

# Improving parameter estimates in generalized linear mixed models

Citation for published version (APA):

Ouwens, J. N. M. (2002). *Improving parameter estimates in generalized linear mixed models*. [Doctoral Thesis, Maastricht University]. Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20020125jo>

## Document status and date:

Published: 01/01/2002

## DOI:

[10.26481/dis.20020125jo](https://doi.org/10.26481/dis.20020125jo)

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

mbdun

# Improving parameter estimates in generalized linear mixed models

by

J.N.M. Ouwens

J.N.M. Ouwens

© Ouwens, J.N.M. (2011)

The work presented in this thesis was performed at the University of Groningen and Wageningen University. The work was supervised by Prof. Dr. A.C. Nisbet and Prof. Dr. J.N.M. Ouwens. The work was supported by the Dutch Research Council (Vernieuwingsimpuls) and the Dutch Research Council (Vernieuwingsimpuls).

Printed on environmentally friendly paper

**nutrim**

The Graduate School



Improving parameter estimates in  
generalized linear mixed models

K.M. Ouwens

© Ouwens, Johannes Nicolaas Martinus, januari 2002

The study presented in this thesis was performed at the Nutrition and Toxicology Research Institute Maastricht (NUTRIM), which participates in the Graduate School VLAG-2 (Food Technology, Agrobiotechnology, Nutrition and Health Sciences) accredited by the Royal Netherlands Academy of Arts and Sciences.

Production: Universitaire Pers Maastricht

# Improving parameter estimates in Generalized Linear Mixed Models

## PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Maastricht,  
op gezag van de Rector Magnificus, Prof. dr. A.C. Nieuwenhuijzen Kruseman  
volgens het besluit van het College van Decanen, in het openbaar te verdedigen  
op vrijdag 25 januari 2002 om 12.00 uur

door

J.N.M. Ouwens



## Contents

<b>1. Methods to improve the estimation process in Generalized Linear Mixed Models.....</b>	<b>1</b>
1.1 Introduction.....	3
1.2 Mixed Models.....	5
1.3 Optimal designs.....	12
1.4 Influence Measures.....	16
1.5 Relationship between D-optimal design criterion and influence measures.....	18
1.6 Outline of dissertation.....	20
<b>2. Maximin D-optimal designs for Linear Mixed Models.....</b>	<b>25</b>
2.1 Introduction.....	27
2.2 Model specification and D-optimality.....	28
2.3 Relationship between the efficiency of designs for positively and negatively correlated random slopes and random intercepts.....	30
2.4 The time period and parameter values.....	32
2.5 Local D-optimal designs for the first and second degree polynomial model.....	34
2.6 Maximin designs for the first and second degree polynomial model.....	42
2.7 Summary and Discussion.....	46
Appendix.....	48
<b>3. Maximin designs in Generalized Linear Mixed Models with covariates.....</b>	<b>53</b>
3.1 Introduction.....	55
3.2 Model specification.....	56
3.3 Optimality criteria.....	60
3.4 Balancedness of maximin designs.....	61
3.5 Parameter space.....	62
3.6 Optimization algorithm.....	63
3.7 Results.....	64
3.8 Discussion.....	66
Appendix.....	68
<b>4. Detection of influential data in Linear Mixed Models.....</b>	<b>73</b>
4.1 Introduction.....	75
4.2 The London Growth Study: Data description and problem formulation.....	76
4.3 Model specification.....	78
4.4 Cook's Distance.....	80
4.5 Orthogonal decomposition of Cook's Distance and Conditional Cook's Distance.....	84
4.6 The London Growth Study: An analysis for influential observations.....	89
4.7 Discussion.....	91
Appendix.....	92

<b>5. Detection of influential data in Generalized Linear Mixed Models.....</b>	<b>97</b>
5.1 Introduction.....	99
5.2 Model specification.....	100
5.3 Local influence.....	101
5.4 An illustration.....	107
5.5 Discussion.....	114
Appendix.....	115
<b>6. Maximin designs and local influence in a health education intervention study.....</b>	<b>119</b>
6.1 Introduction.....	121
6.2 Teeth-brushing on the elementary school.....	122
6.3 Model specification.....	124
6.4 Optimal design.....	126
6.5 Influential schools and pupils.....	131
6.6 Discussion.....	135
<b>Epilogue.....</b>	<b>141</b>
<b>Summary.....</b>	<b>143</b>
<b>Samenvatting in het Nederlands (Dutch summary).....</b>	<b>147</b>
<b>Curriculum Vitae.....</b>	<b>157</b>

## Acknowledgements

This dissertation is the result of four years of work at the department of Methodology and Statistics of Maastricht University. First of all, I thank my supervisor Martijn Berger and co-supervisor Frans Tan. Martijn, your many comments on my English writing and our lively conversations during our meetings focussed me to improve my thesis. Frans, without you, I would never have finished my PhD. Your support was so brilliant that even I believed you! The days of brainstorming and your help with writing papers were very important.

I thank my roommate Marinus Spreen, who coached me at my first consultations. Marinus, thanks for your positive support, not only with the consultancies, but also at the difficult moments of my PhD period. I thank Hubert Schouten and Math Candel, not only for their willingness to correct my thesis.

Mirjam Moerbeek and Valeria Lima Passos worked both on PhD projects closely related to mine. Mirjam, your advices about programming and freely available software accelerated the computation process and our conversations were very inspiring. Valeria, your many comments on my English writing were very welcome. I thank Carel Thijs, who gave me the data for Chapter 6 and gave helpful comments on that chapter. Carel, your enthusiasm about the applicability of my work in practice was very inspiring. I thank all my colleagues for their creation of a stimulating environment.

I thank my brother Henk, for the opportunity to learn Factor analysis, LISREL and many other packages within a practical framework. Henk, my work for your company Consultance Groep Nederland was a very nice pastime during the evenings. It was nice to do research with you. Not only, because you asked my advice, even before the data were sampled. This created a welcome learning environment.

I am very grateful to my parents, without whom I would even not have finished my pre-university education. I want to thank my friends and family especially for the great weekends in Vaassen and I hope to be able to enjoy these weekends for many years in the future.





## **Abstract**

This chapter discusses methods to improve the estimation process in generalized linear mixed models and serves as an introduction for this thesis. Two ways to improve the estimation process are considered, namely the optimization of the choice of the design and the detection of influential data.

The most frequently used optimality criteria are introduced and the preference for the D-optimality criterion is explained. The solutions of the optimality criteria are only local optimal. The most commonly used approaches to deal with local optimality are discussed. It is explained why the maximin approach based on the relative efficiency is most appealing.

As methods for influence assessment, Cook's Distance and local influence measures are discussed. Cook's Distance is a member of the class of case deletion methods, while the local influence measures are members of the class of likelihood based influence measures. It is shown that the choice of the design and the influence of data are closely related.

*Key words:* LMM, GLMM, estimation process, optimal design, D-optimality criterion, maximin approach, Cook's Distance, local influence, potential influence

## 1.1 Introduction

In medical and health sciences, many studies discuss processes of change within subjects. Snijders and Bosker (1999, p.182), for example, studied the growth of children who visited a paediatrician-endocrinologist because of growth retardation. For this study, the average growth trend of the children was of interest.

Figure 1.1 shows the growth trends of a random sample of five children, which are sampled from the estimated model given by Snijders and Bosker (1999, p. 183). The bold line in Figure 1.1 is the expected growth trend over all children, the thin lines give the growth trends for each particular child and the dotted lines connect the observations from each child. Note that the children grow almost linearly in time between 5 and 10 years of age (Goldstein,1979).

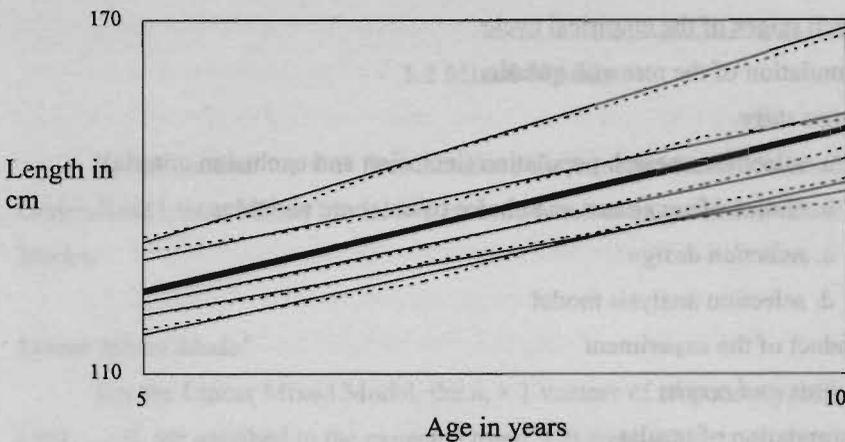


Figure 1.1. Growth of children between 5 and 10 years of age.

The differences among the growth trends of the children can be taken into account by the use of Generalized Linear Mixed Models (Diggle, Liang and Zeger, 1994, ch. 9). For these models, the expected growth trend over all children is modelled by the fixed effects (cf. bold line), while the deviations of the growth trends of the children from the

expected growth trend are modelled by the use of additional child-specific parameters (cf. deviations of thin lines from bold line), which are often called random effects. Finally, for each child, the deviations of the observations from the growth trend of the child are modelled by the addition of a measurement error term (cf. deviations of dotted lines from corresponding thin lines).

In many studies, the effect of a treatment is of interest. The significance of the treatment effect is tested by relating the estimated parameter to its dispersion or variance. Thus, a higher quality of the tests of the treatment effect can be obtained by the improvement of the quality of the estimated parameters and their variances. In this dissertation we will consider methods for the improvement of the quality of the estimated parameters and their variances for stage 2 and 4 of the research stages of the empirical cycle (Bouter and van Dongen, 1988, p.91):

#### Research stages of the empirical cycle

##### 1. Formulation of the research question

##### 2. *Design stage*

a. selection research population (inclusion and exclusion criteria)

b. choice of covariates and choice of outcome variables

c. *selection design*

d. selection analysis model

##### 3. Conduct of the experiment

##### 4. *Analysis and results*

##### 5. Interpretation of results

Large variances of the estimated parameters may be caused by the choice of an inefficient design in stage 2 of the empirical cycle (Atkinson and Donev, 1996, Bunke and Bunke, 1986, and Pukelsheim, 1993). If, for example, the measurements are taken at time points, which are too close to each other, then the estimated time effect will have a relatively large variance. One way to reduce the variance of the estimates is therefore

to use a more efficient design.

The estimates of the model parameters and their variances based on all data may be very different from the estimates of the model parameters based on the data without so called influential subjects and observations. The detection of influential subjects and observations in stage 4 of the empirical cycle can therefore improve the estimation process.

The optimal design process and the detection of influential subjects and observations will be discussed in this thesis for both the Linear Mixed Model and the Generalized Linear Mixed Model. These models are introduced in section 1.2. In section 1.3, the optimal design concept will be explained. The assessment of influence is discussed in more detail in section 1.4. A relation between the  $D$ -optimal design criterion and the influence measures is given in section 1.5. Finally, the topics of the chapters will be briefly introduced in section 1.6.

## 1.2 Mixed Models

In this section we will first introduce the Linear Mixed Model. Thereafter, the Generalized Linear Mixed Model will be introduced as an extension of the Linear Mixed Model.

### *Linear Mixed Model*

For the Linear Mixed Model, the  $n_i \times 1$  vectors of responses  $y_i$  of subject  $i$ ,  $i = 1, \dots, N$ , are ascribed to the expected trend  $X_i\beta$  over subjects, to the subject-specific deviations  $Z_i b_i$  from this expected trend  $X_i\beta$  and to the deviations  $\epsilon_i$  of the observations from the subject-specific profile ( $X_i\beta + Z_i b_i$ ) as follows:

$$y_i = X_i\beta + Z_i b_i + \epsilon_i. \quad (1.1)$$

In Figure 1.1, the expected trend  $X_i\beta$  is presented by the bold line, the subject-

specific profiles ( $X_i\beta + Z_i b_i$ ) are presented by the thin lines and the observations  $y_{ij}$  of subject  $i$ ,  $i = 1, \dots, N$ , at measurement  $j$ ,  $j = 1, \dots, n_i$ , are connected by the dotted lines. In general, the  $n_i \times p$  design matrices  $X_i$  consist of functions of the time points and time varying and time independent covariates. For longitudinal data the columns of the  $n_i \times q$  design matrices  $Z_i$  are often functions of the time points. In this thesis the design matrices  $Z_i$  are considered to be submatrices of the design matrices  $X_i$ .

No subscript is used for the fixed  $p \times 1$  regression parameter vector  $\beta$ , because the expected trend  $X_i\beta$  does not depend on the subjects in the sample. The use of the subscript  $i$  for the  $q \times 1$  random effects regression parameter vector  $b_i$  indicates that the subject-specific deviations  $Z_i b_i$  may be different for each subject  $i$ . The vectors  $b_i$  are supposed to be independently and normally distributed with mean zero and variance-covariance matrix  $D$ .

The  $n_i \times 1$  vectors  $\epsilon_i$  are supposed to be measurement errors (or disturbances), which are normally distributed with mean zero and variance-covariance matrices  $\Psi_i$ . Two structures of the variance-covariance matrices  $\Psi_i$  will be considered. The first structure corresponds to the situation that the measurement errors are independent of each other. The matrix  $\Psi_i$  is then equal to  $\sigma^2 I_i$ ,  $i = 1, \dots, N$ , with  $I_i$  the identity matrix of rank  $n_i$  and  $\sigma^2$  the common variance of the measurement errors. The second structure is the AR(1) structure, for which the correlation between two measurement errors  $j$  and  $j'$  of the same subject is equal to  $\rho^{|t_{ij} - t_{ij'}|}$ , where  $\rho \in [0, 1]$  is the serial correlation coefficient and  $t_{ij}$  and  $t_{ij'}$  are the time points at which the measures  $j$  and  $j'$  of subject  $i$  are taken. In other words, the correlation is a decreasing function of the time between the two measures.

If we interpret the subject-specific deviations  $Z_i b_i$  from the expected profile  $X_i\beta$  as nuisance, then the variance-covariance matrix of the responses of subject  $i$  is ascribed to the random effects and to the measurement errors, i.e. the variance-covariance matrix of the responses is equal to  $V_i = \text{Var}(y_i) = \{ [Z_i \text{Var}(b_i) Z_i'] + \text{Var}(\epsilon_i) \} = [Z_i D Z_i'] + \Psi_i$  (Diggle, Liang and Zeger, 1994, p. 81, Verbeke and Molenberghs, 2000, p. 41).

### Generalized Linear Mixed Model

In model (1.1) the response is a linear function of the design matrices  $X_i$  and  $Z_i$  and the vectors of measurement errors  $\epsilon_i$ . Other functions can also be considered. For the logistic random effects model, for example, the logit of the conditional probability  $\pi_{ij}$  of the response  $y_{ij}$  is modelled, where the probability is conditioned on the random effects  $b_i$ . The logit is ascribed to the expected trend  $X_i\beta$  and the subject-specific deviations  $Z_i b_i$  from this expected trend  $X_i\beta$  as follows:

$$\text{logit}(\pi_{ij}) = X_{ij}\beta + Z_{ij}b_i, \quad (1.2)$$

The Generalized Linear Mixed Model can analogously be defined. Conditional on the random effects  $b_i$ , the responses for the Generalized Linear Mixed Model are obtained from an exponential density, such that

$$h(\mu_{ij}) = X_{ij}\beta + Z_{ij}b_i, \quad (1.3)$$

where  $h$  is the link function of the conditional expectation  $\mu_{ij}$  of the response  $y_{ij}$  of subject  $i$  at measurement  $j$ .

Three link functions  $h$  are used in this thesis and are given in Table 1.1.

Table 1.1 Three models with their link functions and conditional variances

Model	Type of response	$h(\mu_{ij})$	Conditional variance
Linear Mixed Model	Continuous	$\mu_{ij}$	$\sigma^2$
Logistic Mixed Model	Dichotomous	$\text{logit}(\mu_{ij}) = \text{logit}(\pi_{ij})$	$\mu_{ij}(1 - \mu_{ij}) = \pi_{ij}(1 - \pi_{ij})$
Poisson Mixed Model	Count	$\log(\mu_{ij})$	$\mu_{ij}$

The link functions correspond to the Linear Mixed Model, the Logistic Mixed Model and the Poisson Mixed Model. The models differ in the type of the response. The response is continuous for the Linear Mixed Model, is dichotomous for the Logistic Mixed Model and is a count for the Poisson Mixed Model. For the Logistic Mixed Model, it is assumed that the response is a 0/1 dichotomous variable, so that the conditional expectation  $\mu_{ij}$  is equal to the conditional probability  $\pi_{ij}$ .

In chapter 3 of this thesis, we will also consider an Overdispersed Poisson Mixed Model, which is an extension of the Poisson Mixed Model and can be used if the conditional variance is unequal to the conditional expectation.

### *Estimation method*

For ease of presentation, the stacked vector of all regression parameters, all random effect variance and covariance parameters and all other variance and covariance parameters is denoted by  $\theta$ . The space of all possible model parameter vectors  $\theta$  is denoted by  $\Theta$ .

The model parameter vector  $\theta$  is estimated by means of the maximum likelihood method, so that the estimate  $\hat{\theta}$  of  $\theta$  maximizes the likelihood, i.e. maximizes the joint density of the observed data set as a function of the model parameters. Note that the likelihood is a measure of how "likely" or "probable" the data set is with respect to the evaluated distribution.

For the considered Linear Mixed Models and Generalized Linear Mixed Models, the subjects are assumed to be independently distributed. The log-likelihood function is therefore equal to  $\log L(\theta) = \sum_{i=1}^N l_i(\theta)$ , where  $l_i(\theta)$  is the logarithm of the likelihood of the set of observations of subject  $i$  as function of  $\theta$ .

*Linear Mixed Model.*

The log likelihood function  $\log L(\theta)$  for the Linear Mixed Model is equal to

$$\begin{aligned} \log L(\theta) &= \sum_{i=1}^N l_i(\theta) \\ &= \sum_{i=1}^N \log \left\{ (2\pi)^{-\frac{n_i}{2}} \text{Det}(V_i)^{-\frac{1}{2}} \times \exp \left[ -\frac{1}{2} (y_i - X_i \beta)' V_i^{-1} (y_i - X_i \beta) \right] \right\}, \end{aligned} \quad (1.4)$$

where  $\text{Det}(V_i)$  is the determinant of the variance-covariance matrix  $V_i$ . Conditional on  $V_i$ , an estimate of  $\beta$  can be obtained by maximizing expression (1.4). The maximum likelihood estimate and its variance are given by

$$\hat{\beta} = \left( \sum_{i=1}^N X_i' V_i^{-1} X_i \right)^{-1} \sum_{i=1}^N X_i' V_i^{-1} y_i \quad (1.5)$$

and

$$\text{Var}(\hat{\beta}) = \left( \sum_{i=1}^N X_i' V_i^{-1} X_i \right)^{-1}. \quad (1.6)$$

When the variance and covariance parameters are unknown, and when an estimate of  $V_i$  is available, then we can estimate  $\beta$  by replacing  $V_i$  by its estimate.

The estimate of the vector of subject-specific parameters  $\hat{b}_i$  in model (1.1) cannot be obtained by maximizing expression (1.4). Instead, a Bayesian approach is used, to estimate the subject-specific parameters based on both the prior information about the common normal distribution of the  $\hat{b}_i$ 's as well as the deviation of the observations of the subject from the estimated expected profile  $(y_i - X_i \beta)$ :

$$\hat{b}_i = D Z_i' V_i^{-1} (y_i - X_i \beta). \quad (1.7)$$

The estimates of  $b_i$ , for which  $D$  and  $V_i$  are replaced by their estimates, are called empirical Bayes estimates. The Maximum Likelihood estimate for  $V_i$  is obtained by

maximizing expression (1.4), in which  $\beta$  is replaced by  $\hat{\beta}$ , i.e.  $V_i$  and  $\beta$  are estimated by an iterative procedure.

### Generalized Linear Mixed Model.

The log-likelihood function  $\log L(\theta)$  for the Generalized Linear Mixed Model is given by

$$\log L(\theta) = \sum_{i=1}^N \log \int_{-\infty}^{\infty} \left( \prod_{j=1}^{n_i} f(y_{ij} | \mathbf{b}_i, \beta, \phi) \right) g(\mathbf{b}_i | D) d\mathbf{b}_i \quad (1.8)$$

where the density  $f$  is the conditional density of the response  $y_{ij}$ , given the vector of random effects  $\mathbf{b}_i$ , the vector of fixed regression parameters  $\beta$ , and the dispersion parameter  $\phi$ , and the density  $g$  is the density of the vector of random effects  $\mathbf{b}_i$ , given the variance-covariance matrix of the random effects  $D$ .

Expression (1.8) shows that the log-likelihood function  $\log L(\theta)$  is the logarithm of the marginal density of the response obtained by integrating the joint density of the responses and random effects with respect to the random effects. For the Linear Mixed Model with conditionally independent responses, the log-likelihood function (1.8) reduces to expression (1.4). For most models, however, numerical integration methods are required for its evaluation.

The variance of the estimated model parameters is asymptotically equal to the inverse of the Fisher information matrix. The log-likelihood function  $\log L(\theta)$  will be used within the optimal design context and will also be used for the development of influence measures.

### Ellipsoids

In this thesis, the optimal design process is studied and influence measures are developed. Both the considered optimal design criteria as well as the proposed influence measures use characteristics of ellipsoids. We therefore briefly introduce the concept of an ellipsoid in this section. An example of an ellipsoid is plotted in Figure 1.2 (Johnson

and Wichern, 1988, p. 180). The ellipse in Figure 1.2 is given for a set of regression parameter vectors  $\beta = (\beta_1, \beta_2)'$ , such that  $(\beta - \hat{\beta})' \text{Var}(\hat{\beta})^{-1} (\beta - \hat{\beta}) \leq \gamma$  for a specific  $\gamma$ ,  $\hat{\beta}$  and  $\text{Var}(\hat{\beta})$ .

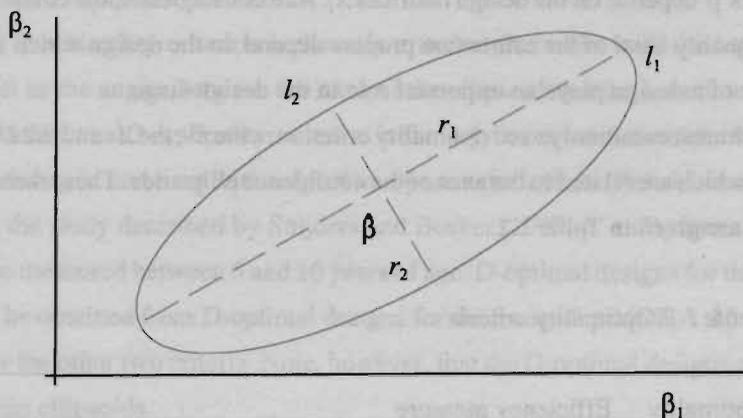


Figure 1.2. Example of an ellipsoid.

The ellipse of Figure 1.2 is plotted in a two-dimensional space with axes represented by the parameter  $\beta_1$  and the parameter  $\beta_2$ . The dashed lines  $l_1$  and  $l_2$  are the longest and smallest axis of the ellipse, respectively. The intersection of the dashed lines  $l_1$  and  $l_2$  is equal to the estimate  $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)'$ . The difference in length of  $l_1$  and  $l_2$  is caused by the variance of  $\hat{\beta}$ , which is larger in the direction of  $l_1$  than in the direction of  $l_2$ . The directions of the axes  $l_1$  and  $l_2$  are related to the covariance between  $\hat{\beta}_1$  and  $\hat{\beta}_2$ . If  $\hat{\beta}_1$  and  $\hat{\beta}_2$  are uncorrelated, then  $l_1$  and  $l_2$  lie along the co-ordinate axes.

To assess whether a  $p$ -dimensional ellipsoid is large, we can use its volume  $V \propto \pi \prod_{i=1}^p r_i$ , its periphery  $P \propto \pi \sum_{i=1}^p r_i$  or the size of its largest axis  $M = \max_{i=1, \dots, p} r_i$ . All three measures depend on the lengths  $r_i$  of the axes of the ellipsoid.

The most commonly used optimal design criteria are also related to the lengths  $r_i$  of the axes of the ellipsoid.

### 1.3 Optimal designs

For the Linear Mixed Model, the variance of the vector of estimated regression parameters  $\hat{\beta}$  depends on the design matrices  $X_i$ . As a consequence, the costs of obtaining a certain quality level of the estimation process depend on the design which is used. The efficiency of a design plays an important role in the design stage.

The most commonly used optimality criteria are the  $E$ -, the  $A$ - and the  $D$ -optimality criterion, which are related to the axes of the confidence ellipsoids. These three efficiency measures are given in Table 1.2.

Table 1.2 Optimality criteria

Optimality	Efficiency measure
$E$ -optimality	Proportional to the inverse of the square of the length of the largest axis of the ellipsoid $1/\max_i r_i^2$ .
$A$ -optimality	Proportional to the inverse of the sum of the squares of the lengths of all axes of the ellipsoid $1/\sum_{i=1}^p r_i^2$ .
$D$ -optimality	Proportional to the inverse of the product of the squares of the lengths of all axes of the ellipsoid $1/\prod_{i=1}^p r_i^2$ .

The  $E$ -optimality criterion minimizes the square of the length of the largest axis  $M^2 = \max_i r_i^2$ . The  $E$ -optimality criterion does not take into account the other axes of the ellipsoid. To take all axes into account, the  $A$ -optimality criterion or the  $D$ -optimality criterion can be used. The  $A$ -optimality criterion minimizes the sum of the squares of the lengths of all axes  $\sum_{i=1}^p r_i^2$ , which is equivalent to minimizing the sum of the variances  $\sum_{i=1}^p \text{Var}(\hat{\beta}_i)$ . The  $D$ -optimality criterion minimizes the product of the squares of the lengths of all axes  $\left(\frac{V}{\pi}\right)^2 = \prod_{i=1}^p r_i^2$  (Johnson and Wichern, 1988, p. 103). Note that the product  $\prod_{i=1}^p r_i^2$  is proportional to the determinant  $\text{Det}[\text{Var}(\hat{\beta})]$  of the variance-

covariance matrix  $\text{Var}(\hat{\beta})$ , which is called the generalized variance (Anderson, 1958, p. 166).

#### *Choice for the D-optimality criterion.*

The D-optimality criterion is the only criterion that takes both the lengths of all axes as well as the angles between the axes of the ellipsoid and the parameter axes into account. Further, the D-optimality criterion is the only criterion for which 'the function induced ordering is invariant with respect to reparameterization' (Pukelsheim, 1993, p. 137). In the study described by Snijders and Bosker (1999, p. 182), for example, the children are measured between 5 and 10 years of age. D-optimal designs for this research period can be obtained from D-optimal designs for the research period  $[-1, 1]$ . This is not possible for the other two criteria. Note, however, that the D-optimal designs may lead to large but thin ellipsoids.

#### *Local Optimality.*

For the Linear Mixed Model, the variance-covariance matrix  $\text{Var}(\hat{\beta})$  of the estimated regression parameters depends on the unknown variance-covariance matrices  $D$  and  $\Psi_i$ . For the Logistic Mixed Model, the variance-covariance matrix  $\text{Var}(\hat{\beta})$  also depends on the unknown vector of regression parameters  $\beta$ . Unfortunately, the D-optimal design for one vector of parameter values can be inefficient for another vector of parameter values. The D-optimal designs are therefore said to be locally optimal. Three approaches to deal with local optimality have been proposed in literature, namely the sequential design approach, the Bayesian design approach and the maximin approach.

*Sequential design approach.* One way to deal with the local optimality of the D-optimal design is to take a small set of observations, estimate the parameters based on this set of observations and iterate the following two steps as long as necessary:

1. optimize the choice of the next observation and select the next observation accordingly.
2. estimate the parameters based on the set with the new observation.

This approach is called the sequential design approach (Wu, 1985 and Wynn, 1970). The

aim of the sequential design approach is to approximate the  $D$ -optimal design on the long run.

*Bayesian design approach.* The second approach is to use a design which is efficient on a weighted average. This is formulated into the Bayesian design criteria. The most frequently used Bayesian design criterion evaluates the expected logarithm of the efficiency  $E_{\theta} \left\{ \log \text{Det}[\text{Var}(\hat{\beta}_{\tau} | \theta)] \right\}$  of the design  $\tau$  over a prior distribution of the model parameter vectors  $\theta \in \Theta$ . The design with the smallest expectation is said to be optimal with respect to the Bayesian design criterion (Chaloner and Larntz, 1989; Chaloner and Verdinelli, 1995).

*Maximin or minimax approach.* There exist two maximin criteria, which are based on the maximin or minimax principle. The first maximin criterion is based on the efficiency of the designs (Berger, King and Wong, 2000, King and Wong, 2000, and Wong, 1992). The efficiency  $\left\{ \text{Det}[\text{Var}(\hat{\beta}_{\tau})] \right\}^{-1}$  of a design  $\tau$  depends on the vector of model parameters  $\theta$ . This implies that for each design  $\tau$ , the efficiency can be minimized over the model parameter space  $\Theta$ . A design for which the *smallest efficiency* is *maximized* is said to be a maximin design. This explains the use of the term 'maximin principle', while the term 'minimax principle' is related to the generalized variance  $\text{Det}[\text{Var}(\hat{\beta}_{\tau})]$ . With respect to the generalized variance  $\text{Det}[\text{Var}(\hat{\beta}_{\tau})]$ , the maximin criterion leads to a design which *minimizes*  $\max_{\theta} \text{Det}[\text{Var}(\hat{\beta})]$ , i.e. minimizes the largest possible generalized variance of the estimates  $\hat{\beta}$ .

The second maximin criterion which is proposed in this dissertation leads to a design for which the *smallest relative efficiency*  $\min_{\theta} \text{eff}(\tau | \tau^*)$  over the model parameter space  $\Theta$  is *maximized*, where the relative efficiency  $\text{eff}(\tau | \tau^*)$  is defined as follows: Given a model parameter vector  $\theta \in \Theta$ , the relative efficiency  $\text{eff}(\tau | \tau^*)$  of a design  $\tau$  is equal to

$$\text{eff}(\tau | \tau^*) = \left\{ \frac{\text{Det}[\text{Var}(\hat{\beta}_{\tau^*})]}{\text{Det}[\text{Var}(\hat{\beta}_{\tau})]} \right\}^{1/p}, \quad (1.9)$$

where  $\hat{\beta}_{\tau}$  and  $\hat{\beta}_{\tau^*}$  are the estimated vector of regression parameters corresponding to the

design  $\tau$  and the  $D$ -optimal design  $\tau^*$ , respectively. The relative efficiency is defined in such a way that if, for example,  $\text{eff}(\tau|\tau^*) = 0.90$  for a particular design  $\tau$ , then about 10% more observations will be needed to obtain the same efficiency as that of the  $D$ -optimal design  $\tau^*$  for the same model.

#### *Choice for the second maximin criterion.*

The sequential approximation of the  $D$ -optimal design can often not be used, due to experimental restrictions like the length of the research period. The Bayesian design criterion and the first maximin criterion share the same disadvantage. For both criteria, the corresponding optimal design may be inefficient for the particular model parameter vector  $\theta$  at hand.

The second maximin criterion is related to the smallest relative efficiency of designs. If the smallest relative efficiency of a design  $\tau$  is large, then the design  $\tau$  is in particular highly efficient for the model parameter vector  $\theta$  at hand. The second maximin criterion, which is related to the relative efficiency of designs, is therefore selected for this thesis.

#### *Coding of the variables.*

An advantage of the  $D$ -optimality criterion is that it is invariant with respect to reparameterization of the variables (Pukelsheim, 1993, p.137). This means that the research period can be assumed, without loss of generality, to start at time point -1 and to end at time point 1. To obtain the  $D$ -optimal designs for other research periods, the time points -1 and 1 must be interpreted as start and end point of the research period, respectively. For the study described by Snijders and Bosker (1999, p.182), for example, the children are measured between 5 and 10 years of age. As a consequence, 5 is mapped onto -1 and 10 is mapped onto 1. The range of the model parameter values is changed, due to the change in the coding of the time variable. If, for example, the linear time effect is equal to 2, then the new value of the linear time effect will become equal to  $2 \cdot (10-5)/(1 - -1) = 2 \cdot 5/2 = 5$ , i.e. it will become equal to the linear time effect times the fraction of the length of the original research period (10-5) and the length of the artificial

time period (1 - -1).

#### 1.4 Influence Measures

The estimation of the parameters and their variances may be improved by the detection of influential subjects and/or influential observations. Two influence measures to detect influential subjects and observations are considered in this thesis, namely Cook's Distance and the local influence measure. Both influence measures are strongly related to ellipsoids. We will discuss this relation for Cook's Distance in the next paragraphs. Then the relation between the ellipsoids and the local influence measure will be given.

##### *Cook's Distance.*

Cook's Distance assesses the difference between the estimates  $\hat{\beta}$  and  $\hat{\beta}_{(d)}$  of  $\beta$ , which are based on the sample with and without subject or observation  $d$ , respectively. Cook's Distance for the  $d$ th subject or observation is defined as

$$C_d = (\hat{\beta}_{(d)} - \hat{\beta})' \text{Var}(\hat{\beta})^{-1} (\hat{\beta}_{(d)} - \hat{\beta}). \quad (1.10)$$

The Cook's Distances can be compared with each other. Subjects and observations with relative large Cook's Distances are said to be influential. The contours of constant Cook's Distance are ellipsoids (Cook and Weisberg, 1982, p. 114). Expression (1.10) looks very similar to the expression  $(\beta - \hat{\beta})' \text{Var}(\hat{\beta})^{-1} (\beta - \hat{\beta}) \leq \gamma$ , which was used to construct the ellipse in Figure 2. Note, however, that in expression (1.10), both  $\hat{\beta}$  and  $\hat{\beta}_{(d)}$  are random.

##### *Influence of case perturbations.*

Cook's Distance measures the amount of change of the estimated parameters, due to the deletion of a subject or observation from the data set. It follows from expression (1.4) and (1.8) that the log-likelihood functions for the data set with and without the  $d$ th subject can both be written as

$$\log L(\theta) = \sum_{i=1, i \neq d}^N l_i(\theta) + w l_d(\theta), \quad (1.11)$$

where  $w = 1$  for the log-likelihood function based on the data set with subject  $d$  and  $w = 0$  for the log-likelihood function based on the data set without subject  $d$ . A value of  $w$  between 0 and 1 implies that the influence of subject  $d$  is downweighted, while a value of  $w$  which is larger than 1 implies that the influence of subject  $d$  is upgraded. Note, for example, that  $w = 2$  implies that an additional subject is measured, for which the log-likelihood is equal to  $l_d(\theta)$ .

*Weighted log-likelihood function.* The weighted log-likelihood function, for which the contribution of each subject to the log-likelihood function is weighted is given by:

$$\begin{aligned} \log L_w(\theta) &= \sum_{i=1}^N w_i l_i(\theta) \\ &= \sum_{i=1}^N w_i \log \int_{-\infty}^{\infty} \left( \prod_{j=1}^{n_i} f(y_{ij} | \mathbf{b}_i, \beta, \phi) \right) g(\mathbf{b}_i | D) d\mathbf{b}_i, \end{aligned} \quad (1.12)$$

*Local influence.* The weighted log-likelihood function depends on the vector of weights  $w = (w_1, \dots, w_N)'$ . The local influence approach is based on the weighted log-likelihood function in (1.12) and assesses the change in the estimates, due to small changes in the vector of weights  $w$ . Analogously to Cook's Distance, the change in the estimates can be compared with the length of the confidence ellipsoid in the direction of the change (Cook and Weisberg, 1982, p. 182, and Cook, 1986).

The local influence measure is discussed in the literature by Cook (1986) and Lesaffre and Verbeke (1998), among others. These papers show how the local influence measure can be defined for subsets of the model parameters, where these local influence measures are defined at subject level. No attention is, however, paid to the development of local influence measures at observation level. An observation-oriented local influence measure will be developed in this thesis in addition to the subject-level influence measures for the Generalized Linear Mixed Models.

*Importance of observation-oriented local influence measures.*

The detection of influential observations in addition to the detection of influential subjects is important for Generalized Linear Mixed Models. Consider, for example, a situation as presented in Figure 1.3. In Figure 1.3, the number of failures made during an examination, is plotted against the time after the beginning of the examination.

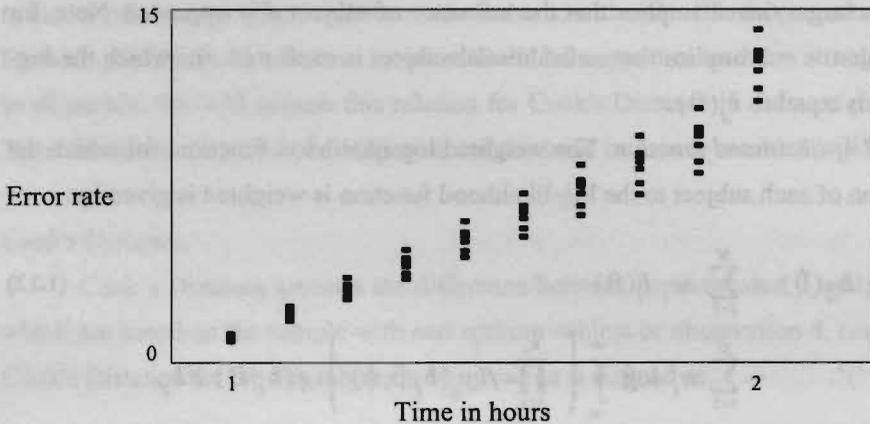


Figure 1.3. Error rate as function of the time of examination.

All subjects show a jump in their error rates curve after 2 hours of examination. This jump can, for example, be explained by loud noises from the environment or by tiredness. Because the jump is obtained for all subjects, the influence of the jump on the estimation process cannot be assessed by the subject-oriented influence measures. This influence, however, can be assessed by observation-oriented influence measures.

### 1.5 Relationship between $D$ -optimal design criterion and influence measures

We have shown that both the optimal design topic as well as the influence topic are related to the characteristics of ellipsoids. Via the ellipsoid, the choice of a design and the influence of particular subjects and observations are related. The  $D$ -optimality criterion and the influence measures can also be connected by means of the linear regression model

$y_i = x_i \beta + \epsilon_i$ , with independent normally distributed responses. The connection is based on the variance of the predicted response at a new vector of independent variables  $x_d$ , i.e. on  $\text{Var}[\hat{y}_d(\tau)] = x_d' \text{Var}(\hat{\beta}_\tau) x_d$ , where the variance is computed from the data of design  $\tau$ , which is not necessarily supported by  $x_d$ .

The equivalence theorem of Kiefer-Wolfowitz (Pukelsheim, 1993, p. 212) states that the  $D$ -optimal design  $\tau^*$  minimizes the largest variance of the predicted response

$\max [x_d' \text{Var}(\hat{\beta}_\tau) x_d]$  over all designs. Further, if the  $D$ -optimal design  $\tau^*$  is used, then the maximum is attained at each support point  $x_d$  of the  $D$ -optimal design  $\tau^*$ .

The expected Cook's Distance  $E(C_d)$  can be written as a product of two increasing functions of the variance of the predicted responses  $\text{Var}[\hat{y}_d(\tau)] = x_d' \text{Var}(\hat{\beta}_\tau) x_d$  under  $\tau$  (Cook and Weisberg, 1982, p. 117, and Weisberg, 1985, p. 127) :

$$E(C_d) = \frac{\text{Var}[\hat{y}_d(\tau)]}{(N-1)p\sigma^2} \left\{ 1 + \frac{\text{Var}[\hat{y}_d(\tau)]}{(N-1)p\sigma^2} \right\}. \quad (1.13)$$

Note that the expectation decreases to 0 if  $N$  increases to infinity. It follows from the previous two paragraphs that the  $D$ -optimal design minimizes the largest expected Cook's Distance. The  $D$ -optimal design criterion therefore leads to a design for which the maximal expected change in the estimate  $\hat{\beta}$ , due to the deletion of an observation is minimized over the regression space. Further, the expected change in the estimate  $\hat{\beta}$ , due to the deletion of an observation is largest, if the observation is obtained from one of the support points of the  $D$ -optimal design.

Cook's Distance is a measure for the influence of the deletion of a subject or observation. In fact, if Cook's Distance for a subject or observation is small, then it does not really matter whether the parameters are estimated based on the data set with or without that subject or observation. On the other hand, if Cook's Distance for a subject or observation is large, it really matters whether the parameters are estimated based on the data set with or without that subject or observation. The  $D$ -optimal design is therefore supported by points, which have large influence on the estimation process. An analogous reasoning can be given for the connection between the  $D$ -optimality criterion and the local

influence measure.

### 1.6 Outline of dissertation

The chapters in this thesis may be read as self-contained papers. A brief introduction to the chapters is given in this subsection.

*Chapter 2.* In this chapter we proceed the work of Berger and Tan (1998) and Tan and Berger (1999). They discussed D-optimal balanced designs in the presence of random intercepts and serial correlated measurement errors. They evaluated numerically the relative efficiency of some commonly used designs. Two extensions will be made in this paper. First, the model which is used in this paper is a random slope model with serial correlation structure. Both balanced designs as well as cohort designs will be studied. The second extension is that a maximin criterion will be used to obtain designs for which the smallest relative efficiency over all parameter values is maximized. We will also search for classes of symmetric designs for which the smallest relative efficiency over the model parameter space is large.

*Chapter 3.* The optimal design criteria discussed in this introduction assume that all independent variables are under experimental control, including all covariates and effect variables. One may, however, want to specify the design for only the treatment variable and the time. A new definition of the design will therefore be given in this chapter. Based on this definition a maximin criterion is proposed to optimize the setting of those variables, for which the distribution is specified by the researcher, taking into account the covariates for which the distribution is not specified by the researcher. We will show that, under certain conditions, the optimal distributions are balanced in the discrete independent variables. The proposed method will be illustrated by a clinical trial where covariates are involved.

*Chapter 4.* The definition of Cook's Distance in expression (1.8) is model independent. It may therefore be expected that this definition of Cook's Distance can also be used for the Linear Mixed Model. In this chapter we will give examples to demonstrate that the defined Cook's Distance may fail to detect or may incorrectly detect influential

observations. A Conditional version of Cook's Distance will be proposed, which can be decomposed in an influence measure for the change in the estimated expected profile and an influence measure for the change in the estimated subject-specific deviations.

*Chapter 5.* The subject-oriented local influence measure is discussed for the Linear Mixed Model by Lesaffre and Verbeke (1998). In this chapter, the subject-oriented local influence measure is generalized to the Generalized Linear Mixed Model. To check whether it is sufficient to use only the subject-oriented local influence measure, the observation-oriented local influence measure is developed. The practical importance of the observation-oriented influence measure is illustrated by the use of a two-treatment multiple period cross-over trial.

*Chapter 6.* In this chapter the maximin criterion which is proposed in Chapter 3 and the local influence measures which are proposed in chapter 5 are illustrated based on the project 'teeth-brushing on the elementary school'. It appeared that the balanced design is highly efficient, irrespective of the uncontrolled conditional distribution and model parameter vector at hand. The detection of influential data lead to an assessment of the stability of the parameter estimates, a change in the evaluated model and the suggestion to sample more schools for the validation of a random effect for the social-economic status.

## References

- Anderson, T.W., (1958). *An introduction to Multivariate Statistical Analysis*. New York: Wiley.
- Atkinson, A.C., and Donev, A.N. (1996). *Optimum experimental designs*. Oxford: Clarendon Press.
- Berger, M.P.F., King, C.Y. J., and Wong, W.K. (2000). Minimax D-optimal designs for item response theory models. *Psychometrika*, 65, 377-390.
- Berger, M.P.F., and Tan, F.E.S. (1998). Optimal Designs for Repeated Measurement Designs. *Kwantitatieve Methoden*, 59, 45-68.

- Bouter, L.M., and van Dongen, M.C.J.M. (1991). *Epidemiologisch onderzoek*. Houten: Bohn Stafleu Van Loghum BV.
- Bunke, H., and Bunke, O. (1986). *Statistical inference in linear models*. New York: Wiley.
- Chaloner, K., and Larntz, K. (1989). Optimal Bayesian design applied to logistic regression experiments. *Journal of Statistical Planning and Inference*, 21, 191-208.
- Chaloner, K., and Verdinelli, I. (1995). Bayesian Experimental Design: A review. *Statistical Science*, 10, 273-304.
- Cook, R.D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society, Series B*, 48, 133-169.
- Cook, R.D., and Weisberg, S. (1982). *Residuals and Influence in Regression*. New York: Chapman and Hall.
- Diggle, P.J., Liang, K.Y., and Zeger, S.L. (1994). *Analysis of longitudinal data*. Oxford: Clarendon Press.
- Goldstein, H. (1979). *The design and analysis of longitudinal studies*, London: academic press.
- Johnson, R.A., and Wichern, D.W. (1988). *Applied multivariate statistical analysis*. New Jersey: Prentice Hall International, Inc., Englewood Cliffs.
- King, J., and Wong, W.K. (2000). Minimax D-optimal designs for the Logistic model. *Biometrics*, 56, 1263-1267.
- Lesaffre, E., and Verbeke, G. (1998). Local influence in linear mixed models, *Biometrics*, 54, 570-582.
- Pukelsheim, F. (1993). *Optimal Design of Experiments*. New York: Wiley.
- Snijders, T.A.B., and Bosker, R.J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modelling*. London: Sage Publications.
- Tan, F.E.S., and Berger, M.P.F. (1999). Optimal allocation of time points for the random effects model, *Communications in Statistics, Simulations and Computations*, 28, 517-540.

- Verbeke, G., and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer Verlag.
- Weisberg, S. (1985). *Applied linear regression*. New York: Wiley.
- Wynn, H.P. (1970). The sequential generation of D-optimum experimental designs. *Annals of Mathematical Statistics*, 41, 1655-1664.
- Wong, W.K. (1992). A unified approach to the construction of minimax designs. *Biometrika*, 79, 611-619.
- Wu, C.F.J. (1985). Efficient sequential designs with binary data. *Journal of the American Statistical Association*, 80, 974-984.



# 2 Maximin D-optimal designs for Linear Mixed Models

Linear mixed models are widely used in many areas of research. In the design of experiments, the goal is to choose a set of experimental conditions that will provide the most accurate estimates of the parameters of the model. This is often done by using a D-optimal design, which is a design that maximizes the determinant of the Fisher information matrix. In this chapter, we consider the problem of finding a maximin D-optimal design for a linear mixed model with correlated measurement errors. We first review the theory of D-optimal designs for linear mixed models, and then discuss the problem of finding a maximin D-optimal design. We also discuss the problem of finding a maximin D-optimal design for a linear mixed model with correlated measurement errors. We conclude with a discussion of the relative efficiency of some commonly used designs.

Two objectives will be made in this chapter. First, the model used in this chapter is a random walk model with AR(1) correlated measurement errors. The D-optimal design for this model is said to be local optimal, i.e. the solution depends on the unknown and nuisance parameters. To deal with the local optimality of the D-optimal design, a maximin criterion is used, which is based on the relative efficiency. The second objective is that we present classes of approximate designs, for which all elements have a large relative efficiency over the whole parameter space. In this chapter, both balanced designs as well as point designs will be studied.

The importance of the polynomial models with random intercept and random slope is, for example, illustrated by Snijders and Bosker (1993, p. 182). The data set they used concerns children who started a professional-vocational program of growth measurement. Snijders and Bosker started the measurements for ages from 5 to 10 years. Children who are relatively short at 5 years grow relatively fast

### Abstract

In this chapter D-optimal cohort designs are computed for the first and second degree polynomial models with random intercept, random slope and first order autoregressive serial correlated measurement errors. To deal with the local optimality of the D-optimal designs, a maximin criterion is used, which is based on the relative efficiency. Large classes of symmetric designs are given, for which each element has a large smallest relative efficiency over the model parameter space.

*Key Words:* Random effects, cohort designs, D-optimality, robust designs, maximin designs.

## 2.1 Introduction

For longitudinal models, the shape of the confidence ellipsoid of the estimated regression parameters depends on the choice of the design (Pukelsheim, 1993). A design which minimizes the volume of the confidence ellipsoid is called D-optimal.

The optimal allocation and selection of time points for longitudinal models is discussed in literature (Atkins and Chang, 1999, Berger and Tan, 1998, Bisschof, 1993, Mentré, Mallet and Baccar, 1997, Tan and Berger, 1999). Atkins and Cheng (1999) discussed optimal quadratic polynomial regression designs in the presence of random intercepts and independent measurement errors. They assumed a constant number of repeated measures across persons. Bischoff (1993) discussed D-optimal designs for linear models under correlated observations. Mentré, Mallet and Baccar (1997) proposed an algorithm for finding a design that maximizes the determinant of the Fisher information matrix for random-effects regression models given a maximal cost. Berger and Tan (1998) and Tan and Berger (1999) discussed D-optimal balanced designs in the presence of random intercepts and serial correlated measurement errors. They evaluated numerically the relative efficiency of some commonly used designs.

Two extensions will be made in this chapter. First, the model used in this chapter is a random slope model with AR(1) correlated measurement errors. The D-optimal solution for this model appears to be local optimal, i.e. the solution depends on the variance and covariance parameters. To deal with the local optimality of the D-optimal designs, a maximin criterion is used, which is based on the relative efficiency. The second extension is that we present classes of symmetric designs, for which all elements have a large smallest relative efficiency over the model parameter space. In this chapter, both balanced designs as well as cohort designs will be studied.

The importance of the polynomial models with random intercept and random slope is, for example, illustrated by Snijders and Bosker (1999, p. 182). The data set they used concerns children who visited a paediatrician-endocrinologist because of growth retardation. Snijders and Bosker studied the measurements for ages from 5 to 10 years. The results show that children who are relatively short at 5 years grow relatively fast

between 5 and 10 years. To obtain this, a random intercept and random slope is used to model the subject specific deviations from the average profile.

An introduction to D-optimality will be given in section 2.2. In section 2.3 the relationship between the efficiency of a design under positively and under negatively correlated random slope and random intercept will be given. In section 2.4 the evaluated parameter values and the evaluated research period will be introduced and motivated and the optimization algorithm will be discussed. The local D-optimal designs for the first and second degree polynomial model will be discussed in section 2.5. In section 2.6 the maximin criterion will be introduced and maximin designs, maximin values and sets of highly efficient designs will be given.

## 2.2 Model specification and D-optimality

Let cohort  $i$  be based on the  $1 \times m$  vector of time points  $t_i' = (t_{i1}, \dots, t_{im}) \in [-1, 1]^m$ ,  $i = 1, \dots, C$ . Then the vector  $y_{ij}' = (y_{ij1}, \dots, y_{ijm})$  of  $m$  measurements of person  $j$ ,  $j = 1, \dots, n_i$ , in cohort  $i$ ,  $i = 1, \dots, C$ , is ascribed to the expected trend  $X_i\beta$  over subjects, to the subject-specific deviations  $Z_i b_{ij}$  from this expected trend  $X_i\beta$  and to the deviations  $\epsilon_{ij}$  of the observations from the subject-specific profile  $(X_i\beta + Z_i b_{ij})$  as follows:

$$y_{ij} = X_i\beta + Z_i b_{ij} + \epsilon_{ij}, \quad (2.1)$$

where the  $p \times 1$  vector  $\beta$  is a vector of fixed regression coefficients and the  $q \times 1$  vector  $b_{ij}$  is the vector of random regression coefficients of subject  $j$  within cohort  $i$ , with mean zero and  $q \times q$  variance-covariance matrix  $D$ . The  $m \times p$  matrices  $X_i$  of explanatory variables and the  $m \times q$  matrices  $Z_i$  of random effect variables consist of polynomial coefficients based on  $t_i$ . In fact, the  $m \times q$  matrices  $Z_i$  are submatrices of the matrices  $X_i$ . The  $m \times 1$  vector  $\epsilon_{ij}$  of measurement errors has mean zero and  $m \times m$  variance-covariance matrix  $\Psi_i$ .

In this chapter, we will confine ourselves to regression models with random intercepts and random slopes and AR(1) correlated measurement errors, i.e. the covariance between two measurements at time points  $t_{ik}$  and  $t_{ik'}$ , conditional on person  $j$  within cohort  $i$  is given by  $\text{Cov}(y_{ijk}, y_{ijk'} | i, j) = \sigma^2 \rho^{|t_{ik} - t_{ik'}|}$ , where  $0 \leq \rho \leq 1$ .

The matrices  $X_i$  and  $Z_i$  consist of polynomial coefficients based on  $t_i$ ,  $i = 1, \dots, C$ . The experimental designs  $\tau$  are therefore of the form

$$\tau = \begin{Bmatrix} t_1', \dots, t_C' \\ w_1, \dots, w_C \end{Bmatrix}, \tag{2.2}$$

where  $\sum_{i=1}^C w_i = 1$  and  $w_i$  is the relative size of cohort  $i$ . The space of all experimental designs is called the design space and is denoted by  $T_{Cm}$ . Further, the matrices  $X_i$ ,  $Z_i$  and  $\Psi_i$  corresponding to a design  $\tau$  are denoted by  $X_{i_\tau}$ ,  $Z_{i_\tau}$  and  $\Psi_{i_\tau}$ , while the vectors of time points  $t_i$  corresponding to a design  $\tau$  are denoted by  $t_{i_\tau}$ ,  $i = 1, \dots, C$ .

The best linear unbiased estimator of  $\beta$  is the generalized version of the Gauss-Markoff estimator (Rao, 1973, p. 234) with variance-covariance matrix equal to:

$$\text{Var}(\hat{\beta}) = \left\{ \sum_i n_i \left[ X_i' (Z_i D Z_i' + \Psi_i)^{-1} X_i \right] \right\}^{-1}. \tag{2.3}$$

*D-optimality criterion*

The volume of the confidence ellipsoid is an increasing function of the determinant  $\text{Det}[\text{Var}(\hat{\beta})]$  of the variance-covariance matrix of the estimated regression parameters, which is also called the generalized variance of the regression parameters (Anderson, 1958, p. 166). If the variance parameters are known, then the generalized variance  $\text{Det}[\text{Var}(\hat{\beta})]$  can be minimized by the choice of a D-optimal design  $\tau^* \in T_{Cm}$ , such that for each design  $\tau \in T_{Cm}$

$$\text{Det}[\text{Var}(\hat{\beta}_{\tau^*})] \leq \text{Det}[\text{Var}(\hat{\beta}_\tau)], \tag{2.4}$$

where  $\hat{\beta}_{\tau^*}$  and  $\hat{\beta}_\tau$  are the estimates of  $\beta$  under  $\tau^*$  and  $\tau$ , respectively.

### Relative efficiency

The relative efficiency can be used to compare the generalized variance of two experimental designs  $\tau$  and  $\tau^*$ . Let  $m_\tau$  and  $m_{\tau^*}$  be the number of repeated measures under design  $\tau$  and the design  $\tau^*$ , respectively. The relative efficiency of  $\tau$  with respect to  $\tau^*$ , conditional on the values of the model parameters, is then equal to:

$$\text{eff}(\tau) = \frac{m_{\tau^*}}{m_\tau} \left\{ \frac{\text{Det}[\text{Var}(\beta_\tau)]}{\text{Det}[\text{Var}(\beta_{\tau^*})]} \right\}^{1/p}, \quad (2.5)$$

where  $p$  is equal to the number of fixed regression coefficients in the model.

Expression (2.5) can be interpreted as follows. If, for example,  $\text{eff}(\tau) = 0.90$  for a particular design  $\tau$ , then 10% less observations will be needed under the design  $\tau^*$  to obtain the same efficiency as that under the design  $\tau$  for the same model. Unless otherwise stated, the relative efficiency is in this chapter computed with respect to the D-optimal design  $\tau^*$ , given the same number of repeated measures  $m$ . Before D-optimal designs are presented, the relationship between the efficiency of designs under positively and negatively correlated random slopes and random intercepts is given.

### 2.3 Relationship between the efficiency of designs for positively and negatively correlated random slopes and random intercepts

To simplify the presentation of the relationship, we define the design  $-\tau$  as follows. If  $\tau$  is equal to  $\begin{Bmatrix} t_1', \dots, t_C' \\ w_1, \dots, w_C \end{Bmatrix}$ , then  $-\tau$  is equal to  $\begin{Bmatrix} -t_1', \dots, -t_C' \\ w_1, \dots, w_C \end{Bmatrix}$ . The relationship is proven in the Appendix of this chapter.

**Theorem.** Given two polynomial random slope models  $M_1$  and  $M_2$  implied by equation (2.1) of the same order  $p - 1$ . Suppose  $\rho \in [0, 1]$  and  $\sigma^2$  are known finite positive values. Let the variance-covariance matrix of the random effects under  $M_1$  be equal to

$D_1 = \begin{pmatrix} D_{11} & D_{12} \\ D_{12} & D_{22} \end{pmatrix}$  and under  $M_2$  be equal to  $D_2 = \begin{pmatrix} D_{11} & -D_{12} \\ -D_{12} & D_{22} \end{pmatrix}$ , for finite values of  $D_{11}$ ,  $D_{12}$  and  $D_{22}$ , such that  $D_1$  and  $D_2$  are positive definite. Let  $\tau \in T_{Cm}$  be such that model  $M_1$  is identified under  $\tau$ , where  $C$  and  $m$  are finite numbers. Further, let  $\text{Det}[\text{Var}(\hat{\beta}_\tau) | M_1]$  and  $\text{Det}[\text{Var}(\hat{\beta}_{-\tau}) | M_2]$  be the generalized variance under design  $\tau$  and model  $M_1$  and the finite generalized variance under design  $-\tau$  and model  $M_2$ , respectively. Then

$$\text{Det}[\text{Var}(\hat{\beta}_\tau) | M_1] = \text{Det}[\text{Var}(\hat{\beta}_{-\tau}) | M_2]. \tag{2.6}$$

Expression (2.6) implies that if  $\tau^*$  is D-optimal for  $M_1$ , then  $-\tau^*$  is D-optimal for  $M_2$ . Further, if  $\tau$  and  $-\tau$  are the same, then

$\text{Det}[\text{Var}(\hat{\beta}_\tau) | M_1] = \text{Det}[\text{Var}(\hat{\beta}_{-\tau}) | M_2]$ . As a consequence, if  $\tau$  is equal to  $-\tau$ , then the relative efficiency of the design  $\tau$  under  $M_1$  with respect to the design  $\tau^*$  is equal to the relative efficiency of the design  $\tau$  under  $M_2$  with respect to the design  $-\tau^*$ . This equality will be used in section 2.6 for the construction of highly efficient designs. For simplicity, designs for which  $\tau$  and  $-\tau$  are the same are said to be symmetric designs. Examples of symmetric designs are

$$\left\{ \begin{pmatrix} -1, & 0.3, & 1 \\ 0.5, & & 0.5 \end{pmatrix}, \begin{pmatrix} -1, & -0.3, & 1 \\ & & 0.5 \end{pmatrix} \right\} \text{ and } \left\{ \begin{pmatrix} -1, & -0.3, & 0.3, & 1 \\ & & & 1 \end{pmatrix} \right\}.$$

Because D-optimal designs are only local optimal, it is difficult to determine which design is D-optimal for a specific research situation. Moreover, the determinant  $\text{Det}[\text{Var}(\hat{\beta}_\tau)]$  is such a complex function, that an analytical solution is hard to realize. In section five, we will show how the D-optimal allocation of time points depends on the parameters of the variance components. Further, the effect of a misspecification of parameter value-combinations will be discussed and a maximin criterion will be introduced to deal with the local optimality of the D-optimal design. The D-optimal

designs are determined numerically. Hence, relevant time points and possible values of the variance components are chosen in the next section, to show how the D-optimal designs depend on these parameters.

### 2.4 The time period and parameter values

The values of the random effects variance-covariance matrices  $D$ , are given in Table 2.1. Without loss of generality, the variance of the random intercept  $D_{11}$  is assumed to be equal to 1.

Table 2.1 Variance-covariance matrices  $D$  of the random effects

$\begin{pmatrix} 1 & 0 \\ 0 & 0.002 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.5\sqrt{0.002} \\ 0.5\sqrt{0.002} & 0.002 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.8\sqrt{0.002} \\ 0.8\sqrt{0.002} & 0.002 \end{pmatrix}$
$\begin{pmatrix} 1 & 0 \\ 0 & 0.0201 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.5\sqrt{0.0201} \\ 0.5\sqrt{0.0201} & 0.0201 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.8\sqrt{0.0201} \\ 0.8\sqrt{0.0201} & 0.0201 \end{pmatrix}$
$\begin{pmatrix} 1 & 0 \\ 0 & 0.21 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.5\sqrt{0.21} \\ 0.5\sqrt{0.21} & 0.21 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.8\sqrt{0.21} \\ 0.8\sqrt{0.21} & 0.21 \end{pmatrix}$
$\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.5\sqrt{3} \\ 0.5\sqrt{3} & 3 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.8\sqrt{3} \\ 0.8\sqrt{3} & 3 \end{pmatrix}$
$\begin{pmatrix} 1 & 0 \\ 0 & 120 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.5\sqrt{120} \\ 0.5\sqrt{120} & 120 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.8\sqrt{120} \\ 0.8\sqrt{120} & 120 \end{pmatrix}$

We choose to evaluate 0, 0.5 and 0.8 as values of the correlation between the random slope and random intercept. Note that, according to the Theorem, local D-optimal designs for  $D_{12} = -0.5\sqrt{D_{11}D_{22}} = -0.5\sqrt{D_{22}}$  and  $D_{12} = -0.8\sqrt{D_{22}}$  can be derived from the corresponding local D-optimal designs for  $D_{12} = 0.5\sqrt{D_{22}}$  and  $D_{12} = 0.8\sqrt{D_{22}}$ .

The variance of the measurement errors  $\sigma^2$  is assumed to be smaller than the

variance of the random intercept  $D_{11}$ , i.e.  $\sigma^2 \leq D_{11} = 1$ . This situation is commonly encountered in longitudinal data (Snijders and Bosker, chapt. 12). We found that the smallest value of  $\sigma^2$  that revealed a meaningful variation was  $\sigma^2 = 0.025$ . Therefore, the three evaluated values of  $\sigma^2$  are  $\sigma^2 = 1, 0.25$  and  $0.025$ . The serial correlation  $\rho$  varies within the interval  $[0, 0.9]$ .

As will be made apparent in section five by numerical computations, the time points of the D-optimal designs are attracted to the location of the smallest dispersion, i.e. to time point  $L = -D_{12} / D_{22}$ . Note that the dispersion at time point  $t$  is given by  $\sqrt{1 + 2tD_{12} + D_{22}t^2 + \sigma^2}$ . The variance of the random slope  $D_{22}$  is chosen so that  $L$  will take a reasonable range of different values. The number of cohorts  $C$  may take the values  $C = 1, 2, 3$  and  $4$ .

The local D-optimal designs are obtained by numerical computations under the restriction that  $t_{ij} \in [-1, 1]^m$  for all  $i, i = 1, \dots, C$ , and  $j, j = 1, \dots, n_i$ . This is motivated by Theorem 8.34 in Bunke and Bunke (1986), which states that the time period can be restricted without loss of generality.

### Optimization Algorithm

The following global search algorithm was used to obtain D-optimal designs for this chapter:

1. Take an initial guess of the D-optimal design
  2. Compute the 'D-optimal solution' based on the local search algorithm BFGS
- Check 1:
3. Take random starting values
  4. Do step 2
  5. Compare with the earlier obtained results
- Check 2:
6. Make a grid of the time points and the weights
  7. Compute the D-optimal solution for the grid
  8. Refine the grid in the neighbourhood of the D-optimal solution and compute the D-optimal solution for the refined grid

The number of starting values for check 1 was set equal to 10. We have used a grid for which each time interval was subdivided into 30 smaller intervals of length  $1/15$  and for which each interval corresponding with the weights of the cohorts was subdivided into 10 smaller intervals of length  $1/10$ . By refining the grid in the neighbourhood of the optimum, we found the same designs as under the random search algorithm.

The random search algorithm seems to work very well and is more than 1000 times faster than the algorithm of evaluating each possible combination of time points. The quasi-Newton algorithm BFGS (Broyden-Fletcher-Goldfarb-Shanno) is described at <http://orion.math.uwaterloo.ca/~hwolkowi/mirror.d/glossary/B.html> and is default in TOMLAB (<http://www.mathtools.com/files/tomlab//tomlabv2.pdf>), which is the optimization toolbox for MATLAB.

## 2.5 Local D-optimal designs for the first and second degree polynomial model

To describe how the local D-optimal designs depend on the model parameters, local D-optimal designs will be given for the first and second degree polynomial model. Then, the effect of some misspecifications on the relative efficiencies will be made apparent. The D-optimal designs which will be discussed in this section correspond with positively correlated random slopes and random intercepts. D-optimal designs for negatively correlated random slopes and random intercepts can be obtained from these designs by the use of the Theorem of section 2.3.

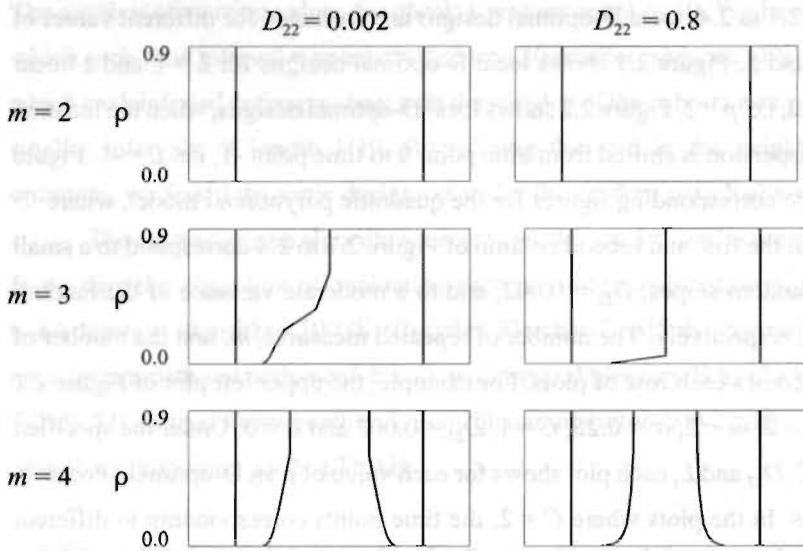
### *Local D-optimal designs*

Note that the optimization problem is reformulated, such that the variance of the random intercepts  $D_{11}$  is equal to 1 and the feasible time points are elements of the interval  $[-1, 1]$ . Taking this into account, the efficiency of a design depends on the serial correlation coefficient  $\rho$ , the variance of the measurement errors  $\sigma^2$ , the variance of the random slopes  $D_{22}$ , the location of the smallest dispersion  $L = -D_{12}/D_{22}$ , the number of repeated measures  $m$ , and the number of cohorts  $C$ .

In Figure 2.1 to 2.4, local D-optimal designs are presented for different values of  $\rho$ ,  $\sigma^2$ ,  $m$ ,  $C$ ,  $D_{22}$  and  $L$ . Figure 2.1 shows local D-optimal designs for  $L = 0$  and a linear polynomial model, i.e.  $p = 2$ . Figure 2.2 shows local D-optimal designs, when the location of the smallest dispersion is shifted from time point 0 to time point -1, i.e.  $L = -1$ . Figure 2.3 and 2.4 are the corresponding figures for the quadratic polynomial model, where  $p = 3$ . The plots in the first and second column of Figure 2.1 to 2.4 correspond to a small variance of the random slopes,  $D_{22} = 0.002$ , and to a moderate variance of the random slopes,  $D_{22} = 0.8$ , respectively. The number of repeated measures,  $m$ , and the number of cohorts,  $C$ , vary across each row of plots. For example, the upper left plot of Figure 2.1 corresponds to  $p = 2$ ,  $m = 2$ ,  $\sigma^2 = 0.25$ ,  $C = 1$ ,  $D_{22} = 0.002$  and  $L = 0$ . Under the specified values of  $\sigma^2$ ,  $m$ ,  $C$ ,  $D_{22}$  and  $L$ , each plot shows for each value of  $\rho$  an D-optimal allocation of the time points. In the plots where  $C = 2$ , the time points corresponding to different cohorts are indicated by different line types. A solid line indicates the first cohort and a dotted line indicates the second cohort.

Several conclusions can be extracted from these figures. From Figure 2.1 and 2.2, it can be concluded that the shift of the location of the smallest dispersion from time point 0 to time point -1 is followed by a shift of the time points of the corresponding local D-optimal designs in the same direction. The local D-optimal designs for  $L = 0$  are symmetric. The symmetry of the local D-optimal designs for  $L = 0$  can be explained by the fact that 0 is the mid time point of the interval  $[-1, 1]$ , where the attraction to the positive time points is as strong as the attraction to the negative time points. The effect of the location of the smallest dispersion also arises for the quadratic polynomial model, but is much smaller than for the linear polynomial model (see Figure 2.3 and 2.4). The figures also show that the effect of the shift in location of the smallest dispersion increases as  $D_{22}$  increases.

$$C = 1, L = 0, \sigma^2 = 0.25, p = 2$$



$$C = 2, L = 0, \sigma^2 = 0.25, p = 2$$

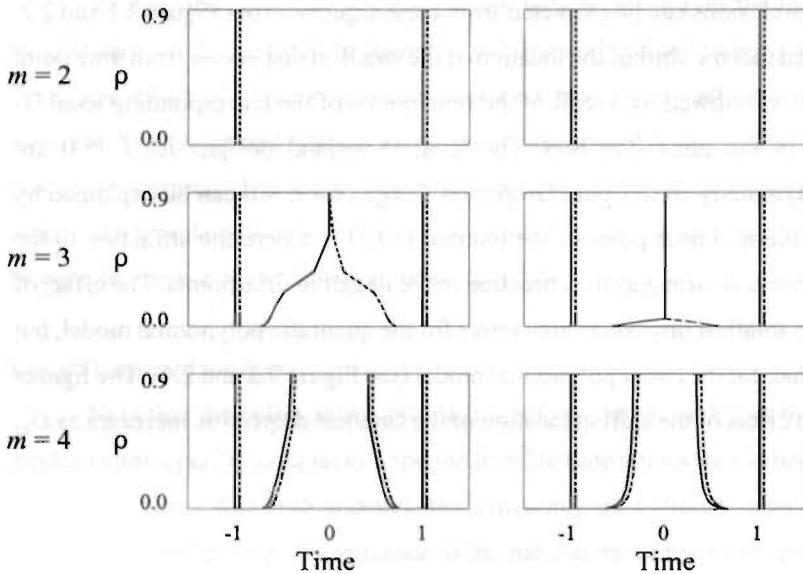
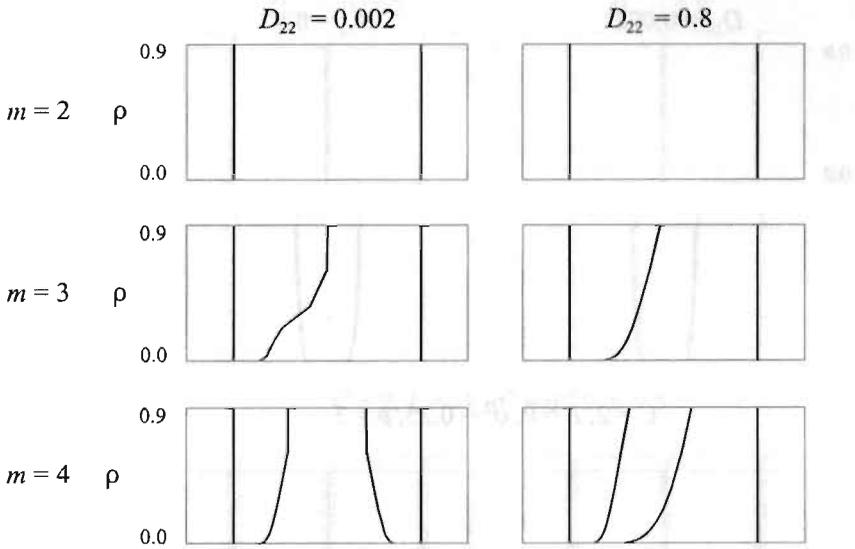


Figure 2.1. Optimal designs for the first degree polynomial model, with  $p = 2$  and  $L = 0$ .

$$C = 1, L = -1, \sigma^2 = 0.25, p = 2$$



$$C = 2, L = -1, \sigma^2 = 0.25, p = 2$$

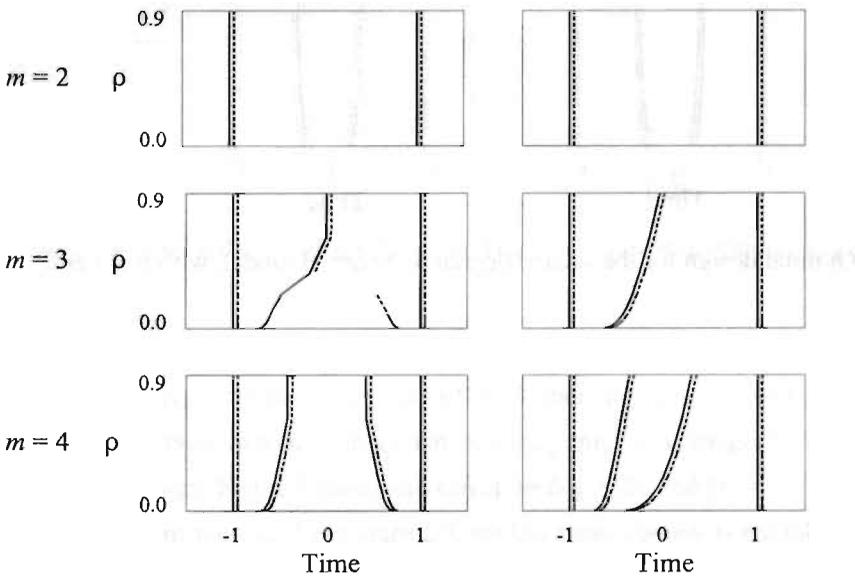


Figure 2.2. Optimal designs for the first degree polynomial model, with  $p = 2$  and  $L = -1$ .

$$C = 1, L = 0, \sigma^2 = 0.25, p = 3$$

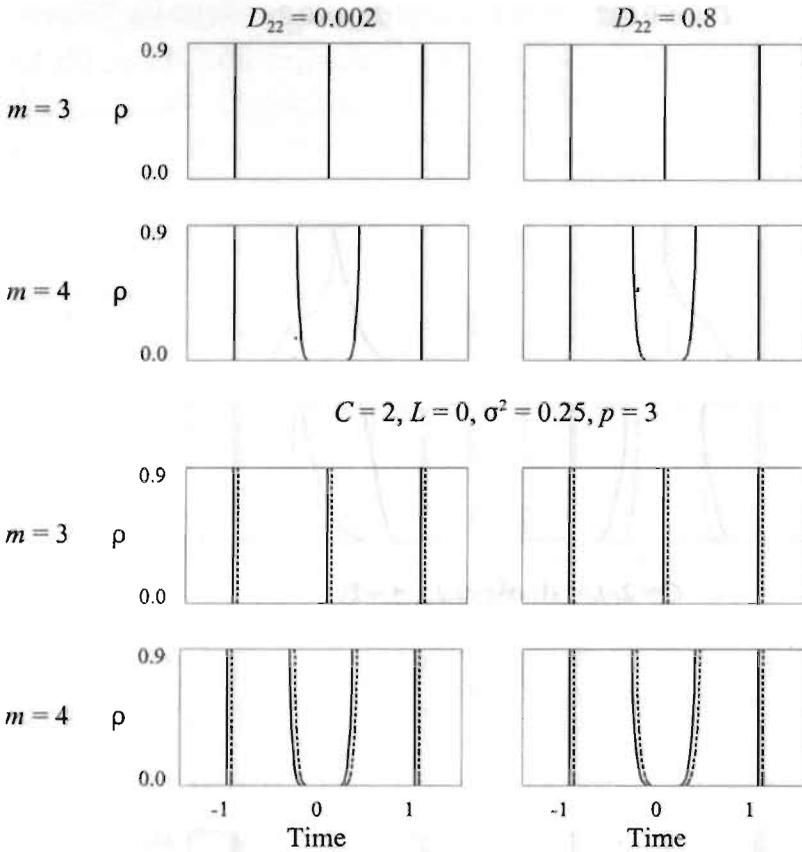


Figure 2.3. Optimal design for the second degree polynomial model, with  $p = 3$  and  $L = 0$ .

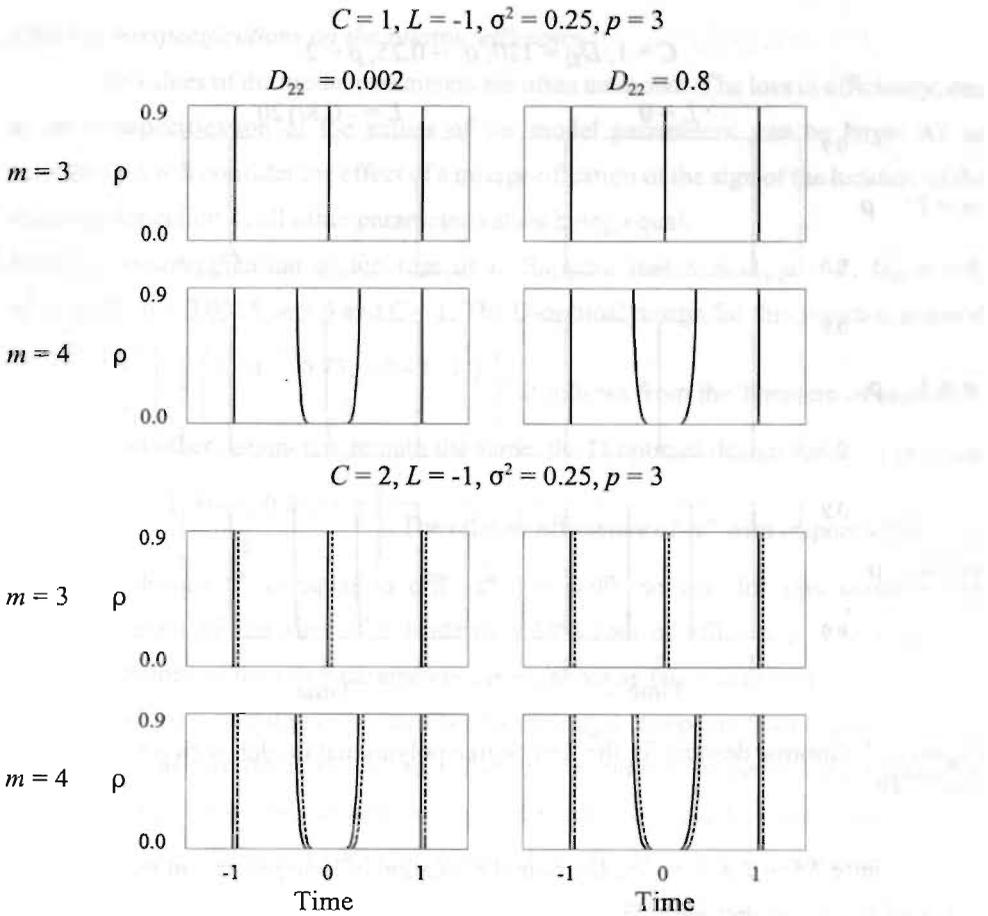


Figure 2.4. Optimal designs for the second degree polynomial model, with  $p = 3$  and  $L = -1$ .

All plotted local D-optimal designs include the time points  $-1$  and  $1$ . However, if  $D_{22}$  is extremely large and  $L$  is smaller than  $0$ , then the influence of the location of the smallest dispersion may be so large that the time point  $1$  is no longer included in the local D-optimal design. Figure 5 shows this effect for  $D_{22} = 120$  and  $D_{12} = 0.8$ . For the local D-optimal designs presented in Figure 2.5, the last measurement is not taken at the end of the research period. We will return to this phenomenon in the discussion section.

$$C = 1, D_{22} = 120, \sigma^2 = 0.25, p = 2$$

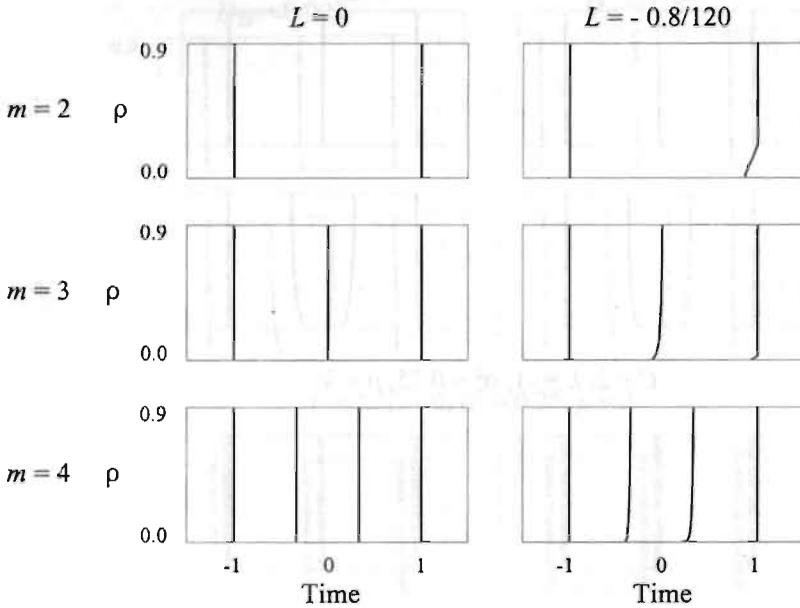


Figure 2.5. Optimal designs for the first degree polynomial model, with  $p = 2$  and  $D_{22} = 120$ .

Figure 2.1 to 2.5 show the D-optimal allocation of time points, but the weights of the cohorts are not included. These weights are equal to  $1/C$  for almost all cases. The weights of the cohorts in the plotted two-cohort designs are only unequal to  $1/C$  for the designs given by the plot of Figure 2.2, corresponding to  $p = 2, m = 3, C = 2, D_{22} = 0.002, L = -1$  and  $\sigma^2 = 0.25$ . For this plot, the weight of the cohort with positive inner time point decreases to zero as  $\rho$  increases to 0.25.

Although the presented local D-optimal designs are based on  $C = 1$  and  $C = 2$  cohorts, local D-optimal designs for  $C = 3$  and  $C = 4$  cohorts were also computed. The local D-optimal designs for  $C = 4$  cohorts coincide with the local D-optimal designs for  $C = 3$  cohorts. The local D-optimal designs for  $C = 3$  cohorts only differ from the local D-optimal designs for  $C = 2$  cohorts, if the quadratic model is evaluated, the number of repeated measures  $m$  is equal to  $m = 4$  and the serial correlation  $\rho$  is very small.

*Effect of misspecifications on the relative efficiency*

The values of the model parameters are often unknown. The loss in efficiency, due to the misspecification of the values of the model parameters, can be large. As an example, we will consider the effect of a misspecification of the sign of the location of the smallest dispersion  $L$ , all other parameter values being equal.

*Effect of misspecification of the sign of  $L$ .* Suppose that  $L = -1$ ,  $p = 2$ ,  $D_{22} = 0.8$ ,  $\sigma^2 = 0.25$ ,  $\rho = 0.0025$ ,  $m = 4$  and  $C = 1$ . The D-optimal design for this situation is equal

to  $\tau^* = \begin{Bmatrix} t_1' \\ w_1 \end{Bmatrix} = \begin{Bmatrix} (-1, -0.73, -0.40, 1) \\ 1 \end{Bmatrix}$ . It follows from the Theorem of section 3.2 that if all other parameters remain the same, the D-optimal design for  $L = 1$  is equal

to  $-\tau^* = \begin{Bmatrix} (-1, 0.40, 0.73, 1) \\ 1 \end{Bmatrix}$ . The relative efficiency of  $-\tau^*$  with respect to the local D-optimal design  $\tau^*$  is equal to  $\text{eff}(-\tau^*) = 0.90$ , so that for this situation the misspecification of the sign of  $L$  leads to a 10% loss of efficiency. The effects of misspecifications of the other parameters can be about as large as that of  $L$ .

An often used design is the equally spaced design. It appears from Figure 2.1 to 2.5 that if the serial correlation is close to 1, then the D-optimal designs are almost equally spaced. In fact, for the first degree polynomial model, the equally spaced design has a smallest relative efficiency of 0.94 for  $m = 3$  and of 0.93 for  $m = 4$ . For the second degree polynomial model, the equally spaced design has a smallest relative efficiency of 0.997 for  $m = 3$  and 0.96 for  $m = 4$ .

It follows from the Theorem of section 2.3 that the relative efficiency of a symmetric design is invariant under a change of the sign of  $L$ , provided that the relative efficiency is determined with respect to the corresponding D-optimal design. It turns out that, evaluating symmetric designs corresponds with the search for maximin designs as will be shown in the next section. In the following section we will show numerically, that there exist maximin designs with a large relative efficiency.

## 2.6 Maximin designs for the first and second degree polynomial model

Let the space  $\Theta$  of optimization parameters be defined by

$$\Theta = \{ \theta: \theta = (\rho, \sigma^2, L, D_{22}, C), \quad (2.7)$$

$$\rho \in [0, 1], \sigma^2 \in [0, 1], D_{22} \geq 0, |L| \leq 1/\sqrt{D_{22}}, C \in \{1, 2, 3, 4\} \}.$$

A maximin design, *MMD*, over a specific set  $S$  of designs is in this chapter defined to be a design in  $S$  for which  $\max_S [ \min_{\Theta} \text{eff}(\tau) ]$  is attained, where  $\text{eff}(\tau)$  is taken with respect to the D-optimal design within the design space  $T_m = \bigcup_C T_{Cm}$ . The corresponding value of  $\max_S [ \min_{\Theta} \text{eff}(\tau) ]$  is called the maximin value, say *MMV*, under  $S$ . It should be noted that the maximin design *MMD* is independent of the values of the variance parameters, due to the minimization over the set  $\Theta$ .

The following corollary of the Theorem given in section 2.3 shows the relationship between maximin designs and symmetric designs.

**Corollary.** Let  $T_m = \bigcup_C T_{Cm}$ . For each design  $\tau \in T_m$  there exists a symmetric design  $\tau_s \in T_m$ , such that  $\min_{\Theta} \text{eff}(\tau_s) \geq \min_{\Theta} \text{eff}(\tau)$ .

The corollary is proven in the appendix. A consequence of the corollary is that for each maximin design  $\tau \in T_m$  there is a maximin design  $\tau_s \in T_m$  which is symmetric. Instead of determining the *MMD* among all symmetric designs, we have decided to evaluate two sets of symmetric designs,  $S_1$ , based on  $C = 2$  cohorts and  $m = 3$  repeated measures, and  $S_2$ , based on  $C = 1$  cohort and  $m = 4$  repeated measures, which are given by:

$$S_1 = \{ \tau: \tau = \left\{ \begin{pmatrix} -1, & -a, & 1 \\ & 0.5 & 0.5 \end{pmatrix} \begin{pmatrix} -1, & a, & 1 \end{pmatrix} \right\}, 0 \leq a \leq 1 \} \quad (2.8)$$

$$S_2 = \{ \tau: \tau = \left\{ \begin{pmatrix} -1, & -a, & a, & 1 \end{pmatrix} \right\}, 0 \leq a \leq 1 \}. \quad (2.9)$$

The smallest relative efficiencies  $\min_{\theta} \text{eff}(\tau)$  are much smaller for models without serial correlation than for models with serial correlation  $\rho \geq 0.0025$ . In this section we will only discuss the results for models with serial correlation  $\rho \geq 0.0025$ , but the results for models without serial correlation are also presented in the figures.

*Highly efficient designs for the first degree polynomial model with serial correlation.* Figure 2.6 and 2.7 show the smallest relative efficiencies of the elements of  $S_1$  and  $S_2$ , respectively, as a function of  $a$ . The solid lines indicate the smallest relative efficiencies for  $\rho \geq 0.0025$ , while the dotted lines indicate the smallest relative efficiencies for  $\rho = 0$ .

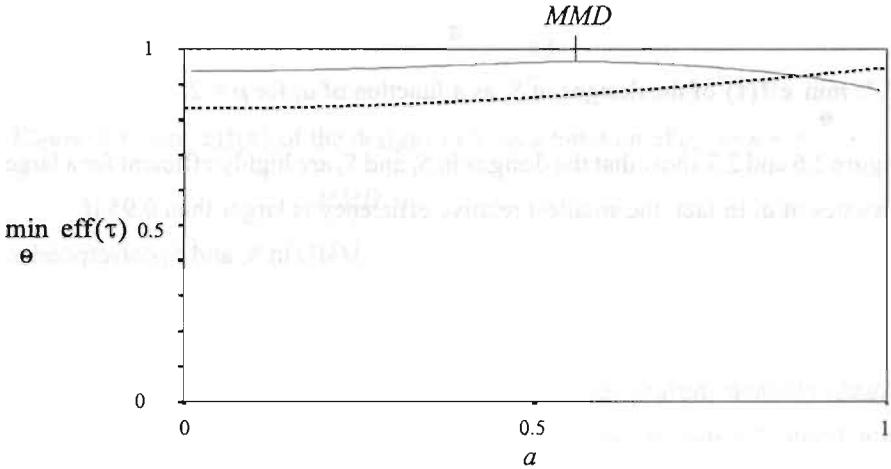


Figure 2.6.  $\min_{\theta} \text{eff}(\tau)$  of the designs in  $S_1$  as a function of  $a$ , for  $p = 2$ .

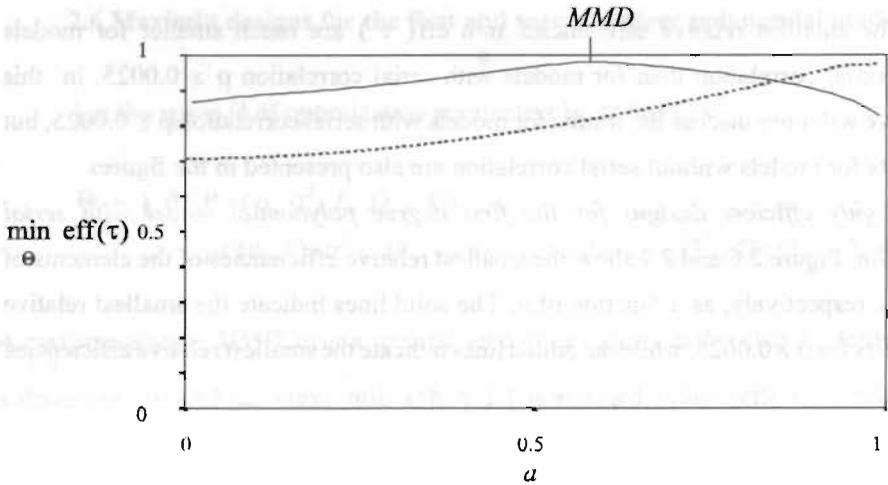


Figure 2.7.  $\min \text{ eff}(\tau)$  of the designs in  $S_2$  as a function of  $a$ , for  $p = 2$ .

Figure 2.6 and 2.7 show that the designs in  $S_1$  and  $S_2$  are highly efficient for a large range of values of  $a$ . In fact, the smallest relative efficiency is larger than 0.95 if  $0.25 < a < 0.80$  for  $S_1$ , and if  $0.45 < a < 0.75$  for  $S_2$ . The *MMD* in  $S_1$  and  $S_2$  correspond to  $a = 0.55$  and  $a = 0.57$ , respectively. The corresponding *MMV* are 0.97 and 0.98, respectively.

*Highly efficient designs for the second degree polynomial model with serial correlation.* Figure 2.8 and 2.9 show the values of  $\min \text{ eff}(\tau)$  for the designs in  $S_1$  and  $S_2$ , respectively, as a function of  $a$ .

As can be seen from Figure 2.8, the design  $\left\{ \begin{matrix} (-1, 0, 1) \\ 1 \end{matrix} \right\}$  is the *MMD* in set  $S_1$ , with *MMV* = 0.997. It should be noted that this design is not D-optimal for all sets of parameter values, just like the design  $\left\{ \begin{matrix} (-1, 1) \\ 1 \end{matrix} \right\}$  is not D-optimal for all sets of parameter values for the first degree polynomial model (see Figure 2.5). The smallest relative efficiency for designs in  $S_1$  is larger than 0.95 for  $a < 0.28$ . Figure 2.9 shows that the smallest relative efficiency for designs in  $S_2$  is larger than 0.95 for  $0.10 < a < 0.38$ . The value of  $a$  corresponding to the *MMD* is equal to  $a = 0.23$ , with *MMV* = 0.995.

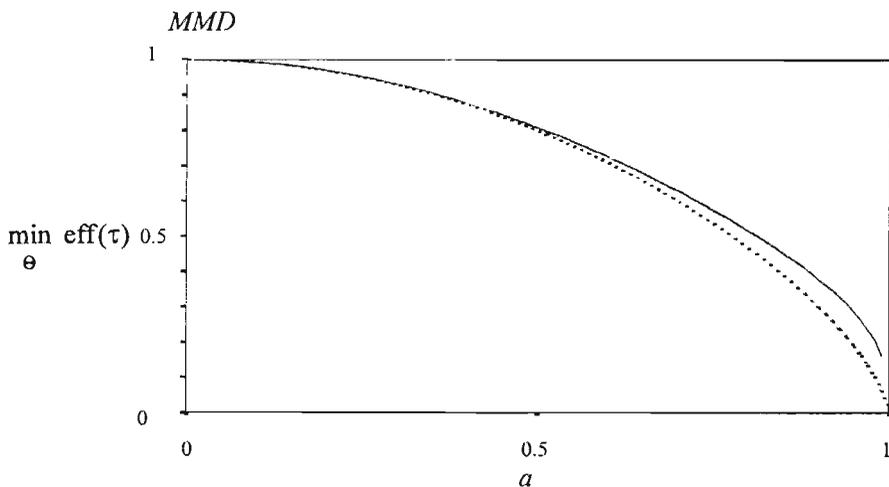


Figure 2.8.  $\min_{\theta} \text{eff}(\tau)$  of the designs in  $S_1$  as a function of  $a$ , for  $p = 3$ .

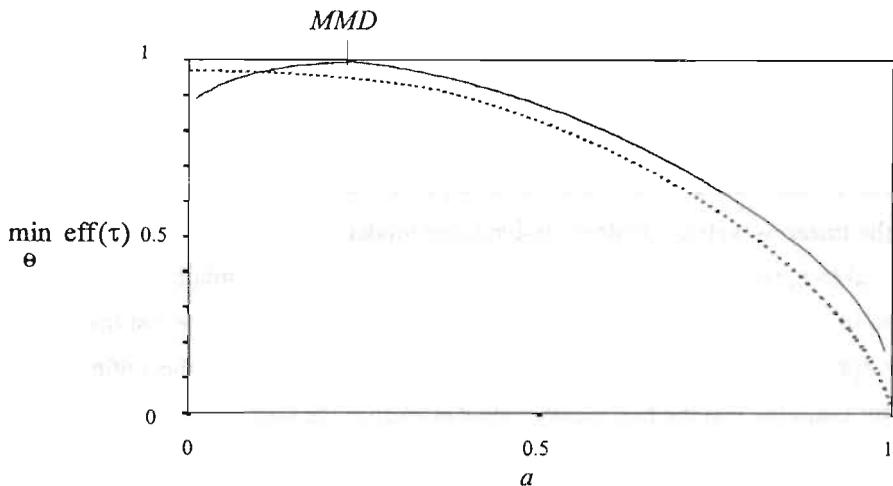


Figure 2.9.  $\min_{\theta} \text{eff}(\tau)$  of the designs in  $S_2$ , as a function of  $a$ , for  $p = 3$ .

*Highly efficient designs for both the first as well as the second degree polynomial model with serial correlation.* The degree of the polynomial is often unknown. In that case it makes sense to use designs, which are highly efficient for both polynomial models.

Figure 2.6 and Figure 2.8 show that the smallest relative efficiency for designs in  $S_1$  is larger than 0.95 for both polynomial models, if  $0.25 < a < 0.28$ . The *MMD* over both polynomial models is the design for which  $a = 0.27$ , with *MMV*  $\approx 0.95$ .

It can be concluded from Figure 2.7 and Figure 2.9 that the smallest relative efficiency for designs in  $S_3$ , is larger than 0.90 for  $0.2 < a < .46$ . The *MMV* over both polynomial models is attained for  $a = 0.39$  and is equal to *MMV* = 0.94. There is no design with smallest relative efficiency above 0.95.

## 2.7 Summary and discussion

In this chapter, local D-optimal designs for the random slope model with serial correlation are studied. The local D-optimal designs were evaluated numerically, because the optimization problem involves finding the roots of high degree polynomials and no analytic solution to this problem seems to exist (Hungerford, 1990, p. 363). Because the D-optimal designs depend heavily on the values of the variance-covariance parameters, the D-optimality criterion is used in combination with a maximin criterion. Designs which are highly efficient under a large range of parameter values are given for both polynomial models. Moreover, it is shown that some of the highly efficient designs are highly efficient for both the linear as well as quadratic polynomial model.

Local D-optimal designs for extremely large  $D_{22}$ ,  $L < 0$  and small values of  $\rho$  are difficult to interpret, because the end point of the time interval is not the last time point of the D-optimal designs (see Figure 2.5). It is possible to restrict the optimization problem by assuming that the first measurement is taken at the start of the research period and the last time point is taken at the end of the research period. This restriction involves a loss of efficiency of at most 0.3%. Although this loss of efficiency is very small, it shows that the claim that measurements have to be taken at the time points -1 and 1 restricts the optimization problem.

It is possible to fit a second degree polynomial model, with  $m = 2$  time points and different cohorts. However, the corresponding D-optimal designs are very inefficient for large values of  $\rho$  and  $D_{22}$ . Numerical computations show that the relative efficiencies of

the local D-optimal designs for  $m = 2$ ,  $D_{22} = 120$ ,  $\sigma^2 = 0.025$  and  $L = 0$ , with respect to the corresponding local D-optimal designs for  $m = 3$ , as function of  $\rho$  is bounded between 0 and 0.45, irrespective of the number of cohorts  $C$ . Further, for all evaluated sets of parameter values, the relative efficiencies of the local D-optimal designs for  $m = 2$  repeated measures, with respect to the corresponding local D-optimal designs for  $m = 3$  repeated measures, are smaller than 1. For the linear polynomial model, the D-optimal number of repeated measures is equal to  $m = 2$ . The *MMD* is  $\left\{ \begin{pmatrix} -1 & 1 \\ 1 & \end{pmatrix} \right\}$ , with  $MMV = 0.997$ , which is very large.

Table 2.2 gives the *MMD* for the set  $S_1$  and  $S_2$  for the first and second degree polynomial model. It indicates that the loss of efficiency, due to the use of the *MMD* is very small. It appeared that the maximin designs have much larger smallest relative efficiencies than the equally spaced designs.

Table 2.2 Relative efficiency of equally spaced designs and maximin designs

Order polynomial	Equally spaced, $m = 3$ $\left\{ \begin{pmatrix} -1, -a, 1 \\ 0.5 & 0.5 \end{pmatrix} \right\}$		Equally spaced, $m = 4$ $\left\{ \begin{pmatrix} -1, -a, a, 1 \\ 1 \end{pmatrix} \right\}$	
First order, $p = 2$	0.94	0.97 ( $a = 0.55$ )	0.93	0.98 ( $a = 0.57$ )
Second order, $p = 3$	0.997	0.997 ( $a = 0$ )	0.96	0.997 ( $a = 0.23$ )
First or second order, $p = 2$ or 3	0.94	0.95 ( $a = 0.28$ )	0.93	0.94 ( $a = 0.39$ )

\* Maximin designs correspond to the value of  $a$  given within brackets

In this chapter the *D*-optimality criterion is extended by the use of a maximin criterion based on the relative efficiency. A topic for further research may be to construct highly efficient designs with respect to other optimality criteria. Finally, the optimization problem was restricted to the situation of no covariates. The model may be extended to polynomial models with time independent and time dependent covariates.

### Appendix

**Theorem.** Given two polynomial random slope models  $M_1$  and  $M_2$  implied by equation (2.1) of the same order  $p - 1$ . Suppose  $\rho \in [0, 1]$  and  $\sigma^2$  are known finite positive values. Let the variance-covariance matrix of the random effects under  $M_1$  be equal to

$D_1 = \begin{pmatrix} D_{11} & D_{12} \\ D_{12} & D_{22} \end{pmatrix}$  and under  $M_2$  be equal to  $D_2 = \begin{pmatrix} D_{11} & -D_{12} \\ -D_{12} & D_{22} \end{pmatrix}$ , for finite values of  $D_{11}$ ,  $D_{12}$  and  $D_{22}$ , such that  $D_1$  and  $D_2$  are positive definite. Let  $\tau \in T_{Cm}$  be such that model  $M_1$  is identified under  $\tau$ , where  $C$  and  $m$  are finite numbers. Further, let  $\text{Det}[\text{Var}(\hat{\beta}_\tau) | M_1]$  and  $\text{Det}[\text{Var}(\hat{\beta}_{-\tau}) | M_2]$  be the generalized variance under design  $\tau$  and model  $M_1$  and the finite generalized variance under design  $-\tau$  and model  $M_2$ , respectively. Then

$$\text{Det}[\text{Var}(\hat{\beta}_\tau) | M_1] = \text{Det}[\text{Var}(\hat{\beta}_{-\tau}) | M_2]. \quad (2.6)$$

**Proof:** It will be shown after the proof of the Theorem that for all  $i, i = 1, \dots, C$ ,

$(Z_{i_\tau} D_1 Z_{i_\tau}' + \Psi_{i_\tau}) = (Z_{i_{-\tau}} D_2 Z_{i_{-\tau}}' + \Psi_{i_{-\tau}})$ . Let  $A$  be the diagonal  $p \times p$  matrix for which  $A_{ii} = 1$ , if  $i$  is odd, and  $A_{ii} = -1$ , if  $i$  is even. Then  $X_{i_\tau} = X_{i_{-\tau}} A$ . Further

$|\text{Det}(A^{-1})| = 1$ . Hence,

$$\text{Det} \left( \left\{ \sum_{i=1}^c [X_{i_\tau}' (Z_{i_\tau} D_1 Z_{i_\tau}' + \Psi_{i_\tau})^{-1} X_{i_\tau}] \right\}^{-1} \right) =$$

$$\text{Det}(A^{-1}) \text{Det} \left\{ \sum_{i=1}^c [X_{i_{-\tau}}' (Z_{i_{-\tau}} D_2 Z_{i_{-\tau}}' + \Psi_{i_{-\tau}})^{-1} X_{i_{-\tau}}] \right\}^{-1} \text{Det}(A^{-1}) =$$

$$\text{Det} \left\{ \sum_{i=1}^c [X_{i_{-\tau}}' (Z_{i_{-\tau}} D_2 Z_{i_{-\tau}}' + \Psi_{i_{-\tau}})^{-1} X_{i_{-\tau}}] \right\}^{-1}.$$

q.e.d.

For all  $i, i = 1, \dots, C ( Z_{i_\tau} D_1 Z_{i_\tau}' + \Psi_{i_\tau} ) = ( Z_{i_{-\tau}} D_2 Z_{i_{-\tau}}' + \Psi_{i_{-\tau}} )$

**Proof:** Let  $\tau \in T_{CM}$ . Then  $Z_{i_{-\tau}} = Z_{i_\tau} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$

$$D_2 = \begin{pmatrix} D_{11} & -D_{12} \\ -D_{12} & D_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} D_{11} & D_{12} \\ D_{12} & D_{22} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

$$Z_{i_{-\tau}} D_2 Z_{i_{-\tau}}' = Z_{i_\tau} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} D_{11} & D_{12} \\ D_{12} & D_{22} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} Z_{i_\tau}' =$$

$$Z_{i_\tau} D_1 Z_{i_\tau}'$$

Moreover  $\Psi_{i_\tau} = \Psi_{i_{-\tau}}$  for all  $i$ .

q.e.d.

**Corollary.** Let  $T_m = \bigcup_C T_{Cm}$ . For each design  $\tau \in T_m$  there exists a symmetric design  $\tau_s \in T_m$ , such that  $\min_{\Theta} \text{eff}(\tau_s) \geq \min_{\Theta} \text{eff}(\tau)$ .

**Proof:** Let  $\tau \in T_m$  be a design. The Theorem states that the relative efficiency of  $\tau$  under

$D = \begin{pmatrix} D_{11} & D_{12} \\ D_{12} & D_{22} \end{pmatrix}$  is equal to the relative efficiency of  $-\tau$  under  $D = \begin{pmatrix} D_{11} & -D_{12} \\ -D_{12} & D_{22} \end{pmatrix}$ , ceteris paribus. Obviously, the domain of the correlation between the random slope and random

intercept, i.e. the domain of  $\frac{D_{12}}{\sqrt{D_{11} D_{22}}}$ , is equal to  $[-1, 1]$ . Consequently, the smallest relative efficiency of  $\tau$  over the model parameter space is equal to the smallest relative efficiency of  $-\tau$  over the model parameter space. Let the design  $(\tau \oplus -\tau)$  be the design for which half of the persons is measured according to  $\tau$  and half of the persons is measured

according to  $-\tau$ . Note that this design is symmetric and an element of  $T_m$ . It can be proven that the information obtained from the design  $(\tau \oplus -\tau)$  is larger than or equal to the minimum of the information obtained from  $\tau$  and the information obtained from  $-\tau$  as follows.

Rao (1973, p.70) states that if  $A$  and  $B$  are real, positive definite and of order  $n$ , then  $\text{Det}(A+B)^{1/n} \geq \text{Det}(A)^{1/n} + \text{Det}(B)^{1/n}$ . Obviously, the right-hand expression is larger than or equal to twice the minimum of  $\text{Det}(A)^{1/n}$  and  $\text{Det}(B)^{1/n}$ . Let  $I$  be the Fisher information of the regression parameters under design  $\tau \in T_m$ . Filling in  $A = \frac{1}{2} I_\tau$  and  $B = \frac{1}{2} I_{-\tau}$ , it follows that the information obtained from the design  $(\tau \oplus -\tau)$  is larger than or equal to the minimum of the information obtained from  $\tau$  and the information obtained from  $-\tau$ . Because this holds for all model parameter vectors, the smallest relative efficiency for the design  $(\tau \oplus -\tau)$  is larger than the minimum of the smallest relative efficiency for the design  $\tau$  and the smallest relative efficiency of the design  $-\tau$ .

q.e.d.

## References

- Atkins, J.E., and Cheng, C.S. (1999). Optimal regression designs in the presence of random block effects, *Journal of Statistical Planning and Inference*, 77, 321-335.
- Berger, M.P.F., and Tan, F.E.S. (1998). Optimal Designs for Repeated Measurement Designs. *Kwantitatieve Methoden*, 59, 45-68.
- Bischoff, W. (1993). On D-optimal designs for linear models under correlated observations with an application to a linear model with multiple response, *Journal of Statistical Planning and Inference*, 37, 69-80.
- Bunke, H. and Bunke, O. (1986). *Statistical Inference in Linear Models*, New York: Wiley.
- Hungerford, T.W. (1990). *Abstract Algebra*, Orlando: Holt, Rinehart and Winston, Inc.
- Mentré, F., Mallet, A. and Baccar, D. (1997). Optimal design in random effect regression models, *Biometrika*, 84, 429-442.
- Pukelsheim, F. (1993). *Optimal Design of Experiments*. New York: Wiley.

- Rao, C.R. (1973). *Linear Statistical Inference and Its Applications*, New York: Wiley.
- Snijders, T.A.B. and Bosker, R.J. (1999). *Multilevel Analysis*, Londen: SAGE Publications.
- Tan, F.E.S. and Berger, M.P.F. (1999). Optimal allocation of time points for the random effects model, *Communications in Statistics, Simulations and Computations*, 28, 517-540.



# **3 Maximin designs in Generalized Linear Mixed Models with covariates**

*Submitted for publication*

### Abstract

A maximin criterion is used to find optimal designs for the logistic random intercept model with discrete covariates. The proposed maximin criterion maximizes the smallest possible relative efficiency not only with respect to all possible values of the model parameters, but also with respect to the unknown conditional joint distribution of the covariates in the model. We show that the maximin design is balanced with respect to the joint distribution of all, experimentally controlled, independent discrete variables in the model. The proposed method will be used to plan a (stratified) clinical trial where covariates are involved.

*Key words:* relative efficiency, maximin design, logistic random effects model, covariates, D-optimal design

### 3.1 Introduction

The search for D-optimal designs has received increasing attention by planners of experimental designs in the last two decades. The main objective is to establish which design has to be chosen, such that the regression parameters of a specified regression model can be estimated most efficiently. For the fixed effects longitudinal logistic model, the optimal allocation and selection of time points is discussed by Abdelbasit and Plackett (1983), Finney (1989) and Meeker and Hahn (1977), among others. For the linear random effects model, optimal designs are discussed by Atkins and Cheng (1999), Berger and Tan (1998), Bischoff (1993), Mentré, Mallet and Baccar (1997) and Tan and Berger (1999).

The D-optimal designs for the logistic random effects model are only locally optimal, i.e. are only optimal for a given set of parameter values. Three methods are proposed in the literature to handle the problem of local optimality, namely the sequential design procedure, the Bayesian method and the maximin method (Atkinson and Donev, 1996). In this chapter, we will concentrate on the maximin method. This method has been advocated by several authors, e.g. Berger, King and Wong (2000), King and Wong (2000), Ouwens, Tan and Berger (2001), Pronzato and Walter (1988) and Wong (1992). The maximin method considered in literature is based on the efficiency. It is, however, preferable to base this method on the relative efficiency.

In this chapter a class of logistic models is considered with random intercept and time-independent discrete variables, which may have interaction with time. For these models, the designs not only consist of the time points at which the measurements are taken, but also of the joint distribution of the discrete independent variables. Unfortunately, it may not be possible to specify the joint distribution of all variables prior to data sampling. In this chapter it is assumed that a marginal joint distribution of some of the independent variables is experimentally controlled, where is marginalised over the strata of the other independent variables. Throughout this chapter, we will denote an independent variable corresponding to the marginal joint distribution as a 'variate'. An independent variable over which is marginalised is called a 'covariate'. A maximin

criterion is proposed that deals with the problem of local optimality, due to both the dependency on the unknown model parameters, as well as to the dependency on the unknown conditional joint distribution of the covariates.

The logistic random effects model is specified in section 3.2. The D-optimality criterion and the maximin criterion will be introduced in section 3.3. Some characteristics of maximin designs for the logistic random intercept model are formulated in section 3.4. In section 3.5, the domain of the parameter values will be introduced. The optimization algorithm is given in section 3.6. In section 3.7, maximin designs are given for some trials with covariates. Finally, the results will be discussed in section 3.8.

### 3.2 Model specification

In this section the logistic random effects model is introduced. The  $1 \times m$  vector  $\mathbf{y}_{is}' = (y_{is1}, \dots, y_{ism})$  is the vector of  $m$  responses of subject  $s$ ,  $s = 1, \dots, n_i$ , in cohort  $i$ ,  $i = 1, \dots, c$ . Conditional on all fixed and random effects in the model, the responses  $y_{isk}$ ,  $k = 1, \dots, m$ , are independently Bernoulli distributed with probabilities  $\pi_{isk}$ , i.e.  $y_{isk} \sim \text{Bernoulli}(1, \pi_{isk})$ . These probabilities are related to the average profile  $X_{isk}'\beta$  and the subject-specific deviations  $Z_{isk}'\mathbf{b}_{is}$  from this average profile as follows:

$$\text{logit}(\pi_{isk}) = \ln \left( \frac{\pi_{isk}}{1 - \pi_{isk}} \right) = X_{isk}'\beta + Z_{isk}'\mathbf{b}_{is}, \quad (3.1)$$

where the  $p \times 1$  vector  $\beta$  is a vector of fixed regression coefficients and the  $q \times 1$  vector  $\mathbf{b}_{is}$  is a vector of random regression coefficients, which is normally distributed with mean zero and variance-covariance matrix  $D$ .

The Maximin procedure proposed in this chapter can be applied to the larger class of logistic models in equation (3.1) with more random effects, higher order polynomials and an arbitrary number of independent variables. Hence, the discussion with respect to the experimental design space, D-optimality- and maximin criterion will be based on the general class of models represented by equation (3.1). We will, however, illustrate the

results in this chapter by the model:

$$\text{logit}(\pi_{isk}) = \beta_0 + \beta_1 t_{ik} + \beta_2 E_{is} + \beta_3 E_{is} t_{ik} + \beta_4 C_{is} + b_{is}, \quad (3.2)$$

where  $\beta_0$  is the intercept,  $t_{ik}$  is the  $k$ -th time measurement of the subjects in cohort  $i$ , the parameters  $\beta_1, \dots, \beta_4$  are fixed regression parameters and  $E_{is}$  and  $C_{is}$  are the values of the dichotomous variables  $E$  and  $C$ , respectively, for subject  $s$  in cohort  $i$ . The time dependent effect of  $E$  is  $(\beta_2 + \beta_3 t_{ik})$ , and  $b_{is}$  is the random intercept, which is assumed to be normally distributed with mean zero and variance  $\delta^2$ .

As a motivating example, consider the clinical trial on ‘epileptic seizures’ described by Diggle, Liang, Zeger (1994, p. 13-15). In this study, the occurrence of epileptic seizures of each patient was recorded during a baseline period before randomization. Patients were then randomized to treatment with the anti-epileptic drug progabide, or to placebo in addition to standard therapy. The dependent variable is the occurrence of a seizure, which were measured repeatedly in time. It appears that there is a strong variation among the patients in the baseline level of seizures, which seems to persist across time. For this example,  $\beta_0$  is the baseline odds of a seizure. The dichotomous variable  $E$  is coded 0/1 and indicates the treatment group ( $E = 1$ ) versus the control group ( $E = 0$ ). Hence, the coefficient  $\beta_1$  is the log ratio of seizure rates between adjacent time points for the control group and the time-dependent coefficient  $(\beta_2 + \beta_3 t_{ik})$  of  $E$  is proportional to the log ratio of seizure rates between the two treatment groups as a function of time. The variable  $C$  could be any dichotomous independent variable, and  $b_{is}$  denotes the deviation of the patient specific baseline seizure count from  $\beta_0$ . In the study at hand, all patients are measured at the same time points, i.e. there is only one cohort ( $i = 1$ ). A total of 31 out of  $n_i = 59$  patients are assigned to the experimental group.

The purpose of this chapter is to establish which design should be chosen, such that all regression parameters can be estimated most efficiently. That is, at which time points the subjects should be measured, and how the joint distribution of  $E$  and  $C$  should be

chosen. When planning an experimental design, some variables may have a known marginal distribution prior to data sampling. In the ‘epileptic seizures’ study, for example, one is able to choose what marginal percentage of the patients should be treated with progabide, where is marginalised over the independent variable  $C$ . In other words, the marginal distribution of the treatment variable  $E$  can be chosen by design. The joint distribution of all independent variables, however, may not be determined in advance. It does not generally make sense to search for an optimal joint distribution of all independent variables. We will denote the independent variables corresponding to the controlled marginal joint distribution as variables which are kept under control, while the independent variables which do not correspond with the controlled marginal joint distribution are denoted as variables which are not kept under control.

To define the experimental design space, consider the case that there are  $q < p$  discrete independent variables (time variables are not included). Suppose that  $S_c = \{X_1, \dots, X_r\}$  is the set of discrete variates, which is kept under control, and let  $S_{nc} = \{X_{r+1}, \dots, X_q\}$  be the set of discrete covariates, which is not kept under control, with  $r \leq q$ . Then the marginal joint distribution of the independent variables in the set  $S_c$  and the time is represented by Table 3.1.

In Table 3.1,  $w_1$  through  $w_c$  are the relative sizes of the  $c$  cohorts. Within each cohort  $i$ ,  $i = 1, \dots, c$ , the subjects are measured at the same time points  $t_{i1}$  through  $t_{im}$ . For each cohort  $i$ ,  $i = 1, \dots, c$ , the number of strata of the cross-classification indicated by the set  $S_c$  is denoted by  $z$  and the corresponding weights are denoted by  $w_{ij}$ . Note that the cross-classification is made marginally over the set  $S_{nc}$  of covariates.

Table 3.1 Experimental design of a longitudinal model with  $c$  cohorts, maximal  $z$  strata within the  $i$ -th cohort and  $m$  repeated measures

Cohort	Weight $w_i$	Time points	Stratum	Weight $w_{ij}$
1	$w_1$	$t_{11}, \dots, t_{1m}$	1	$w_{11}$
			2	$w_{12}$
			$\vdots$	$\vdots$
			$z$	$w_{1z}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$c$	$w_c$	$t_{c1}, \dots, t_{cm}$	1	$w_{c1}$
			2	$w_{c2}$
			$\vdots$	$\vdots$
			$z$	$w_{cz}$

The vector  $(w_{i1}, \dots, w_{iz})'$  of relative sizes of the  $c$  cohorts is an element of the space  $W_c = \{ (w_{i1}, \dots, w_{iz})' : \sum_{i=1}^c w_i = 1, w_i \geq 0, i = 1, \dots, c \}$ . For each cohort  $i$ ,  $i = 1, \dots, c$ , the vector  $(w_{i1}, \dots, w_{iz})'$  of relative sizes of the  $z$  strata indicated by  $S_c$  is an element of  $W_i = \{ (w_{i1}, \dots, w_{iz})' : \sum_{j=1}^z w_{ij} = 1, w_{ij} \geq 0, j = 1, \dots, z \}$  and the time vector  $(t_{i1}, \dots, t_{im})'$  is an element of  $T_{im} = \{ (t_{i1}, \dots, t_{im})' : t_{ik} \in [-1, 1] \}$ . Note that the number of strata  $z$  should be interpreted as the maximal number of strata within the  $c$  cohorts, because some of the  $w_{ij}$ 's may be zero. Based on the previous definitions, the experimental design space  $T$  is defined as

$$T = \bigcup_{c, m} \left[ W_c \times \prod_{i=1}^c (W_i \times T_{im}) \right], \tag{3.3}$$

Note that the maximal number of strata  $z$  is determined by the choice of the set  $S_c$  and remains constant during the search for the maximin design. In this chapter we assume that the maximal number of cohorts and time points are finite to ensure compactness of the design space  $T$ .

### 3.3 Optimality criteria

In this section the D-optimality criterion and the maximin criterion are introduced, which will be used to compare experimental designs of the form mentioned above.

#### *D-optimality criterion and relative efficiency*

Let  $\hat{\beta}$  be the maximum likelihood estimator of  $\beta$ . Then the volume of the confidence ellipsoid of  $\hat{\beta}$  is an increasing function of the determinant  $\text{Det}[\text{Var}(\hat{\beta})]$  of the variance of  $\hat{\beta}$ , i.e. of the generalized variance of  $\hat{\beta}$ . It should be noted that the determinant  $\text{Det}[\text{Var}(\hat{\beta})]$  depends on the joint distribution of all independent variables. Hence, for each design  $\tau \in T$ , the determinant  $\text{Det}[\text{Var}(\hat{\beta})]$  depends not only on the values of the model parameters, but also on the conditional (on the cohorts and  $S_c$ ) joint distribution of the set of covariates  $S_{nc}$ , which is not kept under control. For each vector of model parameters and conditional joint distribution of the covariates in  $S_{nc}$ , the  $\text{Det}[\text{Var}(\hat{\beta})]$  can be minimized by the choice of a D-optimal design  $\tau^* \in T$ , such that for each design  $\tau \in T$

$$\text{Det}[\text{Var}(\hat{\beta}_{\tau^*})] \leq \text{Det}[\text{Var}(\hat{\beta}_{\tau})], \quad (3.4)$$

where  $\hat{\beta}_{\tau^*}$  and  $\hat{\beta}_{\tau}$  are the estimates of  $\beta$  under  $\tau^*$  and  $\tau$ , respectively. The D-optimal design is only locally D-optimal, because the determinant depends on the unknown values of the model parameters and the conditional joint distribution of  $S_{nc}$ .

The relative efficiency can be used to compare the generalized variances of two experimental designs. The relative efficiency of a design  $\tau_1 \in T$  with respect to  $\tau_2 \in T$ , conditional on the values of the model parameters and the conditional joint distribution of  $S_{nc}$ , is given by

$$\text{eff}(\tau_1 | \tau_2) = \left( \frac{\text{Det}[\text{Var}(\hat{\beta}_{\tau_2})]}{\text{Det}[\text{Var}(\hat{\beta}_{\tau_1})]} \right)^{1/p}, \quad (3.5)$$

where  $p$  is equal to the number of fixed regression coefficients. Expression (3.5) can be interpreted as follows. If, for example,  $\text{eff}(\tau_1 | \tau_2) = 0.9$  for a particular design  $\tau_1$ , then about 10% more observations will be needed to obtain the same efficiency as that of design  $\tau_2$  for the same model.

The D-optimal design for model (3.1) depends on the values of the model parameters and the conditional joint distribution of  $S_{nc}$ . An approach to deal with this local optimality is the use of a maximin criterion.

#### *Maximin criterion.*

Both the model parameter vector as well as the conditional joint distribution of the covariates in  $S_{nc}$  must be taken into account. A maximin design  $\tau^{\text{maximin}} \in T$  with respect to the D-optimality criterion is therefore defined to be a design that maximizes, among all designs  $\tau \in T$ , the smallest relative efficiency over both the space  $\Omega$  of model parameter vectors as well as the space  $\mathcal{W}_{\text{cond}}$  of conditional joint distributions of the covariates in  $S_{nc}$ , i.e. the maximin design  $\tau^{\text{maximin}}$  maximizes

$$\min_{\mathcal{W}_{\text{cond}} \Omega} \text{eff}(\tau | \tau^*) \quad (3.6)$$

over the experimental design space  $T$ , where  $\tau^*$  is the local D-optimal design. The corresponding value of expression (3.6) is called the maximin value of  $\tau^{\text{maximin}}$ . In the next section we will give the balancedness characteristic of maximin designs.

### 3.4 Balancedness of maximin designs

The discrete independent variables can, without loss of generality, be assumed to be represented by dummy variables with values 0 and 1. The time period can, without loss of generality, be assumed to be  $[-1, 1]$  (Bunke and Bunke, 1986, theorem 8.34).

Although maximin designs differ per model and can often only be found by numerical computations, some characteristics can be proven analytically. One of these

characteristics is balancedness. A design is called balanced in the strata of a set of independent variables, if the corresponding relative sizes are the same.

**Theorem.** Consider model (3.1). Let  $S_c$  be a set of variates, which is kept under control. The design space  $T$  is defined as in expression (3.3). If the model parameter space is symmetric around 0, then there exists a maximin design which is balanced in  $S_c$ .

The proof of the Theorem is given in the appendix. In the following, we have applied the maximin criterion and the Theorem to the longitudinal model with two additional dichotomous independent variables as specified by model (3.2). Before we proceed with the presentation of the results of the maximin designs and corresponding maximin values for model (3.2), we will restrict the parameter space, because large ranges of the model parameters will evidently lead to relative small maximin values.

### 3.5 Parameter space

We have reduced the parameter space to values which are generally relevant in practice. Each model parameter vector in  $\Omega$  consists of a vector of fixed regression parameters  $\beta$  and a variance of the random intercept  $\delta^2$ . We choose the subspace of  $\Omega$  defined as

$$\Theta = \{ (\beta, \delta^2) \in \Omega : \beta_0 \in [-10, 10], \beta_h \in [-1, 1], h = 1, \dots, 4, \delta^2 \in [0.01, 3] \}.$$

It is additionally required that, for each stratum of  $\{E, C\}$ , the intercept of model (3.2) is an element of  $[-10, 10]$  and the effects of time  $\beta_1$  and  $\beta_1 + \beta_3$  are elements of  $[-1, 1]$ . Note that if  $\beta_1 = 1$  and the research period is  $[-1, 1]$ , then the odds at the end of the research period is  $\exp(2) \approx 7$  times the odds at the start of the research period. The range of  $\beta_0$  is much larger than the ranges of the other parameters, because the probability of getting a disease can be very small, while the change in the probability is not expected to

be large in absolute sense.

### 3.6 Optimization Algorithm

A consequence of the Theorem in section 3.4 is that the search for maximin designs may be restricted to designs which are balanced in all variates that are kept under control. The relative efficiency of each design  $\tau \in T$  is determined relatively to the D-optimal design. Because D-optimal designs do not need to be symmetric, nor balanced, the calculation of the relative efficiencies and maximin values were performed with respect to the whole design space  $T$ .

The following global search algorithm was used to obtain maximin optimal designs for this chapter:

1. take an initial guess of the balanced maximin design,
2. compute the 'maximin solution' based on the local search algorithm BFGS (for maximization), taking into account the minimization of the relative efficiency over  $W_{cond}$  and  $\Theta$  (BFGS is used within the maximization algorithm to obtain the smallest relative efficiency over  $W_{cond}$  and  $\Theta$ ),

Check 1:

3. select randomly balanced design (Use the uniform distribution for selection),
4. do step 2 + compare with earlier computed designs,

Check 2:

5. select balanced designs from the boundary of the design space,
6. do step 2 + compare with earlier computed designs,

Check 3:

7. do step 3 ,
8. compute corresponding smallest relative efficiency and compare with earlier computed designs.

Check 1 and 2 are repeated 10 times and check 3 is repeated 1000 times. The quasi-Newton algorithm BFGS (Broyden-Fletcher-Goldfarb-Shanno) is described at <http://orion.math.uwaterloo.ca/~hwolkowi/mirror.d/glossary/B.html> and is default in TOMLAB (<http://www.mathtools.com/files/tomlab//tomlabv2.pdf>), the optimization toolbox for MATLAB.

### 3.7 Results

The maximin value depends on the independent variables in the set  $S_{nc}$ . Hence, to plan an experimental design for models with covariates, it is necessary to decide in advance, which independent variables will be kept under control (to form the set  $S_c$ ), and which independent variables will not (to form the set  $S_{nc}$ ). In this section maximin designs and maximin values are presented for model (3.2) for the following three cases. The trial on epileptic seizures is used as an example for the first two cases.

*Case 1. The variates  $E$  and  $C$  are both kept under control.*

Suppose that we want to plan a stratified trial, where a specified number of smokers as well as non-smokers will be assigned to the treatment (progabide) and placebo group. Here we have a situation, in which the treatment variable  $E$  as well as the dichotomous Smoking covariate  $C$  are kept under control. The set  $S_c$  of variates, which is kept under control is  $S_c = \{E, C\}$ . The set of covariates, which is not kept under control is empty, i.e.  $S_{nc} = \emptyset$ . According to the Theorem, the maximin design implies a balanced joint distribution of the independent variables in  $S_c$ , i.e. an equal number of patients for each of the strata of  $S_c$ . Furthermore, numerical computations show that the maximin design consists of one cohort with the optimal allocation of time points at -1 and 1 (start and end of the research period). The corresponding maximin value is equal to 0.77.

**Case 2.** *The variate  $E$  is kept under control, but the covariate  $C$  is not.*

A commonly encountered design in many clinical trials, is a design in which only the treatment variate  $E$  and the time are kept under control. A fraction of the patients will then be assigned by design to the progabide group. In this case the set  $S_c$  contains only the variate  $E$ . The set of covariates, which is not kept under control is  $S_{nc} = \{ C \}$ . Thus, the relative efficiency of any design will also depend on the conditional distribution of the covariate  $C$ . According to the Theorem, the maximin design implies that the patients should be equally distributed across the treatment and the control group. Numerical computations show that one cohort is sufficient for the maximin design with the optimal allocation of time points at -1 and 1. The corresponding maximin value is equal to 0.87.

**Case 3.** *Both covariates  $E$  and  $C$  are not kept under control.*

In many observational studies, which involve repeated measures, only the time allocation is planned in advance. Numerical computations show that in this case the maximin design involves one cohort with optimal allocation of time points at -1 and 1. In this case the set  $S_c$  of variates, which is kept under control, is empty, i.e.  $S_c = \emptyset$ . The set of covariates, which is not kept under control is  $S_{nc} = \{ E, C \}$ . The corresponding maximin value is equal to 0.90.

We found, that the relative efficiencies may be small for large values of the regression parameters. The above stated maximin values are calculated for value-combinations that belong to the subspace  $\Theta$  as defined in section 3.5. A further restriction of the parameter space will in general lead to larger maximin values. Table 3.2 shows the effect on the maximin values for the three different cases, with decreasing range of the regression parameters.

Table 3.2 Maximin values as a function of the range of regression parameter values  $\beta_h$ ,  $h = 1, \dots, 4$ . The number of cohorts  $i = 1$

Range of parameter values $\beta_h$ , $h = 1, \dots, 4$	Case 1 $S_c = \{E, C\}$ $S_{nc} = \emptyset$	Case 2 $S_c = \{E\}$ $S_{nc} = \{C\}$	Case 3 $S_c = \emptyset$ $S_c = \{E, C\}$
[-1, 1]	0.77	0.87	0.90
[-0.75, 0.75]	0.82	0.89	0.91
[-0.50, 0.50]	0.88	0.91	0.95
[-0.25, 0.25]	0.94	0.94	0.98

### 3.8 Discussion

It has become increasingly important to search for experimental designs, such that the regression parameters can be estimated as precisely as possible and the power of statistical tests is maximized. It is known from literature (e.g. Rothen, 1998) that a design with equally sized groups does not necessarily lead to estimates of the regression parameters with smallest standard errors. Rothen (1998) found that the stratum with the largest variance should have the largest sample size. In contrast to the unbalancedness of the D-optimal designs, the results in this chapter indicate that the maximin designs imply a balanced design with respect to the marginal joint distribution of all experimentally controlled variates in the model. To plan a clinical trial for the comparison of treatment groups, an equal assignment of patients across the treatment groups is optimal in the maximin sense, irrespective of the number of covariates in  $S_{nc}$ . Furthermore, the time points should be allocated at time points -1 and 1. It seems that the allocation of time points is independent of how many independent variables are kept under control. The same conclusion can be drawn for the second order longitudinal model, with optimal allocation at time points -1, 0 and 1 and one cohort. The corresponding maximin values are slightly larger for the second order model than for the first order model.

Table 3.2 shows that the maximin value tends to be smaller as the number of variates, which is kept under control, increases. It should be noted, that the maximin

values between different cases in Table 3.2 cannot be compared to each other. To obtain these maximin values, the efficiencies of the maximin designs are related to the corresponding efficiencies of different D-optimal designs (expression (3.6)). The D-optimal designs of the first case, for example, are more efficient than the D-optimal designs of the second case, because the first case has more experimentally controlled variates. Hence, the maximin value of the maximin design will tend to be smaller, due to the division of an increasing D-optimal efficiency value as the number of experimentally controlled variates increases.

In practice, the variates are used as stratification variables. One might argue that the maximin value of 0.77 for the first case in Table 3.2 may be too small, and that the situation will even be worse when many experimentally controlled variates are encountered. The use of the maximin criterion to obtain an efficient design would then be questionable and not very useful in practice. In many practical situations, however, additional information about the range of the regression parameters is available or can be obtained by pilot studies. In the epileptic seizures trial, for example, the odds ratio of having a seizure for a subject in the control group relative to a subject in the treatment group may be expected to be less than 2. For the regression parameters, a range of parameter values of  $[-0.35, 0.35]$  is therefore very plausible. As Table 3.2 indicates, the maximin value will increase if the range of the regression parameters decreases. Another reason why the maximin criterion is useful in practice when planning a stratified randomized trial, is that pre-stratification is often desirable for only a few variates (cf. Pocock, 1984, page 81).

In conclusion, the maximin criterion for regression models with covariates leads to designs which are balanced in the covariates with large relative efficiencies. We expect that the optimal allocation of time points will be independent of the number of time-independent covariates in model (3.1) and the number of repeated measures. Further research is needed on this topic and on extensions to other generalized linear mixed models, with more random effects, discrete or continuous covariates, and time varying covariates.

## Appendix

**Theorem.** Consider model (3.1). Let  $S_c$  be a set of variates, which is kept under control. The design space  $T$  is defined as in expression (3.3). If the model parameter space is symmetric around 0, then there exists a maximin design which is balanced in  $S_c$ . Recall that a design is balanced in the strata of a set of independent variables, if the corresponding relative sizes are the same. Within this appendix, the dichotomous variables and the dummies are assumed to be coded -1 and 1, without loss of generality.

**Proof:** Note that there is always at least one maximin design  $\tau^{\text{maximin}} \in T$ , due to the continuity of the relative efficiency, and due to the compactness of the design space  $T$ , the space of conditional distributions of the set of covariates  $S_{nc}$ , and the space of model parameters  $\Omega$ . Let therefore  $\tau^{\text{maximin}} \in T$  be a maximin design.

The Theorem is proven in the following two steps:

1. construction of a maximin design  $-\tau^{\text{maximin}} \in T$  based on the maximin design  $\tau^{\text{maximin}}$ ,
2. proof that  $(\tau^{\text{maximin}} \oplus -\tau^{\text{maximin}})$  is a balanced maximin design, where  $(\tau^{\text{maximin}} \oplus -\tau^{\text{maximin}})$  is the design obtained by assigning half of the subjects to  $\tau^{\text{maximin}}$  and half of the subjects to  $-\tau^{\text{maximin}}$ .

*Step (1): Construction of the maximin design  $-\tau^{\text{maximin}}$ .*

We will describe the construction of  $-\tau^{\text{maximin}}$  for the logistic model (3.2) in the setting that  $C$  is a covariate, i.e. the set of variates is  $S_c = \{ E \}$  and the set of covariates is  $S_{nc} = \{ C \}$ . A generalization to other settings of  $S_c$  and  $S_{nc}$  and higher order polynomial logistic models with an arbitrary number of discrete independent covariates and more random effects can be made analogously.

Let for each design  $\tau \in T$ , the design  $-\tau \in T$  be obtained by interchanging, within each cohort  $i, i = 1, \dots, c$ , the relative sizes  $w_{ij}$  of the strata of  $E$ . Then it can be proven that  $-\tau^{\text{maximin}}$  is also maximin as follows.

Suppose that the relative efficiency of  $-\tau^{\text{maximin}}$  attains its minimum for  $w_{\text{cond}} \in \mathcal{W}_{\text{cond}}$  and  $\theta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \delta^2)' \in \Omega$ . Let  $I(\hat{\beta}_{-\tau})|_{w_{\text{cond}}, \theta}$  denote the Fisher information matrix evaluated in  $w_{\text{cond}}$  and  $\theta$  under  $-\tau$ , which is the inverse of  $\text{Var}(\hat{\beta}_{-\tau})$  evaluated in  $w_{\text{cond}}$  and  $\theta$ . Further, let  $w_{\text{cond}}^*$  be obtained from  $w_{\text{cond}}$  by interchanging within each cohort  $i$ ,  $i = 1, \dots, c$ , the conditional distribution of  $C$  over the strata of  $E$  and let  $\theta^* = (\beta_0, \beta_1, -\beta_2, -\beta_3, \beta_4, \delta^2)'$ . Then to prove that  $-\tau^{\text{maximin}}$  is maximin, it is sufficient to prove that the relative efficiency of  $-\tau^{\text{maximin}}$  under  $w_{\text{cond}}$  and  $\theta$  is equal to the relative efficiency of  $\tau^{\text{maximin}}$  under  $w_{\text{cond}}^*$  and  $\theta^*$ , i.e.

$$\left\{ \frac{\text{Det} \left[ I(\hat{\beta}_{-\tau^{\text{maximin}}})|_{w_{\text{cond}}, \theta} \right]}{\text{Det} \left[ I(\hat{\beta}_{-\tau^*})|_{w_{\text{cond}}, \theta} \right]} \right\}^{1/p} = \left\{ \frac{\text{Det} \left[ I(\hat{\beta}_{\tau^{\text{maximin}}})|_{w_{\text{cond}}^*, \theta^*} \right]}{\text{Det} \left[ I(\hat{\beta}_{\tau^*})|_{w_{\text{cond}}^*, \theta^*} \right]} \right\}^{1/p}, \quad (3.7)$$

where  $\tau^* \in T$  is the locally D-optimal design for  $w_{\text{cond}}^*$  and  $\theta^*$  and  $-\tau^*$  will be proven to be a locally D-optimal design for  $w_{\text{cond}}$  and  $\theta$ .

To prove equation (3.7), equation (3.8) will be used:

$$\forall \tau \in T : \text{Det} \left[ I(\hat{\beta}_{-\tau})|_{w_{\text{cond}}, \theta} \right] = \text{Det} \left[ I(\hat{\beta}_{\tau})|_{w_{\text{cond}}^*, \theta^*} \right] \quad (3.8)$$

Proof of equation (3.8): Equation (3.8) can be proven based on the building blocks for the information matrix (cf. Diggle, Liang and Zeger, 1994, p.173) as follows. For each design  $\tau \in T$  the design matrix  $X_{-\tau}$  under  $-\tau$ ,  $w_{\text{cond}}$  and  $\theta$  can be written as  $X_{\tau}A$  under  $\tau$ ,  $w_{\text{cond}}^*$  and  $\theta^*$ , where  $A$  is the diagonal matrix for which  $A_{ll} = 1$ , for  $l = 1, 2$  and  $5$  and  $A_{ll} = -1$ , for  $l = 3$  and  $4$ . Further,  $\delta^2$  is for  $-\tau$ ,  $w_{\text{cond}}$  and  $\theta$  the same as for  $\tau$ ,  $w_{\text{cond}}^*$  and  $\theta^*$ . Finally, for each random intercept value  $b$ , the conditional probability of a response in stratum  $i$  of  $E$  under  $-\tau$ ,  $w_{\text{cond}}$  and  $\theta$  is equal to the conditional probability of a response in stratum  $(2 - i)$  of  $E$  under  $\tau$ ,  $w_{\text{cond}}^*$  and  $\theta^*$ ,  $i = 1, 2$ , because

$$\begin{aligned} \beta_0 + \beta_1 t + \beta_2 E + \beta_3 E t + \beta_4 C + b &= \\ = \beta_0 + \beta_1 t + (-\beta_2)(-E) + (-\beta_3)(-E)t + \beta_4 C + b. \end{aligned}$$

Hence,  $\text{Det}[ I( \hat{\beta}_{-\tau} ) |_{w_{cond}, \theta} ] = \text{Det}[ A' I( \hat{\beta}_{\tau} ) |_{w_{cond}, \theta} A ] = \text{Det}(A') \text{Det}[ I( \hat{\beta}_{\tau} ) |_{w_{cond}, \theta} ] \text{Det}(A) = \text{Det}[ I( \hat{\beta}_{\tau} ) |_{w_{cond}, \theta} ]$ , which proves equation (3.8).

q.e.d.

A consequence of equation (3.8) is that  $-\tau^*$  is D-optimal for  $w_{cond}$  and  $\theta$ , because

$$\text{Det}[ I( \hat{\beta}_{-\tau^*} ) |_{w_{cond}, \theta} ] = \text{Det}[ I( \hat{\beta}_{\tau^*} ) |_{w_{cond}, \theta} ] = \max_T \text{Det}[ I( \hat{\beta}_{\tau} ) |_{w_{cond}, \theta} ] = \max_T \text{Det}[ I( \hat{\beta}_{-\tau} ) |_{w_{cond}, \theta} ]$$

Equation (3.7) follows from equation (3.8), because

- a)  $\text{Det}[ I( \hat{\beta}_{-\tau^{maximin}} ) |_{w_{cond}, \theta} ] = \text{Det}[ I( \hat{\beta}_{\tau^{maximin}} ) |_{w_{cond}, \theta} ]$ .
- b)  $\text{Det}[ I( \hat{\beta}_{-\tau^*} ) |_{w_{cond}, \theta} ] = \text{Det}[ I( \hat{\beta}_{\tau^*} ) |_{w_{cond}, \theta} ]$ .

Because the relative efficiency of  $-\tau^{maximin}$  is minimal for  $w_{cond}$  and  $\theta$ , this relative efficiency is also obtained by  $\tau^{maximin}$  and  $\tau^{maximin}$  is a maximin design, the design  $-\tau^{maximin}$  is also maximin.

*Step (2): proof that  $( \tau^{maximin} \oplus -\tau^{maximin} )$  is a balanced maximin design.*

Since  $\tau^{maximin}$  and  $-\tau^{maximin}$  are based on the same cohorts,  $( \tau^{maximin} \oplus -\tau^{maximin} )$  is also based on the same cohorts as  $\tau^{maximin}$  and is therefore also an element of  $T$ . The balancedness of the design  $( \tau^{maximin} \oplus -\tau^{maximin} )$  follows by the construction of  $-\tau^{maximin}$ . It remains to proof that the design  $( \tau^{maximin} \oplus -\tau^{maximin} )$  is also maximin.

Take a  $w_{cond} \in W_{cond}$  and a  $\theta \in \Omega$ , for which the relative efficiency of  $( \tau^{maximin} \oplus -\tau^{maximin} )$  is smallest. To proof that  $( \tau^{maximin} \oplus -\tau^{maximin} )$  is also maximin, it is sufficient to proof that

$$\left\{ \frac{\text{Det}[ I( \hat{\beta}_{(\tau^{maximin} \oplus -\tau^{maximin})} ) ]}{\text{Det}[ I( \hat{\beta}_{\tau^*} ) ]} \right\}^{1/p} \geq \min \left( \left\{ \frac{\text{Det}[ I( \hat{\beta}_{\tau^{maximin}} ) ]}{\text{Det}[ I( \hat{\beta}_{\tau^*} ) ]} \right\}^{1/p}, \left\{ \frac{\text{Det}[ I( \hat{\beta}_{-\tau^{maximin}} ) ]}{\text{Det}[ I( \hat{\beta}_{\tau^*} ) ]} \right\}^{1/p} \right), \tag{3.9}$$

where  $\tau^*$  is the D-optimal design under  $w_{cond}$  and  $\theta$ . The inequality (3.9) can be proven based on Rao (1973, p.70) as follows.

$$\begin{aligned} & \left\{ \text{Det} \left[ I( \beta_{(\tau^*_{maximin}, -\tau^*_{maximin})} ) \right] \right\}^{1/p} = \\ & \left\{ \text{Det} \left[ \frac{1}{2} I( \beta_{\tau^*_{maximin}} ) + \frac{1}{2} I( \beta_{-\tau^*_{maximin}} ) \right] \right\}^{1/p} \geq \\ & \left\{ \text{Det} \left[ \frac{1}{2} I( \beta_{\tau^*_{maximin}} ) \right] \right\}^{1/p} + \left\{ \text{Det} \left[ \frac{1}{2} I( \beta_{-\tau^*_{maximin}} ) \right] \right\}^{1/p} = \\ & \frac{1}{2} \left\{ \text{Det} \left[ I( \beta_{\tau^*_{maximin}} ) \right] \right\}^{1/p} + \frac{1}{2} \left\{ \text{Det} \left[ I( \beta_{-\tau^*_{maximin}} ) \right] \right\}^{1/p} \geq \\ & \min \left( \left\{ \text{Det} \left[ I( \beta_{\tau^*_{maximin}} ) \right] \right\}^{1/p}, \left\{ \text{Det} \left[ I( \beta_{-\tau^*_{maximin}} ) \right] \right\}^{1/p} \right), \text{ which proofs} \end{aligned}$$

inequality (3.9).

q.e.d.

### References.

- Abdelbasit, K.M., and Plackett, R.L. (1983). Experimental design for binary data. *Journal of the American Statistical Association*, 78, 90-98.
- Atkins, J.E., and Cheng, C.S. (1999). Optimal regression designs in the presence of random block effects. *Journal of Statistical Planning and Inference*, 77, 321-335.
- Atkinson, A.C., and Donev, A.N. (1996). *Optimum experimental designs*. Oxford: Clarendon Press.
- Berger, M.P.F., King, C.Y.J., and Wong, W.K. (2000). Minimax D-optimal designs for item response theory models, *Psychometrika*, 65, 377-390.
- Berger, M.P.F., and Tan, F.E.S. (1998). Optimal Designs for Repeated Measurement Designs. *Kwantitatieve Methoden*, 59, 45-68.

- Bischoff, W. (1993). On D-optimal designs for linear models under correlated observations with an application to a linear model with multiple response. *Journal of Statistical Planning and Inference*, 37, 69-80.
- Bunke, H., and Bunke, O. (1986). *Statistical inference in linear models*. New York: Wiley.
- Diggle, P.J., Liang, K.Y., and Zeger, S.L. (1994). *Analysis of longitudinal data*. Oxford: Clarendon Press.
- Finney, D.J. (1978). *Statistical methods in biological assay*. New York: Macmillan.
- King, J., and Wong, W.K. (2000). Minimax D-optimal designs for the logistic model. *Biometrics*, 56, 4, *In press*.
- Meeker, W.Q., and Hahn, G.J. (1977). Asymptotically optimum over stress tests to estimate the survival probability at a condition with low expected failure probability. *Technometrics*, 19, 381-399.
- Mentré, F., Mallet, A., and Baccar, D. (1997). Optimal design in random effect regression models. *Biometrika*, 84, 429-442.
- Ouwens, M.J.N.M., Tan, F.E.S., and Berger, M.P.F. (2001). *D-optimal and Maximin designs for the Linear Mixed Model*. Internal report.
- Pocock, S.J. (1984). *Clinical trials. A practical approach*. New York: Wiley.
- Pronzato, L., and Walter, E. (1988). Robust experiment designs via maximin optimization. *Mathematical biosciences*, 89, 161-176.
- Rao, C.R. (1973). *Linear statistical inference and its applications*. New York: Wiley.
- Rochen, J. (1998). Application GEE procedures for sample size calculations in repeated measures experiments. *Statistics in Medicine*, 1643-1658.
- Tan, F.E.S., and Berger, M.P.F. (1999). Optimal allocation of time points for the random effects model. *Communications in Statistics, Simulations and Computations*, 28, 517-540.
- Wong, W.K. (1992). A unified approach to the construction of minimax designs. *Biometrika*, 79, 611-619.

# 4 **Detection of influential data in Linear Mixed Models**

BRUCE L. RANSFORD

Department of Statistics, University of Canterbury, Christchurch, New Zealand

*Journal of Royal Statistical Society D*, 2001, 50 (3), 271-284.

### Abstract

Mixed effects models for longitudinal data with fixed as well as random parameters are often used to describe expected profiles. Influence measures are usually constructed to detect influential subjects and observations for the fixed regression parameters, treating the subject specific parameters as nuisance parameters. One of these measures is the well-known Cook's Distance. It is shown that this statistic may fail to detect or may incorrectly detect influential observations due to the random effect variances and covariances. A conditional version of Cook's Distance is proposed to assess the influence of observations on the estimated regression parameters.

*Key words:* Conditional Cook's Distance, Cook's Distance, influential observations, Random effects

## 4.1 Introduction

In the last two decades many references on longitudinal models with random effects have appeared (Diggle, Liang and Zeger, 1994, Lesaffre & Verbeke, 1998, and Vonesh and Carter, 1987, among others). Despite this growing interest, the development of diagnostic methods remains a somewhat neglected topic. This paper discusses the problem of identifying an observation with a demonstrably large effect on estimated regression parameters. Such observations were called **influential observations** by Belsley, Kuh and Welsch (1980). Chatterjee and Hadi (1986) reviewed several influence measures for linear regression models with uncorrelated error terms. Each of them stresses different aspects of influence on the values of the estimates. These influence measures differ in their intention. Some detect influential observations for the estimated regression parameters, some evaluate the influence on a linear combination of these regression parameters and some focus on the estimated variances of the estimated regression parameters. Extensions to multivariate regression have been suggested by Barrett and Ling (1992), Christensen, Pearson and Johnson, (1992), Cook (1986), De Gruttola, Ware and Louis (1987), Rohlf (1975) and Siontani (1959), among others. One of these measures is Cook's Distance. Although it seems natural to use Cook's Distance as an influence measure for mixed effects regression models, some caution is needed. Christensen et al. (1992) proposed a two-step method to detect influential observations. In the first step a diagnostic tool is used to evaluate the influence on the estimates of the variance components. In the second step they used an extension of Cook's Distance. Banerjee (1998) noticed that the effectiveness of Cook's Distance as an influence measure in the longitudinal data setting is limited. Banerjee and Frees (1997) have applied the concept of partial influence to take into account the effects on subject-specific parameters and to measure the effect of a subject on the population parameters. Ouwens, Tan and Berger (1999, 2000, 2001) demonstrated the necessity to use observation-oriented influence measures in addition to subject-oriented influence measures. They showed that existing subject-oriented influence measures may fail to detect influential subjects, owing to the

relative position of the observations within and across subjects.

In this chapter a conditional version of Cook's Distance is proposed by conditioning on the subjects in the sample. In section 2 the data from the London Growth Study (Tanner, Whitehouse, Marubini and Resele, 1976) are described and the problem of detecting influential observations is motivated. In section 3 the underlying regression model is specified. In section 4 Cook's Distance is investigated in detail. Examples will be given to demonstrate its shortcomings for mixed effects models. A conditional version is proposed in section 5 and some numerical examples are given that show the similarities and discrepancies between the two representations of Cook's Distance. Some general conclusions will also be discussed in detail. Finally, the data of the London growth study will be analysed for influential observations in section 6.

## **4.2 The London Growth Study: Data description and problem formulation**

A more elaborate description and analysis of the London Growth Study can be found in Tanner et al. (1976). Part of this study is known as the Harpenden Growth Study and consists of several hundreds of boys and girls in a children's home in the country just outside London between 1948 and 1972. Each child was observed twice a year until the first signs of puberty, followed by measurements every 3 months until the growth spurt ended, and then each year until they were 20 years old and finally every 5 years. Tanner et al. (1976) analysed 55 boys and 35 girls whose measurements were made regularly until they reached adulthood. Goldstein (1979) selected growth data of 20 pre-adolescent girls between 6 and 10 years old. One of the purposes of the London Growth Study was to describe the mean growth of the girls who had mothers of different height: short mothers ( $< 155$  cm), medium mothers ( $155 - 164$  cm) and tall mothers ( $> 164$  cm). Table 4.1 shows the relationship between height and age of the girls, for the three different groups of mothers. Table 4.1 can also be found in Goldstein (1979).

Table 4.1 London growth study data

	Height ( cm ) at the ages ( years ):				
	6	7	8	9	10
Short mother					
1	111.0	116.4	121.7	126.3	130.5
2	110.0	115.8	121.5	126.6	131.4
3	113.7	119.7	125.3	130.1	136.0
4	114.0	118.9	124.6	129.0	134.0
5	114.5	112.0	126.4	131.2	135.0
6	112.0	117.3	124.4	129.2	135.2
Mean	112.5	116.7	124.0	129.2	135.2
Medium-sized mother					
7	116.0	122.0	126.6	132.6	137.6
8	117.6	123.2	129.3	134.5	138.9
9	121.0	127.3	134.5	139.9	145.4
10	114.5	119.0	124.0	130.0	135.1
11	117.4	123.2	129.5	134.5	140.0
12	113.7	119.7	125.3	130.1	135.9
13	113.6	119.1	124.8	130.8	136.3
Mean	116.2	121.9	127.9	133.2	138.5
Tall mother					
14	120.4	125.0	132.0	136.6	140.7
15	120.2	128.5	134.6	141.0	146.5
16	118.9	125.6	132.1	139.1	144.0
17	120.7	126.7	133.8	140.7	146.0
18	121.0	128.1	134.3	140.3	144.0
19	115.9	121.3	127.4	135.1	141.1
20	125.1	131.8	141.3	146.8	152.3
Mean	120.3	126.7	133.6	139.9	144.9

An inspection of the data reveals that the second observation of the fifth girl is perhaps a transcription error. During the analysis, the influence of this observation on the average rate of change should be evaluated. Furthermore, the influence of the other observations, which might not be apparent in this stage, should also be evaluated.

### 4.3 Model specification

For the Linear Mixed Model, the  $n_i \times 1$  vectors of responses  $y_i$  of subject  $i$ ,  $i = 1, \dots, N$ , are ascribed to the expected trend  $X_i\beta$  over subjects, to the subject-specific deviations  $Z_i b_i$  from this expected trend  $X_i\beta$  and to the deviations  $\epsilon_i$  of the observations from the subject-specific profile ( $X_i\beta + Z_i b_i$ ) as follows:

$$y_i = X_i\beta + Z_i b_i + \epsilon_i, \quad (4.1)$$

where the  $n_i \times 1$  vectors  $\epsilon_1, \epsilon_2, \dots, \epsilon_N$  are supposed to be measurement errors (or disturbances), which are independently normally distributed with mean zero and covariance matrix  $\sigma^2 I_i$ ,  $i = 1, \dots, N$ , with  $I_i$  the identity matrix of rank  $n_i$  and  $\sigma^2$  the common variance of the measurement errors.

No subscript is used for the fixed  $p \times 1$  regression parameter vector  $\beta$ , because the expected trend  $X_i\beta$  does not depend on the subjects in the sample. The use of the subscript  $i$  for the  $q \times 1$  random effects regression parameter vector  $b_i$  indicates that the subject-specific deviations  $Z_i b_i$  may be different for each subject  $i$ . The vectors  $b_i$  are supposed to be independently and normally distributed with mean zero and variance-covariance matrix  $D$ .

In general, the  $n_i \times p$  design matrices  $X_i$  consist of functions of the time points and time varying and time independent covariates. For longitudinal data the columns of the  $n_i \times q$  design matrices  $Z_i$  are often functions of the time points and are considered to be submatrices of the  $X_i$  matrices. This is made possible, without loss of generality, because the random effects parameters  $b_i$  are assumed to have zero mean. The variance of the response  $y_i$  is equal to  $V_i = Z_i D Z_i' + \sigma^2 I_i$ . The block diagonal matrix  $V$  is the matrix with  $V_i$ 's as block diagonal.

The first term  $X_i\beta$  of model (4.1) can be interpreted in two different ways. Each interpretation includes a definition of influential observations. The first interpretation brings into focus the group results. In this case, the expected value of the response  $y_i$  is

equal to the expected profile  $X_i\beta$  with variance-covariance matrix  $V_i$ . An influential observation in this context is supposed to have a large impact on the estimated fixed regression parameters, and the fitted response  $\hat{y}_i$  is equal to the estimated expected profile  $X_i\hat{\beta}$ . The London growth study data also dealt with group results. The main objective was to compare the mean growth of pre-adolescent girls for each of the three groups. The second interpretation brings into focus the subject-specific results. Examples can be found in educational research (Tan, 1994). One of the examples is the analysis of progress-test data. A progress-test intends to measure the progress of the students in a school, and describes and predicts the growth of knowledge of each student. In this case, the expected value of the response  $y_i$  in model (4.1) is equal to the subject-specific profile

$(X_i\beta + Z_i b_i)$  with variance  $\sigma^2 I$ . An influential observation in this context may have a large impact on the estimated subject-specific parameters  $\hat{b}_i$  or on the estimated fixed parameters  $\hat{\beta}$  or on both. No matter which interpretation is used, the estimated average profile  $X_i\hat{\beta}$  can be influenced by a strong effect on  $\hat{b}_i$ . Although the subject-specific parameters may be treated as nuisance parameters, we will show that it makes sense to evaluate influence conditional on the subjects in the sample. Note that the variance of  $y_i$ , conditionally on the subjects in the sample, is equal to the variance of the measurement errors  $\sigma^2 I_i$ . The variance-covariance matrices  $D$  and  $V$  are estimated by the ML method. The estimates of  $\hat{\beta}$  and  $\hat{b}_i$  are given by (Laird and Ware, 1982):

$$\hat{\beta} = \left( \sum_{i=1}^N X_i' \hat{V}_i^{-1} X_i \right)^{-1} \sum_{i=1}^N X_i' \hat{V}_i^{-1} y_i \quad (4.2)$$

and

$$\hat{b}_i = \hat{D} Z_i' \hat{V}_i^{-1} \left( y_i - X_i \hat{\beta} \right), \quad (4.3)$$

respectively.

The estimation process may strongly depend on a small number of observations. A commonly used method to detect such influential observations for the fixed regression

parameters in standard linear regression is Cook's Distance, which will be discussed in the following section.

#### 4.4 Cook's Distance

Cook's Distance assesses the difference between the estimates  $\hat{\beta}$  and  $\hat{\beta}_{(d)}$  of  $\beta$ , which are based on the sample with and without subject or observation  $d$ , respectively.

Cook's Distance for the  $d$ th subject or observation is defined as (Cook, 1977):

$$C_d = \frac{(\hat{\beta} - \hat{\beta}_{(d)})' X' V^{-1} X (\hat{\beta} - \hat{\beta}_{(d)})}{p} \quad (4.4)$$

$$= \frac{(\hat{y} - \hat{y}_{(d)})' V^{-1} (\hat{y} - \hat{y}_{(d)})}{p},$$

where  $\hat{y}$  and  $\hat{y}_{(d)}$  are the fitted values of  $y = (y_1', y_2', \dots, y_N')'$ , using the sample with and without subject or observation  $d$ , respectively, and  $p$  is the number of independent design parameters.

In longitudinal context, influential observations that have an unusually large impact on the subject-specific parameters  $\hat{b}$  cannot always be detected by Cook's Distance and may occasionally lead to an observation being incorrectly detected as influential. These statements will be clarified using a set of 20 subjects, which are generated from the model:

$$y_{ij} = (\beta_0 + b_{0i}) + \beta_1 t_{ij} + \epsilon_{ij}, \quad (4.5)$$

where  $t$  is the time variable,  $y_{ij}$  is the response of observation  $j$  of subject  $i$  at time point  $t_{ij}$ ,  $\beta_0 = 5$  and  $\beta_1 = 2.5$  are the fixed intercept and slope, respectively, and  $b_{0i}$  is the random intercept corresponding with the  $i$ -th subject, which is normally distributed with mean zero and variance  $\delta^2 = 2$ .

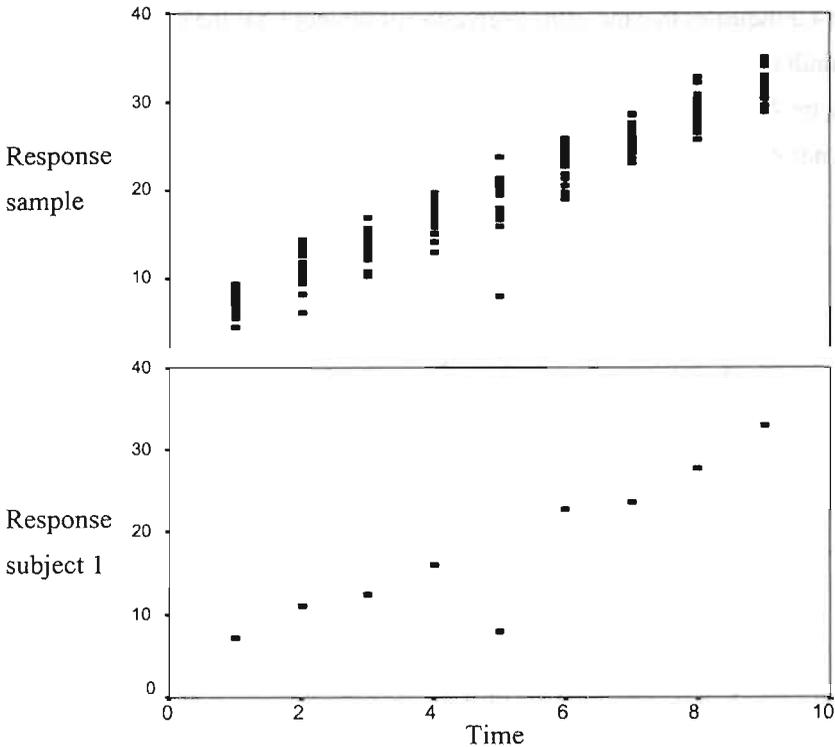


Figure 4.1. Nine repeated measures for twenty subjects including subject 1.

Figure 4.1 shows the relationship between the response  $y_{ij}$  and the time points  $t_{ij}$  for the whole simulated data set as well as for subject 1. Each subject is measured at nine equidistant time points. The fifth observation of subject 1 is taken, such that it has a large impact on the estimated value of  $\hat{\delta}_{01}$ .

Figure 4.2 shows the relative changes in the estimated fixed intercept  $\hat{\beta}_0$ , the estimated fixed slope  $\hat{\beta}_1$  and the estimated random intercepts  $\hat{\delta}_{0i}$ , respectively, due to the deletion of an observation. The relative change in  $\hat{\delta}_{0i}$  is calculated relatively to the sum of the standard deviation and the true value of the fixed intercept, i.e. it is compared to  $\delta + \beta_0$ .

Figure 4.2 indicates that the fifth observation of subject 1 has the largest impact on  $\hat{\beta}_0$ , and the ninth observation of subject 1 has the largest impact on the estimated fixed slope  $\hat{\beta}_1$ . Figure 4.2 also shows that the value of the estimated random intercept  $\hat{b}_{01}$  strongly depends on the fifth observation of subject 1.

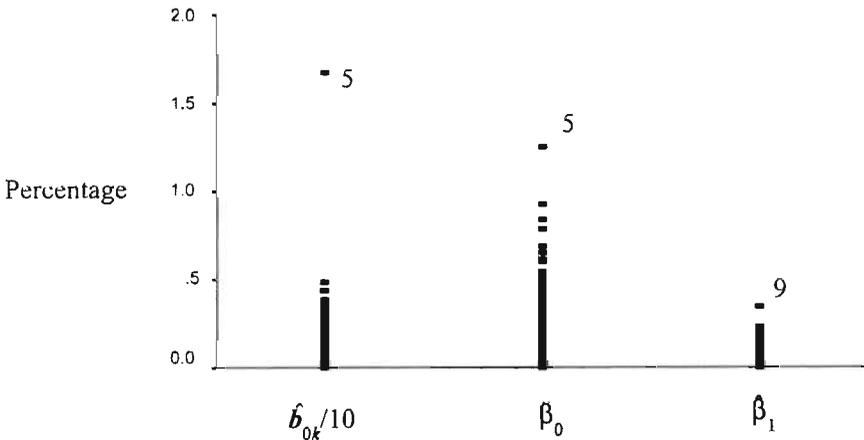


Figure 4.2. Relative changes in  $\hat{b}_{0k}$ ,  $\beta_0$  and  $\beta_1$ .

Figure 4.3 shows the Cook's Distances for the situation presented by Figure 4.1 (CD50) and for the situation that the fifth observation is set equal to the expected value for subject 1 (CD5E). The third diagnostic in Figure 4.3 ( $C_{cond}/2$ ) refers to the conditional Cook's Distance, which will be proposed in the next sections.

Figure 4.3 indicates that the ninth instead of the fifth observation of subject 1 has the largest Cook's Distance (diagram CD50). However, if the fifth observation is set equal to the average value of subject 1 (diagram CD5E), then the Cook's Distance of the ninth observation becomes comparable to the other Cook's Distances.

Obviously, the apparent influence of the ninth observation of subject 1 can be mainly ascribed to the large correlation between the observations of the same subject. We will come back to this point in section five, where we explain why Cook's Distance incorrectly recognizes the ninth observation to be most influential.

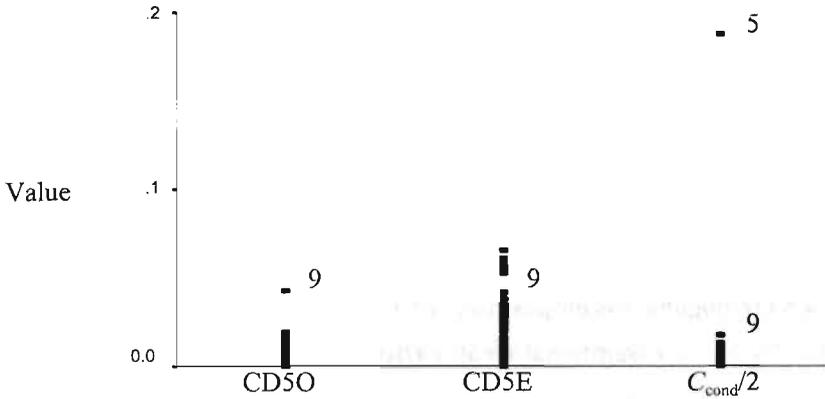


Figure 4.3. Cook's Distance and Conditional Cook's Distance.

Figure 4.4 shows the relationship between the Cook's Distances and the time points at which the measures are taken.

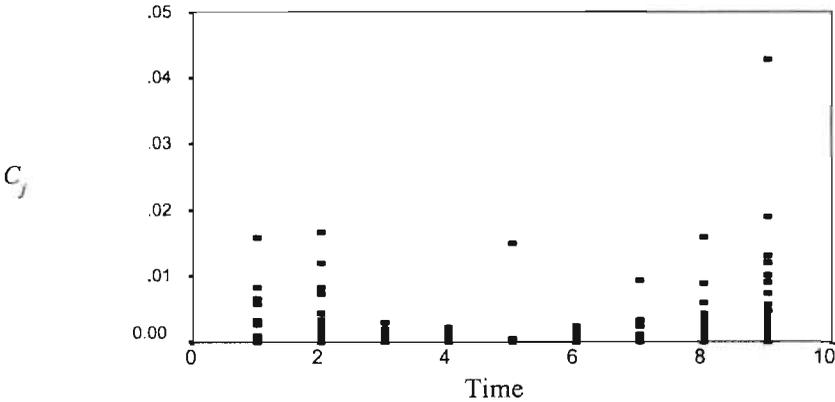


Figure 4.4. Cook's Distance over Time.

Figure 4.4 shows that if none of the observations has a significant impact on the regression parameters and the number of observations is moderate or small, then the Cook's Distances at the boundaries of the time interval will be larger than at the centre.

This is a consequence of the fact that the leverages are largest at the boundaries of the time interval and that Cook's Distance is a monotonic function of these leverages (cf. Weisberg, 1985, pg. 113 and pg. 120). Note that based on Figure 4, the fifth observation of subject 1 is easily detected to be influential.

Another approach, which may be more efficient, is the use of a conditional Cook's Distance, which will be elaborated in the next section.

#### 4.5 Orthogonal decomposition of Cook's Distance and Conditional Cook's Distance

In the previous section, we showed that the undue influence of the ninth observation of subject 1 is mainly caused by the correlation between the ninth and fifth observation of subject 1, together with the large influence of the fifth observation on the estimated subject-specific intercept  $\hat{\beta}_{01}$ . In this section, this result will be further explained and the proposed conditional Cook's Distance will be motivated.

To explain the influence of the ninth observation, model (4.1) can be rewritten as follows. Because the  $Z_i$  matrices can be represented as submatrices of the  $X_i$  matrices, model (4.1) is equivalent with

$$\begin{aligned} y_i &= (Z_i \ A_i)\beta + Z_i b_i + \epsilon_i \\ &= Z_i \beta_1 + A_i \beta_2 + Z_i b_i + \epsilon_i, \end{aligned} \tag{4.6}$$

with  $A_i$  an  $n_i \times (p - q)$  design matrix, and  $\beta = (\beta_1', \beta_2')'$ . Expression (4.6) can be written as

$$y_i = Z_i \beta_1^* + Z_i^\perp \beta_2 + Z_i b_i + \epsilon_i, \tag{4.7}$$

where  $\beta_1^*$  is a function of all fixed parameters, and  $Z_i^\perp$  is the design matrix orthogonal to  $Z_i$ . Applying equation (4.7) to model (4.5) leads to

$$(y_{i1}, \dots, y_{i9})' = (1, \dots, 1)' (\beta_0 + \bar{t} \beta_1 + b_{0i}) + (t_1 - \bar{t}, \dots, t_9 - \bar{t})' \beta_1 + \epsilon_i,$$

with  $\beta_1^* = \beta_0 + \bar{t} \beta_1 = \beta_0 + 5 \beta_1$ , and  $Z_i^+ = (t_1 - \bar{t}, \dots, t_9 - \bar{t})'$ , which is perpendicular to  $(1, \dots, 1)'$ . After transformation, a change in the estimate  $\hat{\beta}_1^*$  represents a shift, and a change in the estimate  $\hat{\beta}_2$  represents a rotation around the centre point of the scattergram. Further, the estimated regression parameters  $\hat{\beta}_1^*$  and  $\hat{\beta}_2$  are uncorrelated, such that the change in the estimate of  $\hat{\beta}_1^*$  not attributed to the change in  $\hat{\beta}_2$  can be evaluated and vice versa.

Figure 4.5 shows that the deletion of the fifth observation of subject 1 caused the largest shift and the deletion of the ninth observation of the same subject caused the largest rotation.

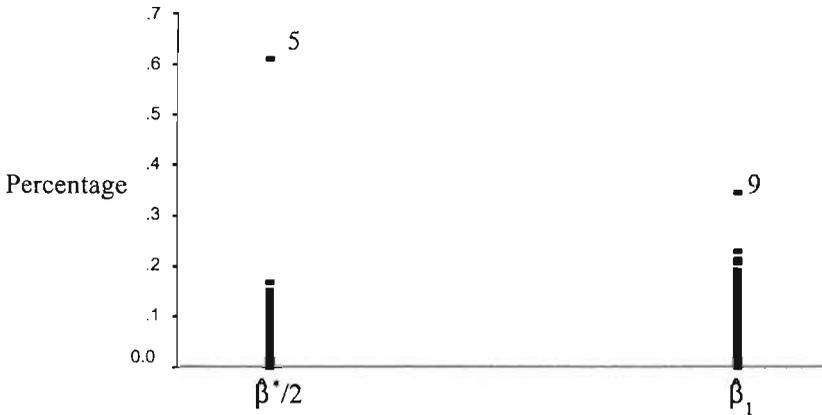


Figure 4.5. Relative change in  $\hat{\beta}_1$  and  $\hat{\beta}_1^*$ .

Figure 4.5 together with Figure 4.4 shows that the sensitivity of the Cook's Distance for changes in the estimates of the random effects is too small. It is shown in the appendix of this chapter that this is caused by the fact that Cook's Distance weighs the change in  $\hat{\beta}_1^*$  by the variance of the random parameters in addition to the variance of the measurement errors, while the change in  $\hat{\beta}_2$  is only weighted by the variance of the measurement errors. In other words, the change in  $\hat{\beta}_1^*$ , due to the deletion of the fifth

observation, is too much weighted down to detect it. This feature of Cook's Distance for mixed effects models can be overcome by approaching the problem conditionally on the subjects in the sample. This conditioning implies that the  $b_i$ 's are treated as being fixed, and we therefore do not have to deal with the variances and covariances of the random parameters. The conditional version of model (4.1) can be expressed as follows:

$$y = U\gamma + \epsilon, \quad (4.8)$$

where  $U$  is of the form

$$U = \begin{bmatrix} X_1 & Z_1 & 0 & 0 & \dots & 0 \\ X_2 & 0 & Z_2 & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ X_N & 0 & \dots & \dots & 0 & Z_N \end{bmatrix},$$

the vector  $\gamma$  is given by

$$\gamma = \left[ \beta' \quad b_1' \quad b_2' \quad \dots \quad b_N' \right]' \quad (4.9)$$

and the vector  $\epsilon = (\epsilon_1', \dots, \epsilon_N')'$  is the vector of measurement errors. The design matrix  $U$  is not of full rank, because the matrices  $Z_i$  are by definition submatrices of the matrices  $X_i$ . As a consequence, the vector of regression parameters  $\gamma$  is not identified. Hence, the unweighted least squares method cannot be applied in this situation. Therefore, we propose to use the Laird and Ware estimators as done in the unconditional case. Moreover, by using the same estimates, Cook's Distance can be compared with the conditional Cook's Distance, which is equal to

$$C_{cond_j} = \sum_{i=1}^N \frac{\left| (X_i\beta + Z_i b_i) - (X_i\beta_{(j)} + Z_i b_{i(j)}) \right|' \left| (X_i\beta + Z_i b_i) - (X_i\beta_{(j)} + Z_i b_{i(j)}) \right|}{\sigma^2[(N-1)q + p]}. \quad (4.10)$$

To evaluate the influence on the subject-specific parameters and fixed regression parameters separately, expression (4.10) can be decomposed as:

$$\begin{aligned}
 C_{cond_j} &= C_{cond_{1j}} + C_{cond_{2j}} + C_{cond_{3j}} = \\
 &= \frac{(\beta - \beta_{(j)})' X' X (\beta - \beta_{(j)})}{k\sigma^2} + \frac{\sum_{i=1}^N (\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{i(j)})' Z_i' Z_i (\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{i(j)})}{k\sigma^2} \\
 &\quad + \frac{2(\beta - \beta_{(j)})' \sum_{i=1}^N X_i' Z_i (\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{i(j)})}{k\sigma^2},
 \end{aligned}$$

where  $k = (N-1)q + p$ . The first term

$$C_{cond_{1j}} = \frac{(\beta - \beta_{(j)})' X' X (\beta - \beta_{(j)})}{k\sigma^2}$$

can be interpreted as a distance measure for  $\beta$ . It can also be interpreted as the change in the estimated average profile. The term  $C_{cond_{1j}}$  is equal to Cook's Distance (4.4), but without normalizing for the random effects variances and covariances (see the corollary in the appendix of this chapter). The second term

$$C_{cond_{2j}} = \sum_{i=1}^N \frac{(\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{i(j)})' Z_i' Z_i (\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{i(j)})}{k\sigma^2}$$

can be interpreted as a distance measure for the change in the estimated vectors of subject-specific regression parameters  $\hat{\mathbf{b}}_i$ . It also assesses the change in the estimated deviations of the subject-specific profiles from the average profile. The third term

$$C_{cond_{3j}} = \frac{2(\beta - \beta_{(j)})' \sum_{i=1}^N X_i' Z_i (\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{i(j)})}{k\sigma^2}$$

is a measure of covariance between a change of the average profile and a change of

position of the subject specific profiles relative to the average profile.

The conditional Cook's Distances for the illustration mentioned above are plotted in Figure 4.3. Figure 4.3 indicates that the fifth observation has a large conditional Cook's Distance. This implies that the conditional Cook's Distance correctly detects the fifth observation to be influential.

Figure 4.6 shows the values of the decomposition terms of equation (4.11).

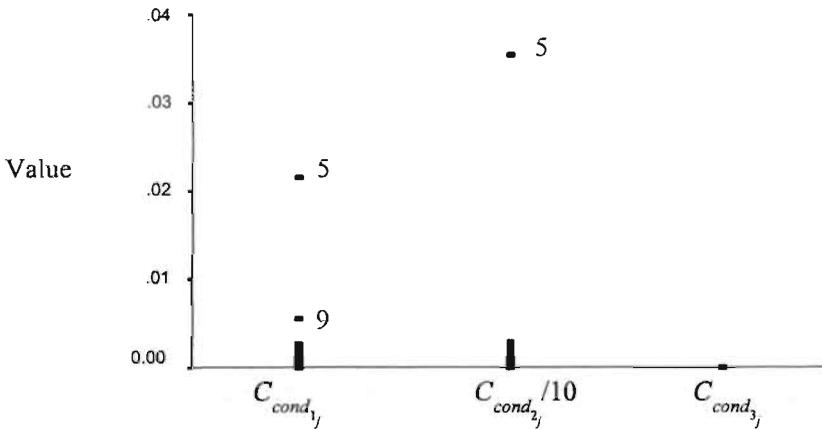


Figure 4.6. Values of the three terms of decomposition (4.11).

Figure 4.6 shows that the fifth observation of subject 1, instead of the ninth observation, has the largest influence value for the first term of equation (4.11). The fifth observation also has the largest effect on the estimated subject-specific regression parameters. The values of the last term of the decomposition are extremely small. Our experience based on a simulation study is that this term remains small in most practical situations.

#### 4.6 The London Growth Study: An analysis for influential observations

The model that we use for the London growth study is a longitudinal random intercept model with group by time interaction. Note that Goldstein (1979) has used a fixed effects model with group by time interaction. The inter-subject variation for each of the three groups is small. The discrepancy between Cook's Distance and its conditional version is mainly due to this variation. Therefore, the first and second group are combined to obtain a larger inter-subject variation. The model is specified as follows.

$$y_i = (\beta_0 + b_{0i}) + \beta_1 G + \beta_2 t + \beta_3 G * t + \epsilon_i, \quad (4.12)$$

where  $G$  is the indicator for tall mothers,  $\beta_j, j = 0, \dots, 3$ , are fixed regression parameters,  $\epsilon_i$  is the error term,  $b_{0i}$  is the intercept of subject  $i$ , and  $y_i$  is the length of girl  $i$  at time point (age)  $t$ . Figure 4.7 shows the Cook's Distances, the conditional Cook's Distances and the values of the three terms of decomposition (4.11).

Cook's Distance  $C_j$  indicates that the fifth observation of the first girl in the second group (girl number 14) is most influential. The conditional Cook's Distance  $C_{cond_j}$ , however, indicates that the second observation of the fifth girl of the first group (girl number 5) is most influential. The same observation is indicated as most influential by the first and the second terms  $C_{cond_{1j}}$  and  $C_{cond_{2j}}$ , respectively. The fifth observation of the first girl of the second group is also indicated by the first term as influential. A closer inspection of the data in Table 4.1 shows, that the growth curve of the height of this girl is flatter than the other curves in her group.

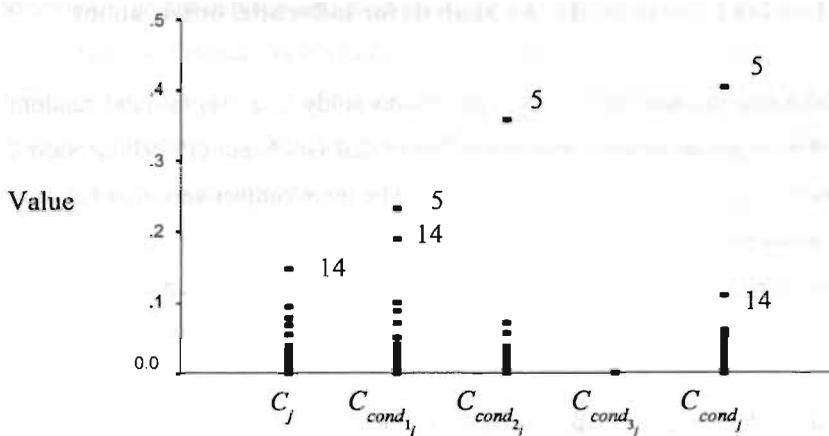


Figure 4.7. Cook's Distances, values of the three terms of the decomposition and Conditional Cook's Distances.

Table 4.2 shows the estimates of the regression parameters based on the sample with and without certain observations. It shows that deleting observation five changes upwards the fixed regression slope estimate by means of  $\beta_3$ . Cook's Distance correctly detects this change. The average growth of the first group remains unchanged. Furthermore, the second observation of the fifth girl of the first group is lower than the first observation of the same person. Deleting this observation changes upwards the estimated intercept  $\beta_0$  as well as the estimated slope  $\beta_1$  of the average growth in this group. The average growth of the second group remains unchanged. Cook's Distance does not detect this change. The conditional Cook's Distance however, points out that the deletion of this observation has the largest impact on both the vector of fixed regression parameters and the subject-specific parameter estimates. This illustration demonstrates that Cook's Distance is less sensitive for observations influential on the fixed parameters associated with the random effects and that the second term of the conditional Cook's Distance is sensitive for observations that are influential for the random effects.

Table 4.2 Estimates and variances of the elements of  $\beta$ 

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
Sample	81.437 (1.218)	1.686 (2.059)	5.506 (2.059)	0.742 (0.173)
Sample without 2nd observation of girl 5	82.050 (1.083)	1.073 (1.828)	5.445 (0.076)	0.804 (0.128)
Sample without 5th observation of girl 14	81.437 (1.191)	0.939 (2.030)	5.506 (0.098)	0.849 (0.170)
Sample without both observations	82.050 (1.050)	0.331 (1.782)	5.445 (0.070)	0.910 (0.120)

#### 4.7 Discussion

In this chapter we argue that influential observations that have a large impact on the subject-specific parameters cannot always be detected by Cook's Distance due to large between subject variation. Since some of the fixed regression parameters are more sensitive to changes of subject-specific parameters, the problem of influential observations should be approached conditionally on the subjects. A conditional version of Cook's Distance is proposed that deals with this problem. The conditional measure can be decomposed into a part that measures the influence on the estimated fixed parameters, a part that measures the influence on the estimated subject-specific parameters, and a part that measures the influence on how the change in the estimated average profile covariates with the change in the subject specific profiles. Based on our own experience with real data and a simulation study, it appears that this last part can be neglected.

In general, the error terms in model (4.1) are correlated due to within individual serial correlations. However, in specific applications the effect of serial correlation may be dominated by the combination of random effects and measurement errors (Diggle et al. 1994, page 88). In this chapter we restrict ourselves to uncorrelated error terms. Note that, if an observation has a large influence on an estimated subject parameter, then deletion of this observation will affect the variances and covariances of the random parameters and hence  $V$ . The second term,  $C_{cond_{2j}}$ , of decomposition (4.11) replaces the variance component analysis of Christensen et al. (1992).

Each influence measure maps multidimensional vectors of estimated parameters onto a one dimensional space, by the use of a measure specific metric. In this paper, we showed that the choice of this metric for the unconditional Cook's Distance may lead to wrong conclusions. We claim that the (unconditional) Cook's Distance may lead to wrong conclusions. We claim that the conditional Cook's Distance, which uses the Euclidian metric, works well in practice. A possible drawback of using the conditional version of Cook's Distance is the fact that we have used empirical bayes estimates for the subject specific parameters. It is not known how large the bias of these estimates can be and the impact of (large) bias on the quality of the conditional Cook's Distance as an influence measure for mixed effects longitudinal models with uncorrelated error terms is unknown. Note that for fixed effects models the unconditional and the conditional Cook's Distance coincide. In that situation we are dealing with the common OLS Cook's Distance.

## Appendix

In this appendix the relationship between Cook's Distance and the first term of the Conditional Cook's Distance is derived. Given model (4.1),  $Z_i$  is a submatrix of  $X_i$ . Consequently the  $X_i$ -space can be decomposed in the  $Z_i$ -space and the subspace of  $X_i$  orthogonal to the  $Z_i$ -space. It follows that for every  $i$  the change in the estimates of  $y_i$  is equal to

$$X_i(\beta - \beta_{(j)}) = Z_i \mathbf{u}_i + Z_i \mathbf{v}_i, \quad (4.13)$$

for some vectors  $\mathbf{u}_i$  and  $\mathbf{v}_i$ , and  $Z_i$  a matrix of full rank for which the columns are vectors in the  $X_i$ -space and are orthogonal to the columns of  $Z_i$ .

**Theorem.** Suppose that  $\{s_{i1}, \dots, s_{ik_i}\}$  is an orthonormal basis for the kernel of  $Z_i D Z_i'$  and  $\{t_{i1}, \dots, t_{i l_{m_i}}\}$  is an orthonormal basis of eigenvectors for the image of  $Z_i D Z_i'$ , with eigenvalues  $\lambda_{im}$ . Let  $\langle \cdot, \cdot \rangle$  be the inproduct.  $p$  times Cook's Distance is then equal to

$$pC_j = \sum_{i=1}^N \left( \sum_{m=1}^{k_i} \frac{\langle s_{im}, Z_i' \mathbf{v}_i \rangle^2}{\sigma^2} + \sum_{m=1}^{l_{m_i}} \frac{\langle t_{im}, Z_i \mathbf{u}_i \rangle^2}{\sigma^2 + \lambda_{im}} \right). \quad (4.14)$$

**Proof:**

$$pC_j = (\hat{\mathbf{y}} - \hat{\mathbf{y}}_{(j)})' V^{-1} (\hat{\mathbf{y}} - \hat{\mathbf{y}}_{(j)}) = \sum_{i=1}^N (\beta - \beta_{(j)})' X_i' V_i^{-1} X_i (\beta - \beta_{(j)}). \quad (4.15)$$

Using the union of  $\{s_{i1}, \dots, s_{ik_i}\}$  and  $\{t_{i1}, \dots, t_{i l_{m_i}}\}$  as a basis for the  $X_i$ -space,  $X_i(\beta - \beta_{(j)})$  can be written as

$$\sum_{m=1}^{k_i} \langle s_{im}, X_i(\beta - \beta_{(j)}) \rangle s_{im} + \sum_{m=1}^{l_{m_i}} \langle t_{im}, X_i(\beta - \beta_{(j)}) \rangle t_{im} \quad (4.16)$$

The  $s_{im}$ 's and  $t_{im}$ 's are eigenvectors of the matrix  $V_i^{-1}$  with eigenvalues  $\sigma^2$  and  $\sigma^2 + \lambda_{im}$ , respectively. Consequently,  $p$  times the Cook's Distance can be written as

$$\sum_{i=1}^N \left[ \sum_{m=1}^{k_i} \frac{\langle s_{im}, X_i(\beta - \beta_{(j)}) \rangle^2}{\sigma^2} + \sum_{m=1}^{l_{m_i}} \frac{\langle t_{im}, X_i(\beta - \beta_{(j)}) \rangle^2}{\sigma^2 + \lambda_{im}} \right]. \quad (4.17)$$

It will be proven below that the  $s_{im}$ 's are orthogonal to the columns of the  $Z_i$  matrix and that the  $t_{im}$ 's are linear combinations of the columns of  $Z_i$ . Expression (4.14) can then be derived from expression (4.17) using expression (4.13) and the statements in the previous sentence, which completes the proof.

$s_{im}$  is orthogonal to the columns of  $Z_i$ , for every  $m$ ,  $m = 1, \dots, k_i$ .

Proof: Take an arbitrary  $i \leq N$  and  $m \leq k_i$ . Then  $0 = Z_i D Z_i' s_{im} = s_{im}' Z_i D Z_i' s_{im}$ . Since  $D$  is of full rank it follows that  $Z_i' s_{im} = 0$ . Consequently  $s_{im}$  is orthogonal to the columns of  $Z_i$ . Because  $i$  and  $m$  were arbitrarily chosen, this holds for all  $i \leq N$  and  $m \leq k_i$ .

The  $t_{im}$ 's are linear combinations of the columns of  $Z_i$ .

Proof: Take an arbitrary  $i \leq N$  and  $m \leq k_i$ . Then  $t_{im}$  can be written as  $t_{im} = Z_i b + d$ , for some  $b$  and  $d$ , where  $d$  is orthogonal to the columns of  $Z_i$ . Furthermore  $t_{im}$  is an eigenvector of  $Z_i D Z_i'$ . Note that  $\lambda_{im}$  is larger than zero. The result can be derived as follows:

$$d' d = d' (Z_i b + d) = d' \left[ \frac{Z_i D Z_i'}{\lambda_{im}} (Z_i b + d) \right] = (d' Z_i) \frac{D Z_i'}{\lambda_{im}} (Z_i b + d) = 0.$$

This implies that  $d$  is zero and thus that  $t_{im}$  is a linear combination of the columns of  $Z_i$ . Because  $i$  and  $m$  were arbitrary, this holds for all  $i$ ,  $i \leq N$  and  $m \leq k_i$ .

q.e.d.

**Corollary.** Let  $k$  be the denominator of the Conditional Cook's Distance

$(p + (N - 1) q)$ .  $k$  times the first term of the Conditional Cook's Distance is then equal to

$$k C_{cond 1_j} = \sum_{i=1}^N \left( \sum_{m=1}^{k_i} \frac{\langle s_{im}, Z_i' v_i \rangle^2}{\sigma^2} + \sum_{m=1}^{l m_i} \frac{\langle t_{im}, Z_i' u_i \rangle^2}{\sigma^2} \right), \quad (4.18)$$

and the relationship between Cook's Distance and the first term of the Conditional Cook's Distance is given by

$$k C_{cond 1_j} = p C_j + \sum_{i=1}^N \sum_{m=1}^{l m_i} \frac{\lambda_{im}}{\lambda_{im} + \sigma^2} \frac{\langle t_{mi}, Z_i' u_i \rangle^2}{\sigma^2}. \quad (4.19)$$

## References

- Banerjee, M. (1998). Cook's Distance in linear longitudinal models, *Communications in Statistics, Theory and Methods*, 12, 2973-2983.
- Banerjee, M., and Frees, E.W. (1997). Influence diagnostics for linear longitudinal models, *Journal of the American Statistical Association*, 92, 999-1005.
- Barrett, B.E., and Ling, R.F. (1992). General Classes of Influence Measures for Multivariate Regression, *Journal of the American Statistical Association*, 87, 184-191.
- Belsley, D.A., Kuh, E., and Welsch, R.E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, New York: Wiley.
- Chatterjee, S., and Hadi, A.S. (1986). Influential Observations, High Leverage Points, and Outliers in Linear Regression, *Statistical Science*, 1, 379-416.
- Christensen, R., Pearson, L.M., Johnson, W. (1992). Case-deletion diagnostics for mixed models, *Technometrics*, 34, 38-45.
- Cook, R.D. (1977). Detection of influential observations in linear regression. *Technometrics*, 19, 15-18.
- Cook, R.D. (1986). Assessment of local influence (with discussion), *Journal of the Royal Statistical Society Sector B*, 48, 133-169.
- De Gruttola, V., Ware, J.H., and Louis, T.A. (1987). Influence Analysis of Generalized Least Squares Estimators, *Journal of the American Statistical Association*, 82, 911-917.
- Diggle, P.J., Liang, K.Y., and Zeger, S.L. (1994). *Analysis of longitudinal data*, Oxford: Clarendon Press.
- Goldstein, H. (1979). *The design and analysis of longitudinal studies*, London: academic press.
- Laird, N.M., and Ware, J.H. (1982). Random-effects models for longitudinal data, *Biometrics*, 38, 963-974.

- Lesaffre, E., and Verbeke, G. (1998). Local influence in linear models, *Biometrics*, 54, 570-582.
- Ouwens, M.J.N.M., Tan, F.E.S., and Berger, M.P.F. (1999). *Local influence for repeated measures generalized linear mixed models*. In H. Friendl, A. Berghold, G. Kauermann (eds.), proceedings of the 14th International Workshop on Statistical Modelling, 308-316, Graz, Austria.
- Ouwens, M.J.N.M., Tan, F.E.S., and Berger, M.P.F. (2000). Detecting influential observations and subjects in Generalized Linear Mixed Models. *Kwantitatieve Methoden*, 64, 71-86.
- Ouwens, M.J.N.M., Tan, F.E.S., and Berger, M.P.F. (2001). Detecting influential observations and subjects in Generalized Linear Mixed Models, *Biometrics*, in press.
- Rohlf, F.J. (1975). Generalization of the Gap Test for the Detection of Multivariate Outliers, *Biometrics*, 31, 93-101.
- Siontani, M. (1959). The Extreme Value of the Generalized Distances of the Individual Points in the Multivariate Normal Sample, *Annals of the Institute of Statistical Mathematics*, 10, 183-208.
- Tan, E.S. (1994). *A stochastic growth model for the longitudinal measure of ability*. Ph.D. Thesis, Maastricht University. Maastricht: Datayse Maastricht.
- Tanner, J.M., Whitehouse, R.H., Marubini, E., and Resele, L.F. (1976). The adolescent growth spurt of boys and girls of the Harpenden Growth Study, *Annals of Human Biology*, 3, 109-126.
- Vonesh, E.F., and Carter, R.L. (1987). Efficient inference for random-coefficient growth curve models with unbalanced data, *Biometrika*, 43, 617-628.
- Weisberg, S. (1985). *Applied linear regression*, New York: Wiley.

# **5** **Detection of influential data in Generalized Linear Mixed Models**

*To appear in Biometrics, 2001, 57, no. 4.*

### Abstract

This chapter discusses the generalization of the local influence measures for normally distributed responses to local influence measures for generalized linear models with random effects. For these models, it is shown that the subject-oriented influence measure is a special case of the proposed observation-oriented influence measure. A two-step diagnostic procedure is proposed. The first step is to search for influential subjects. A search for influential observations is proposed as second step. An illustration of a two-treatment multiple period cross-over trial demonstrates the practical importance of the detection of influential observations in addition to the detection of influential subjects.

*Key words:* influential observations, influential subjects, local influence, GLMM, Poisson regression, random effects, GLM, Likelihood Displacement

## 5.1 Introduction

The interest in generalized linear mixed models (GLMM) has grown steadily during the past decades (Diggle, Liang and Zeger, 1994). Unfortunately, the estimates of the model parameters may heavily depend on a small part of the data set or even on one particular observation or subject. For ease of presentation, a set of observations of one or more subjects is called a data structure. A data structure which has a large impact on the estimated model parameters will be called influential.

An approach to detect influential data structures is to compare the estimates of the model parameters based on the sample with and without the data structures of interest. Diagnostics based on this approach are called deletion diagnostics and are discussed for normally distributed responses by Bannerjee and Frees (1997), Christensen, Pearson and Johnson (1992) and Tan, Ouwens and Berger (1999, 2001). The only influence diagnostics for generalized linear mixed models discussed in literature are the deletion diagnostics of Preisser and Qaqish (1996) in the context of marginal models. For these models the average of the responses for each covariate pattern is modelled.

A second approach to detect influential data structures is based on the curvature of the log-likelihood function. One of the measures based on this approach is local influence and has been developed by Cook (1986). Local influence for generalized linear models is discussed by Thomas and Cook (1989, 1990) and local influence for longitudinal models with normally distributed responses has been discussed by Lesaffre and Verbeke (1998). We will discuss local influence measures for the generalized linear mixed model. For these models it is important to detect influential observations in addition to the detection of influential subjects. The main reason is that influential observations may be distributed across several subjects, such that the subjects themselves are not detected to be influential. On the other hand, when influence is only considered at subject level, no distinction can be made between influence due to subject specific characteristics and influence due to characteristics of

some observations from the subjects. Two classes of models may be applied, namely models with known dispersion parameter, such as the Poisson mixed model, and models with unknown dispersion parameter. It will be shown in this chapter that the proposed observation-oriented local influence measures can give additional information for models with known as well as for models with unknown dispersion parameter.

The detection of local influential observations and subjects can be done stepwise. In the first step the subject-oriented local influence measures are used and in a second step the observation-oriented local influence measures are used. This procedure may be applied iteratively.

In section 5.2 the underlying regression model is specified. The observation-oriented local influence approach is proposed in section 5.3. The additional importance of the observation-oriented influence measures is demonstrated by the analysis of a real data set in section 5.4 and conclusions are drawn in section 5.5.

## 5.2 Model specification

Conditional on the subject specific parameters, the responses  $y_{ij}$  of subject  $i$ ,  $i = 1, \dots, N$ , at time point  $j$ ,  $j = 1, \dots, n_i$ , are independent and drawn from an exponential density function (Diggle, Liang and Zeger, 1994, p. 134) :

$$f ( y_{ij} | \theta_{ij} ) = \exp \left\{ \frac{ \left[ y_{ij} \theta_{ij} - \psi ( \theta_{ij} ) \right] }{ a_{ij} ( \phi ) } + c_{ij} ( y_{ij}, \phi ) \right\}, \quad (5.1)$$

where  $\theta_{ij}$  is the canonical form of the location parameter and is a function of the conditional mean  $\mu_{ij}$ ,  $a_{ij} ( \phi )$  is a known function of the possibly unknown dispersion parameter, or vector of dispersion parameters  $\phi$ ,  $c_{ij}$  is a function of the dispersion parameter and the responses (McCullagh and Nelder, 1996, p. 28), and  $\psi$  is a known function, such that the conditional mean of  $y_{ij}$  is equal to

$\mu_{ij} = E(y_{ij} | \theta_{ij}) = \partial \psi(\theta_{ij}) / \partial \theta_{ij}$ . The subjects are assumed to be independent and the vectors of subject specific parameters  $\mathbf{b}_i$  are outcomes of the normal distribution  $\Phi(0, G)$  where  $G$  is a  $q \times q$  variance-covariance matrix.

In the illustration section we will use the Poisson mixed model with known dispersion parameter  $\phi = 1$ . In some cases, however, the data may be overdispersed and the Poisson mixed model may not be suitable. A commonly used model for overdispersed count data is the negative binomial model (McCullagh and Nelder, 1996, p.199). This model will also be considered in the illustration section.

### 5.3 Local influence

The first part of this section focusses on the detection of influential observations. The relationship between the observation-oriented and the subject-oriented local influence measure is the topic of the second part.

Some specific notation and definitions are used throughout this chapter. The vector of fixed model parameters is denoted by  $\zeta$ . The vector of responses  $\mathbf{y}$  can be written as  $[y_1', \dots, y_N']'$ , where  $y_i$  is the vector of  $n_i$  responses of subject  $i$ ,  $i = 1, \dots, N$ . Finally, the set  $M_d$  is the set of observations (the data structure) of subject  $d$  to be evaluated, the vector  $\mathbf{y}_{dM_d}$  is formed by stacking the responses of subject  $d$  corresponding to the observations in set  $M_d$  and the vector  $\mathbf{y}_{d(M_d^c)}$  is formed by stacking the responses of subject  $d$  corresponding to the observations not in  $M_d$ . It should be noted that  $\mathbf{y}_{d(M_d^c)}$  may be an empty vector.

#### *Observation-Oriented Local Influence.*

In this section an observation-oriented influence measure is derived based on the case-deletion approach, i.e. based on the approach to detect influential data structures by assessing the difference between the maximum likelihood estimates based on the sample with and without the data structure of interest. The difference

between the maximum likelihood estimates can be assessed using the Likelihood Displacement given by Cook and Weisberg (1982):

$$LD_{M_d} = 2[ \log L( \mathbf{y} | \hat{\zeta} ) - \log L( \mathbf{y} | \hat{\zeta}_{(M_d)} ) ], \quad (5.2)$$

where  $\hat{\zeta}$  and  $\hat{\zeta}_{(M_d)}$  are the maximum likelihood estimates based on the complete sample with and without the subset  $M_d$  of observations of subject  $d$ , respectively, and  $\log L$  denotes the log-likelihood function of the complete sample. Expression (5.2) can be Taylored using weighted log-likelihood functions. Cook (1986) called the Taylor expansion the local influence measure.

Since the subjects are independent, the log-likelihood function is the sum of the contributions of the individual subjects to the log-likelihood function, i.e.  $\log L( \mathbf{y} | \zeta ) = \sum_i l_i( \mathbf{y}_i | \zeta )$ . Let  $l_d( \mathbf{y}_d | \zeta )$  and  $l_{d(M_d)}( \mathbf{y}_{d(M_d)} | \zeta )$  be the contributions of the  $d$ -th subject to the log-likelihood functions of the complete sample and of the sample without the observations of subject  $d$  in  $M_d$ , respectively. The log-likelihood functions of the sample with and without the observations of subject  $d$  in  $M_d$  are both of the form

$$\begin{aligned} \log L( \mathbf{y} | \zeta, \omega_{M_d} ) = & \log L( \mathbf{y} | \zeta ) + \\ & + (1 - \omega_{M_d}) \left[ l_{d(M_d)}( \mathbf{y}_{d(M_d)} | \zeta ) - l_d( \mathbf{y}_d | \zeta ) \right], \end{aligned} \quad (5.3)$$

where  $\omega_{M_d}$  is a scalar. If  $\omega_{M_d} = 1$ , expression (5.3) is equal to the log-likelihood function of the complete sample; if  $\omega_{M_d} = 0$ , it is equal to the log-likelihood function of the sample without the observations in  $M_d$ . Expression (5.3) indicates that the weighted log-likelihood function can be defined as:

$$\log L(\mathbf{y} \mid \boldsymbol{\zeta}, \boldsymbol{\omega}) = \log L(\mathbf{y} \mid \boldsymbol{\zeta}) + \quad (5.4)$$

$$+ \sum_{i=1}^N \sum_{M_i} (1 - \omega_{M_i}) \left[ l_{i(M_i)}(\mathbf{y}_{i(M_i)} \mid \boldsymbol{\zeta}) - l_i(\mathbf{y}_i \mid \boldsymbol{\zeta}) \right],$$

where  $\boldsymbol{\omega}$  is the vector of weights  $\omega_{M_i}$ .

Using expression (5.4), expression (5.2) can be written as:

$$LD(\tilde{\boldsymbol{\omega}}) = 2 \left[ \log L(\mathbf{y} \mid \hat{\boldsymbol{\zeta}}_{\tilde{\boldsymbol{\omega}}}, \tilde{\boldsymbol{\omega}}) - \log L(\mathbf{y} \mid \hat{\boldsymbol{\zeta}}_{\tilde{\boldsymbol{\omega}}}, \boldsymbol{\omega}_0) \right], \quad (5.5)$$

where  $\tilde{\boldsymbol{\omega}}$  denotes the vector of weights for which each element is equal to 1, except the weight corresponding to set  $M_i$  which is equal to 0, and  $\boldsymbol{\omega}_0$  is the vector of weights for which each element is equal to 1. The corresponding maximum likelihood estimates of the model parameters are  $\hat{\boldsymbol{\zeta}}_{\tilde{\boldsymbol{\omega}}}$  and  $\hat{\boldsymbol{\zeta}}_{\boldsymbol{\omega}_0}$ . In general the weights may take any value. Different vectors  $\boldsymbol{\omega}$  may be of interest. The evaluation of all possible vectors  $\boldsymbol{\omega}$  of interest, however, may be very labourious. One way to reduce the amount of work is to approximate the likelihood displacement by the second order Taylor expansion around  $\boldsymbol{\omega}_0$ . The likelihood displacement in  $\boldsymbol{\omega}_0$  is zero. Because the value  $\log L(\mathbf{y} \mid \hat{\boldsymbol{\zeta}}_{\boldsymbol{\omega}_0}, \boldsymbol{\omega}_0)$  is for all  $\tilde{\boldsymbol{\omega}}$  larger than or equal to  $\log L(\mathbf{y} \mid \hat{\boldsymbol{\zeta}}_{\tilde{\boldsymbol{\omega}}}, \boldsymbol{\omega}_0)$ , the value of the first derivative in  $\boldsymbol{\omega}_0$  is zero. Hence the second order Taylor expansion depends only on the

second derivative. Let  $\Delta = \frac{\partial^2 \log L(\mathbf{y} \mid \boldsymbol{\zeta}, \boldsymbol{\omega})}{\partial \boldsymbol{\zeta} \partial \boldsymbol{\omega}}$  and  $\mathbf{d}$  be the direction in which the second derivative is evaluated. In general, to evaluate the influence of the observations in  $M_i$  the only element of  $\mathbf{d}$  unequal to zero is the element corresponding to  $\omega_{M_i}$ . The local influence measure in  $\boldsymbol{\omega}_0$  in the direction  $\mathbf{d}$  is defined as twice the absolute value of the second derivative in the direction  $\mathbf{d}$ . Cook (1986) showed that the local influence measure is equal to

$$C_{\mathbf{d}} = 2 \left| \mathbf{d}' \Delta' \left[ \frac{\partial^2 \log L(\mathbf{y} \mid \boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}^2} \right]^{-1} \Delta \mathbf{d} \right|, \quad (5.6)$$

evaluated in  $\zeta = \hat{\zeta}$ . Local influence measures for a particular subset of the model parameters can be obtained by the procedure proposed by Cook (1986).

To calculate the value of the local influence  $C_a$ , the information matrix and  $\Delta$  must be computed. The information matrix of the model parameters is standard output of most computer programs and can be interpreted without knowing the exact analytical expression. Thus, we only need to concentrate on  $\Delta$  to compute the local influence measure in equation (5.6).

Let  $S_i$  and  $S_{i(M_i)}$  be the contributions of subject  $i$  to the score function based on the complete sample with and without set  $M_i$ ,  $i = 1, \dots, N$ , respectively. The column of  $\Delta$  corresponding to weight  $\omega_{M_i}$  is then equal to  $(S_i - S_{i(M_i)})$ . The expressions for  $(S_i - S_{i(M_i)})$  for the exponential density function are given in the appendix of this chapter. The difference in the score functions consists of the subvector of the difference in the score function for the regression parameters, the subvector of the difference in the score function for the parameters specifying the random effects variance-covariance matrix  $G$  and if the overdispersion  $\phi$  is unknown, the subvector of the difference in the score function for the overdispersion parameter  $\phi$ .

The subvector corresponding to the regression parameters for the Poisson mixed model is given by:

$$X_i'(y_i - \eta_i) - X_{i(M_i)}'(y_{i(M_i)} - \eta_{i(M_i)}), \quad (5.7)$$

where  $X_i$  is the design matrix of subject  $i$  with  $j$ th row  $x_{ij}$  and  $X_{i(M_i)}$  is the design matrix of the observations of subject  $i$  not in  $M_i$ . The vector of empirical bayes estimates  $\eta_i$  of the responses of subject  $i$  is based on all  $n_i$  observations of subject  $i$ , and  $\eta_{i(M_i)}$  is the vector of empirical bayes estimates of the responses of subject  $i$  corresponding to the observations not in  $M_i$ . That is, element  $k$  of  $\eta_i$  is equal to

$$\int \mu_{ik} \frac{\left( \prod_{j=1}^{n_i} f_{ij} \right) g_i}{\int \left( \prod_{j=1}^{n_i} f_{ij} \right) g_i db_i} db_i \quad (5.8)$$

and element  $k$  of  $\eta_{i(M)}$  is equal to

$$\int \mu_{ik} \frac{\left( \prod_{j=1, j \notin M_i}^{n_i} f_{ij} \right) g_i}{\int \left( \prod_{j=1, j \notin M_i}^{n_i} f_{ij} \right) g_i db_i} db_i, \quad (5.9)$$

where these expressions are evaluated at the maximum likelihood estimates of the model parameters. The observations are assumed to be conditionally independent. If, for example, serial correlation is present, the above written theory can be applied, but the product of density functions in expression (5.8) and (5.9) will then be replaced by a single joint density function.

Assuming that the random effects variance-covariance matrix is unstructured, the parameters specifying the variance-covariance matrix of the random effects are the elements of the matrix  $G$ . The derivative of the weighted log-likelihood function with respect to the  $jk$ th element of the variance-covariance matrix of the random effects  $G$ ,  $G_{jk}$ , and  $\omega_{M_i}$  for  $j \neq k$  in  $\theta = \hat{\theta}$  is equal to the sum of the  $jk$ th element and the  $kj$ th element of the matrix

$$\frac{1}{2} \hat{G}^{-1} \left[ E(\mathbf{b}_i \mathbf{b}_i' | \mathbf{y}_i, \xi) - E(\mathbf{b}_i \mathbf{b}_i' | \mathbf{y}_{i(M)}, \xi) \right] \hat{G}^{-1}, \quad (5.10)$$

where  $E(\mathbf{b}_i \mathbf{b}_i' | \mathbf{y}_i, \xi)$  and  $E(\mathbf{b}_i \mathbf{b}_i' | \mathbf{y}_{i(M)}, \xi)$  are the empirical bayes estimates of the second moment of the random effects of subject  $i$ , based on the complete sample with and without the observations in set  $M_i$ , respectively. The derivative with respect to  $G_{jj}$  is equal to the  $jj$ th element of the matrix in formula (5.10). The quantities

$$E(\mathbf{b}_i \mathbf{b}_i' | \mathbf{y}_i, \boldsymbol{\zeta}) = \int \mathbf{b}_i \mathbf{b}_i' \frac{\left( \prod_{j=1}^{n_i} f_{ij} \right) \mathbf{g}_i}{\int \left( \prod_{j=1}^{n_i} f_{ij} \right) \mathbf{g}_i d\mathbf{b}_i} d\mathbf{b}_i \quad (5.11)$$

and

$$E(\mathbf{b}_i \mathbf{b}_i' | \mathbf{y}_{i(M)}, \boldsymbol{\zeta}) = \int \mathbf{b}_i \mathbf{b}_i' \frac{\left( \prod_{j=1, j \in M_i}^{n_i} f_{ij} \right) \mathbf{g}_i}{\int \left( \prod_{j=1, j \in M_i}^{n_i} f_{ij} \right) \mathbf{g}_i d\mathbf{b}_i} d\mathbf{b}_i \quad (5.12)$$

are both evaluated at the maximum likelihood estimates.

Note that the larger the difference in contribution of subject  $i$  to the score functions, the more the corresponding maximum likelihood estimates differ. To take the shape of the log-likelihood function into account, expression (5.6) normalizes the difference in contribution using the variance-covariance matrix of the model parameters.

Expression (5.7) is a function of the residuals and the location in the design space where the responses are measured. This supports the development of optimal design criteria and the search for optimal designs (Ouwens, Tan and Berger, 2001). In the next section the relation between the subject-oriented local influence measure and the observation-oriented local influence measure will be discussed.

### Subject-Oriented Local Influence

For the subject-oriented local influence measure each set  $M_i$  will consist of all observations of subject  $i$ , implying that the function  $l_{i(M_i)}(\mathbf{y}_{i(M_i)} | \boldsymbol{\zeta})$  will be equal to zero. Let  $S_i$  be the set of all observations of subject  $i$ ,  $i = 1, \dots, N$ . If we are only interested in the influence of subjects, expression (5.4) can be rewritten as:

$$\log L(\mathbf{y} | \zeta) + \sum_{i=1}^N \sum_{M_i} (1 - \omega_{M_i}) [ l_{iM_i}(\mathbf{y}_{iM_i} | \zeta) - l_i(\mathbf{y}_i | \zeta) ] \quad (5.13)$$

$$= \log L(\mathbf{y} | \zeta) + \sum_{i=1}^N (1 - \omega_{S_i}) [ -l_i(\mathbf{y}_i | \zeta) ] = \sum_{i=1}^N \omega_{S_i} l_i(\mathbf{y}_i | \zeta).$$

The expression  $\sum_{i=1}^N \omega_{S_i} l_i(\mathbf{y}_i | \zeta)$  was also given by Lesaffre and Verbeke (1998, p. 571) for the Linear Mixed Model. For generalized linear mixed models, the subject-oriented influence measures based on expression (5.13) are special cases of the proposed observation-oriented local influence measures based on expression (5.4). In the next section it will be shown that for the analysed dataset and for both GLMM with known dispersion parameter as well as for models with unknown dispersion parameter, the observation-oriented influence measures give additional information.

## 5.4 An illustration

### *Data description.*

Table 5.1 presents the dataset which will be used to illustrate the proposed influence measures (McKnight and Van Den Eeden, 1993). The presented data set is obtained from a two treatment multiple period crossover trial in which the number of headaches per week is repeatedly measured for 27 patients. In the first period, each patient received the placebo. In the other four periods the patients received either the placebo (P) or the aspartame (A), in random order, using the double-blind crossover treatment design. To wash out the effects of the treatment of the foregoing periods the periods are separated by one day. Hedeker (1999) showed that the sequence in which the placebo and aspartame is given did not matter significantly. Furthermore, he assumed only random intercepts. This formed our motivation to assume that the patients only differed in their intercepts and that the sequence in which the observations are taken was not important. The data are grouped in Table 5.1 according to the use of aspartame and the use of placebo. The last column shows the actual

sequence in which the placebo (P) and the aspartame (A) were given and some counts are missing.

Table 5.1 Number of Headaches of the Patients

Patient	Aspartame		Placebo			Sequence
1	0	0	3	3	1	PAPPA
2	5 *		0	2		PAP
3	2	2	2	3	2	PAPPA
4	0	0	0	0	0	PPAAP
5	0	0	3	2	0	PAPAP
6	1	1	1	0	3	PAPPA
7	4	3	1	1	2	PPAPA
8	0	1	1	1	1 (2)	PPAAP
9	2		2	0	1 (5)	PAPP
10	0	0	0	0	0	PPAPA
11	1	1 (3)	0	0	3	PPAPA
12	5 *	0	0	0	0	PPAAP
13	7	6	7	7	7	PAPAP
14	1	2	2	2	0	PPAPA
15	2	1	3	1	0	PPAAP
16	3		1	1		PAP
17	1 (1)		4			PA
18	1	0	0	1	1	PAPPA
19	1	0	0	1	1	PAPAP
20	1 (2)		1	6		PPA
21	2	6*	1	3	3	PAPPA
22	1	1	2	0	0	PAPPA
23	2	3	7 *	3	2	PAPAP
24	1	2	1	0	0	PPAPA
25*	6	7	1	1	0	PAPAP
26			0	1		PP
27	3	2 (4)	3	3	0	PPAPA

Almost all periods spanned seven days, but some periods were smaller. The number of days in the periods smaller than seven days are given within brackets. The asterisks in Table 5.1 indicate the observations and patients which will be detected to be influential by the subject-oriented and observation-oriented influence measures.

*Description of models.*

The measures for local influence may differ in performance for models with known dispersion parameter and for models with unknown dispersion parameter. Therefore, the data will be analysed with the Poisson mixed model with known dispersion parameter and the negative binomial mixed model with unknown dispersion parameter.

For both the Poisson mixed model and the negative binomial mixed model expression (5.14) was fitted:

$$E( y_{ij} | \beta, b_i ) = t_{ij} \exp( \beta_0 + \text{DrugAsp}_{ij} \beta_1 + b_{0i} ), \quad (5.14)$$

where  $\beta_0$  is the fixed intercept,  $\text{DrugAsp}_{ij}$  indicates whether placebo (0) or aspartame (1) is given to patient  $i$  at the  $j$ th time point,  $\beta_1$  is the corresponding fixed regression parameter,  $b_{0i}$  is the random intercept for patient  $i$  and  $t_{ij}$  is the number of days in the period  $j$  for patient  $i$ . The variance of the random intercept  $b_{0i}$  is denoted by  $\delta^2$ .

The Poisson density is a member of the family of density functions of the form of expression (5.1). The Poisson density is equal to

$$\begin{aligned} f( y_{ij} | \mu_{ij} ) &= \frac{\mu_{ij}^{y_{ij}} \exp( -\mu_{ij} )}{y_{ij}!} \\ &= \exp[ y_{ij} \log( \mu_{ij} ) - \mu_{ij} - \log( y_{ij}! ) ] \end{aligned} \quad (5.15)$$

and can be obtained from expression (5.1) by taking  $a_{ij}( \phi ) = 1$ ,  $\theta_{ij} = \log( \mu_{ij} )$ ,  $\psi( \theta_{ij} ) = \exp( \theta_{ij} )$  and  $c( y_{ij}, \phi ) = -\log( y_{ij}! )$ .

The negative binomial density (McCullagh and Nelder, 1996, p.198) is given by

$$f( y_{ij} | \phi, \mu ) = \frac{\Gamma( y + \phi \mu ) \phi^\phi \mu^\mu}{y! \Gamma( \phi \mu ) ( 1 + \phi )^{y + \phi \mu}}, \quad (5.16)$$

where  $\Gamma$  denotes the gamma function. Although this density is not a member of the exponential density function, the above written theory on influence measures can also be applied to the negative binomial mixed effects model.

The computations for the Poisson mixed model are based on Hedekers (1999) program MIXPREG. The computations for the negative binomial mixed model are made using TOMLAB (<http://www.ima.mdh.se/tom>), the optimization toolbox of MATLAB.

*Influence assessment for the Poisson mixed model.*

Figure 5.1 shows the subject-oriented local influence of the patients on the regression parameters  $\beta_0$ ,  $\beta_1$ , the variance of the random intercept  $\delta$ , the vector of regression parameters  $\beta = (\beta_0, \beta_1)'$  and the vector of all model parameters  $(\beta', \delta)'$ , respectively. The symbols in Figure 5.1 are the patient numbers of the corresponding patients in Table 5.1. Patient 25 has a much larger influence on the regression parameters than the other patients. Table 5.1 shows that patient 25 has much more headaches using aspartame than in the placebo condition.

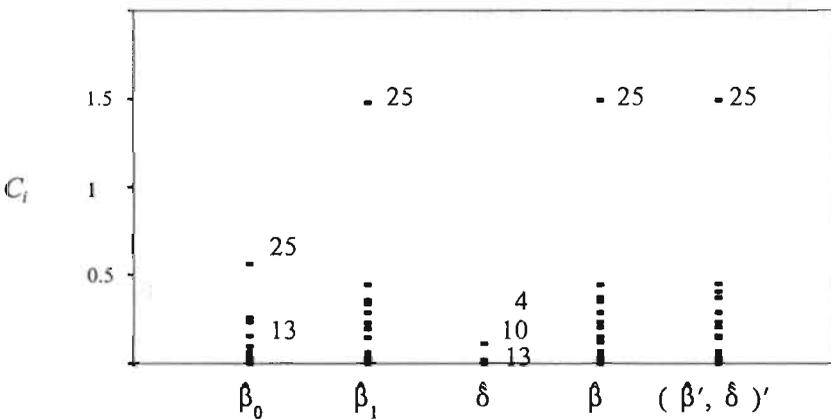


Figure 5.1. Subject-oriented influence measures for five different sets of model parameters for the Poisson random effects model.

After deletion of the observations of patient 25, the subject-oriented influence measures (5.13) did not detect any other influential patients for the regression parameters.

Figure 5.2 shows the values of the observation-oriented influence measures. The symbol  $(a, b)$  in Figure 5.2 refers to the observation in row  $a$  and column  $b$  in Table 5.1. The values of the influence measure for the four displayed influential observations are more than three times larger than the values of the other observations. The influential observations are marked by an asterisk ( $*$ ) in Table 5.1. It can be seen that the influential observations deviate from the subject-specific profiles. For example all observations of patient 12 are equal to 0, except the influential observation itself, which is equal to 5.

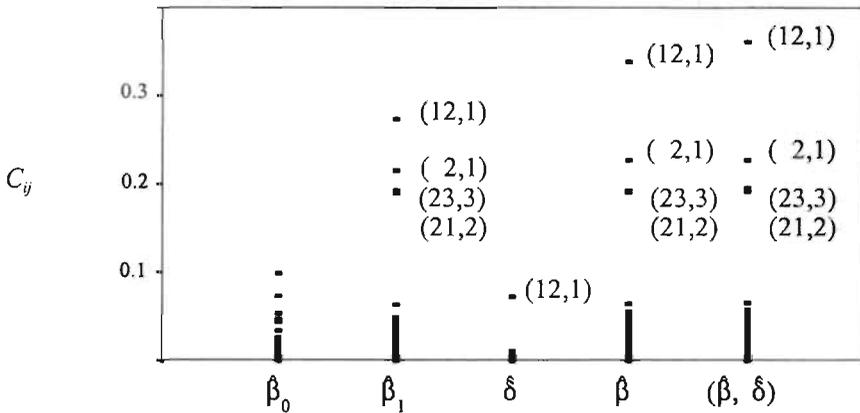


Figure 5.2. Observation-oriented local influence measures for the Poisson random effects model.

To show the effect of the deletion of certain observations and/or patients on the estimated parameters, Table 5.2 displays the estimates of  $\beta_0$ ,  $\beta_1$  and  $\delta$ , based on the sample omitting certain observations and/or patients.

From Table 5.2 it can be obtained that the estimate of  $\beta_1$  remains unstable after the deletion of patient 25. The change in the estimate of  $\beta_1$ , due to the deletion of

patient 25, is 0.13 and is as large as the additional change due to the deletion of the four influential observations. This means that the results still depend heavily on a small part of the data set.

*Table 5.2* Estimated Model Parameters, Based on the Sample with and without Certain Patients or Observations

	$\beta_0$	$\beta_1$	$\delta$
Sample	-1.72	0.28 (0.14)	0.69
Sample without patient 25	-1.70	0.15 (0.14)	0.70
Sample without patient 4,10,13 and 25	-1.61	0.19 (0.16)	0.39
Sample without patient 25 and without the influential observation of patient 12	-1.74	0.09 (0.15)	0.78
Sample without patient 25 and without the influential observation of patient 23	-1.72	0.20 (0.15)	0.68
Sample without patient 25 and without the 3 influential observations of the aspartame group	-1.74	-0.03 (0.16)	0.76
Sample without patient 25 and without the 4 influential observations	-1.77	0.02 (0.16)	0.74

To investigate whether the additional detection of influential observations was partly determined by the fact that the Poisson mixed model assumes a known dispersion, we extended the Poisson mixed model to the negative binomial mixed model and used the subject-oriented influence measures based on the negative binomial mixed model.

*Influence assessment for the negative binomial mixed model.*

The subject-oriented influence values for the estimate of the dispersion parameter  $\phi$ , the estimates of the regression parameters  $\beta_0$  and  $\beta_1$  and the estimate of the variance of the random effects  $\delta$ , separately and together, are given in Figure 5.3.

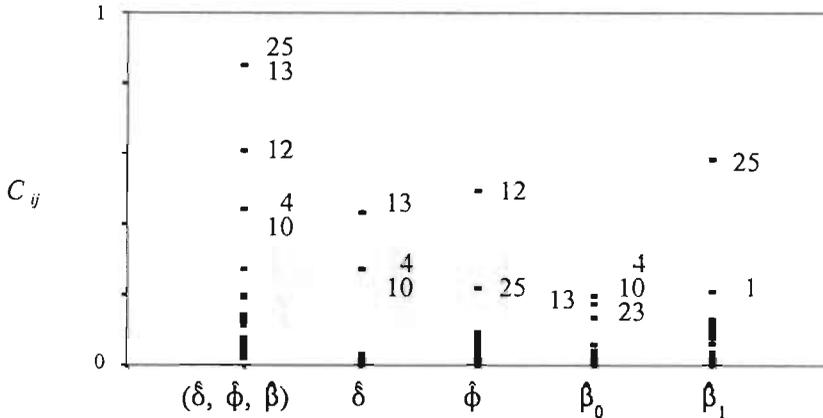


Figure 5.3. Subject-oriented influence measures for the negative binomial random effects model, based on all observations.

Figure 5.3 shows that patients 4, 10 and 13 are again influential for the standard deviation of the random effects  $\delta$ . For the negative binomial mixed model they are also detected to be influential for the estimate of the intercept  $\beta_0$ . Patient 12 is detected to be influential for the estimate of the dispersion parameter  $\phi$ . Patient 23 has a large influence for the estimate of the intercept  $\beta_0$ . Hence, the influential patients 12 and 23, which were not detected by the subject-oriented influence measures under the Poisson mixed model, are detected under the negative binomial mixed model. We deleted sequentially patient 25, patient 13 and the two influential observations of patient 23 and patient 12, but the subject-oriented influence measures did not give any additional information. Figure 5.4 shows the observation-oriented influence measures, after patient 13, patient 25 and the two influential observations of patient 23 and patient 12 are deleted, and it can be seen that again two observations of patient 2 and patient 21

were detected to be influential, while the corresponding patients were not detected at the subject level.

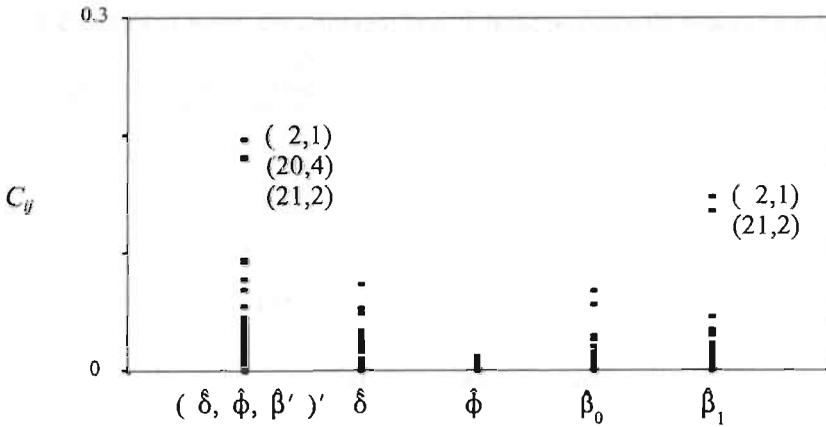


Figure 5.4. Observation-oriented influence measures for the negative binomial random effects model, based on the reduced data set, without patient 13, 25 and observation (12, 1) and (23, 3).

It can be concluded that the observations indicated by an asterisk in Table 5.1 of patient 2 and patient 21 are influential observations, which are not detected to be influential using the subject-oriented influence measures. Therefore, the use of the observation-oriented influence measures can give additional information for models with unknown dispersion parameters.

### 5.5 Discussion

In this chapter the local influence measures are discussed for the generalized linear mixed model. The real data example shows that both the detection of influential observations and the detection of influential subjects are important. It shows that subjects may have a large influence while none of their observations is detected to be influential. On the other hand, the example also shows that influential observations

may be distributed across subjects which are not detected to be influential by the subject-oriented influence measures. If we only use subject-oriented local influence measures the estimates of the parameters may heavily depend on the remaining small number of influential observations. Consequently, if one of these measures is not used, a possible instability of the parameter estimates remains undetected.

Although the proposed procedure to detect influential observations and subjects can be applied to more complex data structures, we only evaluated single subjects or observations. When more complex data structures are evaluated, the number of units in the evaluated data structures should be comparable across the evaluated data structures, because the amount of influence depends on the number of units in the sets. The evaluation of more complex data structures did not give any additional information for our example.

It should be noted that the weights of the weighted log-likelihood function are taken to be equal to 0 or 1. This is a combinatorial restriction. Other weights may also be evaluated. Finally, although models assuming, for example, serial correlations are not discussed, the weighted log-likelihood function in expression (5.4) can also be used for those models.

## Appendix

The expression for the contributions of subject  $i$  to the score functions for the model parameters are given in this appendix.

If the conditional density of the responses of subject  $i$ ,  $i = 1, \dots, N$  is given by

$$f_i \equiv \prod_{j=1}^{n_i} \exp \left[ \frac{y_{ij} \theta_{ij} - \psi(\theta_{ij})}{a_{ij}(\phi)} + c_{ij}(y_{ij}, \phi) \right]$$

and the density of the random effects is given by  $g = \frac{1}{\sqrt{(2\pi)^p |G|}} \exp(-\mathbf{b}_i' G^{-1} \mathbf{b}_i / 2)$ , then the contribution of subject  $i$  to the log-likelihood function based on the complete sample is equal to

$\log \int f_i g db_i$ . Note that the parameters specifying  $f_i$  are not used to specify  $g$  and vice versa, implying that the vector of model parameters  $\zeta$  can be divided into the vector of parameters  $\zeta_1$ , specifying  $G$ , and the vector of parameters  $\zeta_2$ , specifying  $f_i$ . The derivative of the weighted log-likelihood function with respect to the  $jk$ th element of the variance-covariance matrix of the random effects  $G$ ,  $G_{jk}$ , and  $\omega_{M_i}$  for  $j \neq k$  in  $\theta = \hat{\theta}$  is equal to the sum of the  $jk$ th element and the  $kj$ th element of the matrix

$$\frac{1}{2} \hat{G}^{-1} \left[ E( \mathbf{b}_i \mathbf{b}_i' | \mathbf{y}_i, \zeta ) - E( \mathbf{b}_i \mathbf{b}_i' | \mathbf{y}_{i(M_i)}, \zeta ) \right] \hat{G}^{-1}, \quad (5.17)$$

where  $E( \mathbf{b}_i \mathbf{b}_i' | \mathbf{y}_i, \zeta )$  and  $E( \mathbf{b}_i \mathbf{b}_i' | \mathbf{y}_{i(M_i)}, \zeta )$  are the empirical bayes estimates of the second moment of the random effects of subject  $i$ , based on the complete sample with and without the responses in set  $M_i$ , respectively. The derivative with respect to  $G_{jj}$  is equal to the  $jj$ -th element of this matrix (5.17).

Let  $S_i(\zeta_2)$  and  $S_{i(M_i)}(\zeta_2)$  be the contributions of subject  $i$  to the score functions for  $\zeta_2$ .  $S_i(\zeta_2)$  is then equal to:

$$\int \sum_{j=1}^{n_i} \left\{ \frac{y_{ij} - \mu_{ij}}{a_{ij}(\phi)} \frac{\partial \theta_{ij}}{\partial \zeta_2} \right\} \frac{f_i g}{\int f_i g db_i} db_i + \int \sum_{j=1}^{n_i} \left\{ \left[ \frac{\partial c_{ij}(y_{ij}, \phi)}{\partial \phi} - \frac{y_{ij} \theta_{ij} - \psi(\theta_{ij})}{a_{ij}^2(\phi)} \frac{\partial a_{ij}(\phi)}{\partial \phi} \right] \frac{\partial \phi}{\partial \zeta_2} \right\} \frac{f_i g}{\int f_i g db_i} db_i. \quad (5.18)$$

Consequently,  $S_i(\zeta_2)$  is the empirical bayes estimate of

$$\sum_{j=1}^{n_i} \left\{ \frac{y_{ij} - \mu_{ij}}{a_{ij}(\phi)} \frac{\partial \theta_{ij}}{\partial \zeta_2} \right\} + \sum_{j=1}^{n_i} \left[ \frac{\partial c_{ij}(y_{ij}, \phi)}{\partial \phi} - \frac{y_{ij} \theta_{ij} - \psi(\theta_{ij})}{a_{ij}^2(\phi)} \frac{\partial a_{ij}(\phi)}{\partial \phi} \right] \frac{\partial \phi}{\partial \zeta_2}. \quad (5.19)$$

Analogously,  $S_{i(M_i)}(\zeta_2)$  is the empirical bayes estimate of

$$\sum_{j: y_{ij} \in M_i} \left\{ \frac{y_{ij} - \mu_{ij}}{a_{ij}(\phi)} \frac{\partial \theta_{ij}}{\partial \zeta_2} \right\} + \quad (5.20)$$

$$\sum_{j: y_{ij} \notin M_i} \left| \frac{\partial c_{ij}(y_{ij}, \phi)}{\partial \phi} - \frac{y_{ij} \theta_{ij} - \psi(\theta_{ij})}{a_{ij}^2(\phi)} \frac{\partial a_{ij}(\phi)}{\partial \phi} \right| \frac{\partial \phi}{\partial \zeta_2}$$

based on the responses not in  $M_i$ . It should be noted that if the dispersion  $\phi$  is constant, the second term vanishes. This is the case when, for example, Poisson mixed models and logistic mixed models are used.

## References

- Bannerjee, M., and Frees, E. W. (1997). Influence diagnostics for linear longitudinal models, *Journal of the American Statistical Association*, 92, 999-1005.
- Christensen, R., Pearson, L.M., and Johnson, W. (1992). Case deletion diagnostics for mixed models, *Technometrics*, 34, 38-45.
- Cook, R.D. (1986). Assessment of local influence, *Journal of the Royal Statistical Association, Series B*, 48, 133-169.
- Cook, R.D. , and Weisberg, S. (1982). *Residuals and influence in regression*, New York: Chapman and Hall.
- Diggle, P.J., Liang, K.Y., and Zeger, S.L. (1994). *Analysis of longitudinal data*, Oxford: Oxford Science Publications.
- Hedeker, D. (1999). *Mixpreg manual Mixpcm.pdf*,  
<http://www.uic.edu/~hedeker/mixdos.html>.
- Lesaffre, E., and Verbeke, G. (1998). Local influence in linear mixed models, *Biometrics*, 54, 570-582.
- McCullagh, P., and Nelder, J.A. (1996). *Generalized linear models*, London: Chapman & Hall.

- McKnight, B., and van den Eeden, S.K. (1993). A Conditional Analyses for two-treatment multiple-period crossover designs with binomial or poisson outcomes and subjects who drop out, *Statistics in Medicine*, 12, 825-834.
- Ouwens, M.J.N.M., Tan, F.E.S., and Berger, M.P.F. (2001). On the maximin designs for logistic random effects models with covariates. *Proceedings of the 16th International Workshop on Statistical Modelling*, 321-328, Denmark.
- Preisser, J.S., and Qaqish, B.F. (1996). Deletion diagnostics for generalised estimating equations, *Biometrika*, 83, 551-562.
- Tan, F.E.S., Ouwens, M.J.N., and Berger, M.P.F. (1999). On the use of Cook's Distance for longitudinal regression models with random effects. Statistical modelling, Eds. H. Friedl, A. Berghold, G. Kamermann, proceeding of 14th International Workshop on Statistical Modelling, 686-690, Graz, Austria.
- Tan, F.E.S., Ouwens, M.J.N., and Berger, M.P.F. (2001). Detection of influential observations in longitudinal mixed effects regression models, *Journal of the Royal Statistical Society D*, 50, 271-284.
- Thomas, W., and Cook, R.D. (1989). Assessing influence on regression coefficients in generalized linear models, *Biometrika*, 76, 741-749.
- Thomas, W., and Cook, R.D. (1990). Assessing influence on predictions from generalized linear models, *Technometrics*, 32, 59-65.

# **6 Maximin design and local influence in a health education intervention study**

## Abstract

This chapter discusses two methods to improve the estimation process for health education intervention studies with individuals nested within clusters. The first method concerns the choice of the design. A maximin criterion will be used to find a combination of treatment and control schools which maximizes the smallest possible relative efficiency not only with respect to all possible values of the model parameters, but also with respect to the unknown scores on the covariates in the model. The second method concerns the detection of influential data. Based on the detected influential data, we decided to study both the change in the estimated parameters, due to the deletion of schools and pupils, as well as the change in the estimated parameters, due to the addition of a random effect to the model. The previously given two methods to improve the estimation process will be illustrated based on a Dutch school-based intervention study, with pupils nested within schools.

*Key words:* optimal design, maximin, covariates, influential data, local influence

## 6.1 Introduction

In this chapter, the data obtained from the project ‘teeth-brushing on the elementary school’ (Terstegge, Schefman and Thijs, 2000) will be used to illustrate two approaches to improve the estimation process. The objective of the project ‘teeth-brushing on the elementary school’ was to stimulate the pupils to brush their teeth at least twice a day to prevent for caries. The effect of the school-based intervention on the number of times the pupils brush their teeth *at home*, will be considered in this chapter. Additional interest is in the effect of the social-economic status of the pupils.

Two methods to improve the quality of the estimated parameters and their variances will be discussed, namely the use of more efficient designs and the detection of influential subjects and observations.

The choice of an optimal design can lead to an improvement of the quality of the estimated parameters before the data are actually collected. Often referred books on optimal design theory are written by Atkinson and Donev (1996), Bunke and Bunke (1986) and Pukelsheim (1993).

The detection of influential data can also improve the estimation process. The estimates of the model parameters based on the sample may be very different from the estimates of the model parameters based on the sample without the influential data. An introductory book about residuals and influential data for ordinary least squares is written by Cook and Weisberg (1982). A review of influence measures is written by Chatterjee and Hadi (1986). Influence measures for the linear mixed model are discussed by Lesaffre and Verbeke (1998) and Tan, Ouwens and Berger (1999, 2000, 2001). Hosmer and Lemeshow (1989) discuss tools for the exploration of data for the logistic model. Influence measures for the generalized linear mixed model, with the Poisson mixed model and the logistic mixed model as special applications, are given by Ouwens, Tan and Berger (1999, 2000, 2001a).

In section 6.2 the project ‘teeth-brushing on the elementary school’ is described in more detail. The models which are considered in this chapter are given in section 6.3. The optimal design mechanisms are discussed in section 6.4, while methods to assess the influence of schools and pupils on the estimated model parameters are described in section 6.5. Finally, conclusions are drawn in section 6.6.

## 6.2 Teeth-brushing on the elementary school

The objective of the project ‘teeth-brushing on the elementary school’ is to stimulate the pupils to brush their teeth at least *twice* a day. The research units in this project are 733 pupils, who are nested within 18 schools. Each pupil is measured twice, in grade 3 and 5 or in grade 4 and 6. Additional measurements will be taken in grade 7 and 8 and, possibly later, in a follow-up study. The randomization of the treatment is done at school-level, although this seems to be less efficient than randomization at pupil level (Moerbeek, van Breukelen and Berger, 2001a, b). Within the treatment schools, the pupils brush their teeth once a day at school. Within the control schools, the pupils do not brush their teeth at school.

The first measure of each pupil will be used as baseline measure, to indicate whether a pupil already sufficiently brushed his/her teeth at the start of the research period. The dependent variable is based on the second measure and indicates whether a pupil brushes his/her teeth at least twice a day at home in grade 5 or 6. Additional interest is in the difference between the treatment effect for pupils with low social-economic status and the treatment effect for pupils with high social-economic status.

The dependent and independent variables are operationalized as follows:

- The dependent variable PREV is equal to  $PREV = 1$ , if a pupil brushes his/her teeth at least twice a day at home.  $PREV = 0$ , otherwise.
- The independent variable TREAT is equal to  $TREAT = 1$  for the treatment schools.  $TREAT = 0$ , otherwise.

- The independent variable SES (Social-economic Status) is equal to  $SES = 1$  for pupils with high social-economic status.  $SES = 0$ , otherwise.
- The baseline variable AP (Already Prevented) is equal to  $AP = 1$  for pupils who already sufficiently brushed their teeth at the start of the research period.  $AP = 0$ , otherwise.

Table 6.1 gives a summary of the data obtained from the project ‘teeth-brushing on the elementary school’.

Table 6.1 Summary of data

	Treatment schools							Control schools											Mean
$ID_i$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
$n_i$	61	36	38	53	72	69	41	44	49	23	39	31	33	27	21	24	43	18	40 (.16)
$P_{insuff, high_i}$	.20	.36	.24	.24	.21	.13	.27	.23	.14	.04	.23	.29	.33	.33	.23	.17	.34	.11	.23 (.09)
$P_{insuff, low_i}$	.11	.17	.00	.34	.05	.04	.05	.09	.10	.22	.05	.06	.00	.00	.10	.12	.19	.11	.10 (.09)
$P_{suff, high_i}$	.66	.39	.65	.23	.64	.66	.54	.50	.52	.35	.64	.59	.58	.56	.53	.27	.21	.39	.51 (.14)
$P_{suff, low_i}$	.03	.08	.11	.19	.10	.17	.14	.18	.24	.39	.08	.07	.09	.11	.14	.44	.26	.39	.16 (.12)
$P_{1 insuff, high_i}$	.33	.31	.44	.46	.33	.67	.27	.40	.57	1	.44	.33	.36	.44	.60	.50	.53	.00	
$P_{1 insuff, low_i}$	.43	.50	-	.61	.75	1	.50	.50	.80	1	1	.00	-	-	.50	.67	.38	.50	
$P_{1 suff, high_i}$	.83	1	1	.83	.85	.98	.86	1	.88	1	1	1	1	1	1	.94	1	1	
$P_{1 suff, low_i}$	1	1	1	.50	1	.92	.67	1	1	.67	.67	1	1	1	1	1	1	.71	

\*Standard deviation within parentheses

The first row of Table 6.1 gives the identification numbers ID of the schools. The numbers  $n_i$  of pupils sampled from each school are given in the second row. The third through sixth row of Table 6.1 give the proportions  $P_{insuff, high_i}$ ,  $P_{insuff, low_i}$ ,  $P_{suff, high_i}$  and  $P_{suff, low_i}$  corresponding to the strata obtained by cross-classification of the variables SES

(high/low) and AP (insufficient/sufficient). The means of these proportions are given in the last column, where the corresponding standard deviations are given within brackets. The seventh through tenth row of Table 6.1 give the conditional proportions  $P_{1|insuff, high}$ ,  $P_{1|insuff, low}$ ,  $P_{1|suff, high}$  and  $P_{1|suff, low}$  of pupils who brush their teeth at least twice a day at home at the second measurement, given their scores on SES and AP. Finally, Table 6.1 shows that schools 1 through 7 obtained a treatment, while schools 8 through 18 act as control schools. An elaborated description of the project ‘teeth-brushing on the elementary school’ can be found in the report by Terstegge et al. (2000).

### 6.3 Model specification

The full model which we want to consider is given by

$$\begin{aligned} \text{logit}(\pi_{ij}) = & \beta_0 + \beta_1 \text{TREAT}_i + \beta_2 \text{AP}_{ij} + \beta_3 \text{SES}_{ij} + \\ & + \beta_4 \text{TREAT}_i \times \text{AP}_{ij} + \beta_5 \text{TREAT}_i \times \text{SES}_{ij} + \beta_6 \text{AP}_{ij} \times \text{SES}_{ij} + \\ & + \beta_7 \text{TREAT}_i \times \text{AP}_{ij} \times \text{SES}_{ij} + b_{0i} + b_{1i} \text{AP}_{ij}. \end{aligned} \quad (6.1)$$

For model (6.1), the logit of the conditional probability  $\pi_{ij}$  of the response  $\text{PREV} = 1$  for pupil  $j$  of school  $i$ ,  $j = 1, \dots, n_i$ ,  $i = 1, \dots, N$ , is ascribed to the main effects of the independent variables TREAT, AP and SES, to the two-way interactions and the three-way interaction between these variables and to the school-specific deviations. Model (6.1) can be motivated as follows.

The objective of the project ‘teeth-brushing on the elementary school’ is to stimulate the pupils to brush their teeth sufficiently. The treatment supports the behaviour of pupils who already sufficiently brushed their teeth at the start of the research period, while it motivates pupils who did not sufficiently brush their teeth at the start of the research period to brush their teeth more often. The effect of the treatment will therefore

be different for the two strata of AP. Hence, the interaction between the treatment and AP is incorporated in the model.

Because the difference between the treatment effects for the two social-economic strata is of primary interest and the treatment effect is different for the two strata indicated by AP, we decided to study the difference in treatment effect for the two strata of SES per stratum of AP. As a consequence, the three-way interaction  $SES \times AP \times TREAT$  together with the hierarchically lower ordered interactions and main effects are incorporated in model (6.1).

The score of the pupils within each school may be influenced by school-specific learning programs and other school-specific differences. To model these influences, the school-specific parameters  $b_{0i}$  and  $b_{1i}$  are used in model (6.1). Note that model (6.1) is a random effects model, because the school-specific effects  $b_{0i}$  and  $b_{1i}$  are assumed to be random.

Model (6.1) can be written as follows:

$$AP = 1: \tag{6.2}$$

$$\text{logit}(\pi_{ij}) = \beta_{01} + \beta_{11} TREAT_i + \beta_{21} SES_{ij} + \beta_{31} TREAT_i \times SES_{ij} + b_{01i}$$

$$AP = 0:$$

$$\text{logit}(\pi_{ij}) = \beta_{00} + \beta_{10} TREAT_i + \beta_{20} SES_{ij} + \beta_{30} TREAT_i \times SES_{ij} + b_{00i},$$

where  $\pi_{ij}$  is the probability of an increase for pupil  $j$  in school  $i$ ,  $\beta_{01}$  and  $\beta_{00}$  are the intercepts corresponding to  $AP = 1$  and  $AP = 0$ ,  $\beta_{11}$  and  $\beta_{10}$  are the main effects for the treatment,  $\beta_{21}$  and  $\beta_{20}$  are the main effects for the social-economic status of the pupils and  $\beta_{31}$  and  $\beta_{30}$  are the interaction effects between the treatment and the social-economic status, respectively. The vectors  $\mathbf{b} = (b_{01i}, b_{00i})'$  of random intercepts are assumed to be independently normally distributed over schools with mean zero and variance-covariance matrix  $D$ . Based on model (6.2), the vector of all fixed regression parameters  $\beta$  is given by

$$\beta = (\beta_{01}, \beta_{11}, \beta_{21}, \beta_{31}, \beta_{00}, \beta_{10}, \beta_{20}, \beta_{30})'. \tag{6.3}$$

Note that model (6.2) is given conditional on the observed values of the variable AP, i.e. based on the teeth-brushing behaviour of the pupils at the start of the research period. As a consequence, the interactions between the independent variables and AP cannot easily be tested. Because these interactions were not of primary interest, model (6.2) is considered in the following sections.

## 6.4 Optimal design

The variance of the maximum likelihood estimate  $\hat{\beta}$  of the vector of regression parameters  $\beta$  depends on the values of  $\beta$ , the variance-covariance matrix  $D$  of the random effects, the proportion treated schools  $w$  and the joint distribution of the variables SES and AP within the schools (Ouwens, Tan and Berger, 2001b). The proportion treated schools  $w$  is specified by the researcher. The number of pupils  $n_i$  and the proportions  $P_{insuff, low_i}$ ,  $P_{suff, low_i}$ ,  $P_{insuff, high_i}$  and  $P_{suff, high_i}$  within the schools are, however, unknown and cannot easily be fixed in the design stage. As a consequence, we propose to optimize the choice of the proportion treated schools  $w$ , taking both the values of  $\beta$  and  $D$  as well as the number of pupils  $n_i$  and the proportions  $P_{insuff, low_i}$ ,  $P_{suff, low_i}$ ,  $P_{insuff, high_i}$  and  $P_{suff, high_i}$  within the schools into account (Ouwens, Tan and Berger, 2001b). In the following subsection, the commonly used D-optimality criterion and the relative efficiency will be given, which are used to define the maximin criterion.

### *D-optimality criterion.*

For the teeth-brushing project, the D-optimality criterion searches for that proportion of treated schools  $w^*$ , for which the determinant  $\text{Det}[\text{Var}(\hat{\beta})]$  of the variance-covariance matrix of the maximum likelihood estimate  $\hat{\beta}$  is minimized, given  $\beta$ ,  $D$ ,  $n_i$ ,  $P_{insuff, low_i}$ ,  $P_{suff, low_i}$ ,  $P_{insuff, high_i}$  and  $P_{suff, high_i}$ . The D-optimality criterion has a natural interpretation. It searches for that proportion treated schools  $w^*$ , for which the approximated volume of the confidence ellipsoid is minimized, where the confidence ellipsoid is the analogue of the confidence interval for vectors of estimates (Johnson and

Wichern, 1988, p. 103). The determinant  $\text{Det}[\text{Var}(\hat{\beta})]$  is also called the generalized variance of  $\hat{\beta}$  (Anderson, 1958, p. 166). The inverse of Fisher's information matrix is used in this chapter as approximation of the variance matrix  $\text{Var}(\hat{\beta})$ .

Given the values of  $\beta$ ,  $D$ , all numbers of pupils  $n_i$  and all proportions  $P_{insuff, low_i}$ ,  $P_{suff, low_i}$ ,  $P_{insuff, high_i}$  and  $P_{suff, high_i}$ , the generalized variance  $\text{Det}[\text{Var}(\hat{\beta}_w)]$  under a particular proportion  $w$  can be compared with the generalized variance  $\text{Det}[\text{Var}(\hat{\beta}_{w^*})]$  under the D-optimal proportion  $w^*$  by means of the relative efficiency

$$\text{eff}(w | w^*) = \left\{ \frac{\text{Det}[\text{Var}(\hat{\beta}_{w^*})]}{\text{Det}[\text{Var}(\hat{\beta}_w)]} \right\}^{1/p}, \quad (6.4)$$

where  $\hat{\beta}_w$  and  $\hat{\beta}_{w^*}$  are the maximum likelihood estimates of  $\beta$  corresponding to  $w$  and  $w^*$ , respectively. The relative efficiency is defined in such a way that if, for example,  $\text{eff}(w | w^*) = .90$  for a particular proportion  $w$ , then 10% less observations are needed under the D-optimal proportion  $w^*$  to obtain the same generalized variance as under  $w$ , given the same  $\beta$ ,  $D$ , numbers of pupils  $n_i$  and proportions  $P_{insuff, low_i}$ ,  $P_{suff, low_i}$ ,  $P_{insuff, high_i}$  and  $P_{suff, high_i}$ . This relative efficiency is used to define the maximin criterion.

### Maximin criterion.

For the D-optimality criterion, the vector of regression parameters  $\beta$ , the variance-covariance matrix  $D$ , all numbers of pupils  $n_i$  and all proportions  $P_{insuff, low_i}$ ,  $P_{suff, low_i}$ ,  $P_{insuff, high_i}$  and  $P_{suff, high_i}$  are assumed to be known. Unfortunately, these quantities are unknown in advance. We therefore propose to search for a proportion treated schools  $w$ , which is highly efficient under all likely values of  $\beta$ ,  $D$ ,  $n_i$ ,  $P_{insuff, low_i}$ ,  $P_{suff, low_i}$ ,  $P_{insuff, high_i}$  and  $P_{suff, high_i}$ , i.e. we propose to search for a proportion treated schools  $w$  for which the minimal relative efficiency is large. The maximin criterion is based on this principle and is defined as follows.

For each proportion of treated schools  $w$ , the corresponding minimal relative efficiency can be computed over all likely values of  $\beta$ ,  $D$ , numbers of pupils  $n_i$  and proportions  $P_{insuff, low_i}$ ,  $P_{suff, low_i}$ ,  $P_{insuff, high_i}$  and  $P_{suff, high_i}$ . The minimal relative

efficiencies can be compared with each other, to obtain the proportion  $w^{maximin}$  with the largest minimal relative efficiency. Such a proportion  $w^{maximin}$  is said to be maximin. In other words, the maximin criterion leads to a proportion  $w^{maximin}$ , for which the smallest relative efficiency

$$\min \text{eff}(w | w^*) \tag{6.5}$$

over all likely values of  $\beta$ , all likely variance-covariance matrices  $D$  and all likely numbers of pupils  $n_i$  and proportions  $P_{insuff, low_i}$ ,  $P_{suff, low_i}$ ,  $P_{insuff, high_i}$  and  $P_{suff, high_i}$  is maximized. Note that if the corresponding minimal relative efficiency is large, then we have found a proportion of treated schools  $w$ , which is highly efficient, irrespective of the values of  $\beta$ ,  $D$ ,  $n_i$ ,  $P_{insuff, low_i}$ ,  $P_{suff, low_i}$ ,  $P_{insuff, high_i}$  and  $P_{suff, high_i}$ .

The sign and the size of the effect of the school-based intervention project on the teeth-brushing behaviour of the pupils at home are both unknown. Further, the set of likely joint distributions of  $n_i$ , SES and AP for both treatment groups are the same. The theorem of Ouwens, Tan and Berger (2001b) can therefore be applied to state that the maximin optimal choice of the proportion of treated schools  $w$  is equal to .50 or, equivalently, to state that half of the schools should be treated and half of the schools should be selected as controls. To obtain the possible loss in relative efficiency, the different ranges of the parameters and the likely distributions of the independent variables must be specified. To illustrate this loss in relative efficiency, an example will be given in the next subsection.

*Ranges of the parameters and distributions of  $n_i$ ,  $P_{insuff, low_i}$ ,  $P_{suff, low_i}$ ,  $P_{insuff, high_i}$  and  $P_{suff, high_i}$ .*

Based on Terstegge et al. (2000) and personal communication, we assume that for all schools  $n_i = 60$ ,  $P_{insuff, low_i} = .40 \times (1 - p_{high_i})$ ,  $P_{suff, low_i} = .60 \times (1 - p_{high_i})$ ,  $P_{insuff, high_i} = .30 \times p_{high_i}$  and  $P_{suff, high_i} = .70 \times p_{high_i}$ , so that  $p_{high_i}$  is the only proportion which is unknown in advance.

At population level, the proportion pupils with high social-economic status  $p_{high}$  is approximately normally distributed with mean .55 and standard deviation .20 (Terstegge et al., 2000). The sample distribution of  $p_{high_i}$  for the treatment schools and the sample distribution of  $p_{high_i}$  for the control schools are therefore assumed to be normal, with means varying from .50 to .60 and standard deviations varying from .10 to .30. The variance-covariance matrix of the random intercepts  $D$  is taken to be diagonal with elements between 0 and 2. These values 0 and 2 are chosen to cover all values that are likely to occur in practice.

It follows from Terstegge et al. (2000) that  $PREV = 1$  for about half of the pupils for whom  $AP = 0$ . We therefore consider  $\beta_{00} \in [-1, 1]$ ,  $\beta_{20} \in [-1, 1]$ ,  $\beta_{00} + \beta_{10} \in [-1, 1]$  and  $\beta_{20} + \beta_{30} \in [-1, 1]$ . In terms of probabilities and ORs, the ranges of the parameters can be interpreted as:

- The probability that pupils with low social-economic status ( $SES = 0$ ) brush their teeth at least twice a day at the second measurement ( $PREV = 1$ ) lies for all schools between:

$$\min \{ [1 + \exp(-\beta_{00})]^{-1} \} = \min \{ [1 + \exp(-(\beta_{00} + \beta_{10}))]^{-1} \} = [1 + \exp(1)]^{-1} = .27 \text{ and}$$

$$\max \{ [1 + \exp(-\beta_{00})]^{-1} \} = \max \{ [1 + \exp(-(\beta_{00} + \beta_{10}))]^{-1} \} = [1 + \exp(-1)]^{-1} = .73.$$

- The OR for pupils with high social-economic status ( $SES = 1$ ) with respect to pupils with low social-economic status ( $SES = 0$ ) lies for all schools between:

$$\min \exp(\beta_{20}) = \min \exp(\beta_{20} + \beta_{30}) = \exp(-1) = .37 \text{ and}$$

$$\max \exp(\beta_{20}) = \max \exp(\beta_{20} + \beta_{30}) = \exp(1) = 2.7.$$

We expect that only a small group of the pupils who sufficiently brushed their teeth at the beginning of the research period ( $AP = 1$ ) will brush their teeth less than twice a day at home at the second measurement ( $PREV = 0$ ). We therefore assume that  $\beta_{01} \in [2, 4]$ ,  $\beta_{21} \in [-1.5, 1.5]$ ,  $\beta_{01} + \beta_{11} \in [2, 4]$  and  $\beta_{21} + \beta_{31} \in [-1.5, 1.5]$ . In terms of probabilities and ORs, the ranges of parameters can be interpreted as:

- The probability that pupils with low social-economic status ( $SES = 0$ ) brush their teeth at least twice a day at the second measurement ( $PREV = 1$ ) lies for all schools between:

$$\min \{ [ 1 + \exp ( - \beta_{01} ) ]^{-1} \} = \min \{ [ 1 + \exp - ( \beta_{01} + \beta_{11} ) ]^{-1} \} \\ = [1+\exp(-2)]^{-1} = .88 \text{ and}$$

$$\max \{ [ 1 + \exp ( - \beta_{01} ) ]^{-1} \} = \max \{ [ 1 + \exp - ( \beta_{01} + \beta_{11} ) ]^{-1} \} \\ = [1+\exp(-4)]^{-1} = .98.$$

- The OR for pupils with high social-economic status ( $SES = 1$ ) with respect to pupils with low social-economic status ( $SES = 0$ ) lies for all schools between:

$$\min \exp ( \beta_{21} ) = \min \exp ( \beta_{21} + \beta_{31} ) = \exp( -1.5 ) = .22 \text{ and}$$

$$\max \exp ( \beta_{21} ) = \max \exp ( \beta_{21} + \beta_{31} ) = \exp( 1.5 ) = 4.5.$$

*Maximin proportion of treated schools  $w$  and corresponding minimal relative efficiency.*

The maximin proportion of treated schools is equal to  $w^{maximin} = .5$ . Thus it is maximin optimal to take the number of treatment schools equal to the number of control schools. The computed minimal relative efficiency for model (6.2) is equal to

$\min [ \text{eff} ( w^{maximin} | w^* ) ] = .999$ , where for each vector of model parameters  $\theta$  and each value of  $p_{high_i}$ , the proportion  $w^*$  is the  $D$ -optimal proportion treated schools given  $\theta$  and  $p_{high_i}$ .

If the interaction between the variables  $SES$  and  $TREAT$  and the three-way interaction are not incorporated in the model, then the minimal relative efficiency decreases to  $\min [ \text{eff} ( w^{maximin} | w^* ) ] = .94$ . In other words, although it is nearly  $D$ -optimal for model (6.2) to take the number of treatment schools equal to the number of control schools, this does not hold for model (6.2) without the interaction terms.

In the teeth-brushing project, the number of treatment schools is equal to 7, while the total number of schools is equal to 18. The smallest relative efficiency for  $w = 7/18 = .39$  may therefore be of interest and is equal to  $\min [ \text{eff} ( w | w^* ) ] = .95$  for the model with the interaction between the variables  $SES$  and  $TREAT$  and is equal to  $\min [ \text{eff} ( w | w^* ) ] = .79$  for the model without the interaction between the variables  $SES$  and  $TREAT$ . In other words, the maximal loss in relative efficiency for  $w = .39$  is

5% larger for the model with interaction between SES and TREAT and 16% larger for the model without this interaction.

### 6.5 Influential schools and pupils

The estimates of the model parameters based on all data may be very different from the estimates of the model parameters based on the data without the influential schools and pupils. Hence, a second way to improve the estimation process is the detection of influential schools and pupils. The detection of influential schools and pupils may lead to the deletion of them, but may also lead to a change of the model. Moreover, it may lead to a better understanding of the stability of the parameter estimates.

To assess the influence of a set of schools or pupils  $M$ , the parameters can be estimated by means of the data with and without the set  $M$ . The difference between the obtained estimates can be compared with the differences found for other sets. If only a single parameter is of interest, then the differences can be ordered based on their absolute values. If vectors of parameters are of interest, then the differences are more-dimensional and their ordering will have to be based on a more complex criterion (Cook and Weisberg, 1982).

To assess the influence of sets of schools and/or pupils, the Likelihood Displacement can be used (Cook and Weisberg, 1982, Ouwens, Tan and Berger, 1999, 2000, 2001a). This measure compares the likelihood  $L(\hat{\beta}, \hat{D})$  of the sample under  $\beta = \hat{\beta}$  and  $D = \hat{D}$  with the likelihood  $L(\hat{\beta}_{(M)}, \hat{D}_{(M)})$  of the sample under  $\beta = \hat{\beta}_{(M)}$  and  $D = \hat{D}_{(M)}$  and is given by:

$$LD_{(M)} = 2[ \log L(\hat{\beta}, \hat{D}) - \log L(\hat{\beta}_{(M)}, \hat{D}_{(M)}) ], \quad (6.6)$$

where  $\hat{\beta}_{(M)}$  is the maximum likelihood estimate of the vector of regression parameters  $\beta$  based on the sample without the set of schools or pupils  $M$  and  $\hat{D}_{(M)}$  is the maximum likelihood estimate of the variance-covariance matrix of the random effects  $D$  based on

the data without the set of schools or pupils  $M$ . If the Likelihood Displacement in equation (6.6) is large, then the sample is much better fitted under  $\hat{\beta}$  and  $\hat{D}$  than under  $\hat{\beta}_{(M)}$  and  $\hat{D}_{(M)}$ . This is only the case when  $\hat{\beta}$  and  $\hat{D}$  are very different from  $\hat{\beta}_{(M)}$  and  $\hat{D}_{(M)}$ . Hence, a large Likelihood Displacement indicates a large difference between the estimates. The Likelihood Displacement can be modified to accommodate situations in which only a part of the model parameters is of interest (Ouwens, Tan and Berger, 1999, 2000, 2001a).

The evaluation of the Likelihood Displacement may be extremely time consuming. Therefore, a second order Taylor expansion is used as an approximation of the Likelihood Displacement (Ouwens, Tan and Berger, 1999, 2000, 2001a). This approximation is called the local influence measure.

*Results for pupils for whom AP = 0.*

Figure 6.1 shows the influence of the schools on the estimated regression parameters  $\beta_0, \beta_1, \beta_2, \beta_3$  and the estimated vector of regression parameters  $\beta$ .

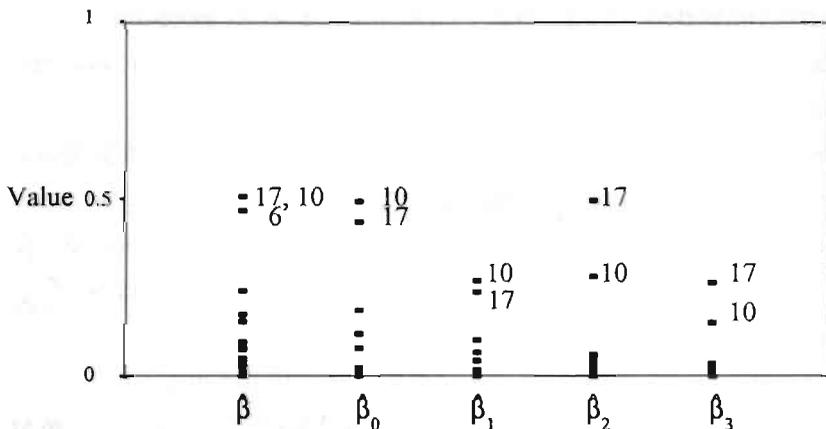


Figure 6.1. Influence of schools.

The symbols in Figure 6.1 are the identification numbers ID of the corresponding schools in Table 6.1. School 10 and 17 are both detected to be influential. These schools may be influential, due to school-specific characteristics, but also due to the influence of

a subgroup of the observations within the schools. Whether the influence of school 10 and 17 can be ascribed to the influence of a subgroup of the pupils within the schools cannot be obtained from Figure 6.1.

To assess whether the influence of school 10 and 17 can be ascribed to a subgroup of the pupils within school 10 and 17, the pupil-oriented influence measures developed by Ouwens, Tan and Berger (1999, 2000, 2001a) can be used. A natural way to subdivide the pupils in subgroups within the schools is to subdivide them based on their social-economic status (cf. Hosmer and Lemeshow, 1989, ch. 5).

It appeared that the influence of school 10 can almost fully be ascribed to the influence of the pupils within the stratum for which  $SES = 1$ . This may be caused by the fact that almost all pupils within school 10 with  $AP = 0$  has a high social-economic status. We therefore decided to evaluate the change in the parameters, due to the deletion of all pupils within school 10. The pupil-oriented influence measures showed that the influence of school 17 can be ascribed to the stratum for which  $SES = 0$ .

Proceeding the detection of influential schools and pupils, we obtained the set of estimates given in Table 6.2. The estimate of the variance of the random intercept  $\delta$  is in all cases equal to  $\hat{\delta} = 0$ .

After the influential schools and strata are detected, we can proceed with the detection of influential pupils. The influence of pupils is related to the estimated probability. If, for example, a pupil scores  $PREV = 1$ , while the corresponding estimated probability is small, than that pupil will have a large influence (Hosmer and Lemeshow, 1989). However, even for small probabilities, there are pupils which will score  $PREV = 1$ , if the data set is large enough. Plotting the values of the influence measures against the probability will give an indication of the influence of pupils (Hosmer and Lemeshow, 1989). For the logistic random effects model, both the conditional probability as well as the marginal probability can be used. For the teeth-brushing data, however, we did not detect any influential pupil.

Table 6.2 Estimates of parameters, based on the sample with and without certain detected schools and strata for  $AP = 0$

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
Sample	.45 (.37)	-.004 (.04)	-.59 (.38)	-.29 (.45)
Sample without low SES of 17	.35 (.37)	.028 (.04)	-.91 (.44)	-.002 (.49)
Sample without low SES of 17 and without school 10	.29 (.37)	.019 (.04)	-.75 (.45)	-.052 (.50)
Sample without low SES of 17 and without 6 and 10.	.26 (.38)	.009 (.04)	-.61 (.45)	-.28 (.51)
Sample without low SES of 17 and without 6, 10 and low SES of school 12	.27 (.38)	-.023 (.04)	-.78 (.46)	-.16 (.52)

It follows from Table 6.2 that the stability of the parameter estimates is relatively large. For the estimates of the parameters for  $AP = 1$ , however, this is not the case.

*Results for pupils for whom  $AP = 1$ .*

Table 6.3 shows the estimates based on the sample with and without the detected influential schools and strata for  $AP = 1$ . It shows that the estimates of the interaction effect  $\hat{\beta}_3$  and the main effect  $\hat{\beta}_2$  remarkably change, if all pupils of school 9 with low social-economic status are deleted. If all pupils of school 16 with low social-economic status are also deleted, then  $p_{suff, high_i} = 1$  for all control schools. This means that, after deletion, model (6.2) is not identified.

Table 6.3 Estimates of parameters, based on the sample with and without certain detected schools and strata for AP = 1

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\delta}$
Sample	2.47 (.53)	-.72 (.69)	1.54 (.45)	-.87 (.71)	.81 (.87)
Sample without low SES of school 9	2.45 (.59)	-.63 (.79)	3.00 (1.19)	-2.34 (1.40)	.97 (1.10)
Sample without low SES of school 9 and low SES of school 16	Model is not identified				

## 6.6 Discussion

In this chapter, the project ‘teeth-brushing on the elementary school’ is used to illustrate two methods to improve the estimation process. The first method concerned the choice of the proportion of treated schools  $w$ . A maximin criterion is used. It appears that  $w = .5$  maximizes the smallest possible relative efficiency not only with respect to all possible values of the model parameters, but also with respect to the unknown scores on the covariates in the model. The second method concerned the detection of influential data. Based on the influence assessment, we obtained that for the model for AP = 0, the deletion of influential data does not lead to large changes in the estimates, while for the model for AP = 1, the estimates are very unstable.

The treatment may be interpreted by the pupils as a replacement of one time teeth-brushing at home. We may therefore expect that during the treatment period, the pupils in the treatment group will brush their teeth less often at home than the pupils in the control schools. As a consequence, the treatment effect is underestimated. Note that if we had modelled the total number of times the pupils brush their teeth, then the treatment effect would be overestimated. In other words, the treatment is associated with the dependent variable such that an unbiased estimate of the treatment effect cannot be obtained from the measurements taken during the treatment period. Hence, it makes sense

to extend the project 'teeth-brushing on the elementary school' with a follow-up study, to obtain an unbiased estimate of the treatment effect.

The distribution of the number of pupils  $n_i$  significantly differs for the two treatment groups. The (conservative) Mann-Whitney U-test statistic is equal to 11.0 and has a  $p$ -value of  $p = .011$ . Further, the overlap between the two distributions is small (between 36 and 49 pupils per school) and only contains 7 schools. It may therefore be concluded that the random treatment assignment to schools did not result in equivalent treatment and control groups.

Under the restrictions given above, the choice of equally sized treatment groups turns out to be optimal according to the theorem of Ouwens, Tan and Berger (2001b). The smallest relative efficiency of the balanced design for model (6.2) is equal to .999. The smallest relative efficiency of the balanced design for model (6.2), when the interactions between SES and TREAT are not incorporated, is equal to .94. Only a subset of the set of possible conditional joint distributions of the number of pupils, the baseline value and the social-economic status of the pupils within the schools is evaluated. Hence, the smallest relative efficiencies that can be encountered in practice may be smaller.

The detection of influential data is used as a tool to explore the data set. In section 6.5 we stated that the stability of the estimates is large for  $AP = 0$ . To assess the stability of the estimates, the range of the estimates is compared with the variance of the estimates. This comparison is, however, somewhat subjective.

The detection of influential schools for  $AP = 0$  is questionable, because the random intercept variance is estimated to be equal to  $\hat{\delta} = 0$ , so that there seems to be no school effect. Nevertheless, the influence of schools is of interest, because the schools are the study units.

The detection of influential schools for  $AP = 1$  leads to the statement that the model is unidentified, when the pupils with low social-economic status of school 9 and 16 are deleted. The model is also unidentified, when the pupils with high social-economic status of school 10, 11 and 18 are deleted. Because the effect of SES is strongly decreasing for school 9 and 16 and strongly increasing for school 10, 11 and 18, we

decided to incorporate a random slope for the variable SES. The influential schools remained influential. For the extended model, the estimates  $\hat{\beta}_0 = 2.82 (.56)$ ,  $\hat{\beta}_1 = -.51 (.81)$ ,  $\hat{\beta}_2 = 1.64 (.80)$ ,  $\hat{\beta}_3 = -1.04 (1.06)$  and  $\hat{D} = \begin{pmatrix} .91 (.86) & -.81 (.99) \\ -.81 (.99) & 1.33 (1.41) \end{pmatrix}$  are obtained. Note that the small amount of schools within the sample may cause the insignificance of the variance of the random slope. More schools may be sampled to assess whether the slope of the variable SES indeed varies over schools.

### Acknowledgements

We wish to thank Terstegge, Schefman and Thijs for their permission to use the data set of the project 'teeth-brushing on the elementary school'.

### References

- Anderson, T.W., (1958). *An introduction to Multivariate Statistical Analysis*. New York:Wiley.
- Atkinson, A.C., and Donev, A.N. (1996). *Optimum experimental designs*. Oxford: Clarendon Press.
- Bunke, H., and Bunke, O. (1986). *Statistical inference in linear models*. New York:Wiley.
- Chatterjee, S., and Hadi, A.S. (1986). Influential Observations, High Leverage Points, and Outliers in Linear Regression, *Statistical Science*, 1, 379-416.
- Cook, R.D., and Weisberg, S. (1982). *Residuals and Influence in Regression*. New York: Chapman and Hall.
- Hosmer, D.W., and Lemeshow, S. (1989). *Applied Logistic Regression*, New York: Wiley.
- Johnson, R.A., and Wichern, D.W. (1988). *Applied multivariate statistical analysis*. New Jersey: Prentice Hall International, Inc., Englewood Cliffs.

- Lcsaffre, E., and Verbeke, G. (1998). Local influence in linear mixed models, *Biometrics*, 54, 570-582.
- Moerbeek, M., van Breukelen, G.J.P., and Berger, M.P.F. (2001a). Design issues for experiments in multilevel populations. *Journal of Education and Behavioral Statistics*, 25, 271-284.
- Moerbeek, M., van Breukelen, G.J.P., and Berger, M.P.F. (2001b). Optimal experimental design for multilevel logistic models. *Journal of the Royal Statistical Society D*, 50, 17-33.
- Ouwens, M.J.N., Tan, F.E.S., and Berger, M.P.F. (1999). Local Influence for Repeated Measures Generalized Linear Mixed Models. *Proceedings of the 14th International Workshop on Statistical Modelling*, 308-317, Austria.
- Ouwens, M.J.N.M., Tan, F.E.S., and Berger, M.P.F. (2000). Detecting influential observations and subjects in Generalized Linear Mixed Models. *Kwantitatieve Methoden*, 64, 71-86.
- Ouwens, M.J.N.M., Tan, F.E.S., and Berger, M.P.F. (2001a). Local influence to detect influential data structures for Generalized Linear Mixed Models. To appear in *Biometrics*, 57 (4).
- Ouwens, M.J.N.M., Tan, F.E.S., and Berger, M.P.F. (2001b). On the maximin designs for logistic random effects models with covariates. *Proceedings of the 16th International Workshop on Statistical Modelling*, 321-328, Denmark.
- Pukelsheim, F. (1993). *Optimal Design of Experiments*. New York: Wiley.
- Tan, F.E.S., Ouwens, M.J.N., and Berger, M.P.F. (1999). On the Use of Cook's Distance for Longitudinal Regression Models with Random Effects. *Proceedings of the 16th International Workshop on Statistical Modelling*. 687-690, Austria.
- Tan, F.E.S., Ouwens, M.J.N., and Berger, M.P.F. (2000). On the use of Cook's Distance for longitudinal mixed effects regression models. *Kwantitatieve methoden*, 63, 25-48.

- Tan, F.E.S., Ouwens, M.J.N.M., and Berger, M.P.F. (2001). Detection of influential observations in longitudinal mixed effect regression models. To appear in *Journal of the Royal Statistical Society D*, 50, 271-284.
- Terstegge, C., Schefman, S., and Thijs, C. (2000) Project tandenpoetsen op de basisschool in Midden-Limburg. Sociaal-economische gezondheidsverschillen en effect op tandenpoetsgedrag na anderhalf jaar.[Teeth brushing at primary school. Socioeconomic differences and effect on teeth brushing behaviour after 18 months of follow-up]. Roermond: GGD Midden-Limburg, augustus 2000.



## **Epilogue**

This thesis discusses methods to improve the estimation process of regression parameters in GLMM. It can be extended in many ways. Some of the possible future research topics are presented in this epilogue.

All estimates presented in this thesis are based on numerical integration procedures. Alternatively, quasi-likelihood procedures may be used, which are much faster but may lead to somewhat different results. The comparison of the numerical integration procedures and the quasi-likelihood procedures deserve much more attention.

In this thesis the minimization of the volume of the confidence ellipsoid is considered. The volume is not the only criterion which is used in the statistical literature. A detailed reconsideration of the several optimality criteria may lead to a synthesis between the several optimal design criteria.

The preference for the maximin approach based on the relative efficiency above both the maximin approach based on the absolute efficiency as well as the Bayesian design criterion is explained in chapter 1. This explanation may be elaborated. Further, a synthesis between the sequential design approach and the maximin approach may be constructed to deal with the situation that the maximin value is too small.

Chapter 1 shows the relationship between the D-optimal design criterion and the expected/potential influence for the linear model with uncorrelated measurement errors. Other models may also be considered. Further, the relationship between the D-optimality criterion and goodness-of-fit, lack-of-fit and goodness-of-link measures may be studied.

Chapter 2 is restricted to polynomial models, i.e. covariates are not incorporated in the evaluated models. Because models with covariates are often encountered in practice, chapter 2 may be extended.

In Chapter 3 a maximin criterion is proposed to optimize the choice of the experimental design. A topic for future research would be to establish necessary and/or sufficient conditions to ensure the existence of balanced maximin solutions. The chapter is limited to marginally controlled designs and conditionally uncontrolled designs. Other

factorizations are also encountered in practice and may be discussed in detail.

In chapter 4, Cook's Distance is defined for the Linear Mixed Model conditionally on the subjects in the sample. This can also be done for the Generalized Linear Mixed Models. The comparison of the Conditional Cook's Distance and the Marginal Cook's Distance can be elaborated. A comparison between the conditional Cook's Distance and the local influence measures may also be of interest.

The influence values do not indicate whether the influence is large enough to be of interest. Moreover, the origin of the influence of subjects and observations is hard to retrieve. Further research on these topics is therefore necessary.

In literature, the concept of local influence is also discussed for other research problems like the drop-out process. The corresponding observation-oriented local influence measures can be developed based on chapter 5. Obviously, the necessity of these measures must be shown.

Chapter 6 illustrates the methods proposed in chapter 3 and 5. The illustration can lead to two tutorials, one about the optimal design approach and one about the influence of data. One may also discuss the relationship between the design which is used and the influence of the data.

The above given research topics are only a sample of the possible future research topics. We want to proceed with the following three topics. First, the maximin criterion will be applied to more complex situations. Second, the balancedness will be proven in a more general framework, for which the conditionally controlled design and the marginally controlled design are special cases and in which subsets of parameter estimates are also involved. Third, the relationships between the choice of the design and the results in the analysis stage will be studied in more detail.

## Summary

The aim of this thesis is to find methods to improve the estimation process of the regression parameters for Generalized Linear Mixed Models. The proposed methods concern the optimization of the choice of the design and the detection of influential data. The methods are discussed for the Linear Mixed Model and the Generalized Linear Mixed Model. The optimality criteria are discussed in chapter 1, 2, 3 and 6, while the influence measures are discussed in chapter 1, 4, 5 and 6.

Chapter 1 starts with the introduction of the Linear Mixed Model and the Generalized Linear Mixed Model. Thereafter, the most frequently used optimality criteria are discussed. One of these criteria is the D-optimality criterion, which is used to find a design for which the volume of the confidence ellipsoid is minimized. The preference for the D-optimality criterion is based on its invariance with respect to the coding of the independent variables.

Which design is D-optimal depends on the values of the model parameters in the Generalized Linear Mixed Model. Because these values are unknown in practice, the sequential design approach, the Bayesian design approach and the maximin approach can be used. In literature, the maximin approach is used to find a design for which the smallest efficiency over the model parameter space is maximized. The choice to let the maximin approach be based on the relative efficiency is explained in this chapter.

The chapter proceeds with the introduction of Cook's Distance and local influence measures to detect influential data. Cook's Distance assesses the change in the estimates, due to case deletion. The local influence measures assess the change in the estimates, due to infinitesimally perturbations of the data set. It is proposed to detect the influential cases by comparing the scores of the cases on the influence measures with each other. The importance of the detection of influential observations in addition to the detection of influential subjects is illustrated.

Chapter 1 ends by explaining the relationship between the D-optimality criterion and the expected/potential influence of observations for the linear regression model with

uncorrelated measurement errors. It is shown that each design point of the D-optimal design has the same expected/potential influence and that there exists no design point with larger expected/potential influence.

Chapter 2 starts with the presentation of D-optimal designs for the first and second degree polynomial model with random intercept, random slope and AR(1) correlated measurement errors. It shows that the solution of the D-optimality criterion depends on the unknown values of the model parameters. As a consequence, the chapter proceeds with the search of designs which are highly efficient for all likely values of the model parameters. Large classes of highly efficient symmetric designs are presented. The evaluation of the relative efficiency leads to the maximin criterion based on the *relative efficiency*.

The optimality criteria discussed in literature assume that *all* independent variables are experimentally controlled. This assumption is often violated in practice. For example, consider the project 'teeth-brushing on the elementary school', which is discussed in more detail in chapter 6. For this project, only the marginal proportion treated schools is experimentally controlled, where the marginals are taken over the joint distribution of the baseline value, the number of pupils within the schools and the social-economic status of the pupils. Because only the marginal proportion treated schools can be experimentally controlled, the existing optimality criteria are inadequate for this situation.

To deal with the previously described situation, the sample design is factorized in chapter 3 in a part which is experimentally controlled and a part which is not. The experimentally controlled part is called the experimental design and corresponds in this chapter with a marginal distribution (cf. marginal proportion treated schools). The other part of the factorization corresponds with a conditional distribution (cf. conditional joint distribution of the baseline value, the number of pupils within the schools and the social-economic status of the pupils). The experimental design should be optimized, taking this unknown conditional distribution into account. Because the conditional distribution is unknown, a maximin criterion is proposed which searches for an experimental design that maximizes the smallest relative efficiency over all likely values of the model

parameters and all likely conditional distributions. It is shown that under certain conditions, the maximin experimental design is balanced in the discrete variables. The proposed maximin criterion is illustrated using a Logistic Mixed Model.

The second method to improve the estimation process is the detection of influential subjects and observations. For the Linear Mixed Model, Cook's Distance is defined marginally over the random effects. Unfortunately, this measure may fail to detect or may incorrectly detect influential observations, due to the random effect variances and covariances. In chapter 4, Cook's Distance is defined conditionally on the subjects in the sample. The conditionally defined Cook's Distance can be partitioned into two influence measures, one measuring the effect on the estimated average profile and one measuring the effect on the estimated subject-specific deviations. The conditionally defined Cook's Distance works better than the marginally defined Cook's Distance.

In chapter 5, we considered likelihood-based influence measures. To accelerate the detection of influential data, these measures are approximated by Taylor expansions, which are called local influence measures. In literature, the local influence measures are defined at subject-level and are only discussed for the Linear Mixed Model. We extended these measures not only to the Generalized Linear Mixed Model, but also to the observation-level. It is shown that the subject-oriented influence measure is a special case of the proposed observation-oriented influence measure. An illustration of a two-treatment multiple period cross-over trial demonstrates the practical importance of the detection of influential observations in addition to the detection of influential subjects.

In chapter 6, the optimal design criterion which is proposed in Chapter 3 and the local influence measures which are proposed in chapter 5 are illustrated based on the project 'teeth-brushing on the elementary school'. It appeared that the balanced design is highly efficient, irrespective of the uncontrolled conditional distribution and model parameter vector at hand. The detection of influential data led to an assessment of the stability of the parameter estimates, a change in the evaluated model and the suggestion to sample more schools for the validation of a random effect for the social-economic status.



## **Dutch Summary**

In dit proefschrift worden methoden voorgesteld, die tot kwalitatief betere schatters van de regressieparameters leiden. De methoden zijn gericht op de optimalisatie van de keuze van het design en de detectie van invloedrijke data en worden besproken voor het lineaire mixed effect model en het gegeneraliseerde lineaire mixed effect model. De optimalisatiecriteria zijn behandeld in hoofdstuk 1, 2, 3 en 6. De invloedsmaten zijn terug te vinden in hoofdstuk 1, 4, 5 en 6. Hieronder volgt een samenvatting van elk hoofdstuk.

Hoofdstuk 1 begint met de introductie van het lineaire mixed effect model en het gegeneraliseerde lineaire mixed effect model. Het hoofdstuk vervolgt met de bespreking van de meest gebruikte optimalisatiecriteria. Een van deze criteria is het D-optimalisatiecriterium. Dit criterium wordt gebruikt voor het bepalen van een design dat het volume van de betrouwbaarheidsellipsoïde minimaliseert. De voorkeur voor dit criterium is gebaseerd op zijn invariantie met betrekking tot de codering van de onafhankelijke variabelen.

Welk design D-optimaal is hangt af van de waarden van de modelparameters. Helaas zijn deze waarden in de praktijk onbekend. Om toch tot een goede keuze van het design te kunnen komen, zijn de sequentiële design benadering, de Bayesiaanse design benadering en de maximin benadering geïntroduceerd. Tot nu toe is de maximin benadering gebruikt voor het bepalen van een design dat de kleinste efficiëntie over de modelparameter ruimte maximaliseert. De keuze om de maximin benadering te baseren op de relatieve efficiëntie is in hoofdstuk 1 gemotiveerd.

Het hoofdstuk gaat verder met de introductie van Cook's Distance en lokale invloedsmaten voor het detecteren van invloedrijke data. Cook's Distance meet de verandering in de regressieparameters door het wegnemen van data. De lokale invloedsmaten meten de verandering in de schatters door infinitesimale verstoringen van de data set. De invloedrijke cases worden gedetecteerd door het onderling vergelijken van de scores van de cases op de invloedsmaten. De toegevoegde waarde van de detectie

van invloedrijke observaties is aangetoond.

Hoofdstuk 1 eindigt met de relatie tussen het D-optimalisatiecriterium en de verwachte/potentiële invloed van observaties voor het lineaire regressie model met ongecorrleerde meetfouten. Het blijkt dat ieder design punt van het D-optimaal design dezelfde verwachte/potentiële invloed heeft en dat geen enkel ander design punt een nog grotere verwachte/potentiële invloed heeft.

Hoofdstuk 2 start met de presentatie van D-optimale designs voor de eerstegraads en tweedegraads polynomen met stochastisch intercept, stochastische helling en AR(1) gecorrleerde meetfouten. Uit deze presentatie blijkt dat de oplossing van het D-optimalisatiecriterium afhangt van de (onbekende) waarden van de modelparameters. Als een gevolg hiervan wordt overgegaan tot het zoeken van designs die voor alle aannemelijke waarden van de modelparameters hoog efficiënt zijn. De evaluatie van de relatieve efficiëntie over de gehele modelparameter ruimte leidt tot de definitie van de maximin benadering op basis van de relatieve efficiëntie.

De in de literatuur behandelde optimalisatiecriteria veronderstellen dat *alle* onafhankelijke variabelen experimenteel gecontroleerd worden. Deze veronderstelling wordt in de praktijk vaak geschonden. Neem bijvoorbeeld het project 'tandenpoetsen op de basisschool', dat uitgebreid behandeld wordt in hoofdstuk 6. In dit project wordt alleen de marginale proportie behandelde scholen experimenteel gecontroleerd, waar de marginalen genomen worden over de verdeling van de beginmeting, het aantal leerlingen per school en de sociaal-economische status van de leerlingen. Omdat alleen de marginale proportie behandelde scholen experimenteel gecontroleerd wordt, zijn de bestaande optimalisatiecriteria voor deze situatie ontoereikend.

Een maximin criterium voor de hiervoor beschreven situatie wordt in hoofdstuk 3 afgeleid door het ontbinden van het steekproef design in dat deel dat experimenteel gecontroleerd wordt en dat deel dat niet experimenteel gecontroleerd wordt. Het experimenteel gecontroleerde gedeelte wordt experimenteel design genoemd en correspondeert in dit hoofdstuk met een marginale verdeling (marginale proportie behandelde scholen). Het andere deel correspondeert met een conditionele verdeling

(conditionele verdeling van de basismeting, het aantal leerlingen per school en de sociaal-economische status van de leerlingen). De optimalisatie van het experimenteel gecontroleerde deel is nu van belang, terwijl er rekening moet worden gehouden met het ongecontroleerde deel. Omdat de conditionele verdeling onbekend is, wordt een maximin benadering voorgesteld, die zoekt naar een experimenteel design waarvoor de kleinste relatieve efficiëntie over alle aannemelijke waarden voor de modelparameters en alle aannemelijke conditionele verdelingen het grootst is. Onder bepaalde omstandigheden blijkt het maximin experimentele design gebalanceerd te zijn in de discrete variabelen. De voorgestelde maximin benadering wordt aan de hand van een logistisch mixed model geïllustreerd.

In hoofdstuk 4 worden twee generalisaties van Cook's Distance voor het lineaire mixed model besproken. Cook's Distance is in het verleden naar het lineaire mixed effect model uitgebreid. Dit werd gedaan via marginalisatie over de stochastische effecten. Helaas detecteert de zo verkregen invloedsmaat de invloedrijke observaties niet altijd en worden soms zelfs de verkeerde observaties gedetecteerd. Dit is voornamelijk toe te wijzen aan de varianties en covarianties van de stochastische effecten. In hoofdstuk 4 wordt Cook's Distance daarom geconditioneerd op de subjecten. De conditioneel gedefinieerde Cook's Distance kan worden ontbonden in twee invloedsmaten, waarbij de ene het effect op het geschatte algemene profiel meet en de andere het effect op de geschatte subject-specifieke afwijkingen. De conditioneel gedefinieerde Cook's Distance werkt beter dan de marginaal gedefinieerde Cook's Distance.

In hoofdstuk 5 behandelen we likelihood-gebaseerde invloedsmaten. Om de detectie van invloedrijke data te versnellen worden de invloedsmaten benaderd door Taylor expansies. Deze Taylor expansies worden lokale invloedsmaten genoemd. In de literatuur worden de lokale invloedsmaten op subject-niveau gedefinieerd en worden zij alleen besproken voor het lineaire mixed effect model. In hoofdstuk 5 worden deze maten uitgebreid tot de gegeneraliseerde lineaire mixed effect modellen en tot het observatie-niveau. De subject-georiënteerde invloedsmaten blijken speciale gevallen van de observatie-georiënteerde invloedsmaten te zijn. De toegevoegde waarde van de

observatie-georiënteerde invloedsmaten wordt aan de hand van een twee-behandelingen meervoudige periode kruisproef studie aangetoond.

In hoofdstuk 6 worden het optimalisatiecriterium van hoofdstuk 3 en de lokale invloedsmaten van hoofdstuk 5 aan de hand van het project 'tandenpoetsen op de basisschool' geïllustreerd. Het blijkt dat het gebalanceerde design hoog efficiënt is, onafhankelijk van de conditionele verdeling en de waarden van de modelparameters. De detectie van invloedrijke data leidt tot een beoordeling van de stabiliteit van de parameterschatters, een verandering in het geëvalueerde model en de suggestie om meer scholen op te nemen in de steekproef voor de validatie van een stochastische helling voor de sociaal-economische status.

## Curriculum Vitae

J.N.M. Ouwens was born on August 25, 1972 in Emst, the Netherlands. In 1984 he attended the M.A.V.O. in Vaassen. He proceeded his pre-university education at the Chr. Lyceum in Apeldoorn from 1985 to 1988, followed L.O.I. courses between 1988 and 1990 and finished his pre-university education in 1992. He became a church musician in 1985 in Vaassen and in 1990 in Apeldoorn and Twello and obtained a qualification in church music (3rd degree) in 1993 in Meppel.

He attended the secondary teacher training in mathematics at the Windesheim in 1992 in Zwolle and studied mathematics at the University of Utrecht from 1993 to 1997. He wrote a thesis on 'Concrete educational tools in practice' at the IVLOS/Freudenthal Institute in Utrecht in 1996 and in 1997 a master's thesis in mathematical statistics, "Robustification of PLS" at Akzo-Nobel in Arnhem.

During the period 1986 to 2000, he helped many students and pupils to prepare their exams in Mathematics, Statistics, History, Dutch and German and founded the homework institute 'the key', for which he was the leading person between 1995 and 1997. From 1993 to 2001, he participated in projects of the Consultance Groep Nederland in Arnhem.

From 1997 to 2001 he worked as a Ph.D. student at the department of Methodology and Statistics at Maastricht University. His career will be continued as researcher at the University of Maastricht.

