quad \begin{pmatrix} 1.0000 & 0.9654 & 0.9697 & 0.9460 & 0.8311 \\ 0.9654 & 1.0000 & 0.9774 & 0.9673 & 0.8742 \\ 0.9697 & 0.9774 & 1.0000 & 0.9853 & 0.9147 \\ 0.9460 & 0.9673 & 0.9853 & 1.0000 & 0.9471 \\ 0.8311 & 0.8742 & 0.9147 & 0.9471 & 1.0000 \end{pmatrix}$$

- This is the key motivation to further extend our mixed model

7.3 The linear mixed model

- One way to extend the random-intercepts model is to also allow the slopes to be subject-specific:

$$Y_{ij} = \begin{cases} (\beta_1 + b_{1i}) + (\beta_2 + b_{2i})t_{ij} + \varepsilon_{ij}, & \text{if short mother} \\ (\beta_3 + b_{1i}) + (\beta_4 + b_{2i})t_{ij} + \varepsilon_{ij}, & \text{if medium mother} \\ (\beta_5 + b_{1i}) + (\beta_6 + b_{2i})t_{ij} + \varepsilon_{ij}, & \text{if tall mother} \end{cases}$$

- As before, the random effects are assumed to be normally distributed with mean zero:

$$\mathbf{b}_i = (b_{1i}, b_{2i})' \sim N(\mathbf{0}, D)$$

- The residuals ε_{ij} are still i.i.d. $N(0, \sigma^2)$, independent of the random effects \mathbf{b}_i .

- D then equals the 2×2 covariance matrix of the random effects:

$$D = \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix}$$

- Interpretation of the parameters:
 - ▷ d_{11} equals the variance of the intercepts b_{1i}
 - ▷ d_{22} equals the variance of the slopes b_{2i}
 - ▷ d_{12} equals the covariance between the intercepts b_{1i} and the slopes b_{2i} .
 - ▷ The correlation between the intercepts and slopes then equals:

$$\text{Corr}(b_{1i}, b_{2i}) = \frac{d_{12}}{\sqrt{d_{11}}\sqrt{d_{22}}}$$

7.4 Analysis in SAS

- SAS program:

```
proc mixed data=growth;
class child group;
model height=age group age*group;
random intercept age / type=un subject=child g gcorr;
run;
```

- As before, fixed effects are to be specified in the MODEL statement, while random effects are specified in the RANDOM statement.
- The option 'type=un' requires an unstructured covariance D , i.e., two variances d_{11} and d_{22} , and one covariance d_{12} , with only restriction that D is positive (semi-)definite.
- The options 'g' and 'gcorr' require the printout of the matrix D (in SAS termed G) and associated correlation matrix.

- Relevant SAS output:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
UN(1,1)	CHILD	7.6028
UN(2,1)	CHILD	-0.4437
UN(2,2)	CHILD	0.1331
Residual		0.4758

Estimated G Matrix

Row	Effect	CHILD	Col1	Col2
1	Intercept	1	7.6028	-0.4437
2	AGE	1	-0.4437	0.1331

Estimated G Correlation Matrix

Row	Effect	CHILD	Col1	Col2
1	Intercept	1	1.0000	-0.4412
2	AGE	1	-0.4412	1.0000

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
AGE	1	17	3572.36	<.0001
GROUP	2	60	0.60	0.5514
AGE*GROUP	2	60	9.23	0.0003

- We still get a highly significant interaction term.
- Covariance parameters:
 - ▷ d_{11} represents the variability in subject-specific intercepts: $\hat{d}_{11} = 7.6028$
 - ▷ d_{22} represents the variability in subject-specific slopes: $\hat{d}_{22} = 0.1331$
 - ▷ d_{12} represents the covariance between subject-specific intercepts and slopes:
 $\hat{d}_{12} = -0.4437$
 - ▷ the correlation between subject-specific intercepts and slopes is estimated as:

$$\widehat{\text{Corr}}(b_{1i}, b_{2i}) = \frac{\hat{d}_{12}}{\sqrt{\hat{d}_{11}}\sqrt{\hat{d}_{22}}} = -0.4412$$

- ▷ σ^2 represents the variability within children: $\hat{\sigma}^2 = 0.4758$

- Note the differences in test results for the fixed effects, when compared to those from the earlier random intercepts model:

NOW

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
AGE	1	17	3572.36	<.0001
GROUP	2	60	0.60	0.5514
AGE*GROUP	2	60	9.23	0.0003

BEFORE

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
AGE	1	77	8385.15	<.0001
GROUP	2	77	0.46	0.6330
AGE*GROUP	2	77	21.66	<.0001

7.5 The hierarchical versus marginal model

- The mixed model can again be viewed as a hierarchical model:

$$Y_{ij} | \mathbf{b}_i \sim \begin{cases} N[(\beta_1 + b_{1i}) + (\beta_2 + b_{2i})t_j, \sigma^2], & \text{if short mother} \\ N[(\beta_3 + b_{1i}) + (\beta_4 + b_{2i})t_j, \sigma^2], & \text{if medium mother} \\ N[(\beta_5 + b_{1i}) + (\beta_6 + b_{2i})t_j, \sigma^2], & \text{if tall mother} \end{cases}$$

- The implied marginal model is again a normal one:
 - ▷ The expectation is the same as under the random intercepts model:

$$E(Y_{ij}) = \begin{cases} \beta_1 + \beta_2 t_j, & \text{if short mother} \\ \beta_3 + \beta_4 t_j, & \text{if medium mother} \\ \beta_5 + \beta_6 t_j, & \text{if tall mother} \end{cases}$$

▷ Variance:

$$\begin{aligned}\text{Var}(Y_{ij}) &= \text{Var}(\beta_1 + b_{1i} + \beta_2 t_j + b_{2i} t_j + \varepsilon_{ij}) \\ &= \text{Var}(b_{1i} + b_{2i} t_j + \varepsilon_{ij}) \\ &= \text{Var}(b_{2i} t_j) + 2\text{Cov}(b_{1i}, b_{2i} t_j) + \text{Var}(b_{1i}) + \text{Var}(\varepsilon_{ij}) \\ &= d_{22} t_j^2 + 2d_{12} t_j + d_{11} + \sigma^2\end{aligned}$$

▷ Covariance between observations from different children i and i^* :

$$\begin{aligned}\text{Cov}(Y_{ij}, Y_{i^*k}) &= \text{Cov}(\beta_1 + b_{1i} + \beta_2 t_j + b_{2i} t_j + \varepsilon_{ij}, \beta_1 + b_{1i^*} + \beta_2 t_k + b_{2i^*} t_k + \varepsilon_{i^*k}) \\ &= \text{Cov}(b_{1i} + b_{2i} t_j + \varepsilon_{ij}, b_{1i^*} + b_{2i^*} t_k + \varepsilon_{i^*k}) \\ &= 0\end{aligned}$$

▷ Covariance between observations j and k from the same child i ($j \neq k$):

$$\begin{aligned}
 & \text{Cov}(Y_{ij}, Y_{ik}) \\
 &= \text{Cov}(\beta_1 + b_{1i} + \beta_2 t_j + b_{2i} t_j + \varepsilon_{ij}, \beta_1 + b_{1i} + \beta_2 t_k + b_{2i} t_k + \varepsilon_{ik}) \\
 &= \text{Cov}(b_{1i} + b_{2i} t_j + \varepsilon_{ij}, b_{1i} + b_{2i} t_k + \varepsilon_{ik}) \\
 &= \text{Cov}(b_{1i}, b_{1i}) + \text{Cov}(b_{1i}, b_{2i} t_k) + \text{Cov}(b_{2i} t_j, b_{1i}) + \text{Cov}(b_{2i} t_j, b_{2i} t_k) \\
 &= \text{Var}(b_{1i}) + \text{Cov}(b_{1i}, b_{2i}) t_k + \text{Cov}(b_{2i}, b_{1i}) t_j + \text{Var}(b_{2i}, b_{2i}) t_j t_k \\
 &= d_{22} t_j t_k + d_{12} (t_j + t_k) + d_{11}
 \end{aligned}$$

▷ Correlation between observations j and k from the same child i ($j \neq k$):

$$\text{Corr}(Y_{ij}, Y_{ik}) = \frac{d_{22} t_j t_k + d_{12} (t_j + t_k) + d_{11}}{\sqrt{d_{22} t_j^2 + 2d_{12} t_j + d_{11} + \sigma^2} \sqrt{d_{22} t_k^2 + 2d_{12} t_k + d_{11} + \sigma^2}}$$

- Note how extending the random intercepts model with random slopes yields a more flexible covariance structure.
- Further extension of the random effects structure would allow for even more flexible variance and correlations functions.
- Note, however, that the covariance structure, implied by the random-effects model, is not necessarily a good description for the data set at hand.
- For example, the fitted variance function for the growth curves equals:

$$\begin{aligned}\widehat{\text{Var}}(Y_{ij}) &= \widehat{d}_{22}t_j^2 + 2\widehat{d}_{12}t_j + \widehat{d}_{11} + \widehat{\sigma}^2 \\ &= 0.1331t_j^2 + 2(-0.4437)t_j + 7.6028 + 0.4758\end{aligned}$$

- In SAS, the fitted covariance and correlation matrices can be obtained from the 'v' and 'vcorr' options in the RANDOM statement:

```
random intercept age / type=un subject=child v vcorr;
```

- Fitted covariance and correlation matrices:

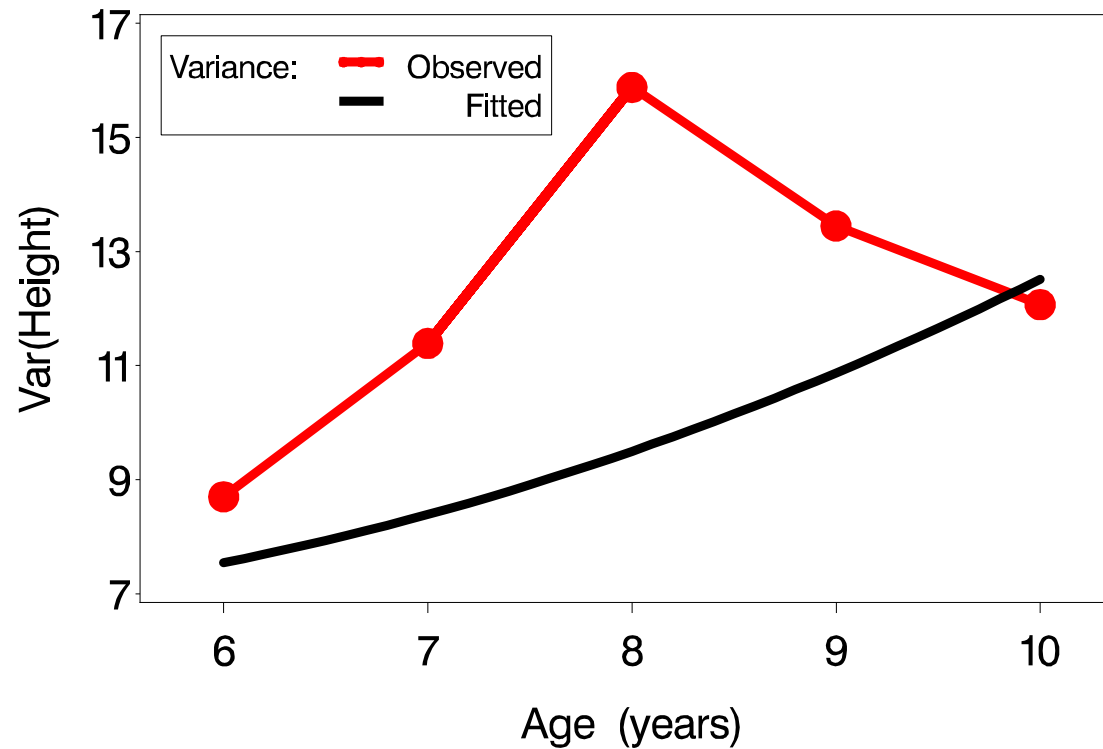
$$\begin{pmatrix} 7.5442 & 7.4230 & 7.7776 & 8.1322 & 8.4869 \\ 7.4230 & 8.3865 & 8.3983 & 8.8860 & 9.3737 \\ 7.7776 & 8.3983 & 9.4949 & 9.6398 & 10.2606 \\ 8.1322 & 8.8860 & 9.6398 & 10.8694 & 11.1474 \\ 8.4869 & 9.3737 & 10.2606 & 11.1474 & 12.5101 \end{pmatrix} \quad \begin{pmatrix} 1.0000 & 0.9332 & 0.9190 & 0.8981 & 0.8736 \\ 0.9332 & 1.0000 & 0.9411 & 0.9307 & 0.9151 \\ 0.9190 & 0.9411 & 1.0000 & 0.9489 & 0.9414 \\ 0.8981 & 0.9307 & 0.9489 & 1.0000 & 0.9560 \\ 0.8736 & 0.9151 & 0.9414 & 0.9560 & 1.0000 \end{pmatrix}$$

- The observed covariance and correlation matrix of the residuals from the ANCOVA model equal:

$$\begin{pmatrix} 8.7041 & 9.6119 & 11.4005 & 10.2351 & 8.5174 \\ 9.6119 & 11.3896 & 13.1437 & 11.9719 & 10.2474 \\ 11.4005 & 13.1437 & 15.8781 & 14.3981 & 12.6611 \\ 10.2351 & 11.9719 & 14.3981 & 13.4490 & 12.0644 \\ 8.5174 & 10.2474 & 12.6611 & 12.0644 & 12.0655 \end{pmatrix} \quad \begin{pmatrix} 1.0000 & 0.9654 & 0.9697 & 0.9460 & 0.8311 \\ 0.9654 & 1.0000 & 0.9774 & 0.9673 & 0.8742 \\ 0.9697 & 0.9774 & 1.0000 & 0.9853 & 0.9147 \\ 0.9460 & 0.9673 & 0.9853 & 1.0000 & 0.9471 \\ 0.8311 & 0.8742 & 0.9147 & 0.9471 & 1.0000 \end{pmatrix}$$

- Graphically:

Observed and fitted variance function



- Obviously, the variance cannot be described by a quadratic function with **positive** curvature.

- One way to further extend the marginal covariance structure is to add random effects to the model, e.g., random coefficients for Age²:

```
random intercept age age*age / type=un subject=child v vcorr;
```

- New fitted covariance matrix, compared to observed covariance from ANCOVA residuals:

$$\begin{pmatrix} 8.2014 & 10.0364 & 10.9352 & 10.6322 & 9.1276 \\ 10.0364 & 12.9310 & 13.8330 & 13.5387 & 11.7826 \\ 10.9352 & 13.8330 & 15.4548 & 15.0042 & 13.2776 \\ 10.6322 & 13.5387 & 15.0042 & 15.2944 & 13.6127 \\ 9.1276 & 11.7826 & 13.2776 & 13.6127 & 13.0531 \end{pmatrix} \quad \begin{pmatrix} 8.7041 & 9.6119 & 11.4005 & 10.2351 & 8.5174 \\ 9.6119 & 11.3896 & 13.1437 & 11.9719 & 10.2474 \\ 11.4005 & 13.1437 & 15.8781 & 14.3981 & 12.6611 \\ 10.2351 & 11.9719 & 14.3981 & 13.4490 & 12.0644 \\ 8.5174 & 10.2474 & 12.6611 & 12.0644 & 12.0655 \end{pmatrix}$$

- The fitted variance function is now 4 degree polynomial:

$$\text{Var}(Y_{ij}) = d_{33}t_j^4 + 2d_{23}t_j^3 + d_{22}t_j^2 + 2d_{13}t_j^2 + 2d_{12}t_j + d_{11} + \sigma^2$$

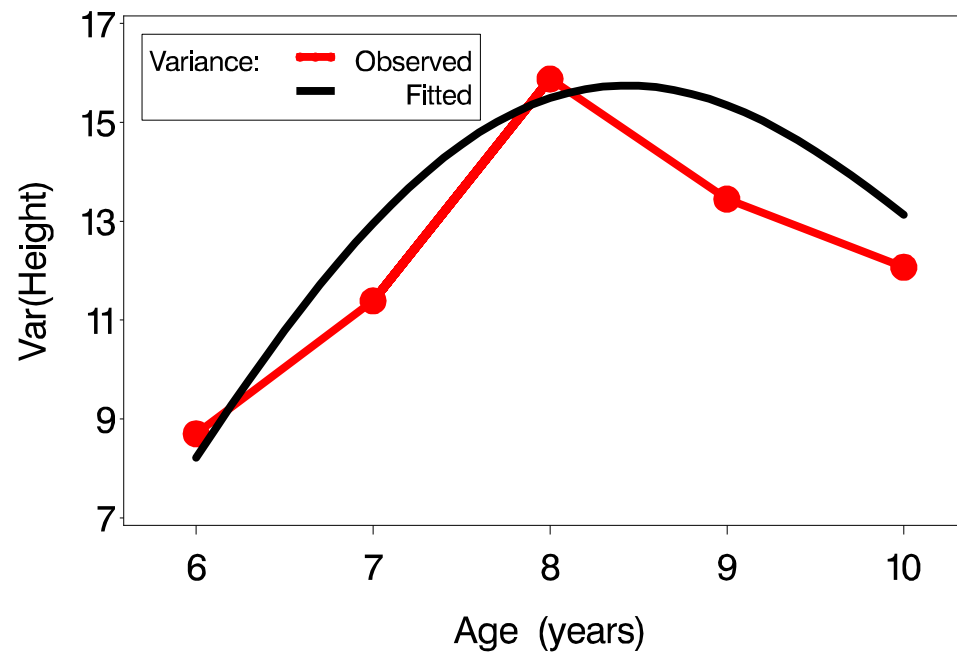
- Estimated random-effects covariance:

Estimated G Matrix

Row	Effect	CHILD	Col1	Col2	Col3
1	Intercept	1	96.3384	-33.4752	2.0725
2	AGE	1	-33.4752	11.5273	-0.7160
3	AGE*AGE	1	2.0725	-0.7160	0.04508

- Graphically:

Observed and fitted variance function



- New fitted correlation matrix, compared to observed correlation from ANCOVA residuals:

$$\begin{pmatrix} 1.0000 & 0.9746 & 0.9713 & 0.9493 & 0.8822 \\ 0.9746 & 1.0000 & 0.9785 & 0.9627 & 0.9069 \\ 0.9713 & 0.9785 & 1.0000 & 0.9759 & 0.9348 \\ 0.9493 & 0.9627 & 0.9759 & 1.0000 & 0.9634 \\ 0.8822 & 0.9069 & 0.9348 & 0.9634 & 1.0000 \end{pmatrix} \quad \begin{pmatrix} 1.0000 & 0.9654 & 0.9697 & 0.9460 & 0.8311 \\ 0.9654 & 1.0000 & 0.9774 & 0.9673 & 0.8742 \\ 0.9697 & 0.9774 & 1.0000 & 0.9853 & 0.9147 \\ 0.9460 & 0.9673 & 0.9853 & 1.0000 & 0.9471 \\ 0.8311 & 0.8742 & 0.9147 & 0.9471 & 1.0000 \end{pmatrix}$$

- Conclusion:

The role of random effects is to model the variance and association structure

- Adding the quadratic random Age effect again implies some changes in the tests for the fixed effects, when compared to those from the previous model with only random intercepts and linear Age effects:

NOW

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
AGE	1	17	3594.53	<.0001
GROUP	2	40	2.21	0.1228
AGE*GROUP	2	40	9.39	0.0005

BEFORE

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
AGE	1	17	3572.36	<.0001
GROUP	2	60	0.60	0.5514
AGE*GROUP	2	60	9.23	0.0003

7.6 Conclusion and terminology

- The linear mixed model is a linear regression model with two sets of regression parameters:
 - ▷ **Fixed effects** β
 - ▷ **Random effects** $b_i \sim N(0, D)$
- The fixed effects are used to model the average outcome
- The random effects are used to model the covariance structure
- All parameters in D , jointly with the residual variance σ^2 , are called **variance components**

Chapter 8

The rat data

- ▷ Example
- ▷ A linear mixed model
- ▷ Fitting the model in SAS

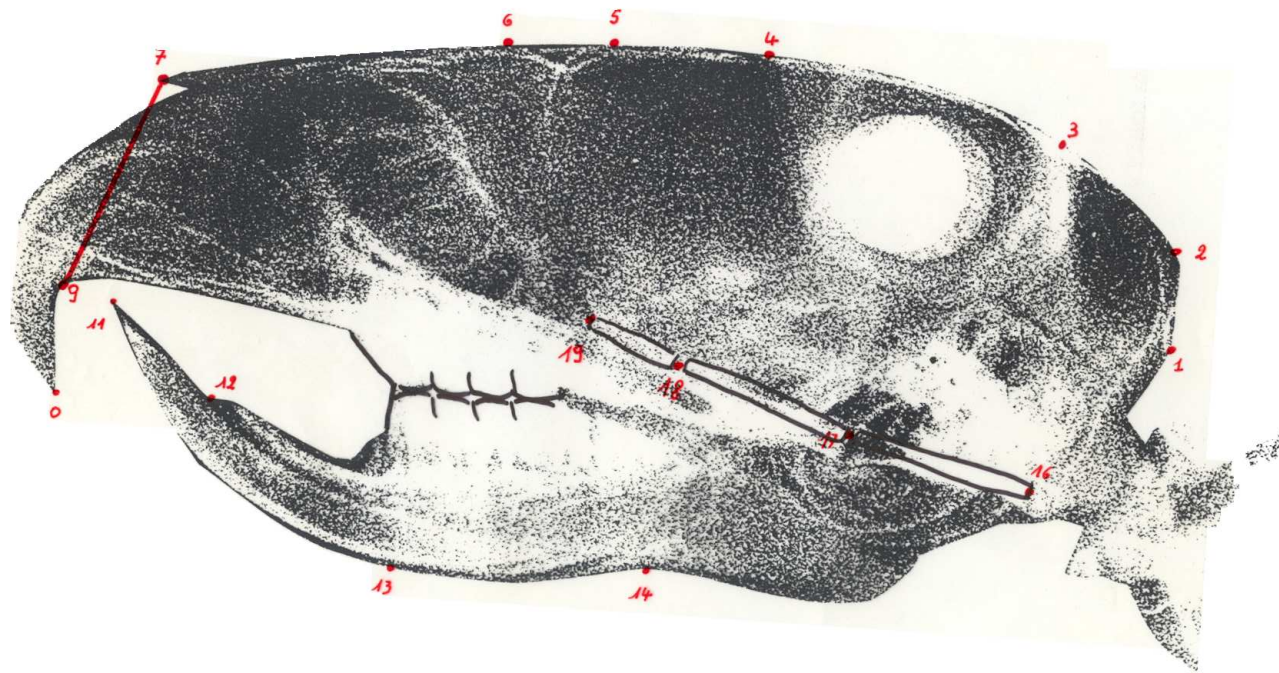
8.1 Example

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

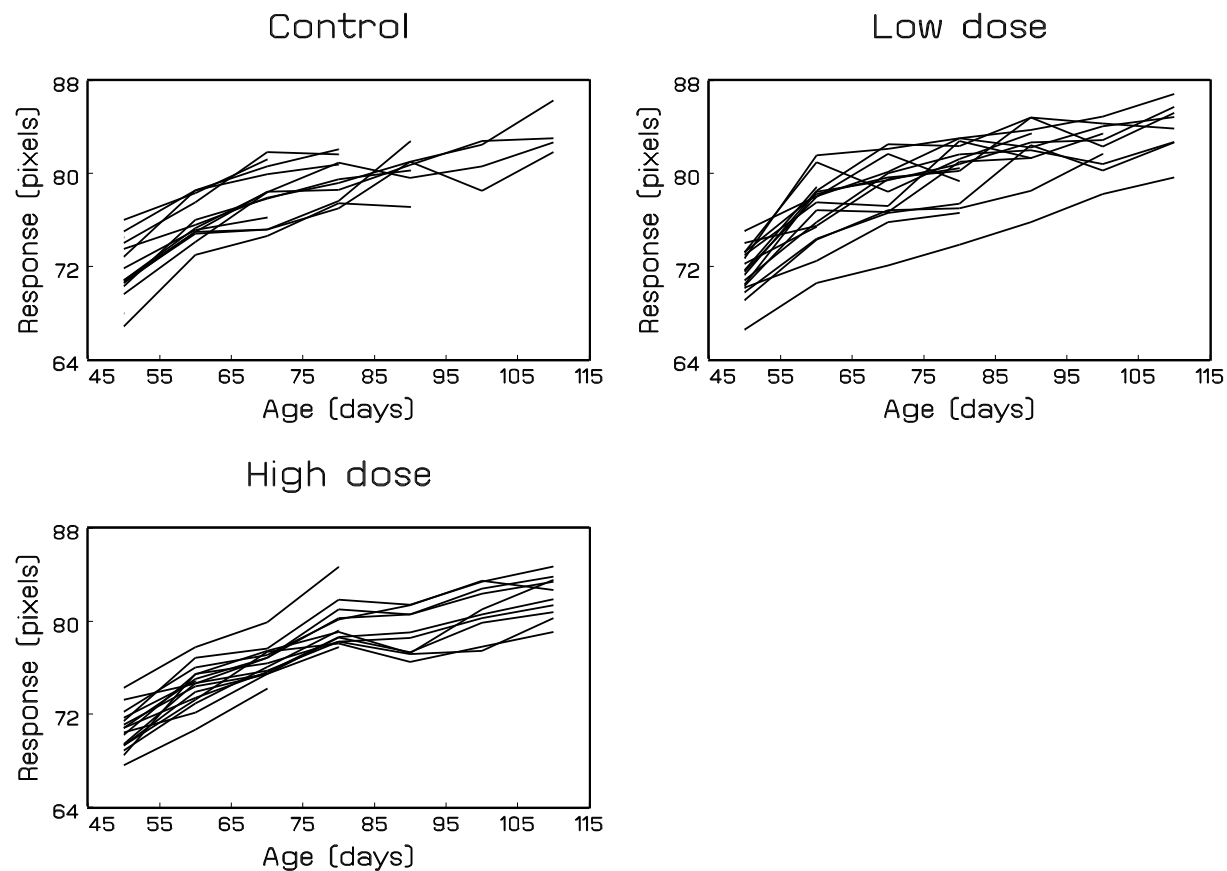
**How does craniofacial growth
depend on testosterone production ?**

- 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

8.2 A linear mixed model

- Since linear mixed models assume a linear regression for each cluster separately, they can also be used for unbalanced data, i.e., data with unequal number of measurements per cluster.
- Note that this was also the case for the lizard data.
- Individual profiles show very similar evolutions for all rats (apart from measurement error)
- This suggests a random-intercepts model
- Non-linearity can be accounted for by using a logarithmic transformation of the time scale:

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

- We then get the following model:

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

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

- 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
- ▷ b_i : subject-specific intercepts

8.3 Fitting the model in SAS

- The following SAS program can be used:

```
data rats;                                proc mixed data = rats ;
set rats;                                  class treat rat;
t=log(1+(age-45)/10);                       model y = treat*t / solution;
run;                                         random intercept / type=un subject=rat g;
                                             contrast 'treatment effect' treat*t 1 -1 0, treat*t 1 0 -1;
                                             run;
```

- Note the parameterization of the fixed effects
- Relevant SAS output:

Covariance Parameter Estimates			Solution for Fixed Effects						
Cov Parm	Subject	Estimate	Effect	TREAT	Estimate	Standard Error	DF	t Value	Pr > t
UN(1,1)	RAT	3.5649	Intercept		68.6074	0.3312	49	207.13	<.0001
Residual		1.4448	t*TREAT	con	7.3138	0.2808	199	26.05	<.0001
			t*TREAT	hig	6.8711	0.2276	199	30.19	<.0001
			t*TREAT	low	7.5069	0.2252	199	33.34	<.0001

Type 3 Tests of Fixed Effects					Contrasts				
Effect	Num DF	Den DF	F Value	Pr > F	Label	Num DF	Den DF	F Value	Pr > F
t*TREAT	3	199	734.11	<.0001	treatment effect	2	199	2.32	0.1013

- Note the difference between the test for 't*treat' and the test for the treatment effect
- A lot of variability between rats, while little variability within rats:
 - ▷ $\sigma_{rat}^2 = 3.565$ represents the variability between rats
 - ▷ $\sigma_{res}^2 = 1.445$ represents the variability within rats
- No significant difference between the treatment groups with respect to the average evolution over time ($p = 0.1013$)
- As before, the variance and correlation structure need to be explored to check model fit.

Chapter 9

The BLSA prostate data

- ▷ Example
- ▷ A linear mixed model
- ▷ Fitting the model in SAS

9.1 Example

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

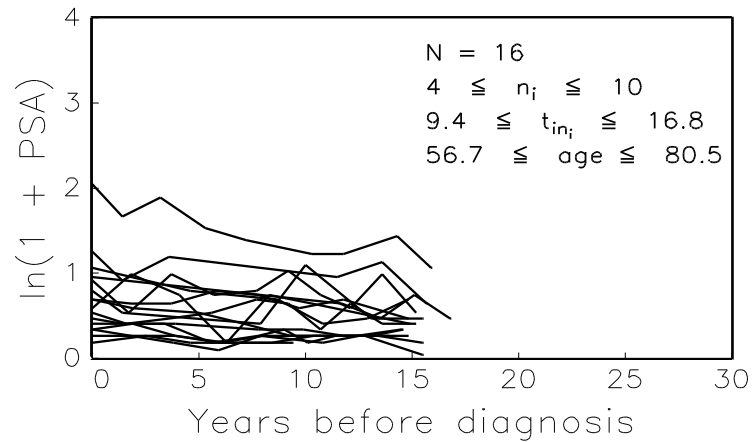
- PSA level is related to the volume of prostate tissue.
- Problem: Patients with **B**enign **P**rostatic **H**yperplasia also have an increased PSA level
- Overlap in PSA distribution for cancer and BPH cases seriously complicates the detection of prostate cancer.
- Research question (hypothesis based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

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

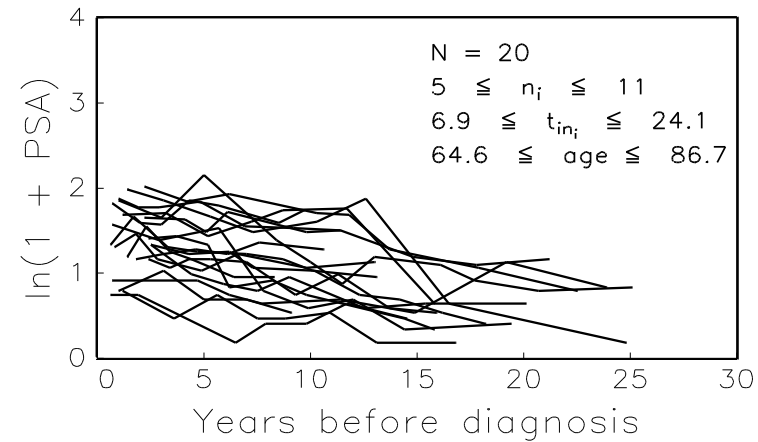
- A retrospective case-control study based on frozen serum samples:
 - ▷ 16 control patients
 - ▷ 20 BPH cases
 - ▷ 14 local cancer cases
 - ▷ 4 metastatic cancer cases
- Complication: No perfect match for age at diagnosis and years of follow-up possible
- Hence, analyses will have to correct for these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

- Individual profiles:

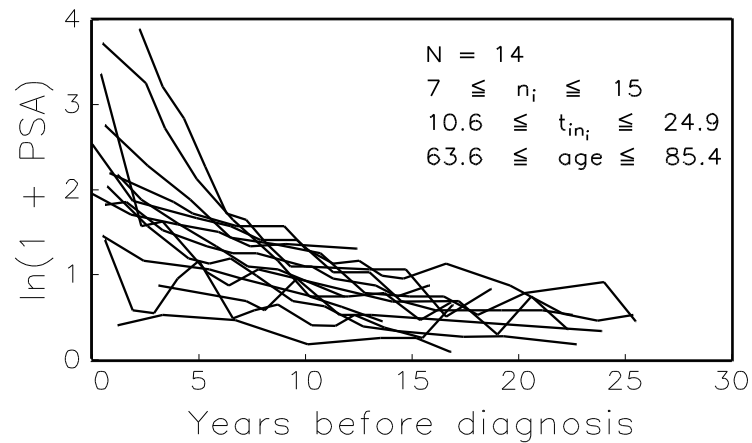
Controls



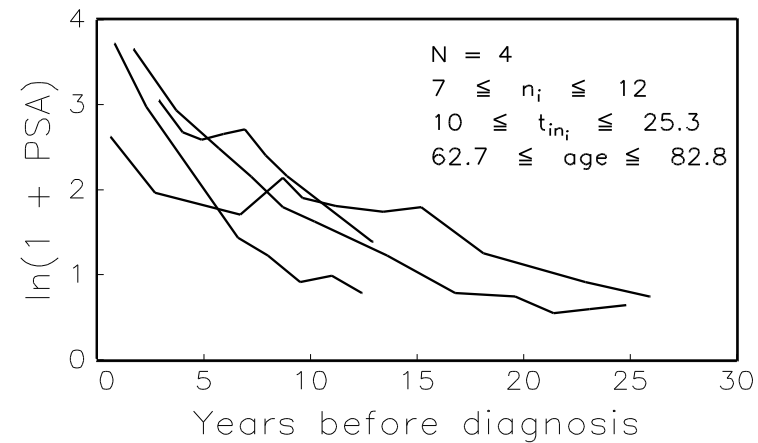
BPH cases



L/R cancer cases



Metastatic cancer cases



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

9.2 A linear mixed model

- A model for the prostate data:

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

- C_i, B_i, L_i, M_i are indicators for the 4 diagnostic groups.
- Parameter interpretation:
 - ▷ Average age-corrected quadratic profiles for all groups, modeled through the fixed effects β
 - ▷ Random effects $b_{1i}, b_{2i},$ and b_{3i} allowing subject-specific evolutions to differ from the average in that diagnostic group, even correcting for age differences

9.3 Fitting the model in SAS

- SAS program:

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

- Note again the particular parameterization for the fixed effects

- Relevant SAS output:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
UN(1,1)	XRAY	0.4518
UN(2,1)	XRAY	-0.5178
UN(2,2)	XRAY	0.9153
UN(3,1)	XRAY	0.1625
UN(3,2)	XRAY	-0.3356
UN(3,3)	XRAY	0.1308
Residual		0.02820

Estimated G Matrix

Effect	XRAY	Col1	Col2	Col3
Intercept	19	0.4518	-0.5178	0.1625
time	19	-0.5178	0.9153	-0.3356
time2	19	0.1625	-0.3356	0.1308

Estimated G Correlation Matrix

Effect	XRAY	Col1	Col2	Col3
Intercept	19	1.0000	-0.8053	0.6686
time	19	-0.8053	1.0000	-0.9700
time2	19	0.6686	-0.9700	1.0000

Solution for Fixed Effects

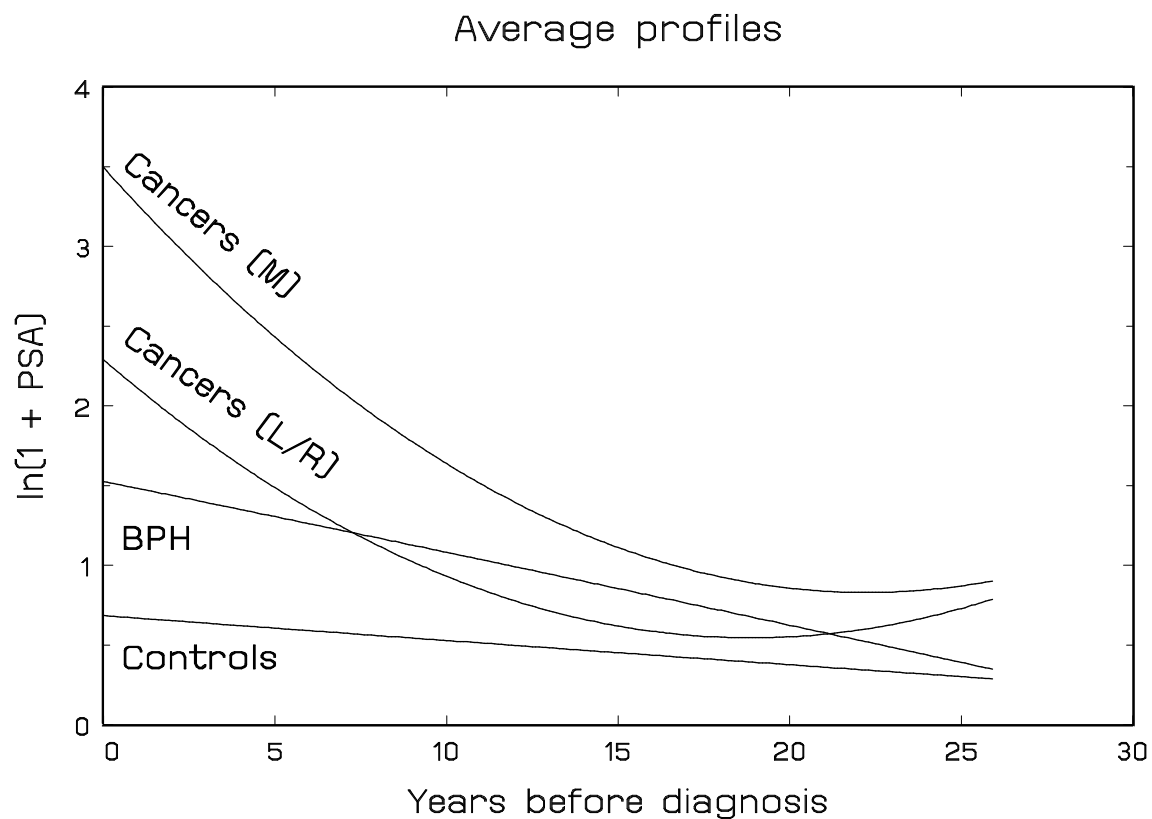
Effect	group	Estimate	Standard Error	DF	t Value	Pr > t
group	1	-1.0984	0.9763	299	-1.13	0.2615
group	2	-0.5228	1.0895	299	-0.48	0.6317
group	3	0.2964	1.0587	299	0.28	0.7797
group	4	1.5494	1.0856	299	1.43	0.1546
AGEDIAG		0.02655	0.01423	299	1.87	0.0631
time*group	1	0.5681	1.4725	299	0.39	0.6999
time*group	2	0.3956	1.6377	299	0.24	0.8093
time*group	3	-1.0359	1.5928	299	-0.65	0.5159
time*group	4	-1.6049	1.6258	299	-0.99	0.3244
AGEDIAG*time		-0.01117	0.02142	299	-0.52	0.6026
time2*group	1	-0.1295	0.6100	299	-0.21	0.8320
time2*group	2	-0.1585	0.6723	299	-0.24	0.8138
time2*group	3	0.3419	0.6563	299	0.52	0.6028
time2*group	4	0.3951	0.6660	299	0.59	0.5535
AGEDIAG*time2		0.002259	0.008829	299	0.26	0.7982

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
group	4	299	15.90	<.0001
AGEDIAG	1	299	3.48	0.0631
time*group	4	299	7.85	<.0001
AGEDIAG*time	1	299	0.27	0.6026
time2*group	4	299	4.44	0.0017
AGEDIAG*time2	1	299	0.07	0.7982

- Note the very strong correlations between random effects
- CONTRAST statements can be used to test for group differences

- Based on the fixed effects, fitted average profiles can be plotted (at median age at diagnosis):



Chapter 10

The Leuven diabetes project

- ▷ Introduction
- ▷ A variety of multilevel models
- ▷ Including covariates at various levels

10.1 Introduction: the DPL project

- The **D**iabetes **P**roject **L**euven
- In Belgium, general practitioners (GP's) cannot rely on structured assistance of dieticians or diabetes nurse educators in their practice.
- The DPL intends to study the effect of implementing a structured model for chronic diabetes care on patients' clinical outcomes.
- GP's will be offered assistance and can redirect patients to the diabetes care team, consisting of a nurse educator, a dietician, an ophthalmologist and an internal medicine doctor.

- In DPL, two programs were implemented and GP's were randomized to one of two groups:
 - ▷ **LIP:** Low Intervention Program (group A)
 - ▷ **HIP:** High Intervention Program (group R)
- We consider the analysis of GP's in the HIP group:
 - ▷ 61 GP's
 - ▷ 1577 patients
 - ▷ number of patients per GP varies between 5 and 138, with a median of 47
- Patients were measured twice:
 - ▷ When the program was initiated (time T0)
 - ▷ After one year (time T1)

- The outcome studied here is HbA1c, glycosylated hemoglobin:
 - ▷ Molecule in red blood cells that attaches to glucose (blood sugar)
 - ▷ High values reflect more glucose in blood
 - ▷ In diabetes patients, HbA1c gives a good estimate of how well diabetes is being managed over the last 2 or 3 months
 - ▷ Non-diabetics have values between 4% and 6%
 - ▷ HbA1c above 7% means diabetes is poorly controlled, implying higher risk for long-term complications.

10.2 A variety of multilevel models

- Let Y_{ijk} be the k th measurement of HbA1, for the j th patient, of the i th GP
- We have 3-level data, hence random effects can enter the models at various levels
- Several models for studying the longitudinal evolutions will be illustrated and compared:
 - ▷ No random effects
 - ▷ Random GP effects
 - ▷ Random patient effects
 - ▷ Random effects for GP and patient

10.2.1 Model 1: No random effects

$$Y_{ijk} = \beta_0 + \beta_1 t_k + \varepsilon_{ijk}, \quad \varepsilon_{ijk} \sim N(0, \sigma_{res}^2)$$

- SAS program:

```
proc mixed data=dpla;  
model hba1c = time / solution;  
run;
```

```
proc glm data=dpla;  
model hba1c = time / solution;  
run;
```

- Relevant output:

Covariance Parameter Estimates		Solution for Fixed Effects					
Cov Parm	Estimate	Effect	Estimate	Standard Error	DF	t Value	Pr > t
Residual	1.2309	Intercept	7.1357	0.02823	2966	252.81	<.0001
		time	-0.3899	0.04076	2966	-9.57	<.0001

10.2.2 Model 2: Random GP effects

$$Y_{ijk} = \beta_0 + \beta_1 t_k + a_i + \varepsilon_{ijk}, \quad a_i \sim N(0, \sigma_{GP}^2), \quad \varepsilon_{ijk} \sim N(0, \sigma_{res}^2)$$

- SAS program:

```
proc mixed data=dpla;  
class mdnr;  
model hba1c = time / solution;  
random intercept / subject=mdnr;  
run;
```

- Relevant output:

Covariance Parameter Estimates			Solution for Fixed Effects					
Cov Parm	Subject	Estimate	Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	mdnr	0.07093	Intercept	7.1695	0.04519	60	158.66	<.0001
Residual		1.1709	time	-0.3873	0.03978	2906	-9.73	<.0001

10.2.3 Model 3: Random patient effects

$$Y_{ijk} = \beta_0 + \beta_1 t_k + b_{j(i)} + \varepsilon_{ijk}, \quad b_{j(i)} \sim N(0, \sigma_{PAT}^2), \quad \varepsilon_{ijk} \sim N(0, \sigma_{res}^2)$$

- SAS program:

```
/* unique patient numbers */  
proc mixed data=dpla;  
class md_patient;  
model hba1c = time / solution;  
random intercept / subject=md_patient;  
run;
```

```
/* patients numbered within GP's */  
proc mixed data=dpla;  
class mdnr patientnr;  
model hba1c = time / solution;  
random intercept / subject=patientnr(mdnr);  
run;
```

- Relevant output:

Covariance Parameter Estimates			Solution for Fixed Effects					
Cov Parm	Subject	Estimate	Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	patientnr(mdnr)	0.6675	Intercept	7.1392	0.02838	1571	251.54	<.0001
Residual		0.5831	time	-0.3785	0.02851	1395	-13.28	<.0001

10.2.4 Model 4: Random GP and patient effects

$$Y_{ijk} = \beta_0 + \beta_1 t_k + a_i + b_{j(i)} + \varepsilon_{ijk},$$
$$a_i \sim N(0, \sigma_{GP}^2), \quad b_{j(i)} \sim N(0, \sigma_{PAT}^2), \quad \varepsilon_{ijk} \sim N(0, \sigma_{res}^2)$$

- SAS program:

```
proc mixed data=dpla;
class mdnr patientnr;
model hba1c = time / solution;
random intercept / subject=mdnr;
random intercept / subject=patientnr(mdnr);
run;
```

```
proc mixed data=dpla;
class mdnr patientnr;
model hba1c = time / solution;
random mdnr patientnr(mdnr);
run;
```

- Relevant output:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
Intercept	mdnr	0.05439
Intercept	patientnr(mdnr)	0.6171
Residual		0.5837

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	7.1668	0.04241	60	169.00	<.0001
time	-0.3780	0.02851	1395	-13.26	<.0001

10.2.5 Summary of results

	Model 1	Model 2	Model 3	Model 4
Parameter	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
Fixed effects:				
β_0	7.1357(0.0282)	7.1695(0.0452)	7.1392(0.0284)	7.1668(0.0424)
β_1	-0.3899(0.0408)	-0.3873(0.0398)	-0.3785(0.0286)	-0.3780(0.0285)
Variance components:				
σ_{GP}^2	_____	0.0709	_____	0.0544
σ_{PAT}^2	_____	_____	0.6675	0.6171
σ_{res}^2	1.2309	1.1709	0.5831	0.5837

- No standard errors reported for variance components, since standard Z -tests do not produce correct tests (see later)
- The various models use different decompositions of the total variability:
 - ▷ Model 1: $\hat{\sigma}^2 = \hat{\sigma}_{res}^2 = 1.2309$
 - ▷ Model 2: $\hat{\sigma}^2 = \hat{\sigma}_{GP}^2 + \hat{\sigma}_{res}^2 = 0.0709 + 1.1709 = 1.2418$
 - ▷ Model 3: $\hat{\sigma}^2 = \hat{\sigma}_{PAT}^2 + \hat{\sigma}_{res}^2 = 0.6675 + 0.5831 = 1.2506$
 - ▷ Model 4: $\hat{\sigma}^2 = \hat{\sigma}_{GP}^2 + \hat{\sigma}_{PAT}^2 + \hat{\sigma}_{res}^2 = 0.0544 + 0.6171 + 0.5837 = 1.2552$
- Inclusion of random effects has little effect on estimation of fixed effects but has severe impact on the standard errors:
 - ▷ Larger standard errors for between-cluster effects (intercept β_0)
 - ▷ Smaller standard errors for within-cluster effects (time β_1)
- There is a significant decrease in HbA1c, under all models

- The models also imply specific correlation structures.
- For example, the marginal association structure implied by Model 4 equals:
 - ▷ Observations from different GP's i and i^* , $i \neq i^*$, are not correlated:

$$\widehat{\text{Corr}}(Y_{ijk}, Y_{i^*j^*k^*}) = 0$$

- ▷ Observations from same GP but different patients j and j^* , $j \neq j^*$, are correlated:

$$\widehat{\text{Corr}}(Y_{ijk}, Y_{ij^*k^*}) = \frac{\hat{\sigma}_{GP}^2}{\hat{\sigma}_{GP}^2 + \hat{\sigma}_{PAT}^2 + \hat{\sigma}_{res}^2} = \frac{0.0544}{1.2552} = 0.0433$$

- ▷ Observations k and k^* , $k \neq k^*$, from same patient are correlated:

$$\widehat{\text{Corr}}(Y_{ijk}, Y_{ijk^*}) = \frac{\hat{\sigma}_{GP}^2 + \hat{\sigma}_{PAT}^2}{\hat{\sigma}_{GP}^2 + \hat{\sigma}_{PAT}^2 + \hat{\sigma}_{res}^2} = \frac{0.0544 + 0.6171}{1.2552} = 0.5350$$

10.3 Including covariates at various levels

- Additional covariates can be added to explain variability at the different levels, or to study what patient and/or GP characteristics are related to time trends.
- We extend Model 4 with the following covariates:
 - ▷ At GP level: Practice form ('one', 'two', 'more')
 - ▷ At patient level:
 - BMI at baseline
 - Whether or not patient is a newly diagnosed diabetic (1: yes, 0: no)
- We will investigate the effect of each covariate separately.
- Obviously, models with multiple covariates are possible as well.

10.3.1 Model 5: Correcting for different practice forms

- SAS program:

```
proc mixed data=dpla;
class mdnr patientnr practice;
model hba1c = practice time time*practice / solution;
random intercept / subject=mdnr;
random intercept / subject=patientnr(mdnr);
run;
```

- Relevant output:

Covariance Parameter Estimates			Type 3 Tests of Fixed Effects				
Cov Parm	Subject	Estimate	Effect	Num DF	Den DF	F Value	Pr > F
Intercept	mdnr	0.05431	practice	2	1393	3.08	0.0461
Intercept	patientnr(mdnr)	0.6152	time	1	1393	162.59	<.0001
Residual		0.5838	time*practice	2	1393	1.28	0.2786

Solution for Fixed Effects

Effect	practice	Estimate	Standard Error	DF	t Value	Pr > t
Intercept		7.1408	0.07505	58	95.14	<.0001
practice	Mor	-0.1158	0.1105	1393	-1.05	0.2948
practice	One	0.1407	0.1001	1393	1.40	0.1603
practice	Two	0
time		-0.3659	0.04976	1393	-7.35	<.0001
time*practice	Mor	0.05052	0.07467	1393	0.68	0.4988
time*practice	One	-0.06162	0.06680	1393	-0.92	0.3565
time*practice	Two	0

- Since Time is included as a continuous covariate, the main effect of Practice reflects differences at baseline between the various practice forms
- We find a significant difference at baseline ($p = 0.0461$) with lower average values of HbA1c the more GP's work together in group practices.
- This difference does not change over time, as the practice form has no significant effect on the change over time ($p = 0.2786$).

10.3.2 Model 6: Correcting for different BMI at baseline

- SAS program:

```
proc mixed data=dpla;
class mdnr patientnr;
model hba1c = bmi0 time time*bmi0 / solution;
random intercept / subject=mdnr;
random intercept / subject=patientnr(mdnr);
run;
```

- Relevant output:

Covariance Parameter Estimates			Type 3 Tests of Fixed Effects				
Cov Parm	Subject	Estimate	Effect	Num DF	Den DF	F Value	Pr > F
Intercept	mdnr	0.05840	bmi0	1	1323	10.55	0.0012
Intercept	patientnr(mdnr)	0.6048	time	1	1323	0.94	0.3320
Residual		0.5468	bmi0*time	1	1323	1.63	0.2023

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	6.6325	0.1654	60	40.09	<.0001
bmi0	0.01750	0.005386	1323	3.25	0.0012
time	-0.1573	0.1621	1323	-0.97	0.3320
bmi0*time	-0.00685	0.005367	1323	-1.28	0.2023

- We find a significantly higher baseline value for HbA1c as the initial BMI is larger ($p = 0.0012$)
- The average time trend is not significantly related to the initial BMI level ($p = 0.2023$).

10.3.3 Model 7: Correcting for new diagnosis

- SAS program:

```
proc mixed data=dpla;  
class mdnr patientnr ;  
model hba1c = new0 time time*new0 / solution;  
random intercept / subject=mdnr;  
random intercept / subject=patientnr(mdnr);  
run;
```

- Relevant output:

Covariance Parameter Estimates			Type 3 Tests of Fixed Effects				
Cov Parm	Subject	Estimate	Effect	Num DF	Den DF	F Value	Pr > F
Intercept	mdnr	0.05561	new0	1	1321	14.67	0.0001
Intercept	patientnr(mdnr)	0.6004	time	1	1321	115.15	<.0001
Residual		0.5385	new0*time	1	1321	52.82	<.0001

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	7.1172	0.04367	60	162.97	<.0001
new0	0.4030	0.1052	1321	3.83	0.0001
time	-0.3148	0.02934	1321	-10.73	<.0001
new0*time	-0.7762	0.1068	1321	-7.27	<.0001

- Since Time is included as a continuous covariate, the main effect of New0 reflects differences at baseline between newly diagnosed diabetics and others.
- We find a significant difference at baseline ($p = 0.0001$) with higher average values of HbA1c for newly diagnosed patients.
- The newly diagnosed patients have a steeper decrease in HbA1c than the others ($p < 0.0001$).

- We can test whether, after one year, both groups are at the same level:

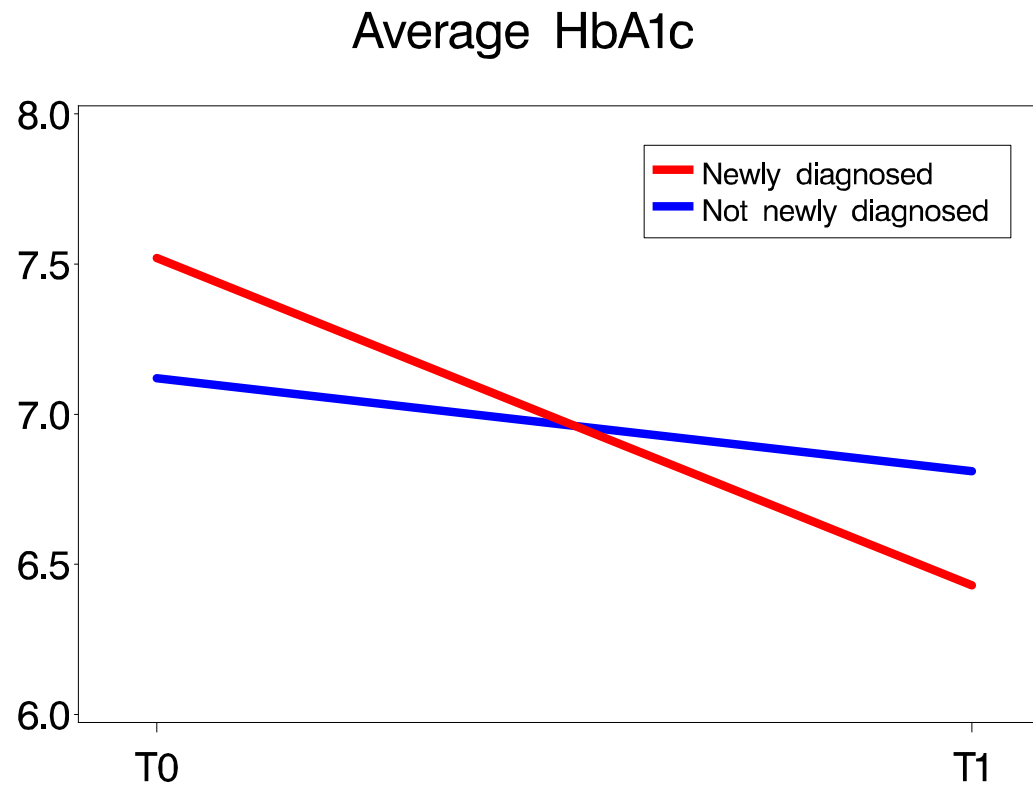
```
estimate 'equal at T1' new0 1 time*new0 1;
```

- Additional output:

	Estimates				
Label	Estimate	Standard Error	DF	t Value	Pr > t
equal at T1	-0.3732	0.1106	1321	-3.38	0.0008

- Hence, after one year, the newly diagnosed diabetics have lower average values for HbA1c ($p = 0.0008$)

- Graphically:



- This can be explained from the fact that the disease gets worse over time, hence HbA1c is more difficult to keep under control.

- We therefore repeat the analysis, correcting for the number of years patients have been diabetic (0 for newly diagnosed patients).

```
proc mixed data=dpla;
class mdnr patientnr ;
model hba1c = duration new0 time time*new0 / solution;
random intercept / subject=mdnr;
random intercept / subject=patientnr(mdnr);
estimate 'equal at T1' new0 1 time*new0 1;
run;
```

- Relevant output:

Covariance Parameter Estimates			Type 3 Tests of Fixed Effects				
Cov Parm	Subject	Estimate	Effect	Num DF	Den DF	F Value	Pr > F
Intercept	mdnr	0.05210	duration	1	1320	67.83	<.0001
Intercept	patientnr(mdnr)	0.5589	new0	1	1320	35.88	<.0001
Residual		0.5401	time	1	1320	114.40	<.0001
			new0*time	1	1320	52.87	<.0001

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	6.8737	0.05181	60	132.67	<.0001
duration	0.03096	0.003759	1320	8.24	<.0001
new0	0.6432	0.1074	1320	5.99	<.0001
time	-0.3141	0.02937	1320	-10.70	<.0001
new0*time	-0.7773	0.1069	1320	-7.27	<.0001

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > t
equal at T1	-0.1340	0.1125	1320	-1.19	0.2338

- The difference after one year, between newly diagnosed diabetics and others is no longer significant ($p = 0.2338$) after correction for the number of years patients have been diabetic.

Chapter 11

Estimation of Random Effects

- ▷ Empirical Bayes estimation
- ▷ Example: Leuven diabetes project
- ▷ Example: Prostate data
- ▷ Average versus cluster-specific prediction

11.1 Empirical Bayes estimation

- Random effects reflect how specific clusters deviate from the population average
- For example, for the Leuven diabetes project, Model 4 for the outcome Y_{ijk} being the k th measurement of HbA1, for the j th patient, of the i th GP, was given by:

$$Y_{ijk} = \beta_0 + \beta_1 t_k + a_i + b_{j(i)} + \varepsilon_{ijk},$$

$$a_i \sim N(0, \sigma_{GP}^2), \quad b_{j(i)} \sim N(0, \sigma_{PAT}^2), \quad \varepsilon_{ijk} \sim N(0, \sigma_{res}^2)$$

- The parameter a_i expresses how the average HbA1c level of patients treated by GP i differs from the overall population average.
- The parameter $b_{j(i)}$ expresses how the average of patient j treated by GP i differs from the average of that specific GP.

- Estimation of the random effects can be helpful for detecting outlying profiles or clusters
- Since these parameters are assumed to be stochastic, Bayesian methods are applied.
- Posterior means:

$$\hat{a}_i = E(a_i \mid Y_{ijk}, \quad \forall j, k)$$
$$\hat{b}_{j(i)} = E(b_{j(i)} \mid Y_{ijk}, \quad \forall k)$$

- The so-obtained estimates are called **Empirical Bayes (EB)** estimates. They are the expected random effects, conditionally on the observed data for that specific cluster
- In practice histograms and/or scatterplots of EB estimates are used to detect outlying clusters

11.2 Example: Leuven diabetes project

- We re-consider Model 4 given by:

$$Y_{ijk} = \beta_0 + \beta_1 t_k + a_i + b_{j(i)} + \varepsilon_{ijk}$$

- The parameters a_i and $b_{j(i)}$ represent GP and patient effects, respectively.
- Histograms and scatterplots will be used to study the EB estimates for a_i and $b_{j(i)}$
- SAS program for calculation of EB estimates:

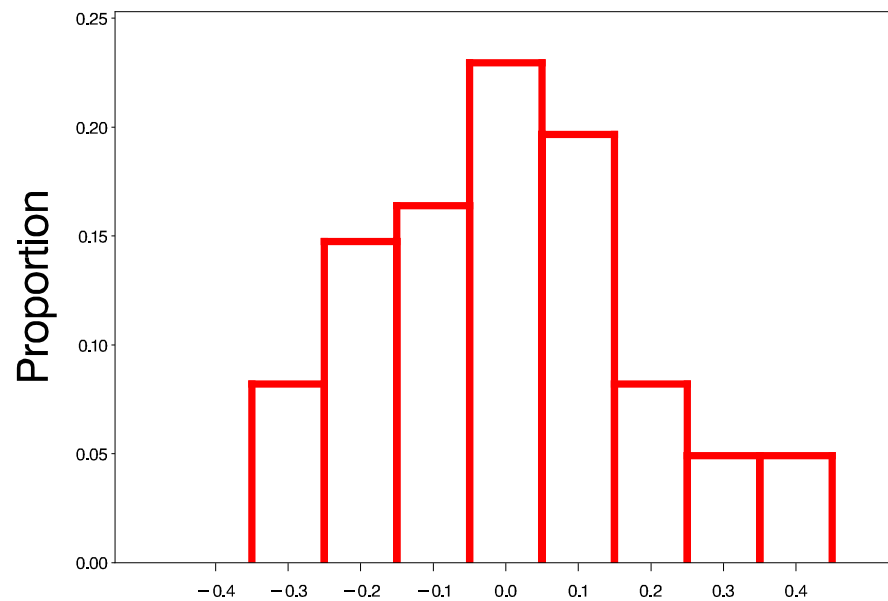
```
proc mixed data=dpla;
class mdnr patientnr;
model hba1c = time / solution;
random intercept / subject=mdnr solution;
random intercept / subject=patientnr(mdnr) solution;
ods listing exclude solutionr;
ods output solutionr=out;
run;
```

- 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.
- EB estimates:

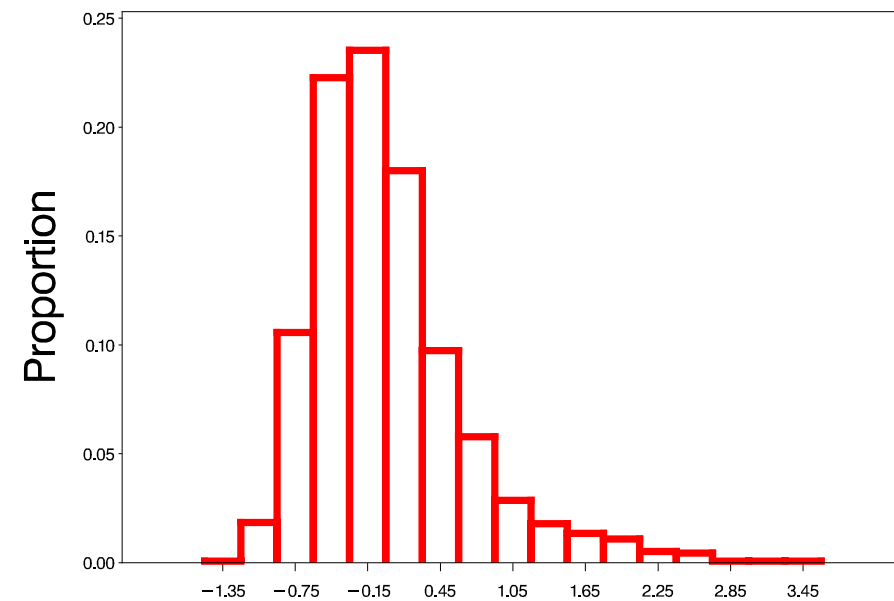
Obs	mdnr	gpeb	Obs	mdnr	patientnr	patienteb
1	2	0.21976	1	2	1	0.00163
2	5	0.02113	2	2	2	-0.03232
3	6	0.18372	3	2	3	2.10624
4	7	0.01447				
5	8	0.08236
6	10	0.15512				
7	11	-0.07875	25	2	25	-0.91490
8	13	-0.25004	26	2	26	0.13741
9	14	-0.06225	27	5	1	-0.06719
			28	5	2	-0.61032
.	29	5	3	0.44199
59	155	-0.01637
60	156	0.26761				
61	165	-0.23516	1569	165	22	-0.43632
			1570	165	23	-1.18312
			1571	165	24	0.17469
			1572	165	25	0.44626

- Histograms of both sets of EB estimates:

EB estimates of GP effects

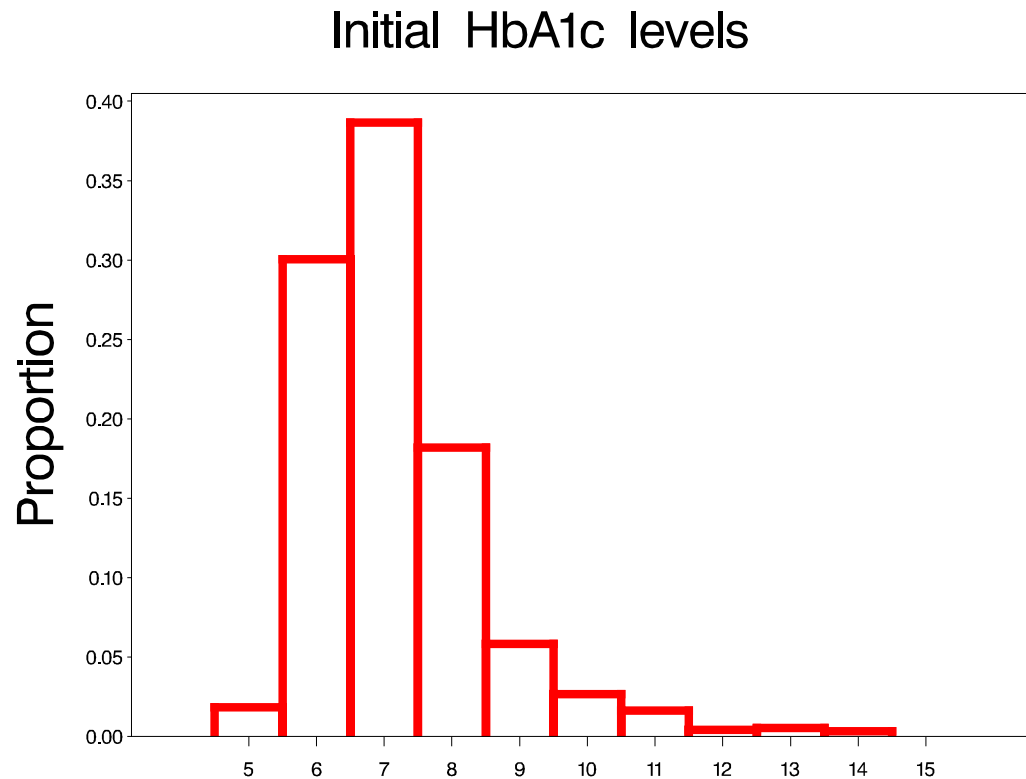


EB estimates of patient effects



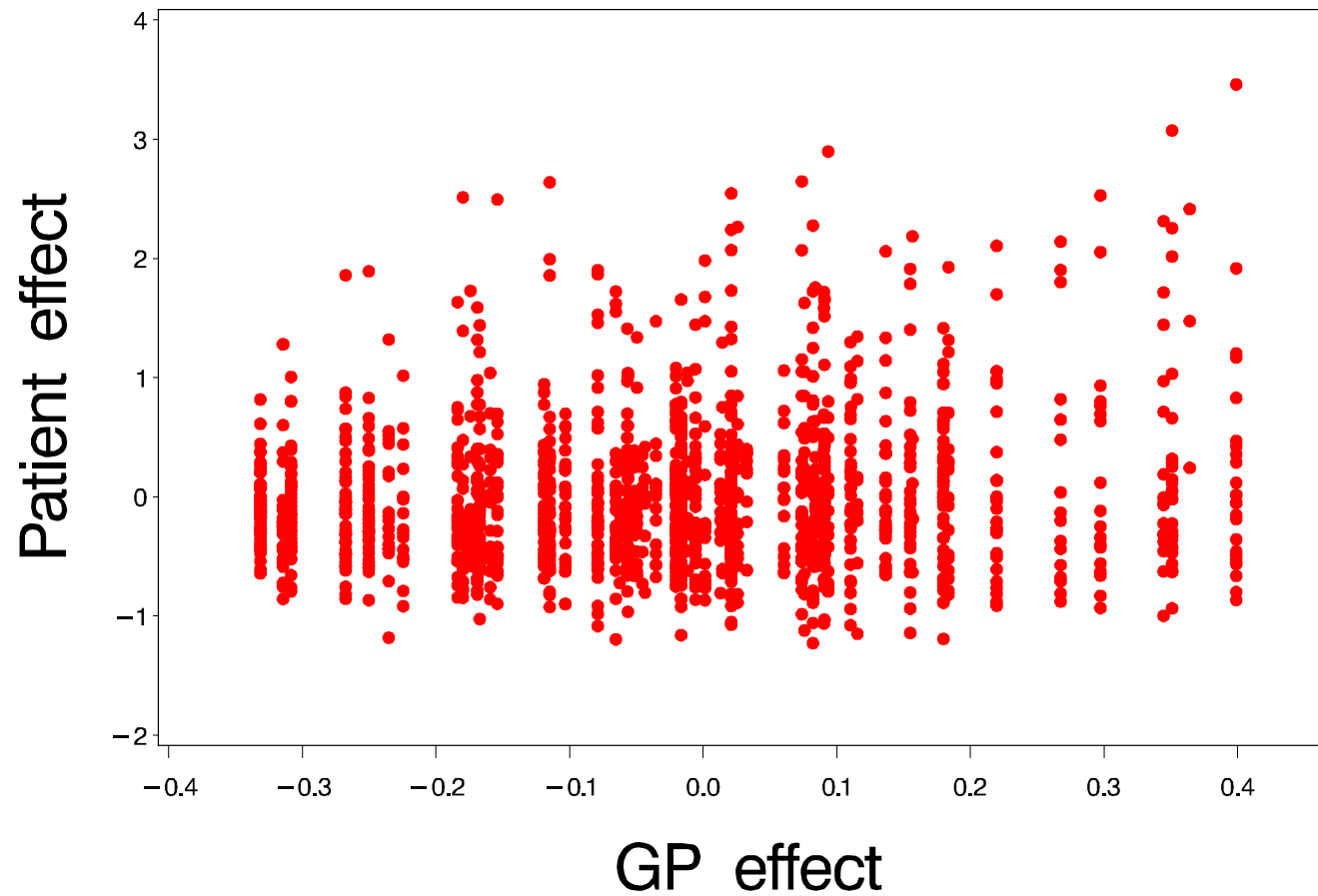
- We notice some patients with extremely large HbA1c values. The largest estimated $b_{j(i)}$ is $\hat{b}_{2(140)} = 3.46$

- This patient was newly diagnosed, with initial BMI equal to 26.40, initial HbA1c equal to 14.3%, and no follow-up measurement after one year
- The initial HbA1c level of 14.3% is extremely high:



- Scatterplot of EB estimates for patients versus GP's:

Scatterplot of EB estimates



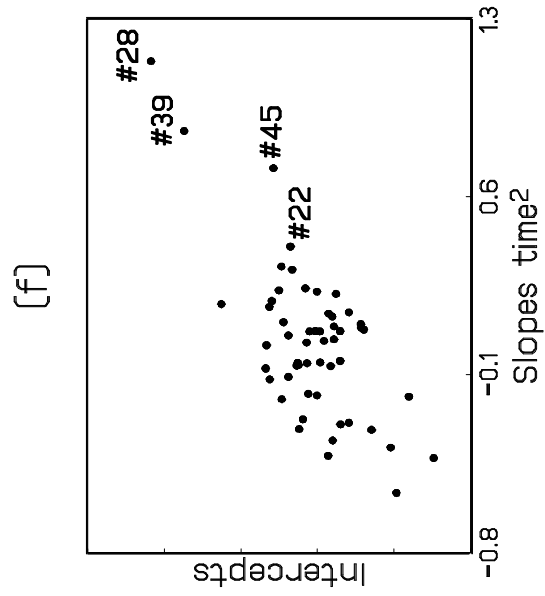
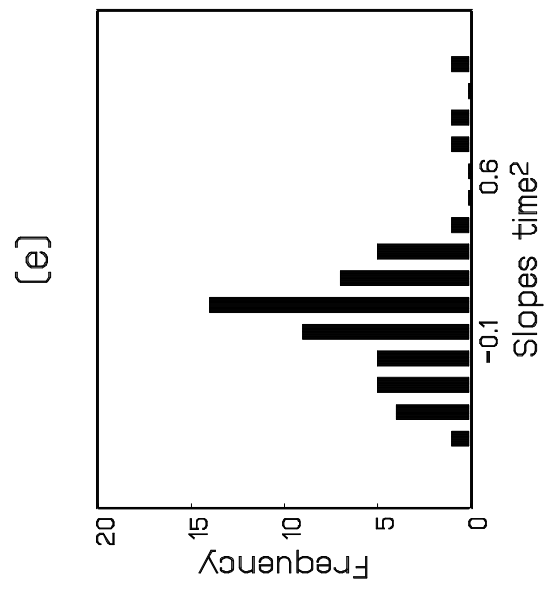
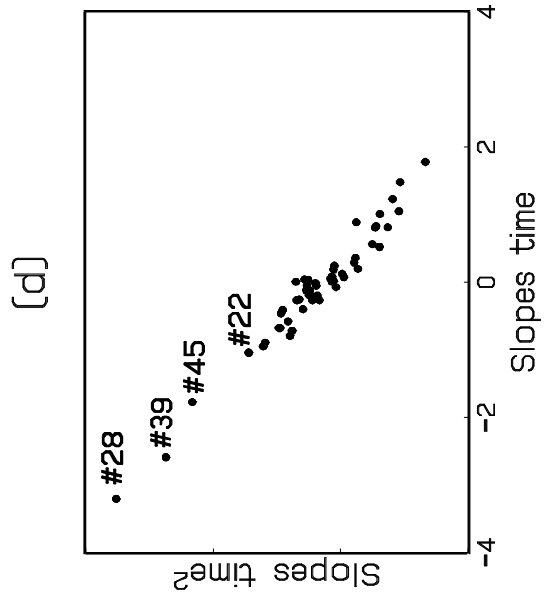
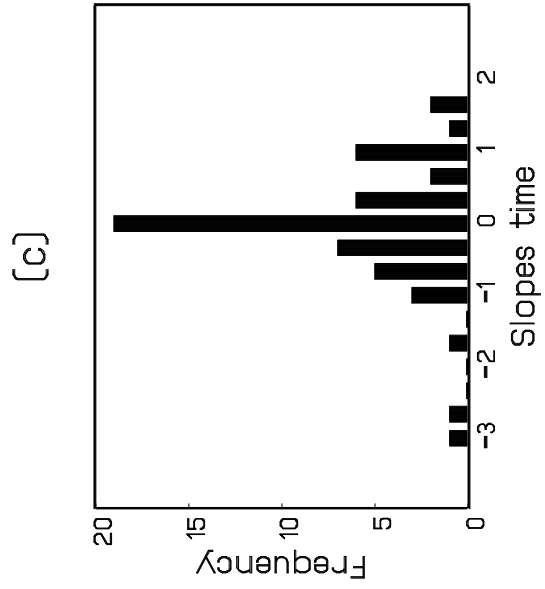
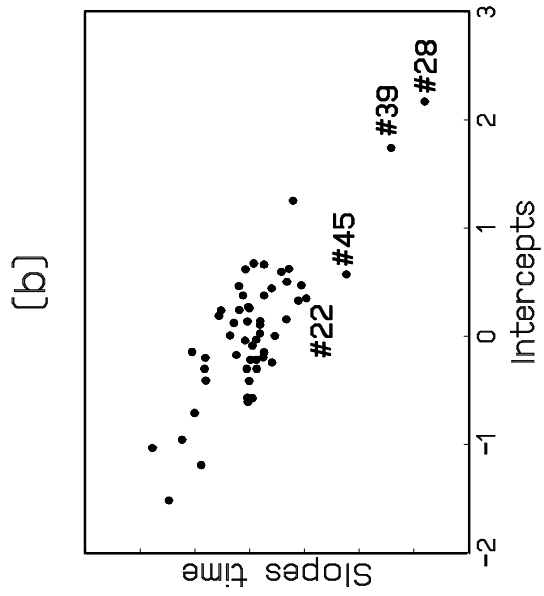
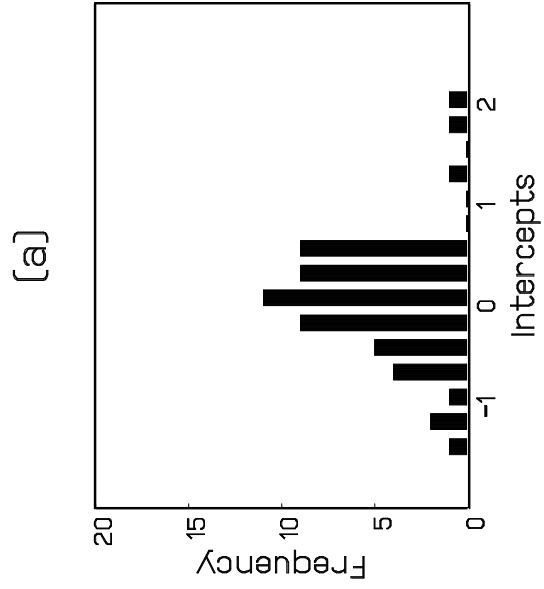
- Note how the patient with the largest estimate for $b_{j(i)}$ was treated by the GP with the largest estimated a_i .
- It is therefore worthwhile to repeat the analysis with this subject removed from the data
- The estimate \hat{a}_{140} drops from 0.40 to 0.27 and four other GP's have now a higher estimate for their effect a_i .
- EB estimates can also be calculated based on other models which include patient or GP characteristics as covariates/factors
- Extreme EB estimates then reflect that a specific GP or patient within GP is outlying, while this extreme behaviour cannot be explained by the covariates in the model.

11.3 Example: Prostate data

- We re-consider the model

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

- Again, histograms and scatterplots of components of $\widehat{\mathbf{b}}_i$ can be used to detect model deviations or subjects with 'exceptional' evolutions over time



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

$$\widehat{D}_{\text{corr}} = \begin{pmatrix} 1.000 & -0.805 & 0.669 \\ -0.805 & 1.000 & -0.970 \\ 0.669 & -0.970 & 1.000 \end{pmatrix}$$

- 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

11.4 Average versus cluster-specific prediction

- Once the EB estimates have been calculated, predictions can be obtained both at the cluster level, as well as on the population average level.

- Re-consider Model 4 for the Leuven diabetes project:

$$Y_{ijk} = \beta_0 + \beta_1 t_k + a_i + b_{j(i)} + \varepsilon_{ijk}$$

- Predictions:

- ▷ On population average level:

$$\widehat{E}(Y_{ijk}) = \widehat{\beta}_0 + \widehat{\beta}_1 t_k$$

- ▷ On cluster level:

$$\widehat{Y}_{ijk} = \widehat{\beta}_0 + \widehat{\beta}_1 t_k + \widehat{a}_i + \widehat{b}_{j(i)}$$

11.5 Example: Leuven diabetes project

- SAS program for predictions under Model 4:

```
proc mixed data=dpla;
class mdnr patientnr;
model hba1c = time / solution outpm=predmean outp=pred;
random intercept / subject=mdnr solution;
random intercept / subject=patientnr(mdnr) solution;
run;

proc print data=predmean;
proc print data=pred;
run;
```

- The option 'predmean' requests calculation of the predicted means
- The option 'pred' requests calculation of predictions at cluster level

- Table of predicted means (option 'predmean'):

Obs	hba1c	mdnr	patientnr	time	Pred
1	6.4	2	1	0	7.16685
2	8.0	2	1	1	6.78883
3	6.7	2	2	0	7.16685
4	7.6	2	2	1	6.78883
5	11.2	2	3	0	7.16685
6	9.4	2	3	1	6.78883
7	6.8	2	4	0	7.16685
8	6.8	2	4	1	6.78883
2965	7.5	165	24	0	7.16685
2966	6.5	165	24	1	6.78883
2967	7.7	165	25	0	7.16685
2968	7.1	165	25	1	6.78883

- Table with predictions on subject level (option 'pred'):

Obs	hba1c	mdnr	patientnr	time	Pred
1	6.4	2	1	0	7.38824
2	8.0	2	1	1	7.01022
3	6.7	2	2	0	7.35429
4	7.6	2	2	1	6.97628
5	11.2	2	3	0	9.49285
6	9.4	2	3	1	9.11484
7	6.8	2	4	0	7.11667
8	6.8	2	4	1	6.73866
2965	7.5	165	24	0	7.10638
2966	6.5	165	24	1	6.72837
2967	7.7	165	25	0	7.37795
2968	7.1	165	25	1	6.99994

- Components needed to calculate predictions:

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	7.1668	0.04241	60	169.00	<.0001
time	-0.3780	0.02851	1395	-13.26	<.0001

Obs	mdnr	gpeb	Obs	mdnr	patientnr	patienteb
1	2	0.21976	1	2	1	0.00163
			2	2	2	-0.03232
			3	2	3	2.10624

- Population average HbA1c values at baseline and after one year:

$$\widehat{E}(Y_{ij1}) = 7.1668 - 0.3780 \times 0 = 7.1668$$

$$\widehat{E}(Y_{ij2}) = 7.1668 - 0.3780 \times 1 = 6.7888$$

- Subject-specific predicted HbA1c values for first three patients treated by GP 2:

$$\widehat{Y}_{211} = 7.1668 + 0.2198 + 0.0016 = 7.3882$$

$$\widehat{Y}_{212} = 6.7888 + 0.2198 + 0.0016 = 7.0102$$

$$\widehat{Y}_{221} = 7.1668 + 0.2198 - 0.0323 = 7.3543$$

$$\widehat{Y}_{222} = 6.7888 + 0.2198 - 0.0323 = 6.9763$$

$$\widehat{Y}_{231} = 7.1668 + 0.2198 + 2.1062 = 9.4929$$

$$\widehat{Y}_{232} = 6.7888 + 0.2198 + 2.1062 = 9.1148$$

Chapter 12

Concluding remarks

- ▷ Introduction
- ▷ Tests for variance components
- ▷ Distributional assumptions for random effects
- ▷ Missing data issues

12.1 Introduction

- Many examples of linear mixed models for longitudinal or clustered data have been discussed
- Most emphasis was on model formulation, SAS implementation, and interpretation of results
- A number of issues have not been discussed:
 - ▷ Estimation methods (ML, REML, ...)
 - ▷ Inference (F -test, t -test, LR test, Wald test, ...)
 - ▷ Model checking
 - ▷ Influence analysis
 - ▷ ...

- These topics are much more difficult and technical than in classical linear models for cross-sectional data, and are therefore outside the scope of this course
- Three illustrations are given:
 - ▷ Tests for variance components
 - ▷ Distributional assumptions for random effects
 - ▷ Missing data issues
- All aspects discussed here equally well apply to generalized linear mixed models and non-linear mixed models.

12.2 Tests for variance components

- In a number of situations, it might be of interest to test whether variance components equal zero.
- For example, consider the Leuven diabetes project, it may be of interest to know whether there is any variability between GP's
- As before, let Y_{ijk} being the k th measurement of HbA1, for the j th patient, of the i th GP.
- Model 4 was given by:

$$Y_{ijk} = \beta_0 + \beta_1 t_k + a_i + b_{j(i)} + \varepsilon_{ijk},$$
$$a_i \sim N(0, \sigma_{GP}^2), \quad b_{j(i)} \sim N(0, \sigma_{PAT}^2), \quad \varepsilon_{ijk} \sim N(0, \sigma_{res}^2)$$

- Absence of any heterogeneity between GP's would be reflected in $\sigma_{GP}^2 = 0$
- It is therefore of interest to test $H_0 : \sigma_{GP}^2 = 0$ versus $H_A : \sigma_{GP}^2 > 0$
- The default output from SAS is:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
Intercept	mdnr	0.05439
Intercept	patientnr(mdnr)	0.6171
Residual		0.5837

- In contrast to, e.g., fixed effects, SAS does not report standard errors, test-statistics, nor p -values
- These can be requested by specifying the 'covtest' option in the PROC MIXED statement:

```
proc mixed data=dpla covtest;
```

- The output for the covariance parameters then becomes:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
Intercept	mdnr	0.05439	0.01858	2.93	0.0017
Intercept	patientnr(mdnr)	0.6171	0.03720	16.59	<.0001
Residual		0.5837	0.02256	25.87	<.0001

- The reported p -values are based on the $N(0, 1)$ approximation to the Z -statistic, which cannot reflect the correct sampling variability in the estimation of the variance components as these are estimated under the restriction of being positive.
- This so-called **boundary** problem requires correction of the classical p -values.
- The correction depends on the model, and sometimes requires simulation methods.
- In the above example, the correction reduces to halving the reported p -values.

- As another example, consider the previous analysis of the growth curves, with random intercepts, and random linear as well as quadratic Age effects.
- The reported tests for the variance components are:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	CHILD	96.3384	58.5931	1.64	0.0501
UN(2,1)	CHILD	-33.4752	17.5533	-1.91	0.0565
UN(2,2)	CHILD	11.5273	5.3639	2.15	0.0158
UN(3,1)	CHILD	2.0725	1.0800	1.92	0.0550
UN(3,2)	CHILD	-0.7160	0.3313	-2.16	0.0307
UN(3,3)	CHILD	0.04508	0.02069	2.18	0.0147
Residual		0.2655	0.05936	4.47	<.0001

- Apart from the boundary problem, the p -value reported for 'UN(3,3)' corresponds to the hypothesis $H_0 : d_{33} = 0$.

- Under H_0 , the random-effects covariance matrix D is of the form:

$$D = \begin{pmatrix} d_{11} & d_{12} & d_{13} \\ d_{12} & d_{22} & d_{23} \\ d_{13} & d_{23} & \mathbf{0} \end{pmatrix}$$

- As a covariance matrix, D needs to be positive (semi-)definite. Hence the only meaningful hypothesis to test would be $H_0 : d_{13} = d_{23} = d_{33} = 0$, implying that D is of the form

$$D = \begin{pmatrix} d_{11} & d_{12} & \mathbf{0} \\ d_{12} & d_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}$$

- Conclusion:

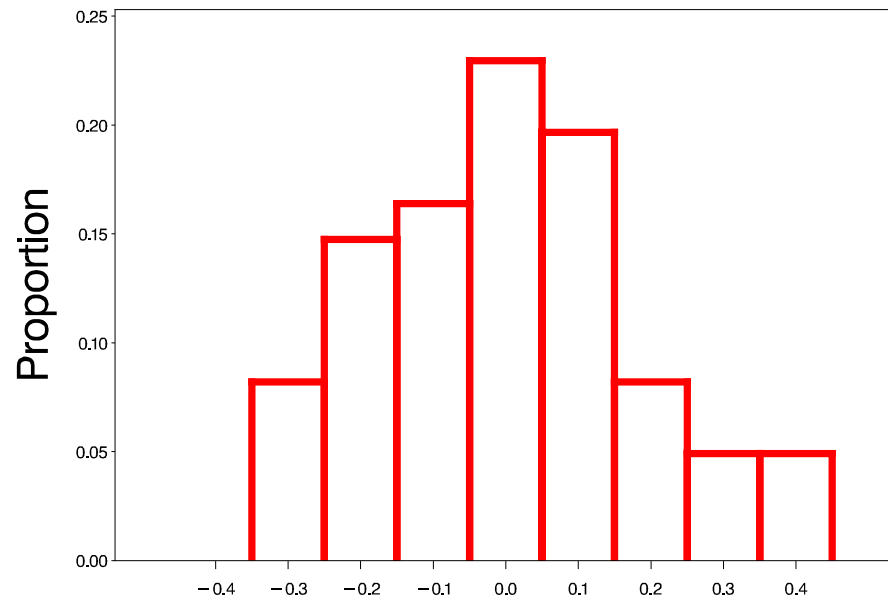
**The default variance components tests often
do not test meaningful hypotheses,
and/or report wrong p -values**

- SAS only reports Wald tests for variance components.
- However the above discussed problems equally well apply to Likelihood Ratio and Score tests, as the three are asymptotically equivalent.

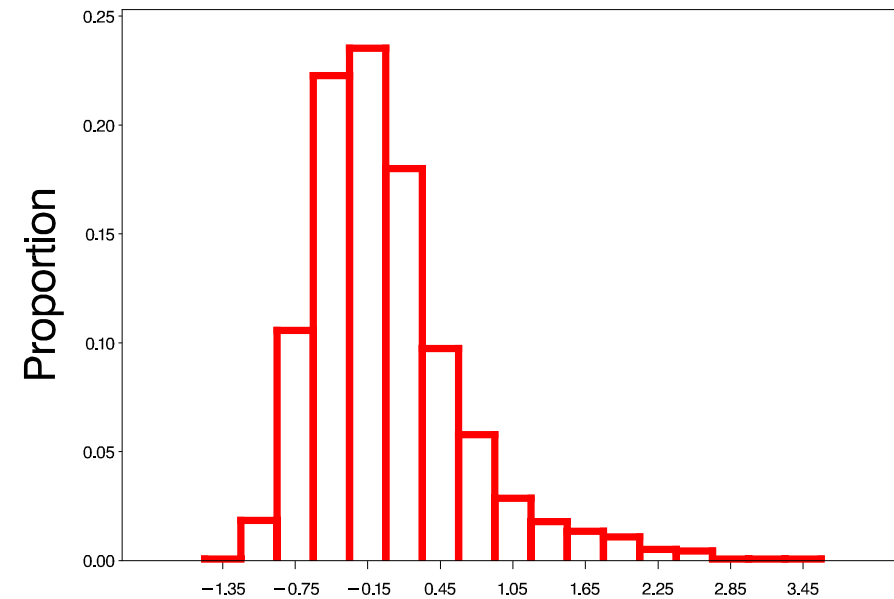
12.3 Distributional assumptions for random effects

- We continue the analysis of the Leuven diabetes project, with Model 4
- Histograms of EB estimates of GP and patient effects were:

EB estimates of GP effects



EB estimates of patient effects



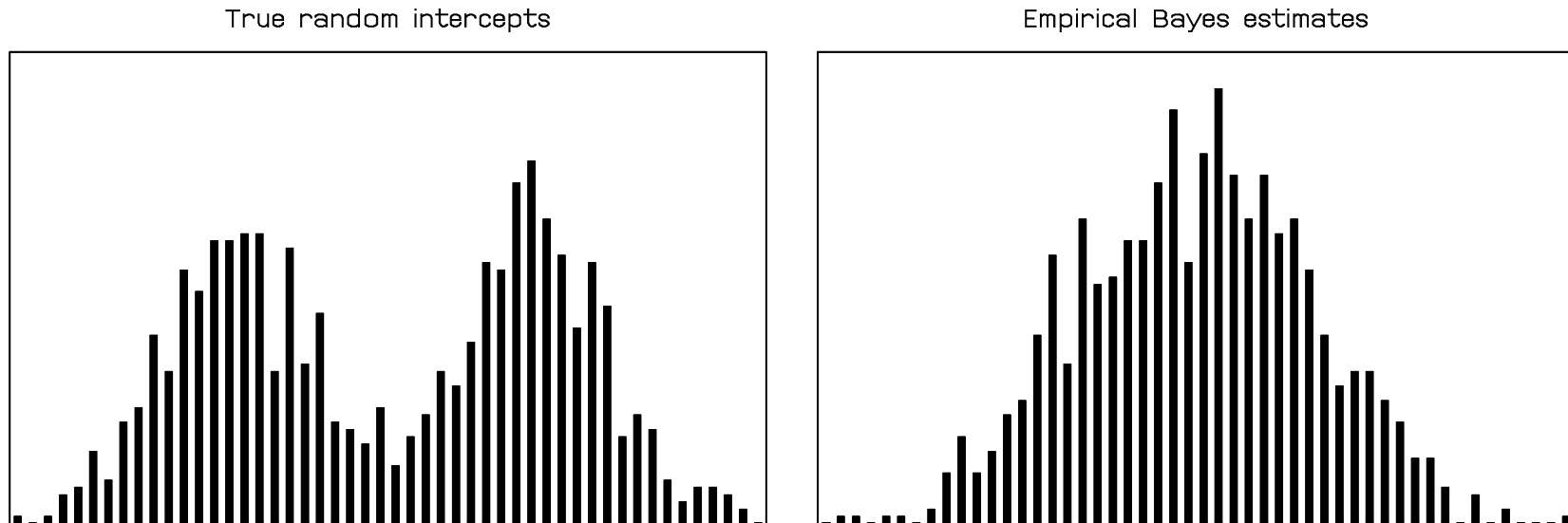
- The histograms seem to suggest that the normality assumption for the random effects a_i and $b_{j(i)}$ is questionable.
- However, one should realize that the precision with which a_i and $b_{j(i)}$ are estimated depends on many aspects, and can vary from patient to patient and from GP to GP
- So, the above histograms do not necessarily reflect non-normality of the random effects a_i and $b_{j(i)}$.
- Outlying EB estimates can be the reflection of a random effect estimated with very little precision.
- The differences in precision can be corrected for by standardizing the EB estimates.

- However, standardized EB estimates still do not necessarily reflect the correct random effects distribution.
- Too illustrate this, consider a small simulation example:
 - ▷ 1000 profiles with 5 measurements, balanced
 - ▷ 1000 random intercepts sampled from

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

- ▷ Error components ε_{ij} with variance $\sigma^2 = 30$
- ▷ Data analysed assuming normality for the intercepts

▷ Histogram of sampled intercepts and empirical Bayes estimates:



- Apparently, the model assumption sometimes forces the estimates to satisfy the assumption.

- Conclusion:

The normality assumption for random effects cannot be tested within the context of the linear mixed model.

Model extensions are needed.

- Fortunately, inferences about the fixed effects are very robust with respect to model deviations, provided the data set contains sufficient **independent** clusters:
 - ▷ Lizard data: sufficient mothers
 - ▷ Rat data: sufficient rats
 - ▷ Growth curves: sufficient children
 - ▷ Leuven diabetes project: sufficient GP's

12.4 Missing data issues

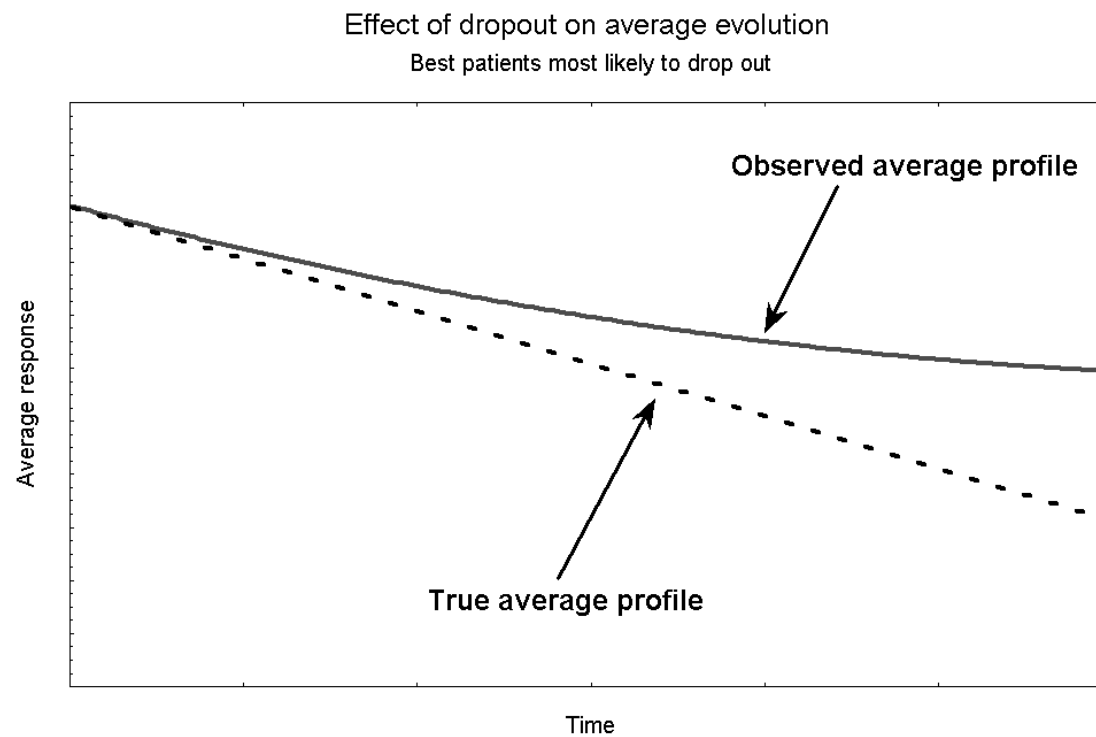
- A key feature of mixed models is that they can be used to model unbalanced data.
- In the context of longitudinal data, this includes situations where not all subjects have the same number of repeated measurements, or where subjects are measured at different time points.
- Mixed models are therefore often used in contexts with missing data, e.g., subjects left the study prematurely.
- However although mixed models can technically handle such unbalanced data sets, the obtained results can be severely biased in cases where missingness is related to the outcome studied.

- General principle:

Dropout related to the outcome can imply biased results

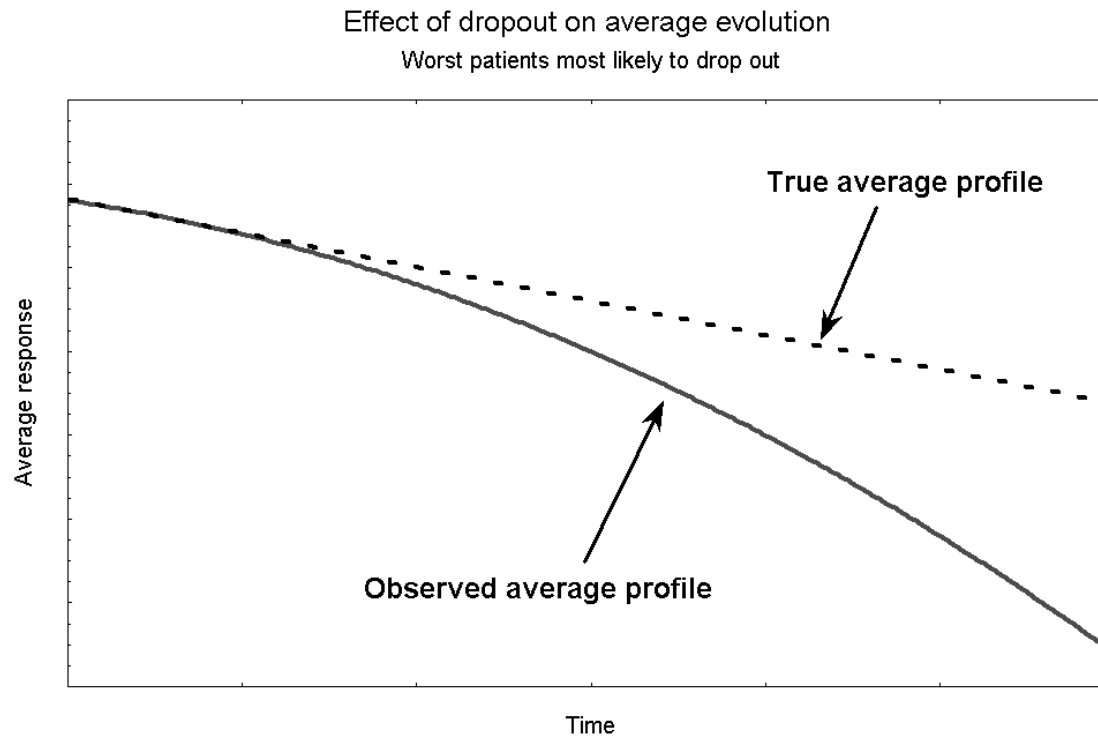
- Unrelated dropout:
 - ▷ Subjects moving
 - ▷ Subjects dying of other causes
 - ▷ Lost blood samples
 - ▷ ...
- If dropout is unrelated to the outcome, the obtained sample can be considered as a random sub-sample, which is still a random sample from the original population

- Related dropout:
 - ▷ 'Best' patients most likely to drop out:



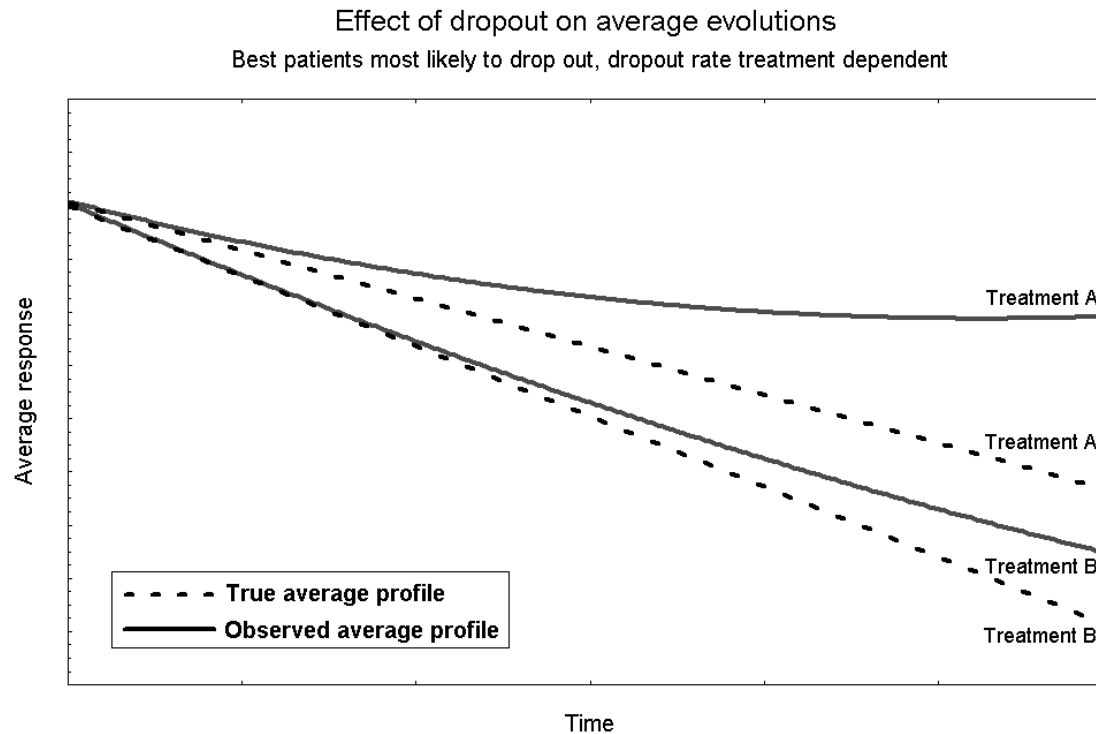
⇒ **Over-pessimistic**

▷ 'Worst' patients most likely to drop out:



⇒ **Over-optimistic**

- ▷ 'Best' patients most likely to drop out, but dropout rate dependent on treatment:



⇒ **Biased estimation of treatment effect**

Part III

Generalized Linear Mixed Models

Chapter 13

The toenail data

- ▷ Example
- ▷ Logistic regression
- ▷ A logistic mixed model
- ▷ Analysis in SAS

13.1 Example

- **T**oenail **D**ermatophyte **O**nychomycosis: 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.
- The multicenter nature will be ignored here. An example will be given later, in the context of the Leuven Diabetes Project.

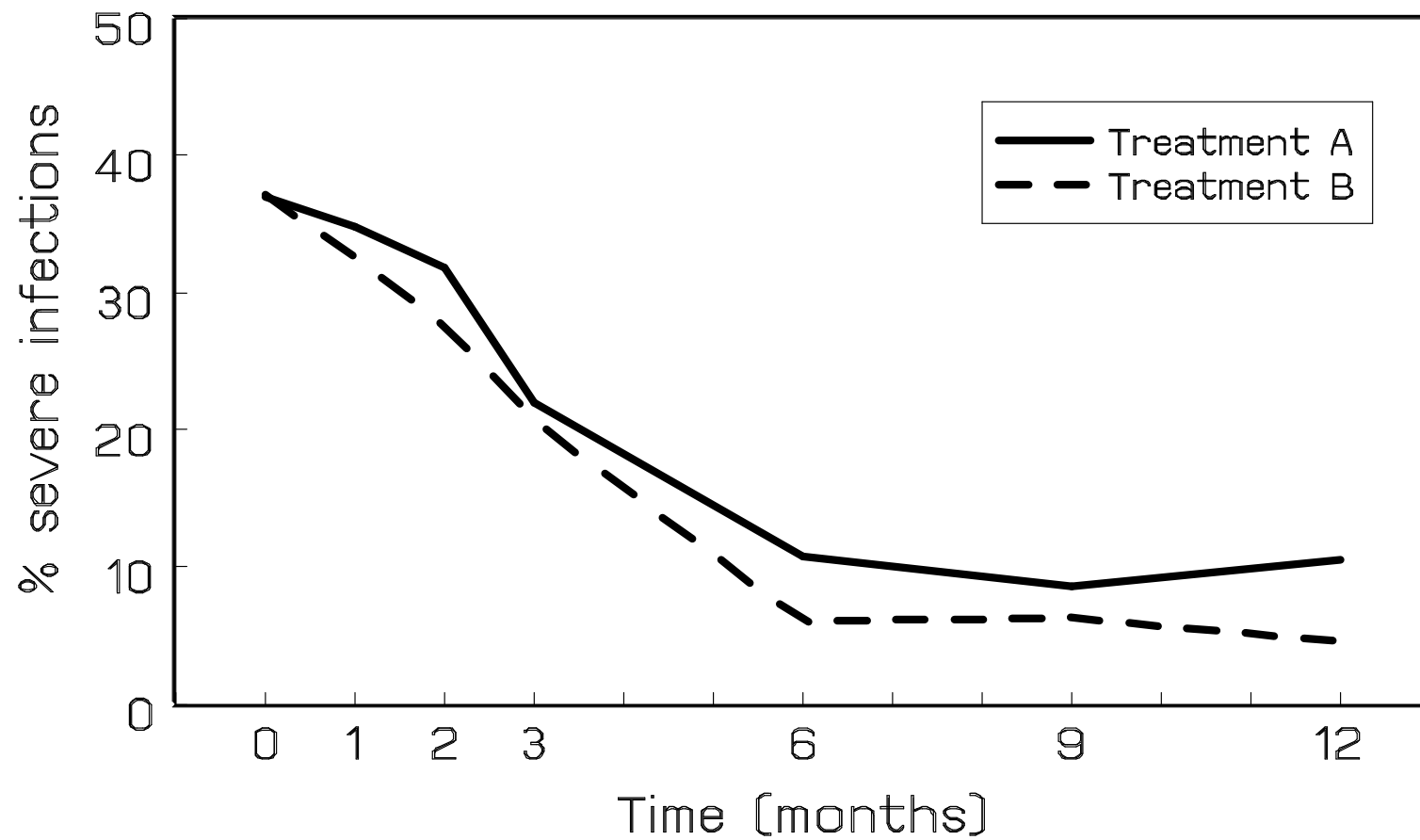
- Research question:

Severity relative to treatment of TDO ?

- 2×189 patients randomized, 36 centers
- Focus here on patients for which the target nail was one of the big toenails
⇒ 150 and 148 patients only
- 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:

Toenail data



- 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

- The toenail data set is an example of a longitudinal study, with unbalanced binary data

13.2 Logistic regression

- As in earlier examples, the toenail data are clustered within study participants
- As before, let Y_{ij} denote the j th measurement taken on the i th patient
- Ignoring the clustering, a typical analysis for studying the relation between Y_{ij} and some known covariates such as time and treatment would be based on logistic regression.
- We then assume a Bernoulli distribution: $Y_{ij} \sim \text{Bernoulli}(\pi_{ij})$
- π_{ij} is the probability for outcome Y_{ij} to be a 'success', i.e., $\pi_{ij} = P(Y_{ij} = 1)$.

- A logistic relation is assumed between π_{ij} and the covariates:

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

- Notation:

▷ T_i : treatment indicator for subject i

▷ t_{ij} : time point at which j th measurement is taken for i th subject

- More complex models can be considered as well (e.g. including polynomial time effects, including covariates, ...).

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

```
proc genmod data=toenail descending;  
model y = treatn time treatn*time / dist=binomial ;  
run;
```

- Selected output:

The GENMOD Procedure

Model Information

Data Set	WORK.TOENAIL
Distribution	Binomial
Link Function	Logit
Dependent Variable	Y
Observations Used	1908

The GENMOD Procedure

Response Profile

Ordered Value	Y	Total Frequency
1	1	408
2	0	1500

PROC GENMOD is modeling the probability that Y='1'.

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1904	1811.8260	0.9516
Scaled Deviance	1904	1811.8260	0.9516
Pearson Chi-Square	1904	1995.2107	1.0479
Scaled Pearson X2	1904	1995.2107	1.0479
Log Likelihood		-905.9130	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	-0.5571	0.1090	-0.7708	-0.3433	26.10	<.0001
treatn	1	0.0240	0.1565	-0.2827	0.3307	0.02	0.8780
time	1	-0.1769	0.0246	-0.2251	-0.1288	51.91	<.0001
treatn*time	1	-0.0783	0.0394	-0.1556	-0.0010	3.95	0.0470
Scale	0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

- We find a significant difference in evolution between the two treatment groups ($p = 0.0470$), where group 1 shows more improvement than treatment group 0.

13.3 A logistic mixed model

- As for the linear models, the clustering can be accounted for by including random effects
- We then explicitly model the belief that not all clusters satisfy the same model with the same parameter values, but intercepts and/or slopes are allowed to be cluster specific.
- For example, a logistic random-intercepts model is obtained as:

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

$$\text{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_0 + \mathbf{b}_i + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij}$$

- As before, the b_i are assumed to be normally distributed: $b_i \sim N(0, \sigma_b^2)$
- A logistic model with random intercepts and random slopes is obtained as:

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

$$\text{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = (\beta_0 + \mathbf{b}_{1i}) + \beta_1 T_i + (\beta_2 + \mathbf{b}_{2i})t_{ij} + \beta_3 T_i t_{ij}$$

- As before, the random effects are assumed to follow a bivariate normal distribution with mean zero:

$$\mathbf{b}_i = (b_{1i}, b_{2i})' \sim N(\mathbf{0}, D)$$

- The logistic mixed model is an example of a **generalized linear mixed model (GLMM)**

- A number of estimation methods is available for fitting GLMM's:
 - ▷ Laplace approximation
 - ▷ Marginal quasi-likelihood (MQL)
 - ▷ Penalized quasi-likelihood (PQL)
 - ▷ (Adaptive) Gaussian quadrature
 - ▷ ...
- Different estimation methods can lead to (strong) differences in the results
- In this course, PQL will be used.
- Next to the estimation of the fixed effects (β parameters) and variance components (elements in D), **empirical Bayes (EB)** estimates for the random effects b_i can be calculated as well.

13.4 Analysis in SAS

- Logistic mixed models can be fitted within the GLIMMIX procedure.
- Up to SAS version 9.1, the GLIMMIX procedure is not part of the standard SAS package. It can be downloaded from the SAS website. Once installed, it remains available for future SAS sessions.

<http://www.sas.com/apps/demosdownloads/setupcat.jsp?cat=SAS%2FSTAT+Software>

- We consider the random-intercepts model:

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

- The model specification in GLIMMIX is very similar to the way linear mixed models were specified in the MIXED procedure.

- SAS program for the random-intercepts model:

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

- Selected output:

Response Profile

Ordered Value	onyresp	Total Frequency
1	0	1500
2	1	408

The GLIMMIX procedure is modeling the probability that onyresp='1'.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Intercept	idnum	4.7116	0.6031

Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	-0.7239	0.2370	292	-3.05	0.0025
treatn	0.000918	0.3363	1612	0.00	0.9978
time	-0.2883	0.03349	1612	-8.61	<.0001
treatn*time	-0.1106	0.05366	1612	-2.06	0.0395

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
treatn	1	1612	0.00	0.9978
time	1	1612	74.10	<.0001
treatn*time	1	1612	4.25	0.0395

- Ignoring the clustering, the difference in slopes between the two treatment groups was estimated as -0.0783 ($p = 0.0470$).
- Under the random-intercepts model this becomes -0.1106 ($p = 0.0395$).

- The variance between subjects is estimated as $\hat{\sigma}_b^2 = 4.7116$
- Our model assumes different fixed intercepts and slopes for both groups.
- Direct estimation of these can be done based on the following reparameterization of the same model:

$$Y_{ij} \sim \text{Bernoulli}(\pi_{ij}), \quad \log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \begin{cases} \beta_1 + b_i + \beta_2 t_{ij}, & \text{Treatment A} \\ \beta_3 + b_i + \beta_4 t_{ij}, & \text{Treatment B} \end{cases}$$

- The corresponding SAS code becomes:

```
proc glimmix data=test;
class idnum treatn;
model onyresp (event='1') = treatn treatn*time / noint dist=binary solution;
random intercept / subject=idnum;
estimate 'difference slopes' treatn*time 1 -1;
run;
```

- The ESTIMATE statement is used to estimate and test the difference between the slopes β_2 and β_4 of both treatment groups.
- Selected output:

Covariance Parameter Estimates				Type III Tests of Fixed Effects				
Cov Parm	Subject	Estimate	Standard Error	Effect	Num DF	Den DF	F Value	Pr > F
Intercept	idnum	4.7116	0.6031	treatn	2	1612	9.26	0.0001
				time*treatn	2	1612	82.30	<.0001

Solutions for Fixed Effects						
Effect	treatn	Estimate	Standard Error	DF	t Value	Pr > t
treatn	0	-0.7239	0.2370	1612	-3.05	0.0023
treatn	1	-0.7230	0.2386	1612	-3.03	0.0025
time*treatn	0	-0.2883	0.03349	1612	-8.61	<.0001
time*treatn	1	-0.3989	0.04193	1612	-9.51	<.0001

Label	Estimates		DF	t Value	Pr > t
	Estimate	Standard Error			
difference slopes	0.1106	0.05366	1612	2.06	0.0395

- Note again that the standard reported F -tests test whether both intercepts and both slopes are equal to zero, respectively.

- Summary of model fit:

Effect	Parameter	Estimate (s.e.)
Intercept group A	β_1	-0.7239 (0.2370)
Intercept group B	β_3	-0.7230 (0.2386)
Slope group A	β_2	-0.2883 (0.0335)
Slope group B	β_4	-0.3989 (0.0419)
Variance random intercepts	σ_b^2	4.7116 (0.6031)

- As an example, a logistic mixed model with uncorrelated random intercepts and slopes can be fitted with the following SAS code:

```
proc glimmix data=test ;
class idnum treatn;
model onyresp (event='1') = treatn treatn*time / noint dist=binary solution;
random intercept timetrans / type=un(1) subject=idnum;
estimate 'difference slopes' treatn*time 1 -1;
run;
```

- The option 'type=un(1)' specifies that the covariance matrix D should be diagonal.
- Selected output:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
UN(1,1)	idnum	4.6285	0.6143
UN(2,1)	idnum	0	.
UN(2,2)	idnum	0.0747	0.0184

Solutions for Fixed Effects

Effect	treatn	Estimate	Standard Error	DF	t Value	Pr > t
treatn	0	-0.6653	0.2375	1325	-2.80	0.0052
treatn	1	-0.6144	0.2396	1325	-2.56	0.0105
time*treatn	0	-0.3528	0.04997	1325	-7.06	<.0001
time*treatn	1	-0.4983	0.05952	1325	-8.37	<.0001

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > t
difference slopes	0.1455	0.07771	1325	1.87	0.0615

- Under the random-intercepts model, the difference in slopes between the two treatment groups was estimated as 0.1106 ($p = 0.0395$).
- Under the model with uncorrelated random intercepts and slopes, this becomes 0.1455 ($p = 0.0615$).

- Note that the estimation of the fixed effects is also affected by including the random slopes:

Effect	Parameter	Random intercepts	Random intercepts & slopes
		Estimate (s.e.)	Estimate (s.e.)
Intercept group A	β_1	-0.7239 (0.2370)	-0.6653 (0.2375)
Intercept group B	β_3	-0.7230 (0.2386)	-0.6144 (0.2396)
Slope group A	β_2	-0.2883 (0.0335)	-0.3528 (0.0450)
Slope group B	β_4	-0.3989 (0.0419)	-0.4983 (0.0596)
Variance random intercepts	d_{11}	4.7116 (0.6031)	4.6285 (0.6143)
Variance random slopes	d_{22}	_____	0.0747 (0.0184)

Chapter 14

The Leuven diabetes project

- ▷ Example
- ▷ A three-level logistic mixed regression model
- ▷ Analysis in SAS

14.1 Example

- Linear mixed models were used earlier to study the evolution of HbA1c in DPL participants, correcting for the clustered nature of the data:
 - ▷ within general practitioners (GP's)
 - ▷ within subjects
- A related outcome of scientific interest is whether the GP is able to keep the HbA1c level under control, i.e., to keep it below 7%
- Hence, the derived outcome of interest is defined as:

$$Y = \begin{cases} 1 & \text{if HbA1c} < 7\% \\ 0 & \text{if HbA1c} \geq 7\% \end{cases}$$

14.2 A three-level logistic mixed regression model

- Let Y_{ijk} be the k th binary outcome measure for patient j of GP i
- Ignoring potential important covariates, a model which accounts for the clustering of the outcomes within patients and GP's is a three-level logistic mixed model:

$$Y_{ijk} \sim \text{Bernoulli}(\pi_{ijk})$$

$$\text{logit}(\pi_{ijk}) = \log\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right) = \beta_0 + \beta_1 t_k + \mathbf{a}_i + \mathbf{b}_{j(i)},$$

$$\mathbf{a}_i \sim N(0, \sigma_{GP}^2), \quad \mathbf{b}_{j(i)} \sim N(0, \sigma_{PAT}^2)$$

- The GP effects $b_{j(i)}$ represent the fact that some GP's are more succesful in controlling the HbA1c level of their patients than others.
- The patient effects a_i represent the fact that controlling the HbA1c level is not equally easy for all patients

14.3 Analysis in SAS

- The SAS code to fit the three-level logistic mixed model equals:

```
proc glimmix data=dpla;
class mdnr patientnr ;
model target (event='1') = time / dist=binary solution;
random intercept / subject=mdnr solution;
random intercept / subject=patientnr(mdnr) solution;
ods listing exclude solutionr;
ods output solutionr=out;
run;
```

- The option 'solution' is added to the RANDOM statement to request calculation of the EB estimates
- As in the MIXED procedure, ODS statements can be used to save the EB estimates into an output data set, rather than print them in the output screen.

- Selected SAS output:

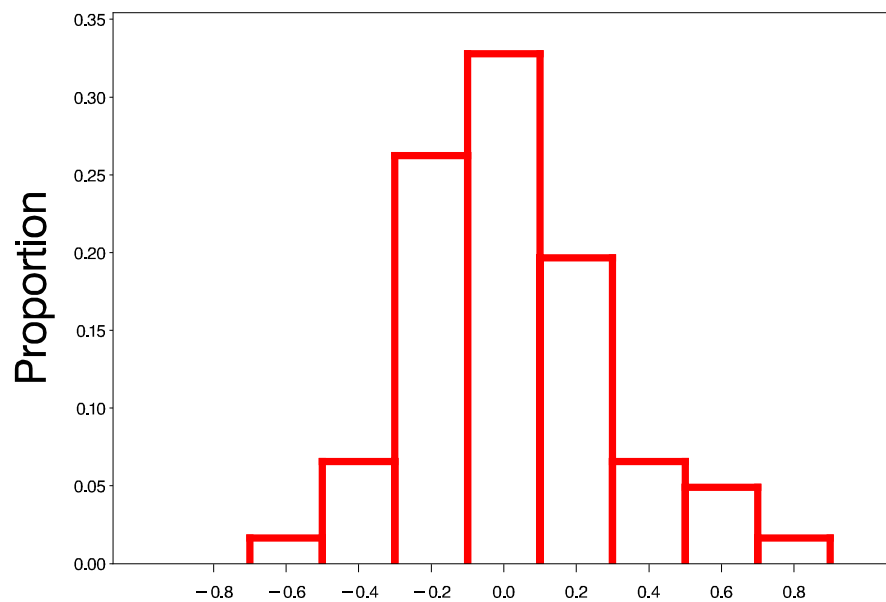
Covariance Parameter Estimates				Type III Tests of Fixed Effects				
Cov Parm	Subject	Estimate	Standard Error	Effect	Num DF	Den DF	F Value	Pr > F
Intercept	mdnr	0.1399	0.05275					
Intercept	patientnr(mdnr)	1.1154	0.1308	time	1	1395	59.07	<.0001

Solutions for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	0.1662	0.07960	60	2.09	0.0410
time	0.6240	0.08119	1395	7.69	<.0001

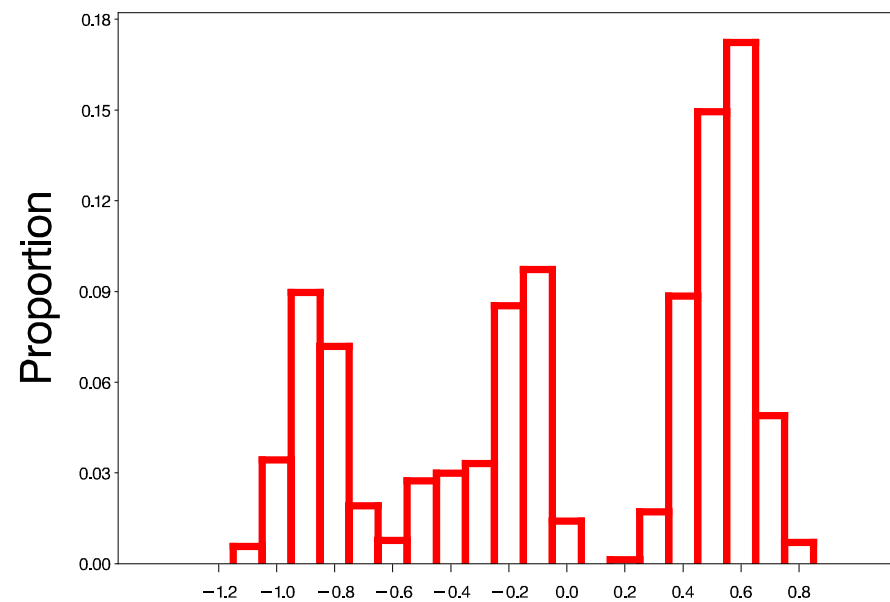
- As for the continuous outcome, we observe far more variability between patients than between GP's;
 - ▷ Between-GP variability: $\hat{\sigma}_{GP}^2 = 0.1399$
 - ▷ Between-patient variability: $\hat{\sigma}_{PAT}^2 = 1.1154$

- Overall, the probability of reaching the target HbA1c ($< 7\%$) increases over time ($p < 0.0001$)
- Histograms of both sets of EB estimates:

EB estimates of GP effects



EB estimates of patient effects



- The histogram of EB estimates $\hat{b}_{j(i)}$ of patient effects suggests three clusters of patients, with approximate cut-offs for $\hat{b}_{j(i)}$ equal to -0.6 and 0.1
- These clusters reflect the possible patient-specific profiles:
 - ▷ Patients with $Y_0 = Y_1 = 0$ are expected to have small predicted probabilities for reaching the target. Their prediction $\hat{b}_{j(i)}$ is expected to be very small (negative).
 - ▷ Patients with $Y_0 = Y_1 = 1$ are expected to have large predicted probabilities for reaching the target. Their prediction $\hat{b}_{j(i)}$ is expected to be very large (positive).
 - ▷ Patients who change $Y_0 = 0 \longrightarrow Y_1 = 1$ or $Y_0 = 1 \longrightarrow Y_1 = 0$ are expected to have intermediate predicted probabilities for reaching the target. Their prediction $\hat{b}_{j(i)}$ is expected to be of a moderate level.

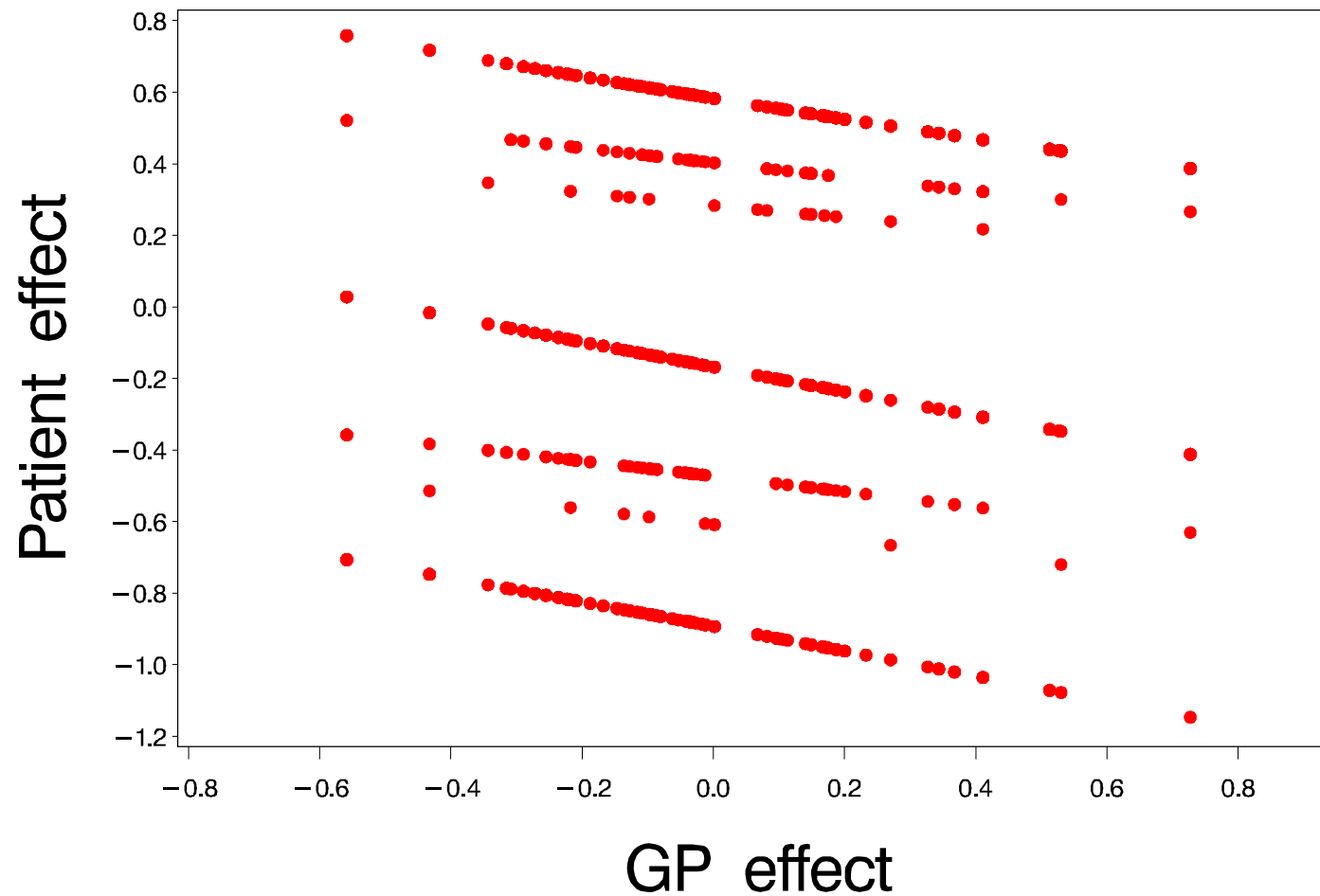
- This can be quantified in the following cross-classification:

Y profile	$\hat{b}_{j(i)} < -0.6$	$-0.6 \leq \hat{b}_{j(i)} < -0.1$	$-0.1 \leq \hat{b}_{j(i)}$
$0 \longrightarrow 0$	345	0	0
$0 \longrightarrow 1$	0	275	0
$1 \longrightarrow 1$	0	0	677

- All patients with HbA1c at target at the start of the study have their HbA1c at target one year later as well.

- Scatterplot of patient effects versus GP effects:

Scatterplot of EB estimates



- For each GP, we observe at most 7 different values for the EB estimates for the patients treated by that GP.
- These 7 values correspond to the 7 different response profiles that can be observed: $0 \longrightarrow 0$, $0 \longrightarrow 1$, $1 \longrightarrow 1$, $0 \longrightarrow \cdot$, $1 \longrightarrow \cdot$, $\cdot \longrightarrow 0$, and $\cdot \longrightarrow 1$.
- The negative trends observed in the scatterplot are also a side effect of the discrete nature of the outcomes.
- Consider two patients, j_1 and j_2 , treated by different GP's, i_1 and i_2 , with the same response profile, e.g., $1 \longrightarrow 1$
- Their subject-specific models are given by:

$$\text{logit}(\pi_{i_1 j_1 k}) = \beta_0 + \beta_1 t_k + a_{i_1} + b_{j_1(i_1)}, \quad \text{for patient } j_1$$

$$\text{logit}(\pi_{i_2 j_2 k}) = \beta_0 + \beta_1 t_k + a_{i_2} + b_{j_2(i_2)}, \quad \text{for patient } j_2$$

- Since both patients have the same data, we expect their predicted probabilities to be the same at all time points, implying

$$a_{i_1} + b_{j_1(i_1)} = a_{i_2} + b_{j_2(i_2)}$$

- Hence, we expect the sum $a_i + b_{j(i)}$ of GP and patient effects to be constant, explaining the strong negative relation between the estimates \hat{a}_i and $\hat{b}_{j(i)}$.

Chapter 15

The Epilepsy data

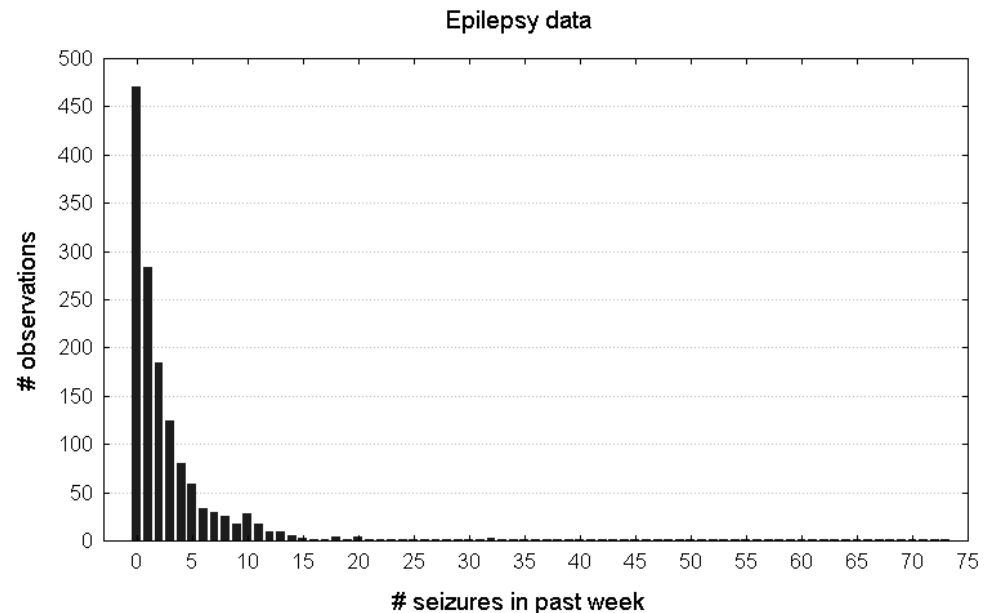
- ▷ Example
- ▷ Poisson regression
- ▷ A Poisson mixed model
- ▷ Analysis in SAS

15.1 Example

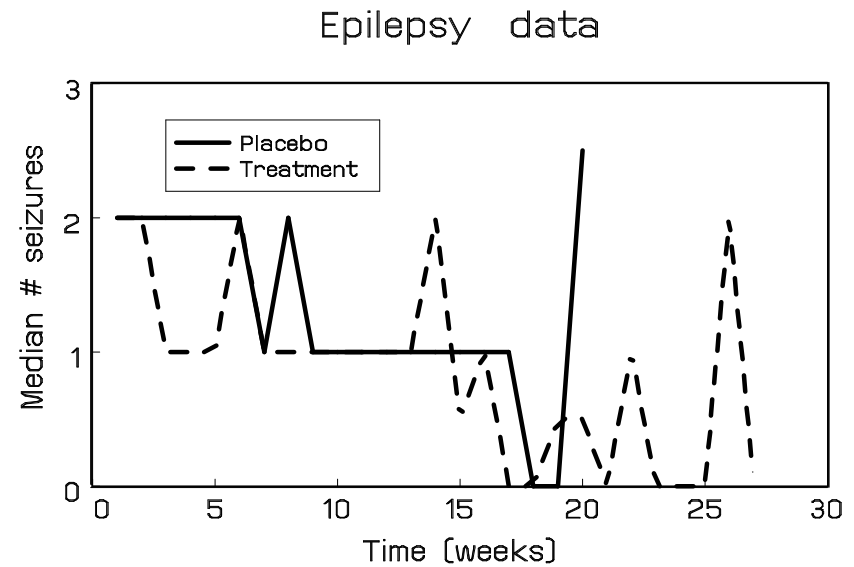
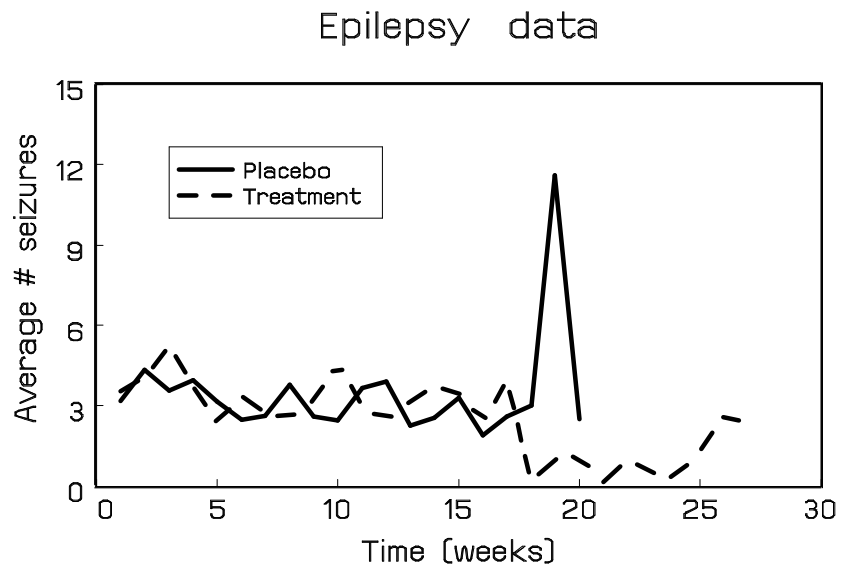
- Randomized, double-blind, parallel group multi-center study for the comparison of placebo with a new anti-epileptic drug (AED), in combination with one or two other (standard) AED's.
- Randomization after a 12-week stabilization period.
- 45 patients in placebo group, 44 in active (new) treatment group
- Double-blind weekly measurements during 16 weeks.
- Afterwards, patients enter a long-term open-extension study, with some patients followed for up to 27 weeks

- The outcome of interest is the number of epileptic seizures experienced during the last week, i.e., since the last time the outcome was measured.
- Number of observations and histogram of the weekly outcome measurements:

Week	# Observations		
	Placebo	Treatment	Total
1	45	44	89
5	42	42	84
10	41	40	81
15	40	38	78
16	40	37	77
17	18	17	35
20	2	8	10
27	0	3	3



- Average and median profiles for both treatments:



- Unstable behavior due to extreme values and few observations past week 20.

15.2 Poisson regression

- As in earlier examples, the data are clustered within study participants
- Ignoring the clustering, a typical analysis for studying the relation between a count outcome Y_{ij} and some known covariates such as time and treatment would consist of Poisson regression.
- We then assume a Poisson distribution: $Y_{ij} \sim \text{Poisson}(\lambda_{ij})$
- The parameter λ_{ij} is the expected (average) count, i.e., $\lambda_{ij} = E(Y_{ij})$.
- A logarithmic relation is assumed between λ_{ij} and the covariates:

$$\log(\lambda_{ij}) = \begin{cases} \beta_1 + \beta_2 t_{ij} & \text{if placebo (group 0)} \\ \beta_3 + \beta_4 t_{ij} & \text{if treated (group 1)}. \end{cases}$$

- More complex models can be considered as well (e.g. including polynomial time effects, including covariates, ...).

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

```
proc genmod data=test;
class trt;
model nseizw = trt trt*time / noint dist=poisson ;
estimate 'slope difference' trt*time 1 -1 ;
run;
```

- The ESTIMATE statement has been added to estimate the difference between the two slopes.

- Selected output:

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald	95% Confidence Limits	Chi-Square	Pr > ChiSq	
Intercept	0	0.0000	0.0000	0.0000	0.0000	.	.	
trt	0	1	1.2662	0.0424	1.1832	1.3493	892.71	<.0001
trt	1	1	1.4531	0.0383	1.3781	1.5282	1439.57	<.0001
time*trt	0	1	-0.0133	0.0043	-0.0218	-0.0048	9.48	0.0021
time*trt	1	1	-0.0328	0.0038	-0.0403	-0.0253	73.86	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000			

Contrast Estimate Results

Label	Estimate	Standard Error	Alpha	Confidence Limits	Chi-Square	Pr > ChiSq	
slope difference	0.0195	0.0058	0.05	0.0081	0.0308	11.34	0.0008

- We find a significant difference in the evolution in the two treatment groups ($p = 0.0008$), where group 1 shows more improvement than treatment group 0.

15.3 A Poisson mixed model

- Correction for the clustered nature of the data can again be based on the inclusion of random effects which model the within-patient correlation.
- For example, consider the random-intercepts model:

$$Y_{ij} \sim \text{Poisson}(\lambda_{ij}), \quad \log(\lambda_{ij}) = \begin{cases} \beta_1 + \mathbf{b}_i + \beta_2 t_{ij} & \text{if placebo (group 0)} \\ \beta_3 + \mathbf{b}_i + \beta_4 t_{ij} & \text{if treated (group 1)}. \end{cases}$$

- As before, the subject-specific intercepts \mathbf{b}_i are assumed to follow a normal distribution $N(0, \sigma_b^2)$.
- Other random effects (slopes) can be introduced as well.

15.4 Analysis in SAS

- SAS code for the random-intercepts model:

```
proc glimmix data=test;
class id trt;
model nseizw = trt trt*time / noint dist=poisson solution;
random intercept / subject=id;
estimate 'slope difference' trt*time 1 -1 ;
run;
```

- Selected SAS output:

Covariance Parameter Estimates				Type III Tests of Fixed Effects				
Cov Parm	Subject	Estimate	Standard Error	Effect	Num DF	Den DF	F Value	Pr > F
Intercept	id	1.1462	0.1835	trt	2	1328	20.64	<.0001
				time*trt	2	1328	9.14	0.0001

Solutions for Fixed Effects

Effect	trt	Estimate	Standard Error	DF	t Value	Pr > t
trt	0	0.8413	0.1668	1328	5.04	<.0001
trt	1	0.6732	0.1692	1328	3.98	<.0001
time*trt	0	-0.01430	0.004404	1328	-3.25	0.0012
time*trt	1	-0.01200	0.004317	1328	-2.78	0.0055

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > t
slope difference	-0.00230	0.006167	1328	-0.37	0.7094

- In contrast to the significant interaction obtained before, ignoring the longitudinal nature of the data ($p = 0.0008$), we no longer find a difference between the two slopes β_2 and β_4 ($p = 0.7094$).
- The between-patient variability is estimated to be $\hat{\sigma}_b^2 = 1.1462$.

- Summary of model fit:

Effect	Parameter	Estimate (s.e.)
Intercept Placebo	β_1	0.8413 (0.1668)
Intercept Active	β_3	0.6732 (0.1692)
Slope Placebo	β_2	-0.0143 (0.0044)
Slope Active	β_4	-0.0120 (0.0043)
Variance random intercepts	σ_b^2	1.1462 (0.1835)

- A mixed model with subject-specific intercepts as well as time effects would be:

$$Y_{ij} \sim \text{Poisson}(\lambda_{ij}), \quad \log(\lambda_{ij}) = \begin{cases} \beta_1 + \mathbf{b}_{1i} + (\beta_2 + \mathbf{b}_{2i})t_{ij} & \text{if placebo (group 0)} \\ \beta_3 + \mathbf{b}_{1i} + (\beta_4 + \mathbf{b}_{2i})t_{ij} & \text{if treated (group 1)}. \end{cases}$$

- As before, the random effects are assumed to follow a bivariate normal distribution with mean zero:

$$\mathbf{b}_i = (b_{1i}, b_{2i})' \sim N(\mathbf{0}, D)$$

- New SAS code:

```
proc glimmix data=test;
class id trt;
model nseizw = trt trt*time / noint dist=poisson solution;
random intercept time / type=un subject=id solution;
estimate 'slope difference' trt*time 1 -1 ;
ods listing exclude solutionr;
ods output solutionr=out;
```

- Selected output:

Covariance Parameter Estimates				Type III Tests of Fixed Effects				
Cov Parm	Subject	Estimate	Standard Error	Effect	Num DF	Den DF	F Value	Pr > F
UN(1,1)	id	1.2577	0.2173					
UN(2,1)	id	-0.01891	0.008784	trt	2	1241	20.86	<.0001
UN(2,2)	id	0.002419	0.000565	time*trt	2	1241	5.08	0.0064

Solutions for Fixed Effects

Effect	trt	Estimate	Standard Error	DF	t Value	Pr > t
trt	0	0.9251	0.1768	1241	5.23	<.0001
trt	1	0.6844	0.1807	1241	3.79	0.0002
time*trt	0	-0.02687	0.009787	1241	-2.75	0.0061
time*trt	1	-0.01616	0.009976	1241	-1.62	0.1056

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > t
slope difference	-0.01071	0.01397	1241	-0.77	0.4436

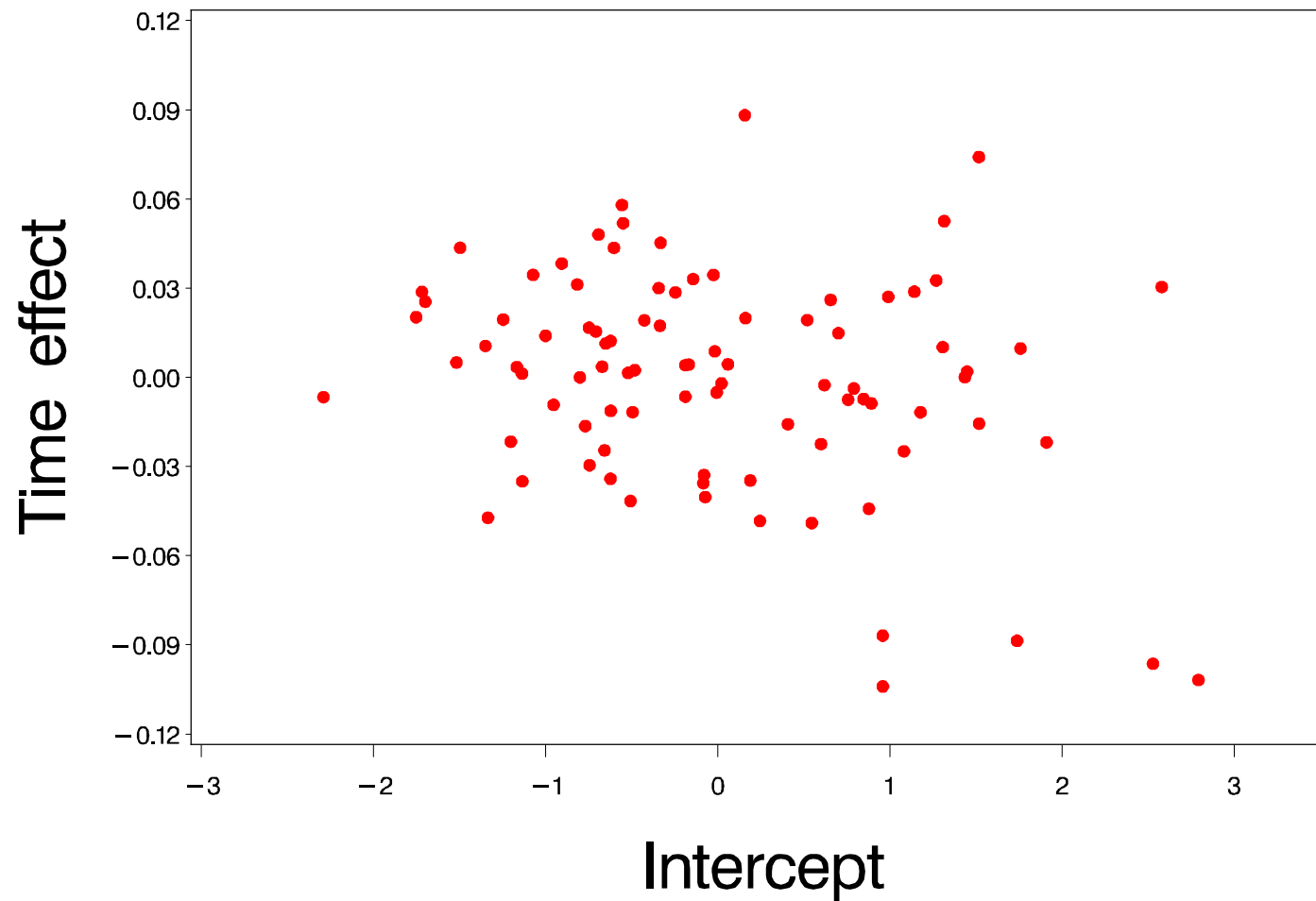
- Under the random-intercepts model, the difference in slopes between the two treatment groups was estimated as -0.0023 ($p = 0.7094$).
- Under the current model, this becomes -0.0107 ($p = 0.4436$).

- Note that the estimation of the fixed effects is also affected by including the random slopes:

Effect	Parameter	Random intercepts	Random intercepts & slopes
		Estimate (s.e.)	Estimate (s.e.)
Intercept Placebo	β_1	0.8413 (0.1668)	0.9251 (0.1768)
Intercept Active	β_3	0.6732 (0.1692)	0.6844 (0.1807)
Slope Placebo	β_2	-0.0143 (0.0044)	-0.0269 (0.0098)
Slope Active	β_4	-0.0120 (0.0043)	-0.0162 (0.0100)
Variance random intercepts	d_{11}	1.1462 (0.1835)	1.2577 (0.2173)
Covariance random intercepts & slopes	d_{12}	_____	-0.0189 (0.0088)
Variance random slopes	d_{22}	_____	0.0024 (0.0006)

- Scatterplot of estimated subject-specific intercepts and slopes:

Scatterplot of EB estimates



Chapter 16

The hierarchical versus marginal model

- ▷ Parameter interpretation in the GLMM
- ▷ Marginalizing the mixed model: The toenail data
- ▷ Marginalizing the mixed model: The epilepsy data

16.1 Parameter interpretation in the GLMM

- Let us re-consider one of the linear mixed models, used before for the model growth curve data:

$$Y_{ij}|b_i \sim \begin{cases} N(\beta_1 + b_i + \beta_2 t_j, \sigma_{res}^2), & \text{if short mother} \\ N(\beta_3 + b_i + \beta_4 t_j, \sigma_{res}^2), & \text{if medium mother} \\ N(\beta_5 + b_i + \beta_6 t_j, \sigma_{res}^2), & \text{if tall mother} \end{cases}$$

- This hierarchical model implied a very specific marginal model, with mean:

$$E(Y_{ij}) = \begin{cases} \beta_1 + \beta_2 t_j, & \text{if short mother} \\ \beta_3 + \beta_4 t_j, & \text{if medium mother} \\ \beta_5 + \beta_6 t_j, & \text{if tall mother} \end{cases}$$

- Hence, the fixed effects have a **subject-specific** interpretation as well as a **population-average** interpretation.
- Let us now consider the logistic random-intercepts model

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

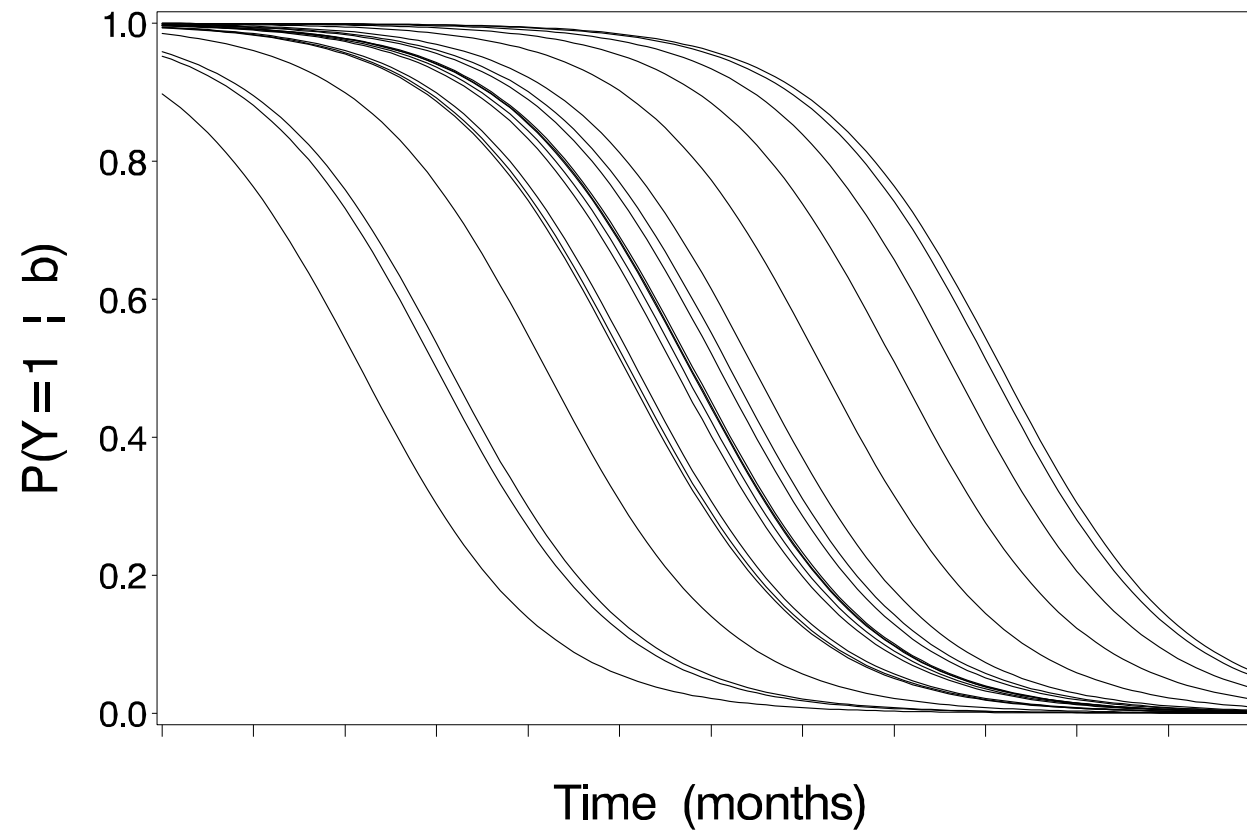
- Equivalently, we have

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

- The above model assumes a logistic evolution of the success probability of each patients, all curves having the same slope β_1 , but different intercepts $\beta_0 + b_i$.

- Graphically:

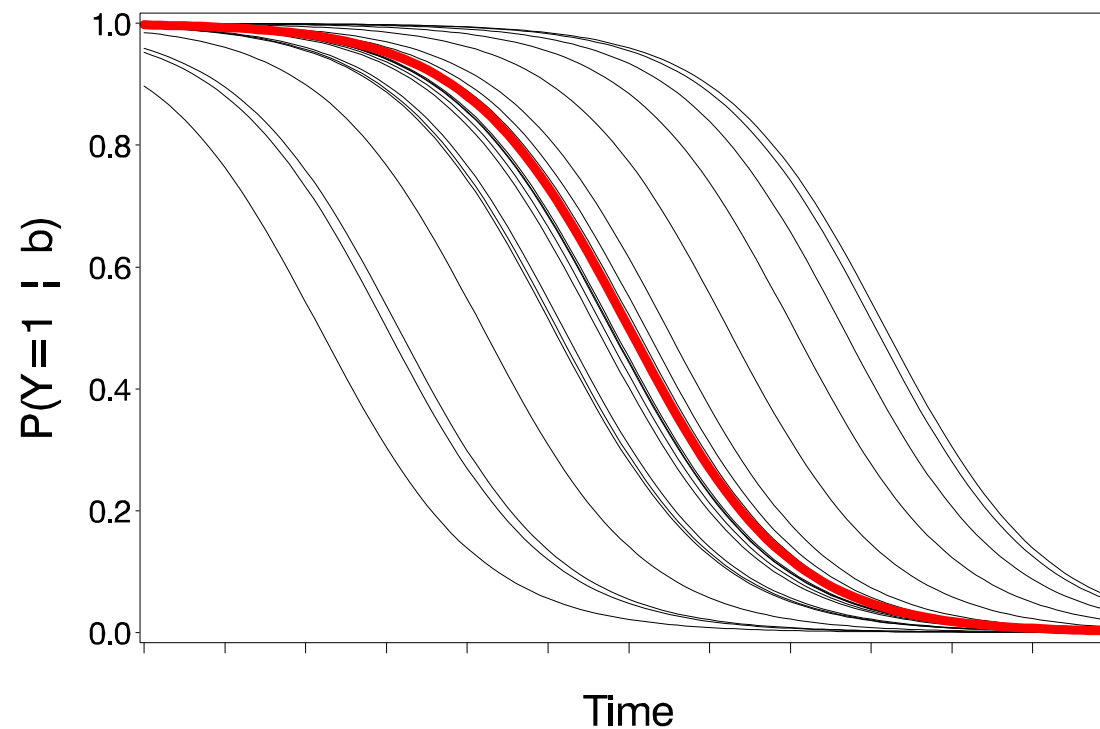
Subject – specific evolutions



- The average subject, i.e., the subject with intercept $b_i = 0$, has success probability given by

$$E(Y_{ij}|b_i = \mathbf{0}) = \frac{\exp(\beta_0 + \mathbf{0} + \beta_1 t)}{1 + \exp(\beta_0 + \mathbf{0} + \beta_1 t)}$$

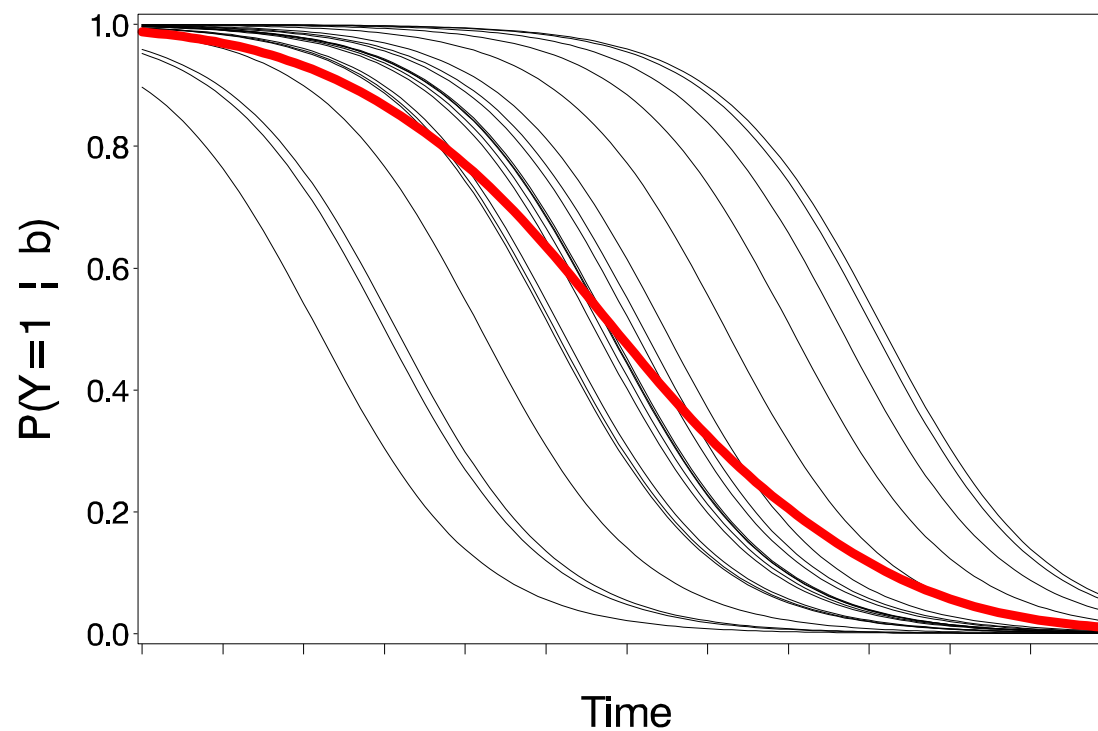
Evolution of average subject



- The marginal population-average evolution is obtained from averaging over the random effects:

$$E(Y_{ij}) = E[E(Y_{ij}|b_i)] = E\left[\frac{\exp(\beta_0 + b_i + \beta_1 t)}{1 + \exp(\beta_0 + b_i + \beta_1 t)}\right] \neq \frac{\exp(\beta_0 + \mathbf{0} + \beta_1 t)}{1 + \exp(\beta_0 + \mathbf{0} + \beta_1 t)}$$

Average evolution



- Conclusion:

Average evolution \neq Evolution average subject

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

- Calculation of the marginal average population requires computation of

$$\begin{aligned} E(Y_{ij}) &= E[E(Y_{ij}|b_i)] = E\left[\frac{\exp(\beta_0 + b_i + \beta_1 t)}{1 + \exp(\beta_0 + b_i + \beta_1 t)}\right] \\ &= \int \frac{\exp(\beta_0 + b_i + \beta_1 t)}{1 + \exp(\beta_0 + b_i + \beta_1 t)} f(b_i) db_i \end{aligned}$$

- This can be done using numerical integration methods, or using sampling based averaging.
- Note that what has been explained here in the context of logistic mixed models equally well applies to every other generalized linear or non-linear model.

16.2 Marginalizing the mixed model: The toenail data

- We re-consider the logistic mixed model with random intercepts.
- The fitted model is given by:

$$Y_{ij} \sim \text{Bernoulli}(\pi_{ij}), \quad \pi_{ij} = \begin{cases} \frac{\exp(-0.7239 + b_i - 0.2883t_{ij})}{1 + \exp(-0.7239 + b_i - 0.2883t_{ij})}, & \text{Treatment A} \\ \frac{\exp(-0.7230 + b_i - 0.3989t_{ij})}{1 + \exp(-0.7230 + b_i - 0.3989t_{ij})}, & \text{Treatment B} \end{cases}$$

- The random effects b_i are normally distributed with mean 0 and variance $\hat{\sigma}_b^2 = 4.7116$.

- The marginal evolution in both groups is obtained from integrating over the random effects $b_i \sim N(0, 4.7116)$:

$$\pi_{ij} = \begin{cases} \int \frac{\exp(-0.7239+b_i-0.2883t_{ij})}{1+\exp(-0.7239+b_i-0.2883t_{ij})} f(b_i) db_i, & \text{Treatment A} \\ \int \frac{\exp(-0.7230+b_i-0.3989t_{ij})}{1+\exp(-0.7230+b_i-0.3989t_{ij})} f(b_i) db_i, & \text{Treatment B} \end{cases}$$

- SAS code:

```
data h;
do treat=0 to 1 by 1;
  do subject=1 to 1000 by 1;
    b=sqrt(4.7116)*rannor(-1) ;
    do t=0 to 12 by 0.1;
      if treat=0 then y=exp(-0.7239 + b -0.2883*t)/(1+ exp(-0.7239 + b -0.2883*t));
        else y=exp(-0.7230 + b -0.3989*t)/(1+ exp(-0.7230 + b -0.3989*t));
      output;
    end;
  end;
end;
end;
```



```

proc sort data=h;
by t treat;
run;

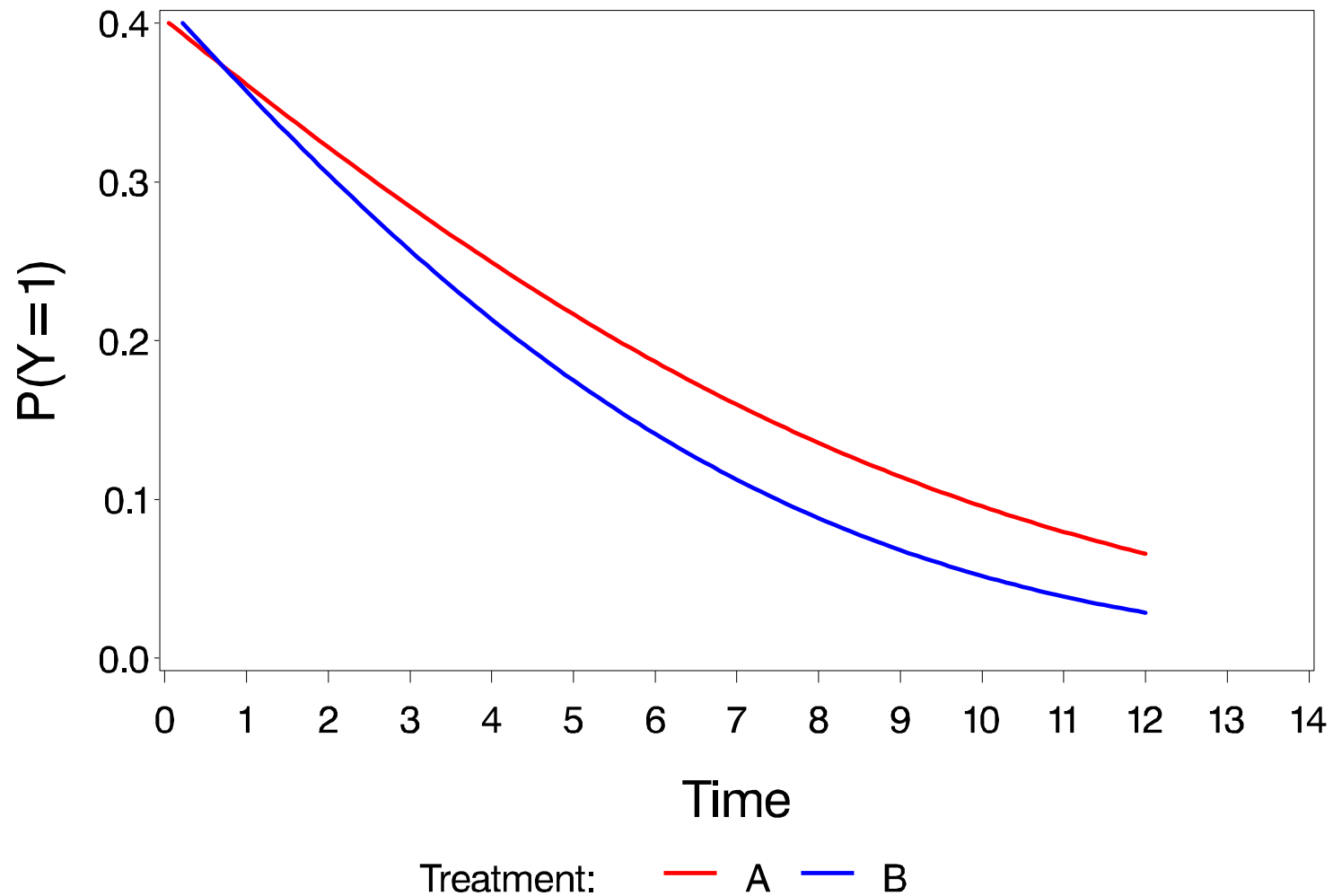
proc means data=h;
var y;
by t treat;
output out=out;
run;

dm "dlgprtsetup orient=L nodisplay";
filename fig 'c:/filename.eps';
goptions reset=all interpol=join ftext=swiss device=pslepsfc
          gsfname=fig gsfmode=replace targetdevice=winprtc;
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=red i=join w=20 l=1 mode=include;
symbol2 c=blue i=join w=20 l=1 mode=include;
where _stat_='MEAN';
run;quit;run;

```

- Result:

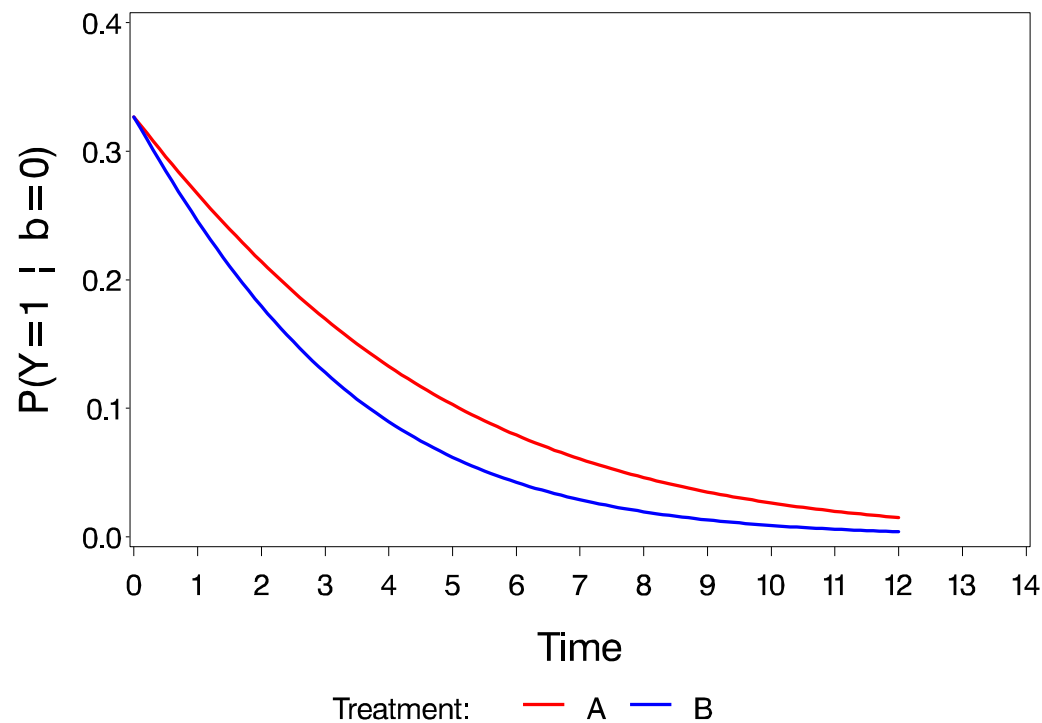
Marginal average evolutions (GLMM)



- The evolution of ‘average’ subjects, i.e., subjects with $b_i = 0$ is given by:

$$\pi_{ij} = \begin{cases} \frac{\exp(-0.7239 + \mathbf{0} - 0.2883t_{ij})}{1 + \exp(-0.7239 + \mathbf{0} - 0.2883t_{ij})}, & \text{Treatment A} \\ \frac{\exp(-0.7230 + \mathbf{0} - 0.3989t_{ij})}{1 + \exp(-0.7230 + \mathbf{0} - 0.3989t_{ij})}, & \text{Treatment B} \end{cases}$$

Evolutions for subjects with random effects zero (GLMM)



16.3 Marginalizing the mixed model: The epilepsy data

- We re-consider the fitted Poisson mixed model with random intercepts and slopes:

$$Y_{ij} \sim \text{Poisson}(\lambda_{ij})$$
$$\lambda_{ij} = \begin{cases} \exp[0.9251 + b_{1i} + (-0.0269 + b_{2i})t_{ij}] & \text{if placebo (group 0)} \\ \exp[0.6844 + b_{1i} + (-0.0162 + b_{2i})t_{ij}] & \text{if treated (group 1)}. \end{cases}$$

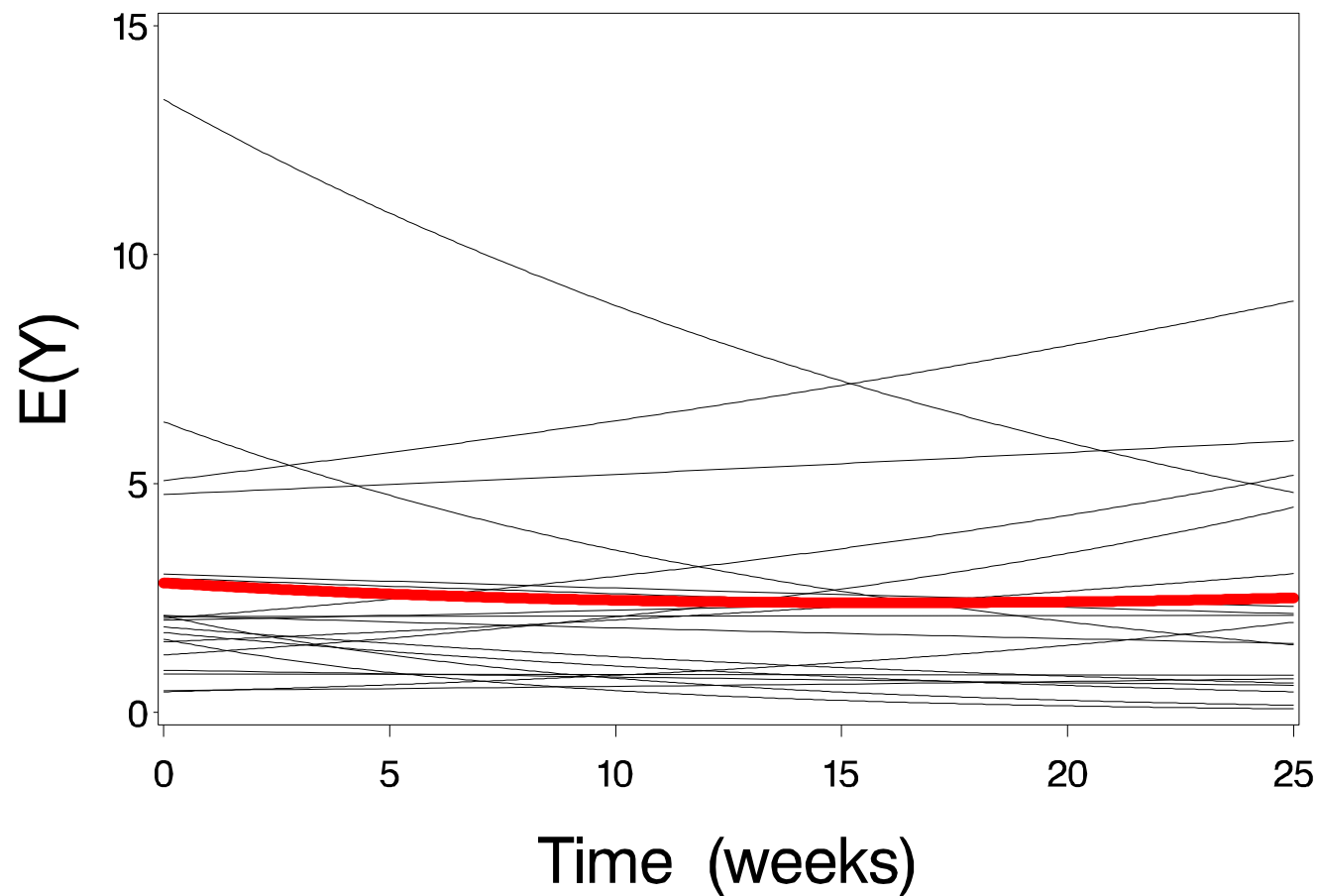
- The random-effects vector $\mathbf{b}_i = (b_{1i}, b_{2i})'$ is $N(\mathbf{0}, D)$ distributed, with fitted D :

$$\widehat{D} = \begin{pmatrix} 1.2577 & -0.0189 \\ -0.0189 & 0.0024 \end{pmatrix}.$$

- The non-linear link function again implies that the parameters only have a subject-specific interpretation.

- Subject-specific profiles for 20 randomly selected subjects, together with their average evolution:

Subject – specific and average evolutions



- The marginal average evolutions are obtained from integrating over the random effects \mathbf{b}_i :

$$E(Y_{ij}) = E [E(Y_{ij}|\mathbf{b}_i)]$$

$$= \begin{cases} \int \int \exp[0.9251 + b_{1i} + (-0.0269 + b_{2i})t_{ij}] f(b_{1i}, b_{2i}) db_{1i} db_{2i} & \text{if placebo (group 0)} \\ \int \int \exp[0.6844 + b_{1i} + (-0.0162 + b_{2i})t_{ij}] f(b_{1i}, b_{2i}) db_{1i} db_{2i} & \text{if treated (group 1).} \end{cases}$$

- As before, the integration can be approximated using sampling based averaging, which requires generating multivariate random vectors $\mathbf{b}_i \sim N(\mathbf{0}, \widehat{D})$
- Most software packages only allow generating univariate standard normals, which can then be transformed using the cholesky decomposition of the covariance \widehat{D}

- The cholesky decomposition L of \widehat{D} is the upper triangular matrix such that $L'L = \widehat{D}$
- In SAS, L can easily be calculated using the IML procedure:

```
proc iml;
d={1.2577 -0.0189 , -0.0189  0.0024};
l=root(d);
print d; print l;
quit;
```

- Output:

D		L	
1.2577	-0.0189	1.1214722	-0.016853
-0.0189	0.0024	0	0.0459998

- Let b_{1i}^* and b_{2i}^* be independent and standard normal distributed, and $\mathbf{b}_i^* = (b_{1i}^*, b_{2i}^*)'$

- We then have that

$$\mathbf{b}_i \equiv L'\mathbf{b}_i^* = \begin{pmatrix} 1.1215 b_{1i}^* + 0 b_{2i}^* \\ -0.0169 b_{1i}^* + 0.0460 b_{2i}^* \end{pmatrix} \sim N(\mathbf{0}, L'IL) = N(\mathbf{0}, \widehat{D})$$

- The SAS code for sampling based averaging:

```

data h;
do treat=0 to 1 by 1;
  do subject=1 to 1000 by 1;
    b1=rannor(-1);
    b2=rannor(-1);
    ranint=1.1215*b1;
    ranslope=-0.0169*b1 + 0.0460*b2;
    do t=0 to 27 by 0.1;
      if treat=0 then y=exp(0.9251+ranint +(-0.0269+ranslope)*t);
        else y=exp(0.6844+ranint +(-0.0162+ranslope)*t);
      output;
    end;
  end;
end;
end;
```



```

proc sort data=h;
by t treat;
run;

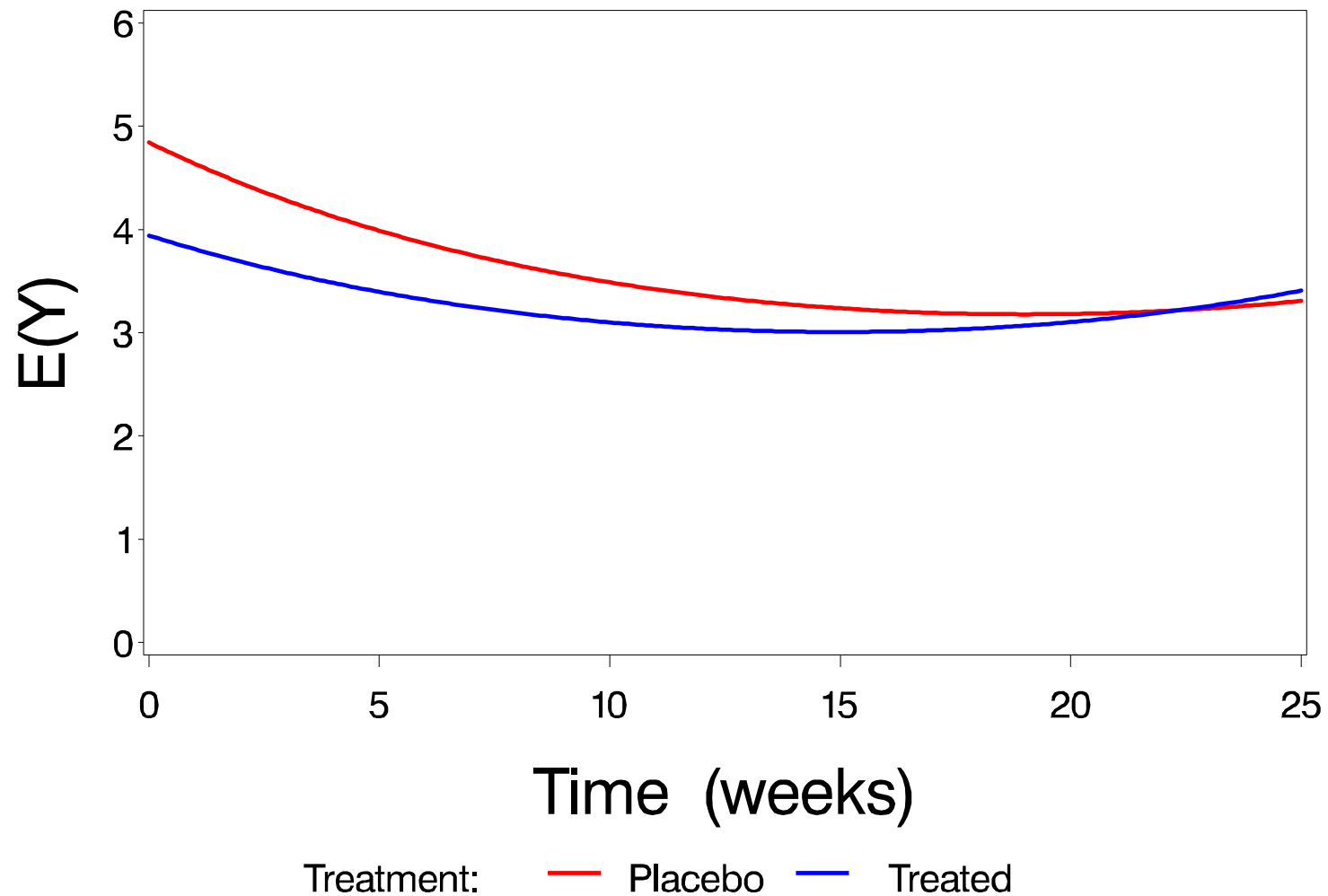
proc means data=h;
var y;
by t treat;
output out=out;
run;

dm "dlgprtsetup orient=L nodisplay";
filename fig 'c:/filename.eps';
goptions reset=all interpol=join ftext=swiss device=pslepsfc
          gsfname=fig gsfmode=replace targetdevice=winprtc ;
proc gplot data=out;
plot y*t=treat / haxis=axis1 vaxis=axis2 legend=legend1;
axis1 label=(h=2.5 'Time (weeks)') value=(h=1.5) order=(0 to 25 by 5) minor=none;
axis2 label=(h=2.5 A=90 'E(Y)') value=(h=1.5) order=(0 to 6 by 1) minor=none;
legend1 label=(h=1.5 'Treatment: ') value=(h=1.5 'Placebo' 'Treated');
title h=3 ' Marginal average evolutions (GLMM)';
symbol1 c=red i=join w=20 l=1 mode=include;
symbol2 c=blue i=join w=20 l=1 mode=include;
where _stat_='MEAN';
run;quit;run;

```

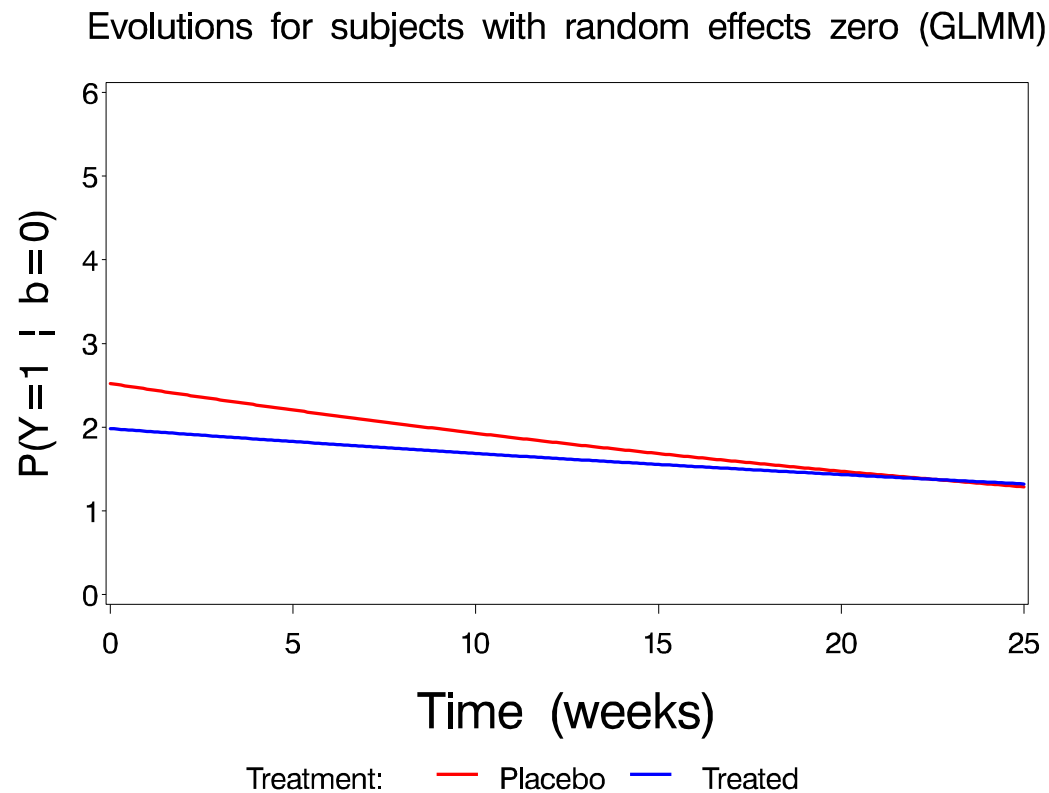
- Result:

Marginal average evolutions (GLMM)



- The evolution of 'average' subjects, i.e., subjects with $b_i = 0$ is given by:

$$\lambda_{ij} = \begin{cases} \exp[0.9251 + \mathbf{0} + (-0.0269 + \mathbf{0})t_{ij}] & \text{if placebo (group 0)} \\ \exp[0.6844 + \mathbf{0} + (-0.0162 + \mathbf{0})t_{ij}] & \text{if treated (group 1)}. \end{cases}$$



Part IV

Non-linear Mixed Models

Chapter 17

Introduction

- ▷ Linear and generalized linear mixed models revisited
- ▷ Non-linear mixed models

17.1 Linear and generalized linear mixed models revisited

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

$$E(Y_{ij}|b_{1i}, b_{2i}) = (\beta_1 + b_{1i}) + (\beta_2 + b_{2i})t_{ij} + \varepsilon_{ij}$$

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

$$E(Y_{ij}|b_{1i}, b_{2i}) = \frac{\exp[(\beta_1 + b_{1i}) + (\beta_2 + b_{2i})t_{ij}]}{1 + \exp[(\beta_1 + b_{1i}) + (\beta_2 + b_{2i})t_{ij}]}$$

- ▷ Count data, for example

$$E(Y_{ij}|b_{1i}, b_{2i}) = \exp[(\beta_1 + b_{1i}) + (\beta_2 + b_{2i})t_{ij}]$$

- In some applications, models are needed, in which the mean is no longer modeled as a function of a linear predictor.
- These are called non-linear mixed models.

17.2 Non-linear mixed models

- In non-linear mixed models, it is assumed that the conditional mean of Y_{ij} , given a vector \mathbf{b}_i of random effects is modeled as:

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

- The vector \mathbf{x}_{ij} contains known covariates
- The vectors $\boldsymbol{\beta}$ and \mathbf{b}_i contain fixed and random effects, respectively
- As before, the random effects are assumed to be normally distributed, with mean $\mathbf{0}$ and covariance D .
- Non-linear mixed models can be fitted within the SAS procedure NLMIXED.

Chapter 18

The orange trees

- ▷ Introduction
- ▷ Analysis in SAS

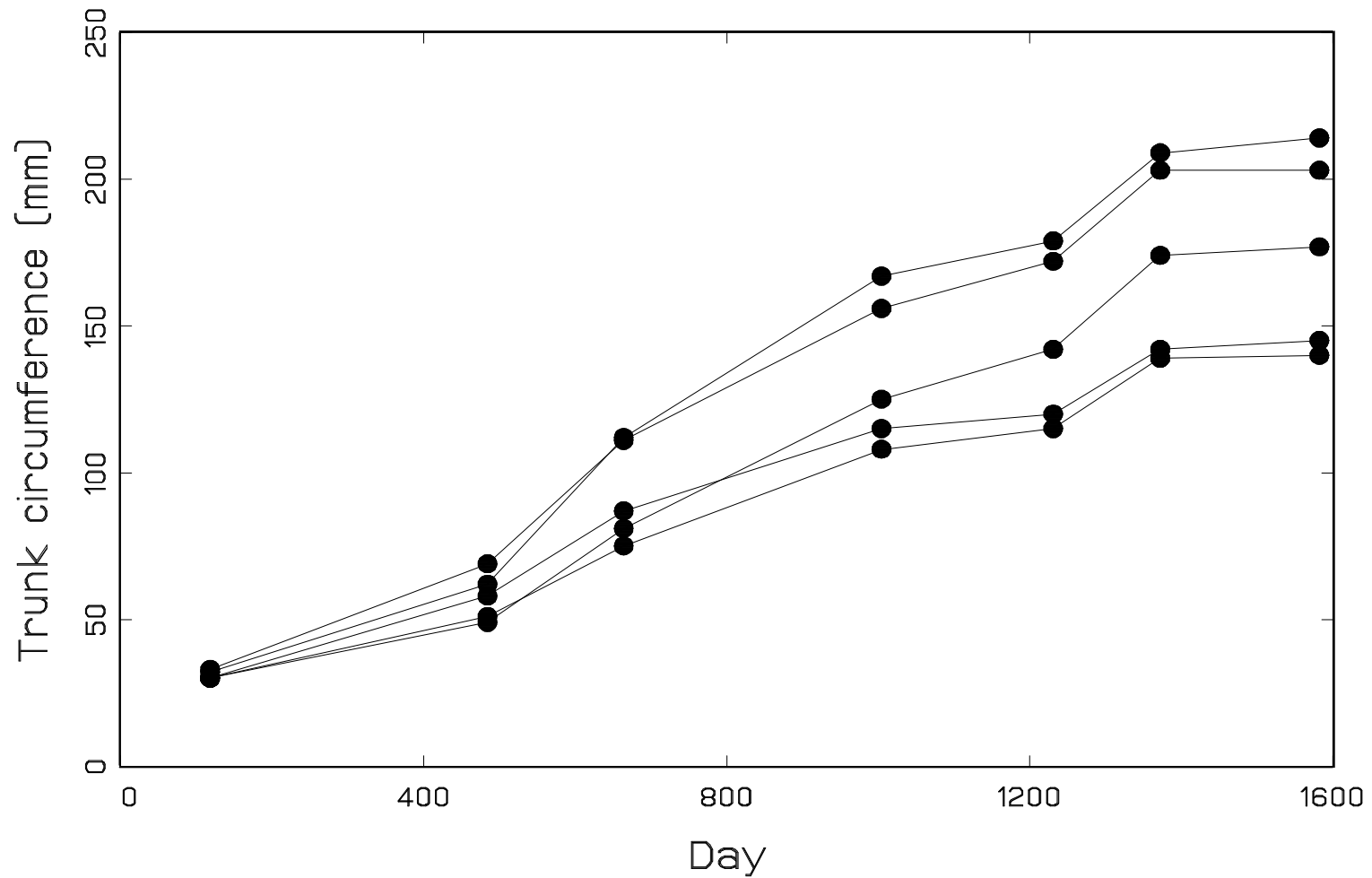
18.1 Introduction

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

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

- Individual profiles:

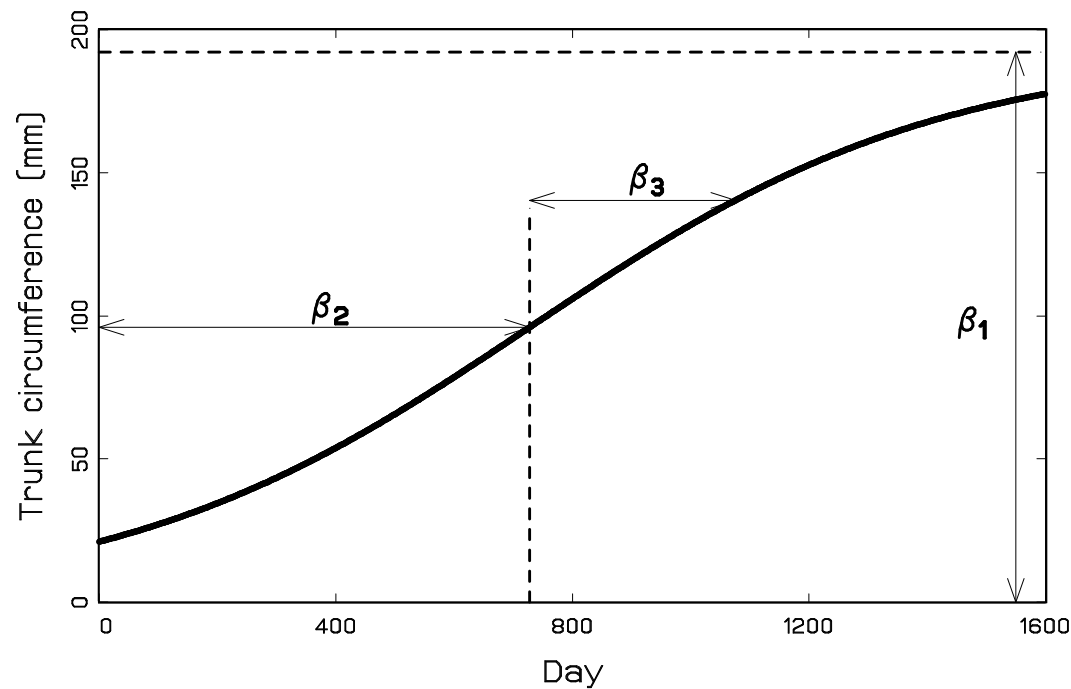
Orange trees



- The following non-linear mixed model has been proposed in the statistical literature:

$$Y_{ij} = \frac{\beta_1 + b_i}{1 + \exp[-(t_{ij} - \beta_2)/\beta_3]} + \varepsilon_{ij}, \quad b_i \sim N(0, \sigma_b^2), \varepsilon_{ij} \sim N(0, \sigma^2)$$

Orange trees: Meaning of parameters



18.2 Analysis in SAS

- In SAS PROC NLMIXED, the model can be fitted using either of the following equivalent programs:

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

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

- Selected output:

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper	Gradient
beta1	192.05	15.6577	4	12.27	0.0003	0.05	148.58	235.53	-6.8E-8
beta2	727.91	35.2487	4	20.65	<.0001	0.05	630.04	825.77	1.383E-8
beta3	348.07	27.0798	4	12.85	0.0002	0.05	272.89	423.26	3.375E-8
sigmab	31.6463	10.2614	4	3.08	0.0368	0.05	3.1560	60.1366	-3.4E-7
sigma	7.8430	1.0125	4	7.75	0.0015	0.05	5.0318	10.6542	-1.92E-6

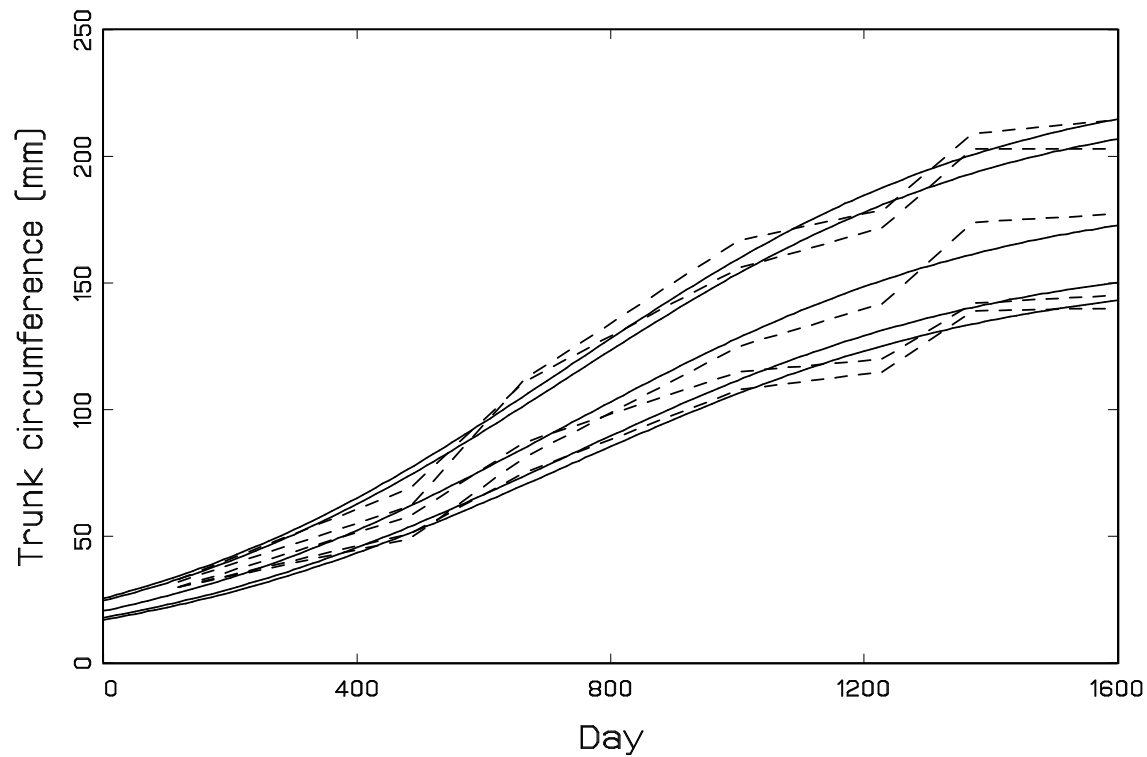
- Empirical Bayes estimates, and subject-specific predictions can be obtained as follows:

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

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

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

Orange trees: Observed and predicted profiles



Chapter 19

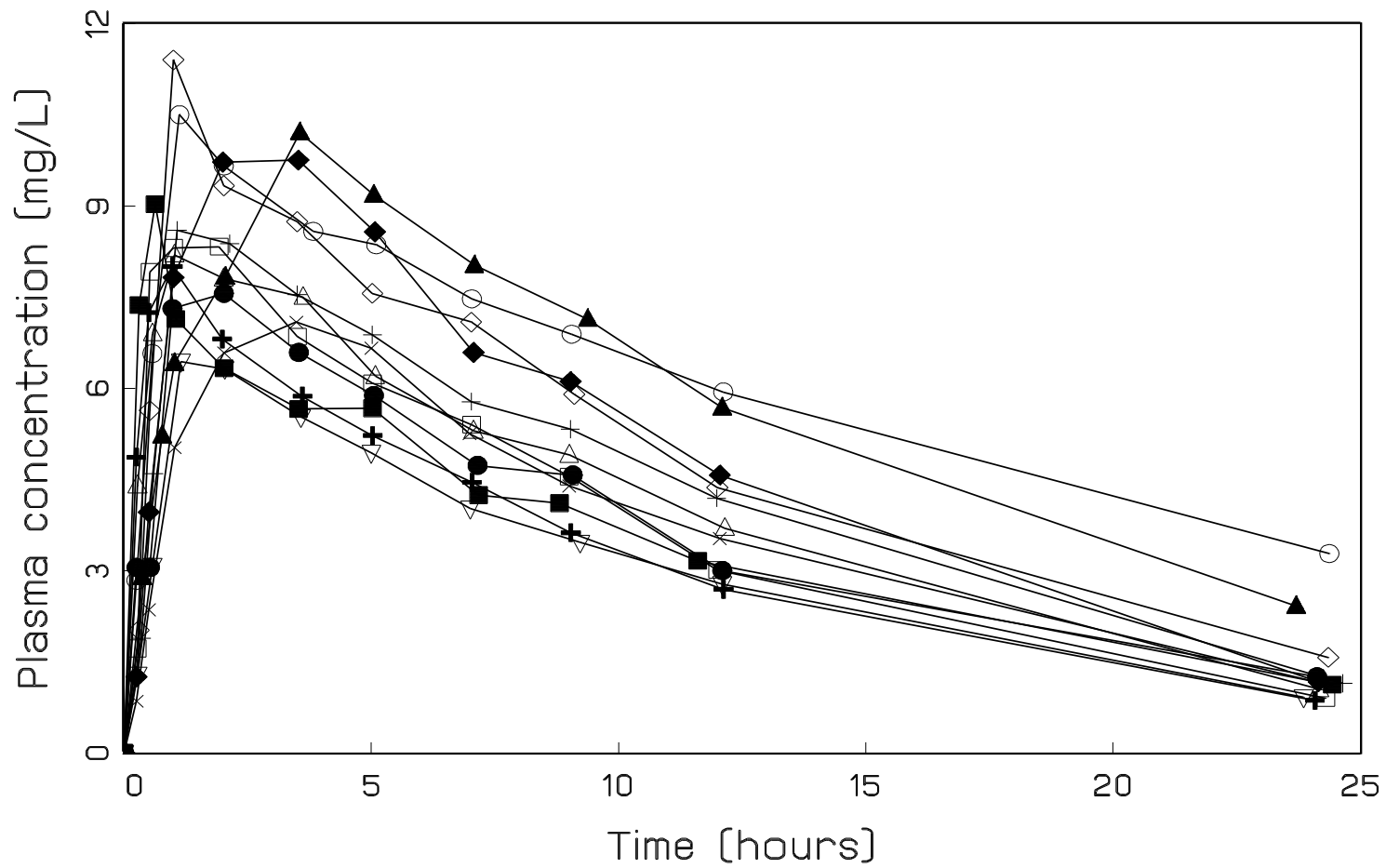
The Theophylline data

- ▷ Introduction
- ▷ Analysis in SAS

19.1 Introduction

- **Pharmacokinetics (PK)** is the study of the time course of a drug concentration in the body, i.e., “what the body does to the drug.”
- **Pharmacodynamics (PD)** is the study of the relationship of the drug concentration and pharmacologic effects, i.e., “what a drug does to the body.”
- We consider the PK study in which longitudinally measured blood concentrations of the anti-asthmatic, orally administered, agent Theophylline are studied
- 12 subjects, dose at $t = 0$
- Blood samples at 10 time points over the following 25 hours
- Outcome of interest: Theophylline concentration

Theophylline Data



- The blood concentration depends on:
 - ▷ **Absorption:** the process of a substance entering the body
 - ▷ **Elimination:** the process of a substance being removed from the body
 - ▷ **Clearance:** the volume of blood cleared of drug, in the kidneys, per unit time
- In the literature, a one-compartment open model with first-order absorption and elimination has been proposed ($t > 0$):

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

- Parameter interpretation:
 - ▷ k_{ai} : fractional absorption rate for subject i
 - ▷ k_{ei} : fractional elimination rate for subject i
 - ▷ Cl_i : clearance for subject i

- In order to restrict k_{ai} , k_{ei} , and Cl_i to be positive, the model is re-parameterized as:

$$Cl_i = \exp(\beta_1 + b_{1i}),$$

$$k_{a,i} = \exp(\beta_2 + b_{2i}),$$

$$k_{e,i} = \exp(\beta_3 + b_{3i}).$$

- b_{1i} , b_{2i} , and b_{3i} are assumed multivariate normal with mean $\mathbf{0}$ and covariance D

19.2 Analysis in SAS

- NLMIXED code:

```
proc nlmixed data=theoph;
parms beta1=-3.22 beta2=0.47 beta3=-2.45
      d11=0.03 d12=0 d22=0.4 d13=0 d23=0 d33=0.03 s2=0.5;
cl = exp(beta1 + b1);
ka = exp(beta2 + b2);
ke = exp(beta3 + b3);
pred = dose*ke*ka*(exp(-ke*time)-exp(-ka*time))/cl/(ka-ke);
model conc ~ normal(pred,s2);
random b1 b2 b3 ~ normal([0,0,0],[d11,d12,d22,d13,d23,d33]) subject=subject;
predict pred out=theopred; run;
```

- Note that very accurate starting values are needed for the various parameters. Otherwise the numerical optimization procedure does not reach convergence
- Starting values can be obtained from fitting non-linear regression models to all subjects separately, from fitting simplified mixed models, or from trying several optimization algorithms.

- Results:

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

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

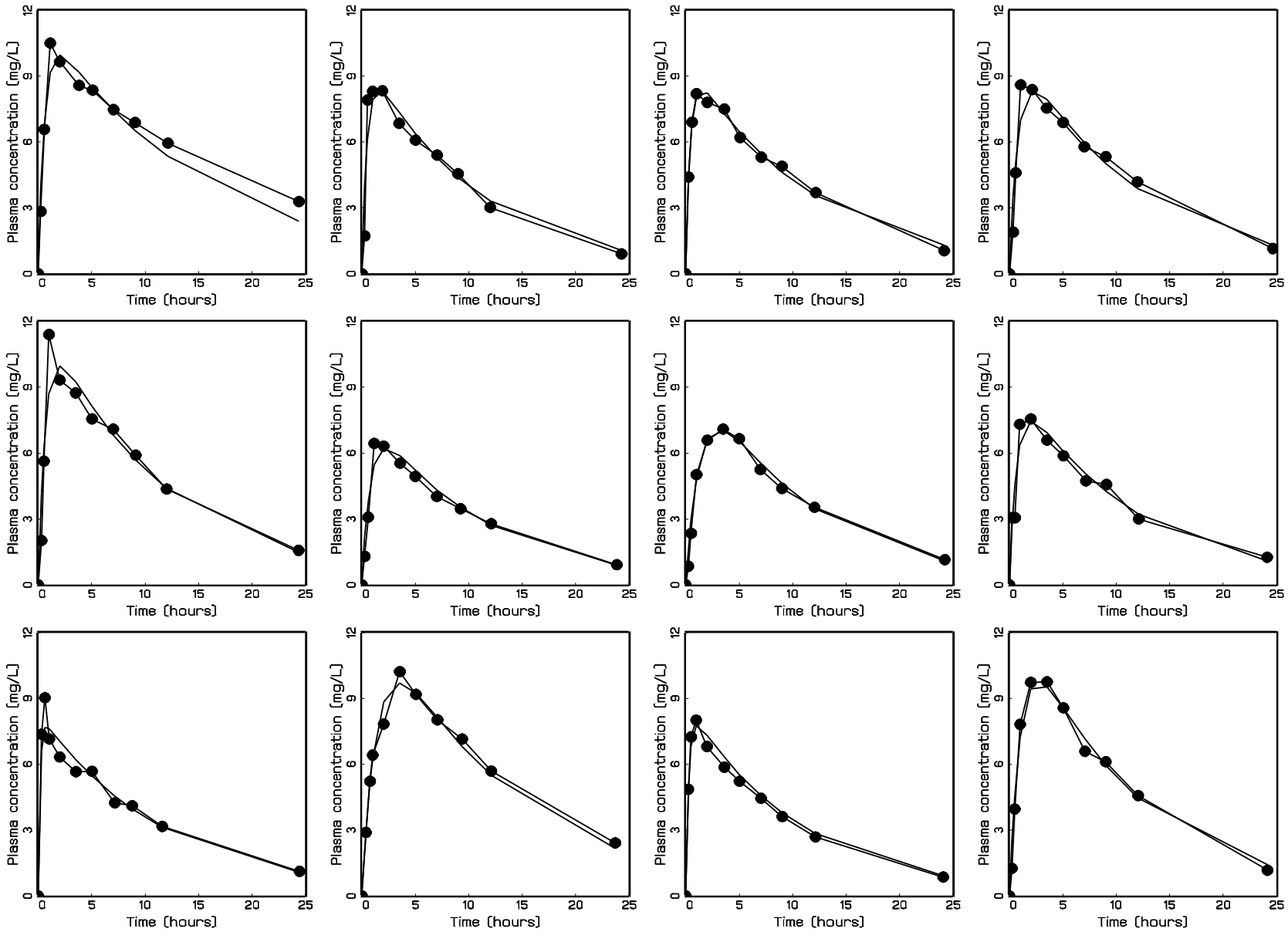
- There seems only weak evidence for correlation between the random effects b_{1i} , b_{2i} , b_{3i}

- A model with uncorrelated random effects can be fitted by replacing the previous PARMs and RANDOM statements by:

```
parms beta1=-3.22 beta2=0.47 beta3=-2.45 d11=0.03 d22=0.4 d33=0.03 s2=0.5;
```

```
random b1 b2 b3 ~ normal([0,0,0],[d11,0,d22,0,0,d33]) subject=subject;
```

- The increase in approximate log-likelihood is only 0.9.
- This supports our initial statement of weak correlation between the various random effects
- As an informal check of our model fit, we can compare the fitted profile for each subject with its observed data.



Chapter 20

Remarks

- ▷ Marginalizing non-linear mixed models
- ▷ Generalized linear mixed models in NLMIXED
- ▷ Generalized non-linear mixed models in NLMIXED

20.1 Marginalizing the mixed model

- It has been discussed that parameters in generalized linear mixed models have subject-specific rather than population-average interpretations.
- The same holds for non-linear mixed models.
- For example, in the Theophylline analysis, the marginal average evolution equals:

$$\begin{aligned} E(Y_{ij}) &= E\{E(Y_{ij}|\mathbf{b}_i)\} \\ &= E\left\{\frac{k_{ai}k_{ei}d_i}{Cl_i(k_{ai} - k_{ei})} \times [\exp(-k_{ei}t_{ij}) - \exp(-k_{ai}t_{ij})] \mid \mathbf{b}_i\right\} \\ &= E\left\{\frac{\exp(\beta_2 + b_{2i}) \exp(\beta_3 + b_{3i})d_i}{\exp(\beta_1 + b_{1i})[\exp(\beta_2 + b_{2i}) - \exp(\beta_3 + b_{3i})]} \right. \\ &\quad \left. \times [\exp(-\exp(\beta_3 + b_{3i})t_{ij}) - \exp(-\exp(\beta_2 + b_{2i})t_{ij})] \mid b_{1i}, b_{2i}, b_{3i}\right\} \end{aligned}$$

- This requires integration over the three-dimensional random-effects distribution of (b_{1i}, b_{2i}, b_{3i}) .
- As before, this can be done using numerical integration methods, or by sample averaging.
- In some special cases, the fixed effects do represent the average evolution.
- For example, for the model used to describe the orange tree data, we have

$$\begin{aligned}
 E(Y_{ij}) &= E\{E(Y_{ij}|b_i)\} = E\left\{\frac{\beta_1 + b_i}{1 + \exp[-(t_{ij} - \beta_2)/\beta_3]} \mid b_i\right\} \\
 &= \frac{\beta_1}{1 + \exp[-(t_{ij} - \beta_2)/\beta_3]}
 \end{aligned}$$

- Whenever the random effects appear in a linear way in the model, no numerical integration methods are needed for marginalizing the non-linear mixed model.

20.2 Generalized linear mixed models in NLMIXED

- Generalized linear mixed models can also be considered as non-linear mixed models: The mean is a non-linear function of the covariates:
 - ▷ Binary data, for example

$$E(Y_{ij}|b_{1i}, b_{2i}) = \frac{\exp[(\beta_1 + b_{1i}) + (\beta_2 + b_{2i})t_{ij}]}{1 + \exp[(\beta_1 + b_{1i}) + (\beta_2 + b_{2i})t_{ij}]}$$

- ▷ Count data, for example

$$E(Y_{ij}|b_{1i}, b_{2i}) = \exp[(\beta_1 + b_{1i}) + (\beta_2 + b_{2i})t_{ij}]$$

- These are special cases as, apart from the link function, there is still a linear predictor of the form $\mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\mathbf{b}_i$

- Generalized linear mixed models can also be fitted using the NLMIXED procedure.
- For example, the above logistic mixed can be fitted as:

```
proc nlmixed data=dataset;  
parms beta1=-3.22 beta2=0.47 d11=0.03 d12=0 d22=0.4;  
teta = beta1 + b1 + beta2*time + b2*time;  
expteta = exp(teta);  
p = expteta/(1+expteta);  
model y ~ binary(p);  
random b1 b2 ~ normal([0,0],[d11,d12,d22]) subject=subject;
```

- For example, the above Poisson mixed can be fitted as:

```
proc nlmixed data=dataset;  
parms beta1=-3.22 beta2=0.47 d11=0.03 d12=0 d22=0.4;  
teta = beta1 + b1 + beta2*time + b2*time;  
lambda=exp(teta);  
model y ~ poisson(lambda);  
random b1 b2 ~ normal([0,0],[d11,d12,d22]) subject=subject;
```

20.3 Generalized non-linear mixed models in NLMIXED

- The linear predictor in generalized linear mixed models can be replaced by any function of the covariates, fixed effects, and random effects.
- For example, in dose-response models, the following logistic mixed model with non-linear predictor is sometimes used:

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

$$\text{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_1 + b_i + \beta_2 \times \text{dose}_i^{\beta_3}$$

- Such models can equally well be fitted using the NLMIXED procedure