Title
Optimal Longitudinal Cohort Designs and Variance Parameter Estimation

Author
Gan, Lu

Publication Date
2011

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Optimal Longitudinal Cohort Designs and Variance Parameter Estimation

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Applied Statistics

by

Lu Gan

March 2012

Dissertation Committee:
Dr. Subir Ghosh, Chairperson
Dr. Barry Arnold
Dr. Jorge Silva-Risso

The Dissertation of Lu Gan is approved:

Committee Chairperson

University of California, Riverside

ACKNOWLEDGMENTS

I owe my deepest gratitude to my supervisor, Dr. Subir Ghosh, who has supported me in every way through the last five years and has influenced not only my graduate studies, but my whole life. I really appreciate his consistent encouragement, sound advice, insightful teaching, and patient guidance. This thesis would not have been possible without him.

I am indebted to Professor Barry Arnold and Professor Jorge Silva-Risso for kindly accepting to serve on my Ph.D. Committee and for their valuable time and advice.

I would like to thank Dr. Daniel Jeske, Dr. Jun Li, Dr. Shizhong Xu, Dr. Linda Penas, and all other faculty members of the Statistics Department of UCR for educating me and providing a stimulating and fun learning environment.

My thanks also go to my friends here at UCR for all the camaraderie, entertainment, and caring they provided.

Lastly, and most importantly, I wish to thank my family. My parents have been a constant source of support – emotional, moral, and of course financial. They deserve a very special word of thanks for always encouraging me and guiding me in a right direction. Without their endless love, unconditional support, and immense faith in me, I could not have achieved anything. My husband Hill has always been my pillar, my joy, and my guiding light, and I thank him with all my heart for being there as the wonderful in my life. I would like to dedicate my thesis to my beloved parents and husband.

ABSTRACT OF THE DISSERTATION

Optimal Longitudinal Cohort Designs and Variance Parameter Estimation

by

Lu Gan

Doctor of Philosophy, Graduate Program in Applied Statistics
University of California, Riverside, March 2012
Dr. Subir Ghosh, Chairperson

Many large scale longitudinal cohort studies have been carried out in different fields of science. Such studies need a careful planning at the design stage to achieve precise estimates of model parameters. This thesis presents the application of optimal design theory in a longitudinal study with two cohorts of n subjects each. For each subject, the observations are taken at three different time points denoted by $(-1, a_i, 1)$, where $-1 < a_i < 1$ ($i = 1, 2$). Our class of longitudinal cohort designs is $\{(-1, a_1, 1)\ (-1, a_2, 1)\}$; $-1 < a_1 \leq a_2 < 1$. Optimal cohort designs for linear mixed effects models with a random intercept and a random slope are computed analytically with respect to the *D*-, *A*-, and *E*-optimality criteria. The results are demonstrated by optimality regions. We also compare cohort designs with equidistant and non-equidistant time points. We have learned that when the covariance of the random effects satisfies certain conditions, the design with equidistant time points is preferred. However, in certain cases, for example, the third case stated in Theorem 3.1, the design with non-equidistant time points is better. We propose a new iterative method for computing the Restricted Maximum likelihood (REML) estimators

of the variance components in the linear mixed effects models using three criterion functions l^*, Δ, and P. Two simulated data sets and one observed facial growth data set are used to illustrate our method.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Longitudinal Data

A dataset is longitudinal if it tracks the same type of information on the same subjects at multiple time points. For example, part of a longitudinal dataset could contain the math test scores of some school children in three successive years.

Table 1.1: School children math test scores

Name	Grade 1 (1995)	Grade 2 (1996)	Grade 3 (1997)
Jasmine	98	92	88
Ben	78	69	81
Bill	82	89	93

The primary advantage of longitudinal data is that they can be used to measure changes. Therefore, the main focus in longitudinal studies is usually on the modeling of responses as a function of time.

Researchers in health science and medicine are often interested in estimating changes over time. In the past, many large-scale longitudinal cohort studies have been set up for this purpose. Some examples of such studies are the New York State cohort study [2], the longitudinal aging study of Amsterdam [9], and the national cancer prevention study [5]. These large-scale studies all have in common that different cohorts of subjects are measured several times over a long period of time.

1.2 Cohort Designs

Cohort designs are divided into three categories: purely longitudinal design, cross-sectional design, and mixed longitudinal design (Tekle, Tan, and Berger, 2008 [28]). Here, cohort is defined as a group of subjects that have a common characteristic (birth year, geographic boundary, age, sex) in a selected time period. In a purely longitudinal design, a single cohort of subjects is measured over the study period. Instead, in a cross-sectional design, two or more cohorts of subjects are selected and each of them is measured at one time point, that is, the number of cohorts is the same as the total number of time points at which the samples are taken. In a mixed longitudinal design, two or more cohorts of subjects are selected and measured at their corresponding time intervals, where the time intervals can be either overlapping or nonoverlapping. These three types of cohort designs are presented in Figure 1.1 to 1.3, respectively.

Figure 1.1: Purely longitudinal design (C = 1 cohort)

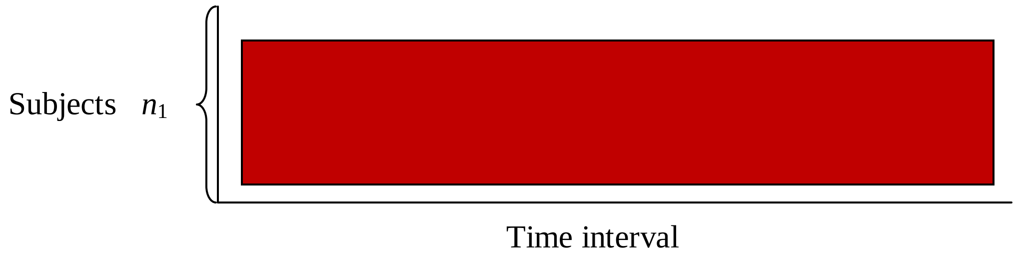


Figure 1.2: Cross-sectional design (C = 2 cohorts)

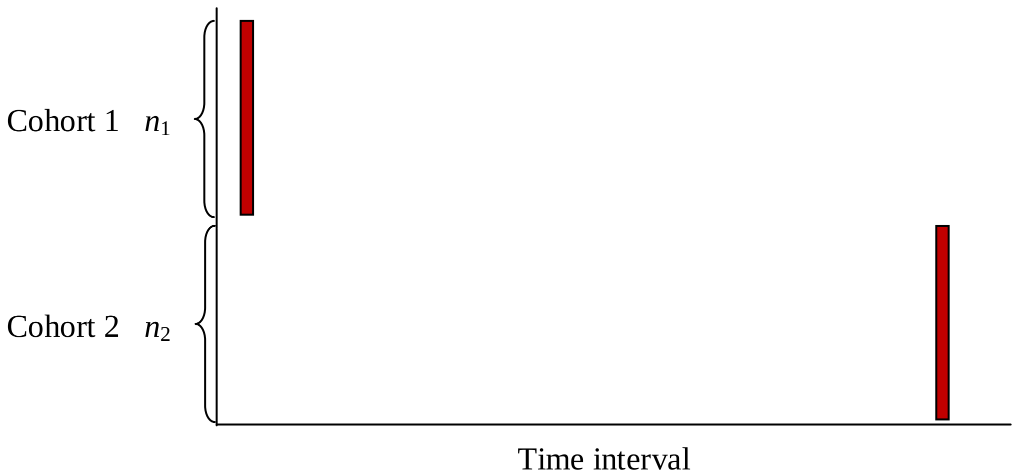
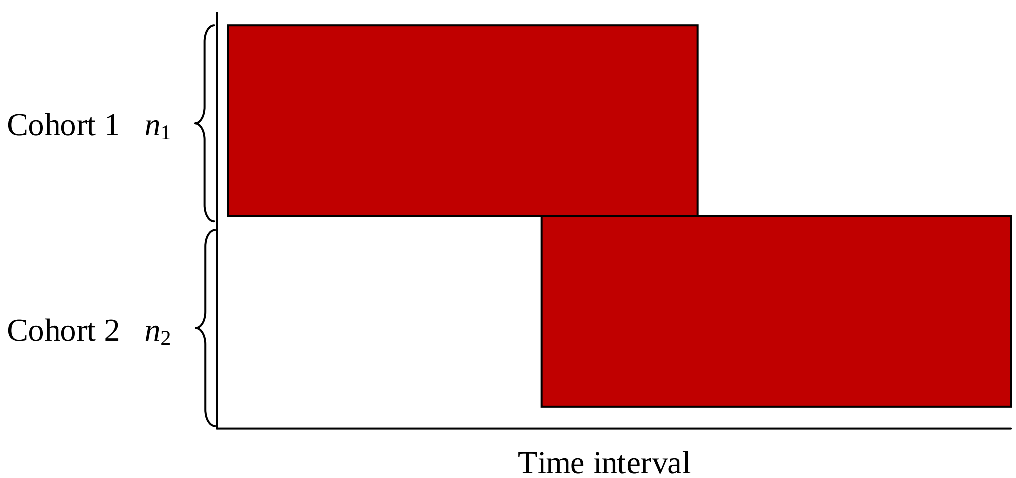


Figure 1.3: Mixed longitudinal design (C = 2 cohorts)



Additionally, a cohort design is characterized by the number of cohorts C, the number of repeated measurements in each cohort m_i, and the number of subjects within each cohort n_i (i = 1,…,C). Note that a purely longitudinal design is a special case of a mixed longitudinal design with C = 1 cohort. In this research, we consider fully overlapping cohorts in a mixed-longitudinal design, that is, the cohorts have the same beginning and ending time points. Moreover, we assume the number of repeated measurements in each cohort is the same and denoted as m.

1.2.1 Notation

The total number of subjects in C cohorts is $N = \sum_{i=1}^{C} n_i$, and the relative size of cohort i (i = 1,…,C) is $\omega_i = n_i / N$ such that $\sum_{i=1}^{C} \omega_i = 1$. The time points for the i^{th} cohort are $\boldsymbol{t}_i = (t_{1(i)}, ..., t_{m(i)})'$. In general, the cohort designs τ are expressed as

$$\tau = \begin{Bmatrix} \boldsymbol{t}_1 & \boldsymbol{t}_2 & \dots & \boldsymbol{t}_C \\ \omega_1 & \omega_2 & \dots & \omega_C \end{Bmatrix}.$$

Suppose that T_{Cm} is the design class of all cohort designs τ with C cohorts and m repeated measures in each cohort. In chapter 2, design classes T_{22} and T_{23} will be discussed in great detail.

1.3 Literature Review

As we mentioned earlier that there have been many large-scale longitudinal cohort studies in health science and medicine. The objective of these longitudinal cohort studies is to identify the change of responses over time. To achieve this objective, we need to estimate the model parameters efficiently. As a result, it is worthwhile to implement optimal design procedures to design these studies. The problems of optimal designs for longitudinal cohort studies have been studied in the literature. The following five papers are relevant to our work.

1. Tan and Berger (1999):

The model considered is a special case of the model (2.1) with C = 1 and $\beta_0 = \beta_1 = \cdots = \beta_{p-1} = 0.$ The random effects model is then

$$y_{kj} = b_{0j} + b_{1j}t_k + \cdots + b_{(q-1)j}t_k^{q-1} + \varepsilon_{kj},$$

where y_{kj} is the k^{th} measurement ($k = 1,\ldots,m$) taken on the j^{th} subject ($j = 1,\ldots,n$) at time point t_k. The above model can also be written in matrix notation as follows

$$\boldsymbol{y}_j = \boldsymbol{Z}\boldsymbol{b}_j + \boldsymbol{\varepsilon}_j,$$

where $\boldsymbol{y}_j = (y_{1j},\ldots,y_{mj})'$ is the $m \times 1$ vector of repeated measurements taken on the j^{th} subject at m time points $\boldsymbol{t} = (t_1,\ldots,t_m)'$, $\boldsymbol{Z}$ is the $m \times q$ matrix of explanatory variables of rank q which consists of polynomial coefficients based on $\boldsymbol{t}$. The $\boldsymbol{b}_j$ is a ($q \times 1$) column vector of random regression coefficients with mean vector $\boldsymbol{b}$ and covariance matrix $\mathbf{D}$. The error $\boldsymbol{\varepsilon}_j$ has mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Psi}$.

Tan and Berger numerically derived D-optimal designs for the next three cases.

Case I: $\mathbf{D} = \mathbf{I}$ and $\mathbf{\Psi} = \mathbf{I}$,

Case II: $\mathbf{D} = \begin{bmatrix} a & & & b \\ & a & & \\ & & \ddots & \\ b & & & a \end{bmatrix}$, i.e. a uniform matrix with constant off-diagonal elements, $(a > 0, b > 0)$, and $\mathbf{\Psi} = \mathbf{I}$,

Case III: $\mathbf{D} = \mathbf{I}$ and $\mathbf{\Psi} = \mathrm{AR}(1)$, i.e. the covariance between two measurement errors for the same subject at time t_k and $t_{k'}$ is $\sigma^2 \rho^{|t_k - t_{k'}|}$ and $0 < \rho < 1$.

Table 1.2 shows the *D*-optimal designs under Case I and Case II corresponding to linear (q = 2), quadratic (q = 3), and cubic (q = 4) polynomial models. The results of *D*-optimal allocation of time points for Case III with serial correlation $0 < \rho < 1$ are illustrated in Table 1.3.

Table 1.2: *D*-optimal designs for Case I and II in time interval [−1, 1]

Degree	Number of repeated measures (m)	Optimal allocation of time points
Linear: q = 2	2	−1, 1
Quadratic: q = 3	3	−1, 0, 1
Cubic: q = 4	4	−1, −0.44, 0.44, 1

Table 1.3: *D*-optimal designs for Case III with $0 < \rho < 1$ in time interval [−1, 1]

Degree	Number of repeated measures (m)	Optimal allocation of time points
Linear: $q = 2$	2	−1, 1
	3	−1, 0, 1
	4	−1, −0.4, 0.4, 1
	5	−1, −0.6, 0, 0.6, 1
	6	−1, −0.6, −0.2, 0.2, 0.6, 1
Quadratic: $q = 3$	3	−1, 0, 1
	4	−1, −0.3, 0.3, 1
	5	−1, −0.5, 0, 0.5, 1
	6	−1, −0.6, −0.2, 0.2, 0.6, 1
Cubic: $q = 4$	4	−1, −0.5, 0.5, 1
	5	−1, −0.65, 0, 0.6, 1
	6	−1, −0.7, −0.3, 0.3, 0.7, 1

2. Ouwens, Tan, and Berger (2002):

The model discussed is the same as the model (2.1) and is given by

$$y_{kj(i)} = \beta_0 + \beta_1 t_{k(i)} + \cdots + \beta_{p-1} t_{k(i)}^{p-1} + b_{0j} + b_{1j} t_{k(i)} + \cdots + b_{(q-1)j} t_{k(i)}^{q-1} + \varepsilon_{kj(i)},$$

where $y_{kj(i)}$ is the k^{th} ($k = 1,\ldots,m$) measurement taken on the j^{th} subject ($j = 1,\ldots,n_i$) in the i^{th} cohort ($i = 1,\ldots,C$) at the time point $t_{k(i)}$. For each subject *j*, the above model can be expressed in a more general matrix notation

$$\boldsymbol{y}_{j(i)} = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{b}_{j(i)} + \boldsymbol{\varepsilon}_{j(i)},$$

where $\boldsymbol{y}_{j(i)} = (y_{1j(i)}, \ldots, y_{mj(i)})'$ is the $m \times 1$ vector of repeated measurements of subject j in cohort i, and $\boldsymbol{t}_{(i)} = (t_{1(i)}, \ldots, t_{m(i)})'$ is the corresponding vector of time points, $\boldsymbol{X}_i$ is the $m \times p$ matrix of explanatory variables of rank p and $\boldsymbol{Z}_i$ is the $m \times 2$ matrix of random effects. Both $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ consist of polynomial coefficients based on $\boldsymbol{t}_{(i)}$. The $p \times 1$ vector $\boldsymbol{\beta}$ is a vector of fixed regression coefficients. The 2×1 vector $\boldsymbol{b}_{j(i)} = (b_{0j(i)}, b_{1j(i)})'$ is the vector of random intercept and random slope with mean **0** and covariance $\mathbf{D} = \begin{bmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{bmatrix}$. The error vector $\boldsymbol{\varepsilon}_{j(i)}$ has mean **0** and covariance $\boldsymbol{\Psi}_i$, where $\boldsymbol{\Psi}_i$ has a first-order autoregressive serial correlation structure, i.e. the covariance between two measurement errors for the same subject at time t_{ik} and $t_{ik'}$ is $\sigma^2 \rho^{|t_{ik} - t_{ik'}|}$, where σ^2 is the common variance for error components and $0.0025 \leq \rho \leq 1$. The random vector $\boldsymbol{b}_{j(i)}$ and the error vector $\boldsymbol{\varepsilon}_{j(i)}$ are independent.

Ouwens, Tan, and Berger defined the optimization parameter space as $\Theta = \left\{ \theta : \theta = (\rho, \sigma^2, L, d_{22},),\ \rho \in [0.0025, 1],\ 0.025 \leq \sigma^2 \leq 1,\ d_{22} \geq 0,\ |L| \leq 1 / \sqrt{d_{22}} \right\}$.

Their numerical study found the *maximin D*-optimal designs corresponding to linear (p = 2) and quadratic (p = 3) polynomial models with random intercepts and random slopes for the following two design classes shown in Table 1.4.

Table 1.4: *maximin* *D*-optimal designs in time interval [−1, 1]

Degree	Design class	Optimal allocation of time point
Linear: p = 2	$\left\{ \begin{matrix} (-1,-a,1) & (-1,a,1) \\ 0.5 & 0.5 \end{matrix} \middle\vert\, 0 \le a \le 1 \right\}$	a = 0.55
Quadratic: p = 3		a = 0
Linear or quadratic: p = 2 or 3		a = 0.28
Linear: p = 2	$\left\{ \begin{matrix} (-1,-a,a,1) \\ 1 \end{matrix} \middle\vert\, 0 \le a \le 1 \right\}$	a = 0.57
Quadratic: p = 3		a = 0.23
Linear or quadratic: p = 2 or 3		a = 0.39

3. Moerbeek (2005):

Here, a fixed effects model is used to model the responses y_{kj} (k = 1,…,m, j = 1,…,n) of subject j at time points t_k

$$y_{kj} = \beta_0 + \beta_1 t_k + \cdots + \beta_{p-1} t_k^{p-1} b_{0j} + \varepsilon_{kj}.$$

The above fixed effects model is a special case of the model (2.1) with C = 1 and $b_{0j} = b_{1j} = \cdots = b_{(q-1)j} = 0$. In matrix notation, we have

$$\boldsymbol{y}_j = \boldsymbol{X}_j \boldsymbol{\beta} + \boldsymbol{\varepsilon}_j,$$

where $\boldsymbol{y}_j = (y_{1j},\ldots,y_{mj})'$ is an (m × 1) vector of repeated measurements taken on the j^{th} subject at m time points $\boldsymbol{t} = (t_1,\ldots,t_m)'$. $\boldsymbol{X}_j$ is an (m × p) design matrix of explanatory variables which consists of polynomial coefficients based on $\boldsymbol{t}$. The p × 1 vector $\boldsymbol{\beta}$ is a vector of fixed regression coefficients. The error $\boldsymbol{\varepsilon}_j$ has mean vector **0** and covariance matrix $\sigma^2\mathbf{V}_j$, where matrix $\mathbf{V}_j$ has a first-order autoregressive serial correlation structure. The element (*i*, *i'*) of $\mathbf{V}_j$ is equal to $\rho^{|t_i - t_{i'}|}$ and $0 \le \rho \le 1$.

The optimal designs for polynomial models in time interval [0, 2] with *m* = *p* and uncorrelated errors, i.e. $\mathbf{V}_j = \mathbf{I}$, were computed numerically and shown in Table 1.5.

Table 1.5: *A*-, *D*-, and *E*-optimal designs in time interval [0, 2]

Degree	*A*-optimal design	*D*-optimal design	*E*-optimal design
Linear: *p* = 2	0, 2	0, 2	0, 2
Quadratic: *p* = 3	0, 1.057, 2	0, 1, 2	0, 1.069, 2
Cubic: *p* = 4	0, 0.528, 1.572, 2	0, 0.553, 1.447, 2	0, 0.526, 1.582, 2

In addition, Moerbeek discovered that the efficiency of a design is generally higher if the assumed order of polynomial is closer to the true order. He showed that the design with the number of time points equal to the number of regression coefficients is optimal.

4. Winkens, Schouten, Breukelen, and Berger (2005):

A mixed effects model with linearly divergent treatment effects is considered

$$y_{kj} = \mu_k + \beta G t_k + b_{0j} + \varepsilon_{kj} + e_{kj},$$

where y_{kj} is the k^{th} measurement (k = 1,…,m) taken on the j^{th} subject (j = 1,…,n) at time point t_k, G is an indicator of treatment, which is equal to 1 if subject j belongs to the treatment group, and 0 otherwise. There are two random error components: ε_{kj} and e_{kj}. When μ_k = 0 and G = 1, the above mixed effects model is a special case of the model (2.1) with C = 1, $\beta_0 = 0$, p = 2, and q = 1. For subject j, the model can be written as

$$\boldsymbol{y}_j = \boldsymbol{X}_j\boldsymbol{\beta} + \boldsymbol{Z}_j\boldsymbol{b}_j + \boldsymbol{\varepsilon}_j + \boldsymbol{e}_j,$$

$$\text{where } \boldsymbol{X}_j\boldsymbol{\beta} = \begin{pmatrix} 1 & 0 & \cdots & 0 & G\times t_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & G\times t_m \end{pmatrix} \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \\ \beta \end{pmatrix} \text{ and } \boldsymbol{Z}_j\boldsymbol{b}_j = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} (b_{0j}),$$

$\boldsymbol{y}_j = (y_{1j},\ldots,y_{mj})'$ is an (m × 1) vector of repeated measurements of subject j at time points $\boldsymbol{t} = (t_1,\ldots,t_m)'$, $\boldsymbol{\beta}$ is the treatment effect parameter. Moreover, $\boldsymbol{\varepsilon}_j$ is a component of autocorrelation and $\boldsymbol{e}_j$ is an extra component of measurement error. The random vector $\boldsymbol{b}_j$ and the error vectors $\boldsymbol{\varepsilon}_j$ and $\boldsymbol{e}_j$ are independent, normally distributed with mean **0** and covariance matrices **D**, $\sigma_\varepsilon^2\mathbf{H}$ and $\sigma_e^2\mathbf{I}$, respectively. An AR(1) correlation structure is considered for **H**, i.e. The element (i, i') of **H** is equal to $\rho^{|t_i - t_{i'}|}$ and $0 \le \rho \le 1$. The covariance of vector $\boldsymbol{y}_j$ is denoted as matrix **V**. They presented D-optimal designs in time interval [0, 1] for three covariance structures of **V** and three numerical values of correlation ρ_{1m} , where ρ_{1m} stands for the correlation between the first and last measurement taken from the same subject.

Table 1.6: *D*-optimal designs for ρ_{1m} = 0.3, 0.5, or 0.7 in time interval [0, 1]

Covariance structure	Number of repeated measures (*m*)	Optimal allocation of time points
'Compound Symmetry' or 'AR(1) + ME'	3	0, 1, 1
	4	0, 1, 1, 1
	5	0, 0^{+}, 1, 1, 1
'AR(1)'	3	0, 0.64, 1
	4	0, 0.5, 0.78, 1
	5	0, 0.42, 0.65, 0.84, 1

$^{+}$The second time point is 1 when the correlation ρ_{1m} is 0.3.

Note that 'Compound Symmetry' indicates a model with random intercept and measurement error $\boldsymbol{e}_j$; 'AR(1) + ME' implies a fixed effects model with serial correlation $\boldsymbol{\varepsilon}_j$ and measurement error $\boldsymbol{e}_j$; and 'AR(1)' represents a fixed effects model with only serial correlation $\boldsymbol{\varepsilon}_j$. It can be seen that a large gain in efficiency is obtained by adding repeated measures at the end of the study, if covariance structure is either 'Compound Symmetry' or 'AR(1) + ME'.

5. Tekle, Tan and Berger (2008):

The model considered is the same as the model (2.1)

$$y_{kj(i)} = \beta_0 + \beta_1 t_{k(i)} + \cdots + \beta_{p-1} t_{k(i)}^{p-1} + b_{0j} + b_{1j} t_{k(i)} + \cdots + b_{(q-1)j} t_{k(i)}^{q-1} + \varepsilon_{kj(i)},$$

where $y_{kj(i)}$ is the k^{th} ($k = 1,\ldots,m_i$) measurement taken on the j^{th} subject ($j = 1,\ldots,n_i$) in the i^{th} cohort ($i = 1,\ldots,C$) at the time point $t_{k(i)}$. For each subject j, the above model can be expressed in a more general matrix notation

$$\boldsymbol{y}_{j(i)} = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_{j(i)} + \boldsymbol{\varepsilon}_{j(i)},$$

where $\boldsymbol{y}_{j(i)} = (y_{1j(i)},\ldots,y_{m_i j(i)})'$ is the $m_i \times 1$ vector of repeated measurements for person j in cohort i observed at m_i time points $\boldsymbol{t}_{(i)} = (t_{1(i)},\ldots,t_{m_i(i)})'$. The $\boldsymbol{X}_i$ is an $m_i \times p$ matrix of explanatory variables and $\boldsymbol{Z}_i$ is the $m_i \times 2$ submatrix of $\boldsymbol{X}_i$. Both $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ consist of polynomial coefficients based on $\boldsymbol{t}_{(i)}$. The $p \times 1$ vector $\boldsymbol{\beta}$ is a vector of fixed regression coefficients. The 2×1 vector $\boldsymbol{b}_{j(i)} = (b_{0j(i)}, b_{1j(i)})'$ is the vector of random intercept and random slope with mean $\mathbf{0}$ and covariance $\mathbf{D}$. The $m_i \times 1$ error vector $\boldsymbol{\varepsilon}_{j(i)}$ has mean $\mathbf{0}$ and covariance $\sigma^2\boldsymbol{R}_i$, where σ^2 is the common variance for error components, $\boldsymbol{R}_i$ is the error correlation matrix which has an AR(1) structure, i.e. the element ($s(i)$, $s'(i)$) of $\boldsymbol{R}_i$ is equal to $\rho^{|t_{s(i)} - t_{s'(i)}|}$ and $0 \le \rho \le 1$.

Taking into account the cost of the study, Tekle, Tan and Berger constructed D-optimal designs for a mixed longitudinal study with three nonoverlapping cohorts. The total number of repeated measurements is defined by $M = \sum_{i=1}^{3} m_i$, and ω_i is the relative size of cohort i ($i = 1, 2, 3$) such that $\sum_{i=1}^{3}\omega_i = 1$. The set of time points considered is $\{1, 2,\ldots, 15\}$.

With $\mathbf{D} = \begin{bmatrix} 1 & 0.25 \\ 0.25 & 0.25 \end{bmatrix}$, for a given value of M and serial correlation ρ, the optimal weights ω_i were derived numerically for each possible combination of time points using the Broyden-Fletcher-Goldfarb and Shanno (BFGS) algorithm, and then the *D*-optimal design was selected out of all those possible cohort designs. The results of *D*-optimal designs with optimal weights are shown in Table 1.7.

Table 1.7: *D*-optimal designs for C = 3 cohorts in time interval [1, 15]

Degree	M	Cohort	Optimal allocation of time points	Optimal weights		
				$\rho = 0$	$\rho = 0.4$	$\rho = 0.8$
Linear: p = 2	4	1st	1, 13	0.9139	0.9131	0.9227
		2nd	14	0.0860	0.0868	0.0772
		3rd	15	0.0001	0.0001	0.0001
Quadratic: p = 3	5	1st	1, 7, 13	0.9059	0.9570	0.9575
		2nd	14	0.0940	0.0429	0.0424
		3rd	15	0.0001	0.0001	0.0001
Cubic: p = 4	6	1st	1, 4, 10, 13	0.8826	0.9568	0.9846
		2nd	14	0.1173	0.0431	0.0153
		3rd	15	0.0001	0.0001	0.0001

Notice that the time point in the third cohort has very little weight. Tekle, Tan and Berger also showed that for a given number of cohorts $C \leq 4$ and number of fixed regression parameters $p \leq 4$ in the polynomial models, the optimum number of time points per subject in most cases of the design is $C + (p - 1)$.

It should be pointed out that most the previous works have been focused on either purely longitudinal designs or mixed longitudinal designs with non-overlapping cohorts. Not much research, however, has been done on mixed longitudinal designs with overlapping cohorts. Moreover, in the literature the optimal designs were all found numerically.

1.4 Thesis Contribution

In this thesis, we consider the linear mixed effects models for describing the data of longitudinal cohort studies. The optimal design approach is used in finding the optimal allocation of time points to maximize the information for the estimation of the parameters in the model. Three situations that have not been considered earlier are presented in this thesis. The first situation is that mixed longitudinal designs with two fully overlapping cohorts are considered. We are interested in finding the optimal cohort designs under certain structures of **D** and $\boldsymbol{R}_i$ (i = 1, 2). The second situation is that optimal designs for longitudinal cohort studies are derived analytically. The *D*-, *A*-, and *E*- optimal design for the design class T_{23} with $\boldsymbol{R}_i = \mathbf{I}$ and $\mathbf{D} = \mathbf{I}$ is the design with time points $(-1, -\sqrt{a_0}, 1)$ for the 1st cohort and $(-1, \sqrt{a_0} 1)$ for the 2nd cohort, where $a_i^2 \le a_0$ $(0 < a_0 < 1)$, $i = 1,$ 2. The third situation is that we present conditions on the covariance of the random effects so that the design with equidistant time points is better than the design with non-equidistant time points. We find that the *D*-optimal design for the design class $T_{23(a)}$ with $\boldsymbol{R}_i = \mathbf{I}$ and a general **D** is the design with equidistant time points (−1, 0, 1) for both cohorts, if $D_f^{(1)} > 0$. We also find that the design class $T_{23(-a,\, a)}$ is preferred over $T_{23(a)}$ with respect to both *D*- and *A*-optimality criteria, if **D** is a diagonal matrix. Furthermore, we obtain the estimators for the variance components using the restricted maximum likelihood estimation (REML) procedure. The standard REML procedure uses numerical methods for solving the estimation equations, for example, the SAS PROC MIXED procedure. We propose a new method of estimating the variance components by using three criteria: function Δ, the log-likelihood l^*, and function P. The most accurate final solution of REML is obtained when the numerical value of Δ is minimum, the numerical value of l^* is maximum, and the numerical value of P is relatively small. Our method is computer intensive and comparable with the standard methods of estimation used by SAS.

1.5 Thesis Description

In Chapter 2, we discuss the theory of the mixed effects models. We then introduce the random intercept and slope models and two design classes, T_{22} and T_{23}. In Chapter 3, we go over the concepts of *D*- and *A*-optimality criteria and the relative efficiency. In addition, we present optimal cohort designs for the design class $T_{23(a)}$ with correlation matrix $\boldsymbol{R}_i = \mathbf{I}$ (i = 1, 2) under the *D*- and *A*-optimality criteria. The results of optimal designs are illustrated by optimality regions. However, we know repeated measurements from the same subject are often correlated. We then in Chapter 4 consider the first-order autoregressive correlation structure, i.e. AR(1), for the error correlation matrix $\boldsymbol{R}_i$. We present general results with their applications in comparison of design classes $T_{23(a)}$ and $T_{23(-a,\, a)}$, with respect to *D*- and *A*-optimality criteria. Chapter 5 considers another error correlation structure: compound symmetric (CS). The *D*-optimal cohort designs are obtained for design class $T_{23(a)}$ under the linear mixed effects model (2.3) with $\boldsymbol{R}_i$ = CS. The comparison between cohort designs with equidistant and non-equidistant time points is also presented. In Chapter 6, we obtain *D*-, *A*-, and *E*-optimal designs analytically for design class T_{23} with covariance matrix $\mathbf{D} = \mathbf{I}$ and error correlation matrix $\boldsymbol{R}_i = \mathbf{I}$. In Chapter 7, we estimate the variance components using the restricted maximum likelihood estimation (REML) procedure and then present a method of solving REML estimation using three criterion functions. Conclusions and discussion are given in Chapter 8.

Chapter 2
Linear Mixed Effects Model for Cohort Designs

2.1 Introduction

In this chapter, we present the general mixed effects model for longitudinal studies, followed by the introduction of random intercept and slope models and two design classes: T_{22} and T_{23}. An illustrated example of cohort design is given. Longitudinal studies, often called repeated measurements in health science and medicine, arise when subjects provide responses on multiple time points. The objective of longitudinal studies is to identify the change of responses over time. There are several textbooks and research papers on the analysis of longitudinal studies, for example: Silvey (1980) [25], and Tekle, Tan, and Berger (2008) [28].

2.2 Mixed Effects Model

We often assume that the observations are drawn independently from the populations considered. However, for longitudinal data the independence assumption is not generally true. Mixed effects models are used for analyzing longitudinal data with a complex and multilevel structure. Mixed effects models for longitudinal data have been discussed by Diggle, Liang and Zeger (1994) [8] and Verbeke and Molenberghs (2000) [31], among others. Mixed effects model for longitudinal data can be described as follows.

Suppose a researcher is interested in studying the dependence of response over time. Let $y_{kj(i)}$ be the k^{th} ($k = 1,\ldots,m$) measurement taken on the j^{th} subject ($j = 1,\ldots,n_i$) in the i^{th} cohort ($i = 1,\ldots,C$) at the time point $t_{k(i)}$, where a cohort is defined as a group of subjects experiencing some event (birth, age, geographic boundary, sex) in a selected time period. The general mixed effects model is given by

$$y_{kj(i)} = \beta_0 + \beta_1 t_{k(i)} + \cdots + \beta_{p-1} t_{k(i)}^{p-1} + b_{0j} + b_{1j} t_{k(i)} + \cdots + b_{(q-1)j} t_{k(i)}^{q-1} + \varepsilon_{kj(i)}, \qquad (2.1)$$

where time point $t_{k(i)}$ is the explanatory variable, $y_{kj(i)}$ is the response variable, the p parameters $\beta_0, \beta_1, \ldots, \beta_{p-1}$ are fixed parameters that describe the overall effects, the q coefficients $b_{0j}, b_{1j}, \ldots, b_{(q-1)j}$ are random (subject-specific) parameters, which describe the variation between subjects.

For subject j in cohort i, model (2.1) can be rewritten as

$$\boldsymbol{y}_{j(i)} = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{b}_{j(i)} + \boldsymbol{\varepsilon}_{j(i)},$$

where $\boldsymbol{y}_{j(i)} = (y_{1j(i)}, \ldots, y_{mj(i)})'$ is an (m × 1) vector of repeated measurements taken on the j^{th} subject in the i^{th} cohort, the corresponding time points are $\boldsymbol{t}_{(i)} = (t_{1(i)}, \ldots, t_{m(i)})'$. The $\boldsymbol{\beta}$ is a (p × 1) vector of fixed regression coefficients. The $\boldsymbol{X}_i$ is an (m × p) matrix of explanatory variables of rank p and $\boldsymbol{Z}_i$ is an (m × q) submatrix of $\boldsymbol{X}_i$ of random effect variables. For example, for mixed effects models with linear polynomials, p = 2. The $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are design matrices that consist of polynomial coefficients on the vector of time points $\boldsymbol{t}_{(i)} = (t_{1(i)}, \ldots, t_{m(i)})'$. The m × 1 vector $\boldsymbol{\varepsilon}_{j(i)}$ has mean **0** and an (m × m) covariance matrix $\sigma^2 \boldsymbol{R}_i$, where σ^2 is the common variance for error components, $\boldsymbol{R}_i$ is an (m × m) correlation matrix of the error vector. The q × 1 vector $\boldsymbol{b}_{j(i)}$ consists random regression coefficients of subject j within cohort i, with mean **0** and q × q covariance matrix $\sigma^2 \mathbf{D}$. Furthermore, it is assumed that the vectors $\boldsymbol{b}_{j(i)}$ and vector $\boldsymbol{\varepsilon}_{j(i)}$ are independent. Note that the m time points $\boldsymbol{t}_{(i)}$ are assumed to be the same for each subject in the i^{th} cohort. Subjects are nested within cohorts. Observations taken from any two subjects either in the same or different cohorts are uncorrelated. In the next section, we focus on a special case of the general mixed effects model, namely the random intercept and slope model.

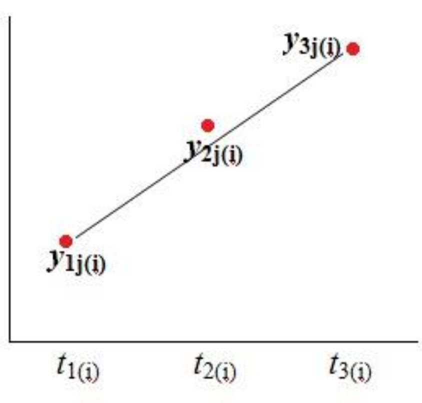
2.2.1 Random intercept and slope model

Consider, as an example, a longitudinal study to investigate the effect of daily calcium supplementation on bone gain in adolescent men and women. The response variable is the total bone mineral density (TBBMD, gr/cm^2). The cohorts are defined by gender and each cohort has n subjects. The total number of subjects is $N = 2n$. Suppose that $m = 3$ repeated measurements are taken from each subject, the following table illustrates the observations taken on the n subjects in both cohorts.

Table 2.1: Observations $y_{kj(i)}$ ($k = 1, 2, 3, j = 1,\ldots,n, i = 1, 2$)

	Subject	Time Point				Subject	Time Point		
		$t_{1(1)}$	$t_{2(1)}$	$t_{3(1)}$			$t_{1(2)}$	$t_{2(2)}$	$t_{3(2)}$
	1	$y_{11(1)}$	$y_{21(1)}$	$y_{31(1)}$		1	$y_{11(2)}$	$y_{21(2)}$	$y_{31(2)}$
	:	:	:	:		:	:	:	:
Cohort 1	j	$y_{1j(1)}$	$y_{2j(1)}$	$y_{3j(1)}$	*Cohort 2*	j	$y_{1j(2)}$	$y_{2j(2)}$	$y_{3j(2)}$
(Male)	:	:	:	:	(Female)	:	:	:	:
	n	$y_{1n(1)}$	$y_{2n(1)}$	$y_{3n(1)}$		n	$y_{1n(2)}$	$y_{2n(2)}$	$y_{3n(2)}$

For fixed cohort i and subject j, suppose the TBBMD profile picture of the j^{th} subject within the i^{th} cohort indicates a linear trend as follows:



Moreover, suppose that both the intercept and trend due to time vary by individuals, that is, each TBBMD profile picture has its own intercept and slope, then a special case of model (2.1) is given by:

$$y_{kj(i)} = \beta_0 + \beta_1 t_{k(i)} + b_{0j(i)} + b_{1j(i)} t_{k(i)} + \varepsilon_{kj(i)}. \qquad (2.2)$$

It should be noted that model (2.2) consists of two parts: one with a fixed intercept and fixed slope β_1, and the other with a random intercept and random slope. The random intercept $b_{oj(i)}$ and random slope $b_{1j(i)}$ represent subject-specific variation from the overall population intercept β_0 and the overall population slope β_1 for the j^{th} subject (j = 1,…,n) in the i^{th} cohort (i = 1, 2), respectively.

For each subject j in the i^{th} cohort, model (2.2) can be expressed in a more general matrix notation

$$\boldsymbol{y}_{j(i)} = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{b}_{j(i)} + \boldsymbol{\varepsilon}_{j(i)}, \qquad (2.3)$$

where $\boldsymbol{y}_{j(i)} = (y_{1j(i)}, y_{2j(i)}, y_{3j(i)})'$, i = 1, 2, j = 1,…,n,

$$\boldsymbol{X}_i = \boldsymbol{Z}_i = \begin{pmatrix} 1 & t_{1(i)} \\ 1 & t_{2(i)} \\ 1 & t_{3(i)} \end{pmatrix}, \qquad (2.4)$$

$\boldsymbol{\beta} = (\beta_0, \beta_1)'$, the (2 × 1) vector $\boldsymbol{b}_{j(i)} = (b_{0j(i)}, b_{1j(i)})'$ is *i.i.d* with mean **0** and a covariance matrix $\sigma^2 \mathbf{D} = \sigma^2 \begin{bmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{bmatrix}$, $\boldsymbol{\varepsilon}_{j(i)} = (\varepsilon_{1j(i)}, \varepsilon_{2j(i)}, \varepsilon_{3j(i)})'$ has mean **0** and a covariance matrix $\sigma^2 \boldsymbol{R}_i$, $\boldsymbol{b}_{j(i)}$ and $\boldsymbol{\varepsilon}_{j(i)}$ are independent.

The covariance matrix of $\boldsymbol{y}_{j(i)}$ is given by

$$\mathrm{var}(\boldsymbol{y}_{j(i)}) = \sigma^2(\boldsymbol{Z}_i\mathbf{D}\boldsymbol{Z}_i' + \boldsymbol{R}_i) = \boldsymbol{\Sigma}_i, \qquad (2.5)$$

which is determined by the structure of the correlation matrix of the error vector $\boldsymbol{R}_i$, and by the covariance of the random effects **D**. In addition, the expectation of $\boldsymbol{y}_{j(i)}$ is

$$\mathrm{E}(\boldsymbol{y}_{j(i)}) = \boldsymbol{X}_i\boldsymbol{\beta}. \qquad (2.6)$$

2.2.2 Examples of D matrix

The covariance matrix of the estimated fixed effects $\hat{\boldsymbol{\beta}}$ in (2.8) depends on the elements of the matrix $\boldsymbol{\Sigma}_i$, namely the $\boldsymbol{R}_i$ and **D** matrices. For illustration, let us consider the following cases for **D**:

(1) **D** = 0, which implies that there are no random effects. The model in (2.1) reduces to a fixed-effects model with the random error component.

(2) $\mathbf{D} = \begin{bmatrix} 1 & 0 \\ 0 & 0.2 \end{bmatrix}$, i.e., a random intercept and random slope model with variances for the random intercept d_{11} = 1 and random slope d_{22} = 0.2. The correlation between random intercept and random slope is assumed zero.

(3) $\mathbf{D} = \begin{bmatrix} 1 & 0.5\sqrt{0.2} \\ 0.5\sqrt{0.2} & 0.2 \end{bmatrix}$, which indicates a random intercept and random slope model with correlation between the random intercept and random slope r = 0.5. Note that the covariance between the random intercept and random slope $d_{12} = r\sqrt{d_{11}\, d_{22}}$.

2.2.3 Covariance matrix

For the random intercept and slope model in (2.3), the vector of all observations are denoted by $\underline{\boldsymbol{y}} = (\boldsymbol{y}'_{1(1)},...,\boldsymbol{y}'_{n(1)},\boldsymbol{y}'_{1(2)},...,\boldsymbol{y}'_{n(2)})'$ with its expectation and covariance as follows

$$\mathrm{E}(\underline{\boldsymbol{y}}) = \begin{bmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \mathbf{X}\boldsymbol{\beta}, \qquad \mathrm{Var}(\underline{\boldsymbol{y}}) = \left[\begin{array}{ccc|ccc} \boldsymbol{\Sigma}_1 & & 0 & 0 & \cdots & 0 \\ & \ddots & & \vdots & \ddots & \vdots \\ 0 & & \boldsymbol{\Sigma}_1 & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & \boldsymbol{\Sigma}_2 & & 0 \\ \vdots & \ddots & \vdots & & \ddots & \\ 0 & \cdots & 0 & 0 & & \boldsymbol{\Sigma}_2 \end{array}\right] = \boldsymbol{\Psi}. \qquad (2.7)$$

We note that the ($6n \times 6n$) covariance $\boldsymbol{\Psi}$ has block-diagonal form. The $\boldsymbol{\Sigma}_i$ (i = 1, 2) is the covariance matrix of $\boldsymbol{y}_{j(i)}$ defined in (2.5).

For known $\boldsymbol{\Psi}$, the best linear unbiased estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Psi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Psi}^{-1}\mathbf{Y} = \left[\sum_{i=1}^{2}\sum_{j=1}^{n}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i\right]^{-1}\left[\sum_{i=1}^{2}\sum_{j=1}^{n}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{y}_{j(i)}\right],$$

which has the covariance matrix

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \left[\sum_{i=1}^{2} n\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i\right]^{-1} = \left[n\boldsymbol{X}_1'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{X}_1 + n\boldsymbol{X}_2'\boldsymbol{\Sigma}_2^{-1}\boldsymbol{X}_2\right]^{-1}. \qquad (2.8)$$

The $\mathrm{Var}(\hat{\boldsymbol{\beta}})$ is important for finding the optimal cohort designs that will be discussed in Section 3.2. For instance, the determinant of $\mathrm{Var}(\hat{\boldsymbol{\beta}})$, which is often referred to as the generalized variance, will be used for finding the *D*-optimal cohort designs.

2.3 Design Class

Now we present two design classes: one with two repeated measurements on each subject, and the other with three repeated measurements on each subject.

2.3.1 Design class T_{22}

The response variable $y_{kj(i)}$ in (2.2) is a linear function of time, so the minimum number of repeated measurements required is two. One measurement is taken at the beginning of the study, and the other measurement is taken at the end of the study. If the time period [−1, 1] is used and C = 2 cohorts of equal sample size are considered, then the design class with m = 2 repeated measurements is given by

$$T_{22} = \begin{Bmatrix} (-1,1) & (-1,1) \\ 0.5 & 0.5 \end{Bmatrix}. \qquad (2.9)$$

In some cases, however, more than two repeated measurements are needed to accurately model the longitudinal trend of a continuous response variable over time and to detect any departure from the linear trend. Willett, Singer and Martin (1998) [34] gave a rule of thumb and recommended including at least one more time point (repeated measurement) than the number of fixed-effects parameters in the model. Vickers (2003) [32] argued that although increase the number of repeated measurements to three or four measurements, will increase the power of a test, the benefit of an additional repeated measurement rapidly decreases at the number of measurements rises. His results support the conclusion that it is not very efficient to include too many repeated measurements in a longitudinal study. As a result, the design class T_{23} is considered in this dissertation and is defined in (2.10).

2.3.2 Design class T_{23}

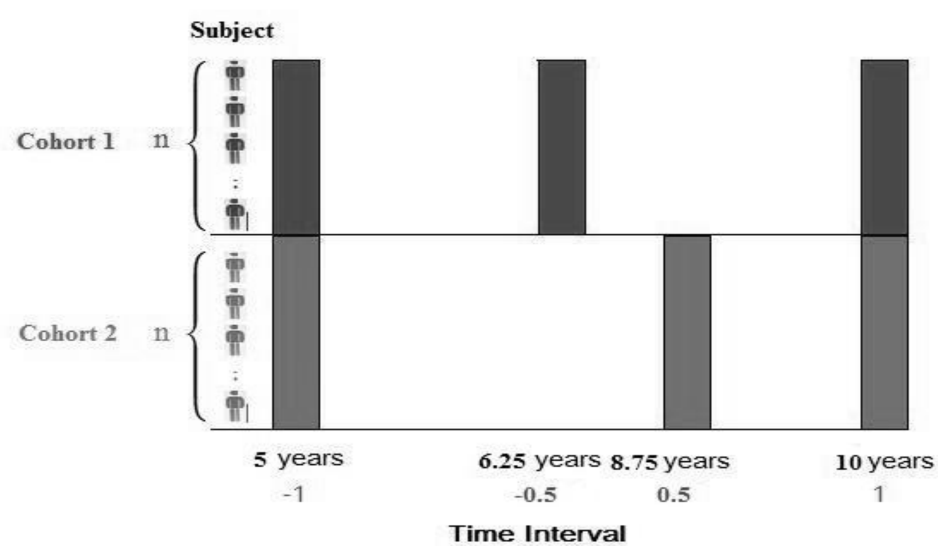
Considering cohort designs with m = 3 repeated measurements obtained from C = 2 cohorts with n subjects each, define the class of cohort designs T_{23} such that

$$T_{23} = \left\{ \begin{matrix} (-1, a_1, 1) & (-1, a_2, 1) \\ 0.5 & 0.5 \end{matrix} \middle| -1 < a_1 \le a_2 < 1 \right\}, \qquad (2.10)$$

and a cohort design $\tau = \begin{pmatrix} (-1, a_1, 1) & (-1, a_2, 1) \\ 0.5 & 0.5 \end{pmatrix}$ for some specified values of a_1 and a_2 satisfying $-1 < a_1 \le a_2 < 1$, where $(-1, a_i, 1)$ are the time points for the i^{th} cohort (i = 1, 2), and 0.5 on the second row is the relative sample size indicating that the two cohorts are equally sized. Note that the first measurement for each subject is taken at $t_{1(i)} = -1$, and the total duration of the study period is fixed and rescaled to [−1, 1], i.e. $t_{3(i)} = 1$.

2.4 Examples

As a motivating example for illustrating some of the issues of cohort designs, let us consider the following longitudinal study. Suppose we are interested in describing children's growth pattern over time, then it would be possible to design a study in which the height of a cohort of children is assessed at different time points. Here, cohorts can be defined by birth. For example, a group of children who were born in a particular period of year, say 1995, form a birth cohort. In our example, suppose there are children in two cohorts based on birth year. The study period is five years and three measures of height are taken from each child. For the 1st cohort (born in 1995), we take observations from those children at their years of age: 5 yrs, 6 yrs and 3 month (6.25), and 10 yrs. For the 2nd cohort (born in 2000), we take observations from those children at their years of age: 5 yrs, 8 yrs and 9 month (8.75), and 10 yrs. We assume that we have complete data for our analysis. The objective of this study is to find how growth pattern changes over time within a cohort and whether growth pattern changes differently in the 1st cohort compared with that in the 2nd cohort. The following figure shows how the data look like. Each bar represents n measurements taken at a particular time point. The height of bars does not have any special meaning.



The time interval [5, 10] is rescaled to [−1, 1] by applying a linear transformation:

$$\frac{x-\frac{5+10}{2}}{\frac{10-5}{2}}=\frac{2(x-7.5)}{5},$$ where x is the age of the child ($5 \le x \le 10$).

Now we can write the aforementioned cohort design as

$$\tau_1 = \begin{Bmatrix} (-1,-0.5,1) & (-1,0.5,1) \\ 0.5 & 0.5 \end{Bmatrix}, \qquad (2.11)$$

where $\boldsymbol{t}_1 = (-1,-0.5,1)'$ and $\boldsymbol{t}_2 = (-1,0.5,1)'$ are the time points for the 1st and 2nd cohort, respectively. Note that $\tau_1 \in \mathrm{T}_{23}$, and $\omega_1 + \omega_2 = 0.5 + 0.5 = 1$.

Suppose we have n = 50 children in each cohort. For the j^{th} child (j = 1,…,50) within the i^{th} cohort (i = 1, 2), the model is

$$\boldsymbol{y}_{j(i)} = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_{j(i)} + \boldsymbol{\varepsilon}_{j(i)},$$

where

$$\mathbf{X}_1 = \boldsymbol{Z}_1 = \begin{bmatrix} 1 & -1 \\ 1 & -0.5 \\ 1 & 1 \end{bmatrix}, \ \mathbf{X}_2 = \boldsymbol{Z}_2 = \begin{bmatrix} 1 & -1 \\ 1 & 0.5 \\ 1 & 1 \end{bmatrix}.$$

For illustration, assuming $\mathbf{D} = \begin{bmatrix} 50 & 30 \\ 30 & 80 \end{bmatrix}$, $\boldsymbol{R}_i = \mathbf{I}$ (i = 1, 2), and $\sigma^2 = 1$, the covariance matrix of parameter estimators can be found as

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}_{\tau_1}) = \left[n\sigma^2 \sum_{i=1}^{2} \{\boldsymbol{X}_i'(\boldsymbol{Z}_i\mathbf{D}\boldsymbol{Z}_i' + \mathbf{I})^{-1}\boldsymbol{X}_i\} \right]^{-1} = \begin{bmatrix} 0.5035 & 0.3 \\ 0.3 & 0.8046 \end{bmatrix},$$

with its determinant $\det\{\mathrm{Var}(\hat{\boldsymbol{\beta}}_{\tau_1})\} = 0.3151$, where $\hat{\boldsymbol{\beta}}_{\tau_1}$ is the estimator of $\hat{\boldsymbol{\beta}}$ under the design τ_1. Clearly, the numerical value of $\det\{\mathrm{Var}(\hat{\boldsymbol{\beta}})\}$ varies across different cohort designs. Based on D-optimality criterion that we will explain in the next chapter, the smaller the numerical value of $\det\{\mathrm{Var}(\hat{\boldsymbol{\beta}})\}$, the more efficient the cohort design. Now the following questions might be raised: (1) is $\det\{\mathrm{Var}(\hat{\boldsymbol{\beta}}_{\tau_1})\}$ calculated under the cohort design τ_1 small enough? (2) Does τ_1 give the optimal allocation of time points? To answer those questions, the concept of optimal cohort designs is first discussed in Chapter 3.

Chapter 3

Optimal Cohort Designs for $T_{23(a)}$ with $\boldsymbol{R_i} = \mathbf{I}$

3.1 Introduction

To find optimal cohort designs that yield precise estimation of the fixed parameters in model (2.2), we have to decide on the optimal allocation of time points. The 'optimal' allocation of time points depends on the specified optimality criterion, which is a function of the covariance matrix in (2.8). In this chapter, we briefly introduce *D*- and *A*-optimality criteria and the relative efficiency. Then *D*- and *A*-optimality criteria are used in finding the optimal cohort designs for the design class $T_{23(a)}$ defined in (3.1) with correlation matrix $\boldsymbol{R_i} = \mathbf{I}$. Moreover, we present conditions on the covariance matrix $\mathbf{D}$ so that the cohort design with equidistant time points is *D*-optimal. The results are illustrated by *D*-optimality regions.

3.2 Optimality Criteria

The concept of optimal designs was introduced by Wald (1943) [33]. The optimal design τ^* is the design, among the design class T_{23} (2.10), for which Var($\hat{\boldsymbol{\beta}}$) in (2.8) is minimized. Since matrices cannot be minimized in a unique way, Kiefer (1959) [11] proposed several meaningful functions as optimality criteria for studying optimal designs. Two optimality criteria are used in this chapter, namely *D*- and *A*-optimality. The most often used *D*-optimality criterion seeks to minimize the determinant of the covariance matrix of the parameter estimates. This is equivalent to minimizing the volume of the confidence ellipsoid for the estimated regression parameters. Obviously, a smaller volume implies better estimation. The *A*-optimality seeks to minimize the trace of the covariance matrix, i.e. the average variance of the parameter estimates. So, A in the name of this criterion stands for average.

In the present chapter, $\boldsymbol{R_i}$ = **I** correlation structure is considered, i.e. the errors are assumed to be uncorrelated. For simplicity, we first consider a special case of the design class T_{23} (2.10) by assuming the time points for both cohorts are identical, i.e. $a_1 = a_2 = a$. We define

$$\mathrm{T}_{23(a)} = \left\{ \begin{matrix} (-1,a,1) & (-1,a,1) \\ 0.5 & 0.5 \end{matrix} \middle| -1 < a < 1 \right\}. \qquad (3.1)$$

We use the random intercept and slope model that we have discussed earlier in modeling the longitudinal data. For the j^{th} subject in the i^{th} cohort, the model is given by

$$\boldsymbol{y}_{j(i)} = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_{j(i)} + \boldsymbol{\varepsilon}_{j(i)}, \qquad j = 1,\ldots,n,\ i = 1, 2,$$

where

$$\mathbf{X}_i = \boldsymbol{Z}_i = \mathbf{X} = \boldsymbol{Z} = \begin{bmatrix} 1 & -1 \\ 1 & a \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \underline{\boldsymbol{j}} & \underline{\boldsymbol{a}} \end{bmatrix},$$
$$\underline{\boldsymbol{a}} = (-1, a, 1)', \text{ and } \underline{\boldsymbol{j}} = (1,1,1)'.$$

Therefore, the covariance matrix of $\boldsymbol{y}_{j(i)}$ is homogeneous across subjects and cohorts, i.e. $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$ for both i (i = 1, 2). It follows from (2.8) that the variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ is

$$\text{Var}(\hat{\boldsymbol{\beta}}) = \begin{bmatrix} 2n\underline{\boldsymbol{j}}'\boldsymbol{\Sigma}^{-1}\underline{\boldsymbol{j}} & 2n\underline{\boldsymbol{j}}'\boldsymbol{\Sigma}^{-1}\underline{\boldsymbol{a}} \\ 2n\underline{\boldsymbol{j}}'\boldsymbol{\Sigma}^{-1}\underline{\boldsymbol{a}} & 2n\underline{\boldsymbol{a}}'\Sigma^{-1}\underline{\boldsymbol{a}} \end{bmatrix}^{-1}. \qquad (3.2)$$

Consequently, the inverse of $\frac{2n}{\sigma^2}\text{Var}(\hat{\boldsymbol{\beta}})$ is given by

$$\left[\frac{2n}{\sigma^2}\text{Var}(\hat{\boldsymbol{\beta}})\right]^{-1} = \begin{bmatrix} \underline{\boldsymbol{j}}'(\mathbf{ZD}\boldsymbol{Z}'+\mathbf{I})^{-1}\underline{\boldsymbol{j}} & \underline{\boldsymbol{j}}'(\mathbf{ZD}\boldsymbol{Z}'+\mathbf{I})^{-1}\underline{\boldsymbol{a}} \\ \underline{\boldsymbol{j}}'(\mathbf{ZD}\boldsymbol{Z}'+\mathbf{I})^{-1}\underline{\boldsymbol{a}} & \underline{\boldsymbol{a}}'(\mathbf{ZD}\boldsymbol{Z}'+\mathbf{I})^{-1}\underline{\boldsymbol{a}} \end{bmatrix}, \qquad (3.3)$$

which will be used for finding the *D*-optimal cohort designs in section 3.2.1. Here, the symmetric 2 × 2 matrix in (3.3) only depends on three model parameters: d_{11}, d_{12}, and d_{22} and one design parameter a. The range of those parameters is shown in Table 3.1.

Table 3.1: The range of parameters

a	d_{11}	d_{22}	d_{12}
(−1, 1)	> 0	> 0	$\left(-\sqrt{d_{11}d_{22}}, \sqrt{d_{11}d_{22}}\right)$

Notice that (1) since we are considering cohort designs with three distinct time points, we assume the additional measurement cannot be taken at the beginning ($a \neq -1$) or at the end ($a \neq 1$) of the study; (2) both d_{11} and d_{22} are greater than zero, as they are variances for the random intercept and random slope, respectively; (3) the absolute value of d_{12} is less than $\sqrt{d_{11}d_{22}}$ to make sure that matrix **D** is positive-definite.

3.2.1 *D*-optimality

The *D*-optimality criterion minimizes the determinant of the covariance matrix $\mathrm{Var}(\hat{\boldsymbol{\beta}})$ by choosing a *D*-optimal cohort design τ_D such that, for each design $\tau \in \mathrm{T}_{23(a)}$,

$$\det\left\{\mathrm{Var}(\hat{\boldsymbol{\beta}}_{\tau_D})\right\} \leq \det\left\{\mathrm{Var}(\hat{\boldsymbol{\beta}}_{\tau})\right\},$$

where $\hat{\boldsymbol{\beta}}_{\tau_D}$ and $\hat{\boldsymbol{\beta}}_{\tau}$ are estimators of $\hat{\boldsymbol{\beta}}$ under design τ_D and τ, respectively. Minimizing $\det\left\{\mathrm{Var}(\hat{\boldsymbol{\beta}})\right\}$ is equivalent to maximizing $\det\left\{[\mathrm{Var}(\hat{\boldsymbol{\beta}})]^{-1}\right\}$.

We denote $\det\left\{\left[2n\mathrm{Var}(\hat{\boldsymbol{\beta}})/\sigma^2\right]^{-1}\right\}$ as $Q_1(a)$, where

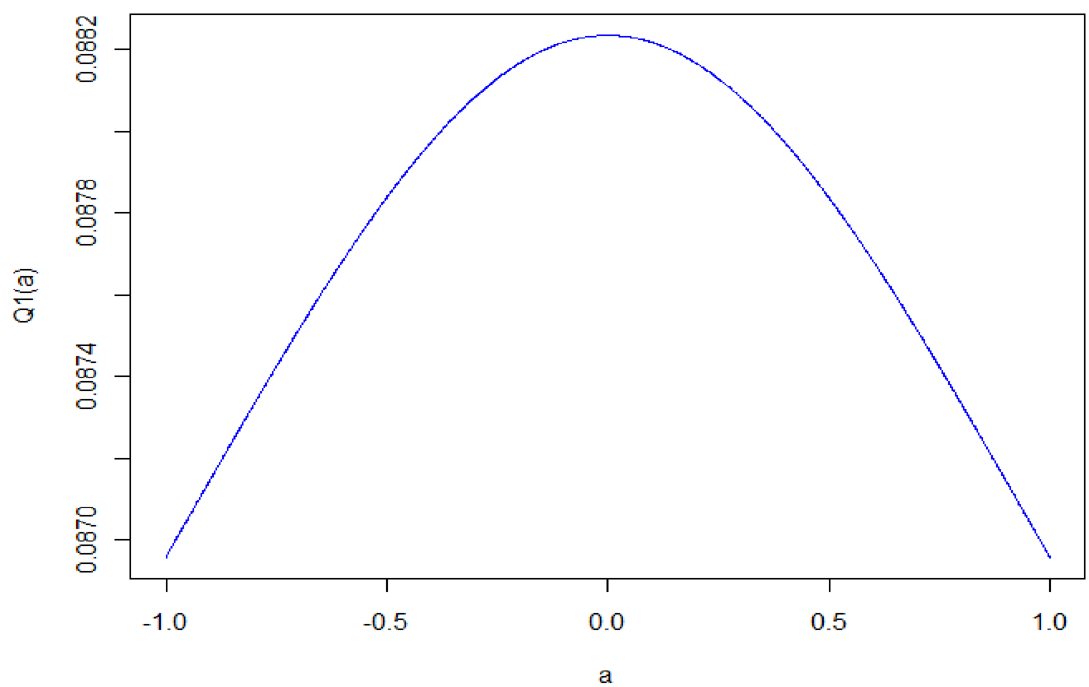
$$Q_1(a) = \left(\underline{\boldsymbol{j}}'(\mathbf{ZDZ}'+\mathbf{I})^{-1}\underline{\boldsymbol{j}}\right)\left(\underline{\boldsymbol{a}}'(\mathbf{ZDZ}'+\mathbf{I})^{-1}\underline{\boldsymbol{a}}\right) - \left(\underline{\boldsymbol{j}}'(\mathbf{ZDZ}'+\mathbf{I})^{-1}\underline{\boldsymbol{a}}\right)^2. \quad (3.4)$$

Here, we divide by σ^2 to consider a criterion that is scale free. Of course, n is the number of subjects within each cohort.

The design $\tau_D \in \mathrm{T}_{23(a)}$ with the maximum value of $Q_1(a)$ in (3.4) is considered *D*-optimal.

Consider again, the example of longitudinal study on children's growth pattern in Section 2.4. For three repeated measurements taken at time points (-1, a, 1) from each child in both cohorts, the *D*-optimal cohort design is obtained by maximizing $Q_1(a)$ as a function of time point 'a', if the **D** matrix is assumed known. Suppose $\mathbf{D} = \begin{bmatrix} 1 & 0 \\ 0 & 8 \end{bmatrix}$, numerical values of $Q_1(a)$ are calculated and plotted against the time point a ($-1 < a < 1$) in Figure 3.1.

Figure 3.1: Plot of $Q_1(a)$ given $d_{11} = 1$, $d_{12} = 0$, and $d_{22} = 8$



It can be seen that $Q_1(a)$ reaches its maximum 0.0882 when a is 0, which indicates that a = 0 is the best choice with respect to D-optimality criterion. Therefore, the D-optimal design among design class $T_{23(a)}$ is $\tau_D = \begin{Bmatrix} (-1,0,1) & (-1,0,1) \\ 0.5 & 0.5 \end{Bmatrix}$ given $\mathbf{D} = \begin{bmatrix} 1 & 0 \\ 0 & 8 \end{bmatrix}$.

3.2.2 *A*-optimality

The *A*-optimality minimizes sum of the variances of the estimated parameters, which is the sum of the diagonal elements of $\mathrm{Var}(\hat{\boldsymbol{\beta}})$ in (3.2). Denote $\mathrm{tr}\{\mathbf{M}\}$ the trace of matrix **M**. An *A*-optimal design τ_A is a design such that, for each design $\tau \in \mathrm{T}_{23(a)}$,

$$\mathrm{tr}\left\{\mathrm{Var}(\hat{\boldsymbol{\beta}}_{\tau_A})\right\} \le \mathrm{tr}\left\{\mathrm{Var}(\hat{\boldsymbol{\beta}}_{\tau})\right\},$$

where $\hat{\boldsymbol{\beta}}_{\tau_A}$ and $\hat{\boldsymbol{\beta}}_{\tau}$ are estimators of $\hat{\boldsymbol{\beta}}$ under design τ_A and τ, respectively. Without loss of generality, *A*-optimal designs is found by minimizing $\mathrm{tr}\left\{2n\mathrm{Var}(\hat{\boldsymbol{\beta}})\big/\sigma^2\right\}$ where

$$\frac{2n}{\sigma^2}\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \begin{bmatrix} \underline{\boldsymbol{j}}'(\mathbf{ZDZ}'+\mathbf{I})^{-1}\underline{\boldsymbol{j}} & \underline{\boldsymbol{j}}'(\mathbf{ZDZ}'+\mathbf{I})^{-1}\underline{\boldsymbol{a}} \\ \underline{\boldsymbol{j}}'(\mathbf{ZDZ}'+\mathbf{I})^{-1}\underline{\boldsymbol{a}} & \underline{\boldsymbol{a}}'(\mathbf{ZDZ}'+\mathbf{I})^{-1}\underline{\boldsymbol{a}} \end{bmatrix}^{-1}, \tag{3.5}$$

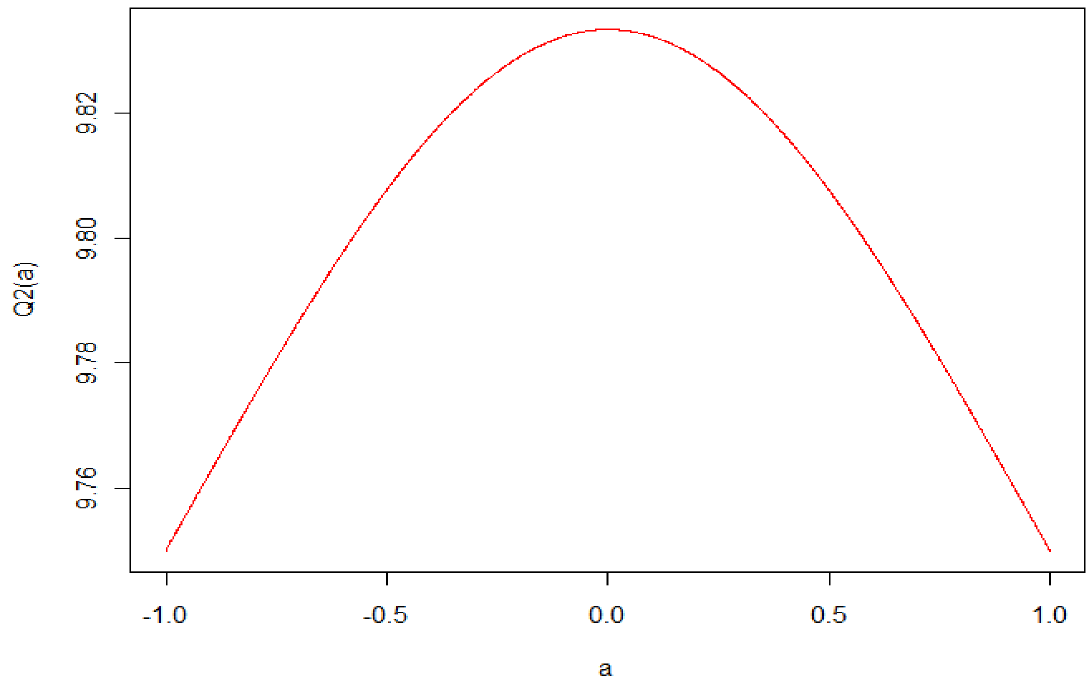
Define $\mathrm{tr}\left\{2n\mathrm{Var}(\hat{\boldsymbol{\beta}})\big/\sigma^2\right\}$ as $Q_2(a)$.

It can be verified that

$$Q_2(a) = \frac{\underline{\boldsymbol{j}}'(\mathbf{ZDZ}'+\mathbf{I})^{-1}\underline{\boldsymbol{j}} + \underline{\boldsymbol{a}}'(\mathbf{ZDZ}'+\mathbf{I})^{-1}\underline{\boldsymbol{a}}}{Q_1(a)}, \tag{3.6}$$

where $Q_1(a)$ has been defined in (3.4). The design $\tau_A \in \mathrm{T}_{23(a)}$ with the minimum value of $Q_2(a)$ in (3.6) is considered *A*-optimal. Revisit the example that we have talked about in Section 3.2.1. Under the same condition $\mathbf{D} = \begin{bmatrix} 1 & 0 \\ 0 & 8 \end{bmatrix}$, *A*-optimal designs are constructed by minimizing $Q_2(a)$, where $Q_2(a)$ is simplified to a function of *a* only and shown in Figure 3.2.

Figure 3.2: Plot of $Q_2(a)$ given d_{11} = 1, d_{12} = 0, and d_{22} = 8



In Figure 3.2, the curve represents the function $Q_2(a)$. It can be observed that $Q_2(a)$ decreases as a moves further away from the center point '0'. If we put a threshold on a ($-1 < a < 1$) such that $a^2 \le 0.5$ (Note: choosing a^2 very close to 1 should be avoided, as design class $T_{23(a)}$ involves three distinct time points), then the designs with $a = \pm\sqrt{0.5}$ are considered *A*-optimal. In other words, given $\mathbf{D} = \begin{bmatrix} 1 & 0 \\ 0 & 8 \end{bmatrix}$, the *A*-optimal designs are

$$\tau_A = \begin{Bmatrix} (-1,-\sqrt{0.5},1) & (-1,-\sqrt{0.5},1) \\ 0.5 & 0.5 \end{Bmatrix} \text{ and } \tau_A = \begin{Bmatrix} (-1,\sqrt{0.5},1) & (-1,\sqrt{0.5},1) \\ 0.5 & 0.5 \end{Bmatrix}.$$

3.2.3 The Relative Efficiency

To compare the efficiencies of two cohort designs τ_1 and τ_2, the relative efficiency can be used. The *D*- and *A*-relative efficiency of a design τ_2 compared with a design τ_1 are given by the ratio (Atkinson and Donev, 1992 [1]):

$$D\text{-Eff}(\tau_2 \mid \tau_1) = \frac{\det\{\mathrm{Var}(\hat{\boldsymbol{\beta}}_{\tau_2})\}}{\det\{\mathrm{Var}(\hat{\boldsymbol{\beta}}_{\tau_1})\}}, \qquad (3.7)$$

$$A\text{-Eff}(\tau_2 \mid \tau_1) = \frac{\mathrm{tr}\{\mathrm{Var}(\hat{\boldsymbol{\beta}}_{\tau_2})\}}{\mathrm{tr}\{\mathrm{Var}(\hat{\boldsymbol{\beta}}_{\tau_1})\}}, \qquad (3.8)$$

respectively. $\hat{\boldsymbol{\beta}}_{\tau_1}$ and $\hat{\boldsymbol{\beta}}_{\tau_2}$ are estimators of $\hat{\boldsymbol{\beta}}$ under design τ_1 and τ_2, correspondingly. If the ratio is less than 1, then design τ_2 is more efficient than design τ_1.

To illustrate, let us compare the efficiencies of *D*- and *A*-optimal cohort designs found in Section 3.2.1 and 3.2.2. We present the numerical values of *D*- and *A*-criterion functions for those three designs in Table 3.2.

Table 3.2: Optimality criterion functions

Design	Determinant $\det\{2n\mathrm{Var}(\hat{\boldsymbol{\beta}})/\sigma^2\}$	Trace $\mathrm{tr}\{2n\mathrm{Var}(\hat{\boldsymbol{\beta}})/\sigma^2\}$
$\tau_D = \begin{Bmatrix} (-1,0,1) & (-1,0,1) \\ 0.5 & 0.5 \end{Bmatrix}$	11.3333	9.8333
$\tau_A = \begin{Bmatrix} (-1,-\sqrt{0.5},1) & (-1,-\sqrt{0.5},1) \\ 0.5 & 0.5 \end{Bmatrix}$	11.4286	9.7857
$\tau_A = \begin{Bmatrix} (-1,\sqrt{0.5},1) & (-1,\sqrt{0.5},1) \\ 0.5 & 0.5 \end{Bmatrix}$	11.4286	9.7857

It is not surprising to see that both the determinant and trace of $2n\text{Var}(\hat{\boldsymbol{\beta}})/\sigma^2$ remain the same for $a = -\sqrt{0.5}$ and $a = \sqrt{0.5}$, since from Figure 3.1 and Figure 3.2, respectively, we observe that $Q_1(a)$ and $Q_1(a)$ are symmetric about $a = 0$ given $\mathbf{D} = \begin{bmatrix} 1 & 0 \\ 0 & 8 \end{bmatrix}$.

The D-efficiencies of A-optimal designs τ_A relative to the D-optimal design τ_D is given by: $D\text{-Eff}(\tau_D \mid \tau_A) = \dfrac{\det\{\text{Var}(\hat{\boldsymbol{\beta}}_D)\}}{\det\{\text{Var}(\hat{\boldsymbol{\beta}}_A)\}} = \dfrac{11.3333}{11.4286} = 0.9917$. The result implies that, under D-optimality criterion, adding the additional time point in the middle of the study ($a = 0$) reduces the variance of the parameter estimators by about 1%, compared to placing the time point at a distance of $\sqrt{0.5}$ away from 0. The reduction in variance is small due to the facts that the assumed **D** matrix has small numerical values for d_{11}, d_{12}, and d_{22}, and the time interval $[-1, 1]$ is short.

3.3 *D*-optimal Designs

The D-optimal design that we found in Section 3.2.1 indicates $a = 0$ is the best choice under the condition $\mathbf{D} = \begin{bmatrix} 1 & 0 \\ 0 & 8 \end{bmatrix}$. If $a = 0$, the time points are equally spaced. However, in reality, equidistant time points do not always yield efficient estimation of parameters. Therefore, we now raise the question as to which design, equidistant time points or non-equidistant time points, should be implemented under what condition.

3.3.1 Comparison of $Q_1(0)$ with $Q_1(a)$

Recall that a *D*-optimal design is found by maximizing $Q_1(a)$ in (3.4). To evaluate the designs with $a = 0$ vs. $a \neq 0$, we compare $Q_1(0)$ with $Q_1(a)$ for $a \neq 0$ and $-1 < a < 1$.

The exact expression of $Q_1(a)$ for $-1 < a < 1$ is

$$Q_1(a) = \frac{\begin{pmatrix} d_{22}(2a^4+10a^2+12)+d_{11}(d_{22}(4a^4+24a^2+36)+6a^2+18) \\ +d_{12}(4a^3+12a)+2a^2-d_{12}^2(4a^4+24a^2+36)+6 \end{pmatrix}}{\begin{pmatrix} 3d_{11}+2d_{22}+2ad_{12}+6d_{11}d_{22}+a^2d_{22} \\ -6d_{12}^2-2a^2d_{12}^2+2a^2d_{11}d_{22}+1 \end{pmatrix}^2} = \frac{\mathrm{Num}_1(a)}{\mathrm{Den}_1(a)}, \qquad (3.9)$$

which is a function of a, d_{11}, d_{22}, and d_{12}. Notice that $\mathrm{Den}_1(a)$ is positive for all a, d_{11}, d_{22}, and d_{12}. We consider the difference between $Q_1(0)$ and $Q_1(a)$ as follows

$$Q_1(0)-Q_1(a) = \frac{\mathrm{Num}_1(0)}{\mathrm{Den}_1(0)} - \frac{\mathrm{Num}_1(a)}{\mathrm{Den}_1(a)}. \qquad (3.10)$$

With $\mathrm{Den}_1(0) > 0,\ \mathrm{Den}_1(a) > 0,$ we know

$$Q_1(0)-Q_1(a) > 0 \ \Leftrightarrow \mathrm{Num}_1(0)\mathrm{Den}_1(a) - \mathrm{Num}_1(a)\mathrm{Den}_1(0) > 0.$$

Define $\mathrm{D}_f^{(1)} = \mathrm{Num}_1(0)\mathrm{Den}_1(a) - \mathrm{Num}_1(a)\mathrm{Den}_1(0),$ such that

$$\begin{aligned}\mathrm{D}_f^{(1)} = {}& -2a(a-6d_{12}+3ad_{11}-ad_{22})(3d_{11}-6d_{12}^2+2d_{22}+6d_{11}d_{22}+1) \\ & (3d_{11}+2d_{22}+2ad_{12}+6d_{11}d_{22}+a^2d_{22}-6d_{12}^2-2a^2d_{12}^2+2a^2d_{11}d_{22}+1).\end{aligned}$$

Denote

$c^{(0)} = -2a,$

$$c^{(1)} = a - 6d_{12} + 3ad_{11} - ad_{22} = a(1 + 3d_{11} - d_{22}) - 6d_{12},$$

$$c^{(2)} = 3d_{11} + 2d_{22} + 6d_{11}d_{22} - 6d_{12}^2 + 1 = 3d_{11} + 2d_{22} + 6(d_{11}d_{22} - d_{12}^2) + 1 > 0,$$

$$\begin{aligned} c^{(3)} &= 3d_{11} + 2d_{22} + 2ad_{12} + 6d_{11}d_{22} + a^2 d_{22} - 6d_{12}^2 - 2a^2 d_{12}^2 + 2a^2 d_{11}d_{22} + 1 \\ &= (3d_{11} + 2d_{22} + 6d_{11}d_{22} - 6d_{12}^2 + 1) + a^2(d_{22} + 2d_{11}d_{22} - 2d_{12}^2) + 2ad_{12}, \end{aligned}$$

$$\begin{aligned} &\text{where } 3d_{11} + 2d_{22} + 6d_{11}d_{22} - 6d_{12}^2 + 1 = 3d_{11} + 2d_{22} + 1 + 6(d_{11}d_{22} - d_{12}^2) > 0, \\ &\qquad d_{22} + 2d_{11}d_{22} - 2d_{12}^2 = d_{22} + 2(d_{11}d_{22} - d_{12}^2) > 0. \end{aligned} \qquad (3.11)$$

Then, $\mathrm{D}_f^{(1)}$ can be rewritten as the product of $c^{(0)}$, $c^{(1)}$, $c^{(2)}$, and $c^{(3)}$

$$\mathrm{D}_f^{(1)} = c^{(0)} c^{(1)} c^{(2)} c^{(3)}. \qquad (3.12)$$

We note that when $a = 0$, $c^{(0)} = 0$ so that $\mathrm{D}_f^{(1)} = 0$. Because it is unnecessary to compare $Q_1(0)$ with $Q_1(a = 0)$, we evaluate the signs of $\mathrm{D}_f^{(1)}$ for $a \neq 0$ and $-1 < a < 1$ in the following nine cases.

Table 3.3: Analysis of $\mathrm{D}_f^{(1)}$ in nine cases

	$c^{(1)} > 0$	$c^{(1)} < 0$	$c^{(1)} = 0$
$d_{12} > 0$	I	II	III
$d_{12} < 0$	IV	V	VI
$d_{12} = 0$	VII	VIII	IX

3.3.2 The Situation with $-1 < a < 0$

We present the signs of $c^{(0)}$, $c^{(1)}$, $c^{(2)}$, $c^{(3)}$, and $\mathrm{D}_f^{(1)}$ for the aforementioned nine cases in Table 3.4.

Table 3.4: Analysis of $D_f^{(1)}$ for $-1 < a < 0$

Case	$c^{(0)}$	$c^{(1)}$	$c^{(2)}$	$c^{(3)}$	$D_f^{(1)}$
I	> 0	> 0	> 0	> 0	> 0
II	> 0	< 0	> 0	> 0, if $6(d_{11}d_{22}-d_{12}^2)+2a^2(d_{11}d_{22}-d_{12}^2)+(1+3d_{11}+2d_{22}+a^2d_{22})>-2ad_{12}>0$ < 0, if $0<6(d_{11}d_{22}-d_{12}^2)+2a^2(d_{11}d_{22}-d_{12}^2)+(1+3d_{11}+2d_{22}+a^2d_{22})<-2ad_{12}$	< 0 > 0
III	> 0	= 0	> 0	> 0	= 0
IV	> 0	> 0	> 0	> 0	> 0
V	> 0	< 0	> 0	> 0	< 0
VI	> 0	= 0	> 0	> 0	= 0
VII	> 0	> 0	> 0	> 0	> 0
VIII	> 0	< 0	> 0	> 0	< 0
IX	> 0	= 0	> 0	> 0	= 0

When $D_f^{(1)} > 0$, the design with $a = 0$ is a *D*-optimal design. When $D_f^{(1)} = 0$, the designs with $a = 0$ and $a \neq 0$ are indistinguishable with respect to *D*-optimality criterion. Furthermore, when $D_f^{(1)} < 0$, the design with $a = 0$ is not a *D*-optimal design.

Theorem 3.1

When $-1 < a < 0$, we have

1. $D_f^{(1)} \geq 0$ for all a, d_{11}, d_{22}, and d_{12} satisfying $c^{(1)} \geq 0$,

2. $D_f^{(1)} > 0$ for all a, d_{11}, d_{22}, and $d_{12} > 0$ satisfying $c^{(1)} < 0$ and $c^{(3)} < 0$,

3. $\mathrm{D}_f^{(1)} < 0$ for all a, d_{11}, d_{22}, and (i) $d_{12} \leq 0$ or (ii) $d_{12} > 0$ and $c^{(3)} > 0$ satisfying $c^{(1)} < 0$.

Proof.

1. Suppose a, d_{11}, d_{22}, and d_{12} satisfying $c^{(1)} = 0$. In (3.12), $\mathrm{D}_f^{(1)} = c^{(0)}c^{(1)}c^{(2)}c^{(3)} = 0$.

   When a $(-1 < a < 0)$, d_{11}, d_{22}, and d_{12} satisfying $c^{(1)} > 0$, from (3.11)

   $c^{(1)} = a(1+3d_{11}-d_{22})-6d_{12} > 0,$ so $\dfrac{a^2(1+3d_{11}-d_{22})}{3} < 2ad_{12}$ which implies

   $\dfrac{1}{3}a^2d_{22}+2ad_{12} > \dfrac{1}{3}a^2+a^2d_{11} > 0.$ Hence, from (3.11)

   $$c^{(3)} = 1+6(d_{11}d_{22}-d_{12}^2)+2a^2(d_{11}d_{22}-d_{12}^2)+(3d_{11}+2d_{22}+\frac{2}{3}a^2d_{22})$$
   $$+(\frac{1}{3}a^2d_{22}+2ad_{12}) > 0.$$

   Since $c^{(0)} = -2a\sigma^4 > 0,$ and $c^{(2)} = 3d_{11}+2d_{22}+6(d_{11}d_{22}-d_{12}^2)+1 > 0,$

   we have $\mathrm{D}_f^{(1)} = c^{(0)}c^{(1)}c^{(2)}c^{(3)} > 0$ and the result follows.

2. Suppose a, d_{11}, d_{22}, and $d_{12} > 0$ satisfying $c^{(1)} < 0$ and $c^{(3)} < 0.$ For $-1 < a < 0$,

   $c^{(0)} = -2a\sigma^4 > 0$ and $c^{(2)} = 3d_{11}+2d_{22}+6(d_{11}d_{22}-d_{12}^2)+1 > 0.$ Then from (3.12),

   we have $\mathrm{D}_f^{(1)} = c^{(0)}c^{(1)}c^{(2)}c^{(3)} > 0$ and the result follows.

3. (i) Suppose a, d_{11}, d_{22}, and $d_{12} \leq 0$ satisfying $c^{(1)} < 0.$ Then from (3.11),

   $c^{(1)} = a(1+3d_{11}-d_{22})-6d_{12} < 0,$ so $\dfrac{a^2(1+3d_{11}-d_{22})}{3} > 2ad_{12} \geq 0.$

   $$c^{(3)} = (3d_{11}+2d_{22}+6d_{11}d_{22}-6d_{12}^2+1)+a^2(d_{22}+2d_{11}d_{22}-2d_{12}^2)+2ad_{12} > 0.$$

   Since $c^{(0)} = -2a\sigma^4 > 0,$ and $c^{(2)} = 3d_{11}+2d_{22}+6(d_{11}d_{22}-d_{12}^2)+1 > 0,$

we have $\mathrm{D}_f^{(1)} = c^{(0)}c^{(1)}c^{(2)}c^{(3)} < 0$ and the result follows.

(*ii*) Suppose a, d_{11}, d_{22}, and d_{12} satisfying $c^{(1)} < 0$ and $c^{(3)} > 0$. For $-1 < a < 0$,

$c^{(0)} = -2a\sigma^4 > 0$ and $c^{(2)} = 3d_{11} + 2d_{22} + 6(d_{11}d_{22} - d_{12}^2) + 1 > 0$.

From (3.12), we have $\mathrm{D}_f^{(1)} = c^{(0)}c^{(1)}c^{(2)}c^{(3)} < 0$ and the result follows.

Theorem 3.2

When $-1 < a < 0$,

1. The condition that all a, d_{11}, d_{22}, and d_{12} satisfying $c^{(1)} \geq 0$ implies that d_{12} must satisfy

$$-\sqrt{d_{11}d_{22}} < d_{12} \leq \min\left\{\sqrt{d_{11}d_{22}},\ \frac{a}{6}(1+3d_{11}-d_{22})\right\} \text{ for all } a,\ d_{11}, \text{ and } d_{22}.$$

2. The condition that all a, d_{11}, d_{22}, and $d_{12} > 0$ satisfying $c^{(1)} < 0$ and $c^{(3)} < 0$ implies

$$\max\left\{0,\ \frac{a}{6}(1+3d_{11}-d_{22})\right\} < d_{12} < \sqrt{d_{11}d_{22}} \text{ and}$$
$$0 < 6(d_{11}d_{22} - d_{12}^2) + 2a^2(d_{11}d_{22} - d_{12}^2) + (1 + 3d_{11} + 2d_{22} + a^2 d_{22}) < -2ad_{12}$$

for all a, d_{11}, and d_{22}.

3. (*i*) The condition that all a, d_{11}, d_{22}, and $d_{12} \leq 0$ satisfying $c^{(1)} < 0$ implies that d_{12} must satisfy

$$\max\left\{-\sqrt{d_{11}d_{22}},\ \frac{a}{6}(1+3d_{11}-d_{22})\right\} < d_{12} \leq 0 \text{ for all } a,\ d_{11}, \text{ and } d_{22}.$$

(*ii*) The condition that all a, d_{11}, d_{22}, and $d_{12} > 0$ satisfying $c^{(1)} < 0$ and $c^{(3)} > 0$

implies that

$$\max\left\{0,\ \frac{a}{6}(1+3d_{11}-d_{22})\right\}<d_{12}<\sqrt{d_{11}d_{22}} \text{ and}$$
$$6(d_{11}d_{22}-d_{12}^2)+2a^2(d_{11}d_{22}-d_{12}^2)+(1+3d_{11}+2d_{22}+a^2d_{22})>-2ad_{12}>0$$

for all a, d_{11}, and d_{22}.

Proof.

1. Suppose a, d_{11}, d_{22}, and d_{12} satisfying $c^{(1)}\geq 0$. Then from (3.11),

   $c^{(1)}=a(1+3d_{11}-d_{22})-6d_{12}\geq 0,$ which implies $d_{12}\leq\frac{a}{6}(1+3d_{11}-d_{22}).$

   Since $-\sqrt{d_{11}d_{22}}<d_{12}<\sqrt{d_{11}d_{22}},$ we have

   $-\sqrt{d_{11}d_{22}}<d_{12}\leq\min\left\{\sqrt{d_{11}d_{22}},\ \frac{a}{6}(1+3d_{11}-d_{22})\right\}$ and the result follows.

2. Suppose a, d_{11}, d_{22}, and $d_{12}>0$ satisfying $c^{(1)}<0$ and $c^{(3)}<0.$ Then from (3.11),

   $c^{(1)}=a(1+3d_{11}-d_{22})-6d_{12}<0,$ which implies $\frac{a}{6}(1+3d_{11}-d_{22})<d_{12}.$

   Since $0<d_{12}<\sqrt{d_{11}d_{22}},$ we have $\max\left\{0,\ \frac{a}{6}(1+3d_{11}-d_{22})\right\}<d_{12}<\sqrt{d_{11}d_{22}}.$

   $c^{(3)}=6(d_{11}d_{22}-d_{12}^2)+2a^2(d_{11}d_{22}-d_{12}^2)+(1+3d_{11}+2d_{22}+a^2d_{22})+2ad_{12}<0.$

   Hence, $6(d_{11}d_{22}-d_{12}^2)+2a^2(d_{11}d_{22}-d_{12}^2)+(1+3d_{11}+2d_{22}+a^2d_{22})<-2ad_{12}$

   and the result follows.

3. (i) Suppose a, d_{11}, d_{22}, and $d_{12} \le 0$ satisfying $c^{(1)} < 0$. From (3.11),

$c^{(1)} = a(1+3d_{11} - d_{22}) - 6d_{12} < 0$, which implies $\frac{a}{6}(1+3d_{11} - d_{22}) < d_{12}$.

Since $-\sqrt{d_{11}d_{22}} < d_{12} \le 0$, we have $\max\left\{-\sqrt{d_{11}d_{22}},\ \frac{a}{6}(1+3d_{11} - d_{22})\right\} < d_{12} \le 0$

and the result follows.

(ii) Suppose a, d_{11}, d_{22}, and $d_{12} > 0$ satisfying $c^{(1)} < 0$ and $c^{(3)} > 0$. From (3.11),

$c^{(1)} = a(1+3d_{11} - d_{22}) - 6d_{12} < 0$, which implies that $\frac{a}{6}(1+3d_{11} - d_{22}) < d_{12}$.

Since $0 < d_{12} < \sqrt{d_{11}d_{22}}$, we have $\max\left\{0,\ \frac{a}{6}(1+3d_{11} - d_{22})\right\} < d_{12} < \sqrt{d_{11}d_{22}}$.

$c^{(3)} = 6(d_{11}d_{22} - d_{12}^2) + 2a^2(d_{11}d_{22} - d_{12}^2) + (1+3d_{11} + 2d_{22} + a^2 d_{22}) + 2ad_{12} > 0$.

Hence, $6(d_{11}d_{22} - d_{12}^2) + 2a^2(d_{11}d_{22} - d_{12}^2) + (1+3d_{11} + 2d_{22} + a^2 d_{22}) > -2ad_{12}$

and the result follows.

Theorem 3.3

When $-1 < a < 0$, we have

1. $\mathrm{D}_f^{(1)} \ge 0$ for all a, d_{11}, d_{22}, and d_{12} satisfying

   $-\sqrt{d_{11}d_{22}} < d_{12} \le \min\left\{\sqrt{d_{11}d_{22}},\ \frac{a}{6}(1+3d_{11} - d_{22})\right\}$,

2. $\mathrm{D}_f^{(1)} > 0$ for all a, d_{11}, d_{22}, and d_{12} satisfying

   $\max\left\{0,\ \frac{a}{6}(1+3d_{11} - d_{22})\right\} < d_{12} < \sqrt{d_{11}d_{22}}$
   and $0 < 6(d_{11}d_{22} - d_{12}^2) + 2a^2(d_{11}d_{22} - d_{12}^2) + (1+3d_{11} + 2d_{22} + a^2 d_{22}) < -2ad_{12}$,

3. $D_f^{(1)} < 0$ for all a, d_{11}, d_{22}, and d_{12} satisfying

$$(i)\ \max\left\{-\sqrt{d_{11}d_{22}},\ \frac{a}{6}(1+3d_{11}-d_{22})\right\} < d_{12} \le 0,$$

$$(ii)\ \max\left\{0,\ \frac{a}{6}(1+3d_{11}-d_{22})\right\} < d_{12} < \sqrt{d_{11}d_{22}} \text{ and}$$

$$6(d_{11}d_{22}-d_{12}^2)+2a^2(d_{11}d_{22}-d_{12}^2)+(1+3d_{11}+2d_{22}+a^2d_{22}) > -2ad_{12} > 0.$$

Proof.

1. Suppose a, d_{11}, d_{22}, and d_{12} satisfying

$$-\sqrt{d_{11}d_{22}} < d_{12} \le \min\left\{\sqrt{d_{11}d_{22}},\ \frac{a}{6}(1+3d_{11}-d_{22})\right\}.$$

Then, $d_{12} \le \frac{a}{6}(1+3d_{11}-d_{22})$ which implies $c^{(1)} \ge 0$ so, by part 1 of Theorem 3.1,

$D_f^{(1)} \ge 0$ and the result follows.

2. Suppose a, d_{11}, d_{22}, and d_{12} satisfying $\max\left\{0,\ \frac{a}{6}(1+3d_{11}-d_{22})\right\} < d_{12} < \sqrt{d_{11}d_{22}}$,

and $0 < 6(d_{11}d_{22}-d_{12}^2)+2a^2(d_{11}d_{22}-d_{12}^2)+(1+3d_{11}+2d_{22}+a^2d_{22}) < -2ad_{12}$.

Since $\frac{a}{6}(1+3d_{11}-d_{22}) < d_{12}$, we have $c^{(1)} < 0$. From (3.11),

$$6(d_{11}d_{22}-d_{12}^2)+2a^2(d_{11}d_{22}-d_{12}^2)+(1+3d_{11}+2d_{22}+a^2d_{22}) < -2ad_{12},$$

which implies that $c^{(3)} < 0$. Thus, by part 2 of Theorem 3.1, $D_f^{(1)} > 0$

and the result follows.

3. (i) Suppose a, d_{11}, d_{22}, and $d_{12} \le 0$ satisfying

$$\max\left\{-\sqrt{d_{11}d_{22}},\ \frac{a}{6}(1+3d_{11}-d_{22})\right\} < d_{12} \le 0.$$

Then $\frac{a}{6}(1+3d_{11}-d_{22}) < d_{12}$, so that $c^{(1)} < 0$.

Hence, by part 3(i) of Theorem 3.1, we have $D_f^{(1)} < 0$ and the result follows.

(ii) Suppose a, d_{11}, d_{22}, and d_{12} satisfying

$$\max\left\{0,\ \frac{a}{6}(1+3d_{11}-d_{22})\right\} < d_{12} < \sqrt{d_{11}d_{22}} \text{ and}$$

$$6(d_{11}d_{22}-d_{12}^2)+2a^2(d_{11}d_{22}-d_{12}^2)+(1+3d_{11}+2d_{22}+a^2d_{22}) > -2ad_{12} > 0.$$

Since $\frac{a}{6}(1+3d_{11}-d_{22}) < d_{12}$, we have $c^{(1)} < 0$.

From (3.11), $6(d_{11}d_{22}-d_{12}^2)+2a^2(d_{11}d_{22}-d_{12}^2)+(1+3d_{11}+2d_{22}+a^2d_{22}) > -2ad_{12}$ implies that $c^{(3)} > 0$. Hence, by part 3(ii) of Theorem 3.1, $D_f^{(1)} < 0$ and the result follows.

3.3.3 The Situation with $0 < a < 1$

Now we are considering the nine cases shown in Table 3.3 for $0 < a < 1$. We present the signs of $c^{(0)}$, $c^{(1)}$, $c^{(2)}$, $c^{(3)}$, and $D_f^{(1)}$ in Table 3.5. It can be seen that $D_f^{(1)}$ is always positive for case II, V, and VIII.

Table 3.5: Analysis of $D_f^{(1)}$ for $0 < a < 1$

Case	$c^{(0)}$	$c^{(1)}$	$c^{(2)}$	$c^{(3)}$	$D_f^{(1)}$
I	< 0	> 0	> 0	> 0	< 0
II	< 0	< 0	> 0	> 0	> 0
III	< 0	$= 0$	> 0	> 0	$= 0$
IV	< 0	> 0	> 0	> 0, if $6(d_{11}d_{22} - d_{12}^2) + 2a^2(d_{11}d_{22} - d_{12}^2) +$ $(1 + 3d_{11} + 2d_{22} + a^2 d_{22}) > -2ad_{12} > 0$ < 0, if $0 < 6(d_{11}d_{22} - d_{12}^2) + 2a^2(d_{11}d_{22} - d_{12}^2)$ $+(1 + 3d_{11} + 2d_{22} + a^2 d_{22}) < -2ad_{12}$	< 0 > 0
V	< 0	< 0	> 0	> 0	> 0
VI	< 0	$= 0$	> 0	> 0	$= 0$
VII	< 0	> 0	> 0	> 0	< 0
VIII	< 0	< 0	> 0	> 0	> 0
IX	< 0	$= 0$	> 0	> 0	$= 0$

Theorem 3.4

When $0 < a < 1$, we have

1. $D_f^{(1)} \geq 0$ for all a, d_{11}, d_{22}, and d_{12} satisfying $c^{(1)} \leq 0$,

2. $D_f^{(1)} > 0$ for all a, d_{11}, d_{22}, and $d_{12} < 0$ satisfying $c^{(1)} > 0$ and $c^{(3)} < 0$,

3. $D_f^{(1)} < 0$ for all a, d_{11}, d_{22}, and (*i*) $d_{12} \geq 0$ or (*ii*) $d_{12} < 0$ and $c^{(3)} > 0$ satisfying $c^{(1)} > 0$.

Proof. Similar to proofs of Theorem 3.1.

Theorem 3.5

When $0 < a < 1$,

1. The condition that all a, d_{11}, d_{22}, and d_{12} satisfying $c^{(1)} \le 0$ implies that d_{12} must satisfy

$$\max\left\{-\sqrt{d_{11}d_{22}},\ \frac{a}{6}(1+3d_{11}-d_{22})\right\} \le d_{12} < \sqrt{d_{11}d_{22}} \text{ for all } a,\ d_{11}, \text{ and } d_{22}.$$

2. The condition that all a, d_{11}, d_{22}, and $d_{12} < 0$ satisfying $c^{(1)} > 0$ and $c^{(3)} < 0$ implies

$$-\sqrt{d_{11}d_{22}} < d_{12} < \min\left\{\sqrt{d_{11}d_{22}},\ \frac{a}{6}(1+3d_{11}-d_{22})\right\} \text{ and}$$
$$0 < 6(d_{11}d_{22}-d_{12}^2)+2a^2(d_{11}d_{22}-d_{12}^2)+(1+3d_{11}+2d_{22}+a^2d_{22}) < -2ad_{12}$$

for all a, d_{11}, and d_{22}.

3. (i) The condition that all a, d_{11}, d_{22}, and $d_{12} \ge 0$ satisfying $c^{(1)} > 0$ implies that d_{12} must satisfy

$$0 \le d_{12} < \min\left\{\sqrt{d_{11}d_{22}},\ \frac{a}{6}(1+3d_{11}-d_{22})\right\} \text{ for all } a,\ d_{11}, \text{ and } d_{22}.$$

(ii) The condition that all a, d_{11}, d_{22}, and $d_{12} < 0$ satisfying $c^{(1)} > 0$ and $c^{(3)} > 0$

implies that

$$-\sqrt{d_{11}d_{22}} < d_{12} < \min\left\{0,\ \frac{a}{6}(1+3d_{11}-d_{22})\right\} \text{ and}$$
$$6(d_{11}d_{22}-d_{12}^2)+2a^2(d_{11}d_{22}-d_{12}^2)+(1+3d_{11}+2d_{22}+a^2d_{22}) > -2ad_{12} > 0$$

for all a, d_{11}, and d_{22}.

Proof. Similar to proofs of Theorem 3.2.

Theorem 3.6

When $0 < a < 1$, we have

1. $\mathrm{D}_f^{(1)} \geq 0$ for all a, d_{11}, d_{22}, and d_{12} satisfying

$$\max\left\{-\sqrt{d_{11}d_{22}},\ \frac{a}{6}(1+3d_{11}-d_{22})\right\} \leq d_{12} < \sqrt{d_{11}d_{22}},$$

2. $\mathrm{D}_f^{(1)} > 0$ for all a, d_{11}, d_{22}, and d_{12} satisfying

$$-\sqrt{d_{11}d_{22}} < d_{12} < \min\left\{\sqrt{d_{11}d_{22}},\ \frac{a}{6}(1+3d_{11}-d_{22})\right\}$$

and $0 < 6(d_{11}d_{22}-d_{12}^2)+2a^2(d_{11}d_{22}-d_{12}^2)+(1+3d_{11}+2d_{22}+a^2d_{22}) < -2ad_{12}$,

3. $\mathrm{D}_f^{(1)} < 0$ for all a, d_{11}, d_{22}, and d_{12} satisfying

$$(i)\ \ 0 \leq d_{12} < \min\left\{\sqrt{d_{11}d_{22}},\ \frac{a}{6}(1+3d_{11}-d_{22})\right\}$$

$$(ii)\ \ -\sqrt{d_{11}d_{22}} < d_{12} < \min\left\{0,\ \frac{a}{6}(1+3d_{11}-d_{22})\right\} \text{ and}$$

$$6(d_{11}d_{22}-d_{12}^2)+2a^2(d_{11}d_{22}-d_{12}^2)+(1+3d_{11}+2d_{22}+a^2d_{22}) > -2ad_{12} > 0.$$

Proof. Similar to proofs of Theorem 3.3.

Theorem 3.1 to 3.6 state the general results of comparing $Q_1(0)$ with $Q_1(a)$ for $a \neq 0$ and $-1 < a < 1$. For d_{11}, d_{12}, and d_{22} satisfying certain conditions so that $\mathrm{D}_f^{(1)} > 0$, we find that $Q_1(a)$ is maximized at $a = 0$ and the design with equidistant time points is D-optimal.

3.3.4 *D*-optimality Region

To demonstrate how results from the previous two sections can be used, we apply Theorem 3.1 and 3.4 for determining the optimality region within which a = 0 is the *D*-optimal design. Table 3.6 presents such region with conditions on a, d_{11}, d_{12}, and d_{22}.

Table 3.6: *D*-optimality region

a	d_{11}	d_{12}	d_{22}
$(-1, 1)$	$d_{11} > 0$	$\left(-\sqrt{d_{11}d_{22}}, \sqrt{d_{11}d_{22}}\right)$	$d_{22} > \max\left\{0, 1+3d_{11} - \frac{6d_{12}}{a}\right\}$

From Table 3.6, we know that the design with $a = 0$ is *D*-optimal for all $-1 < a < 1$, $d_{11} > 0$, $-\sqrt{d_{11}d_{22}} < d_{12} < \sqrt{d_{11}d_{22}}$, and $d_{22} > \max\left\{0, 1+3d_{11} - \frac{6d_{12}}{a}\right\}$.

We can examine the *D*-optimality regions using 3-D plots. However, one parameter needs to be fixed first. For illustration, we choose an arbitrary value between −1 and 1 for a, i.e. $a = -0.5$. Then, for d_{11} and d_{12} bounded by $0 < d_{11} < 8$ and $-2 < d_{12} < 2$, respectively, the region within which the design with $a = 0$ is *D*-optimal is the area such that $d_{22} > \max\left\{\frac{d_{12}^2}{d_{11}}, 1+3d_{11} + \frac{6d_{12}}{0.5}\right\}$. Figure 3.3 demonstrates the region of $d_{22} > d_{12}^2/d_{11}$ by the area above the surface. Furthermore, the area above the flat surface in Figure 3.4 illustrates the region of $d_{22} > 1+3d_{11} + \frac{6d_{12}}{0.5}$. The intersection of the preceding two regions is the *D*-optimality region where choosing $a = 0$ is *D*-optimal and such region is shown in

Figure 3.5. Note that in the following three 3-D plots, X-axis is d_{11} ranged from 0 to 8, Y-axis is d_{12} ranged from −2 to 2, and Z-axis is d_{22}.

Figure 3.3: Plot of $d_{22} > d_{12}^2/d_{11}$ for $0 < d_{11} < 8$ and $-2 < d_{12} < 2$

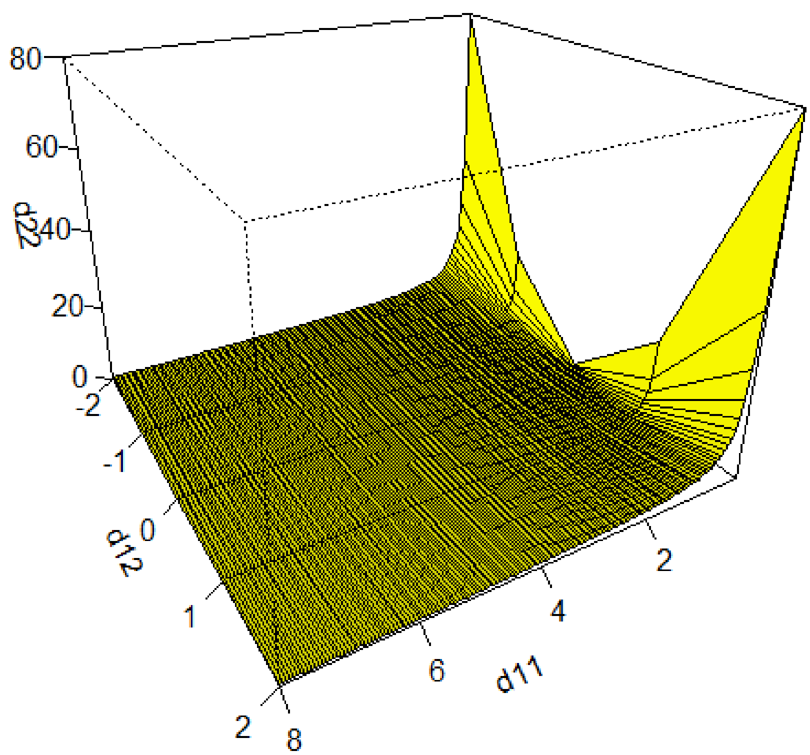
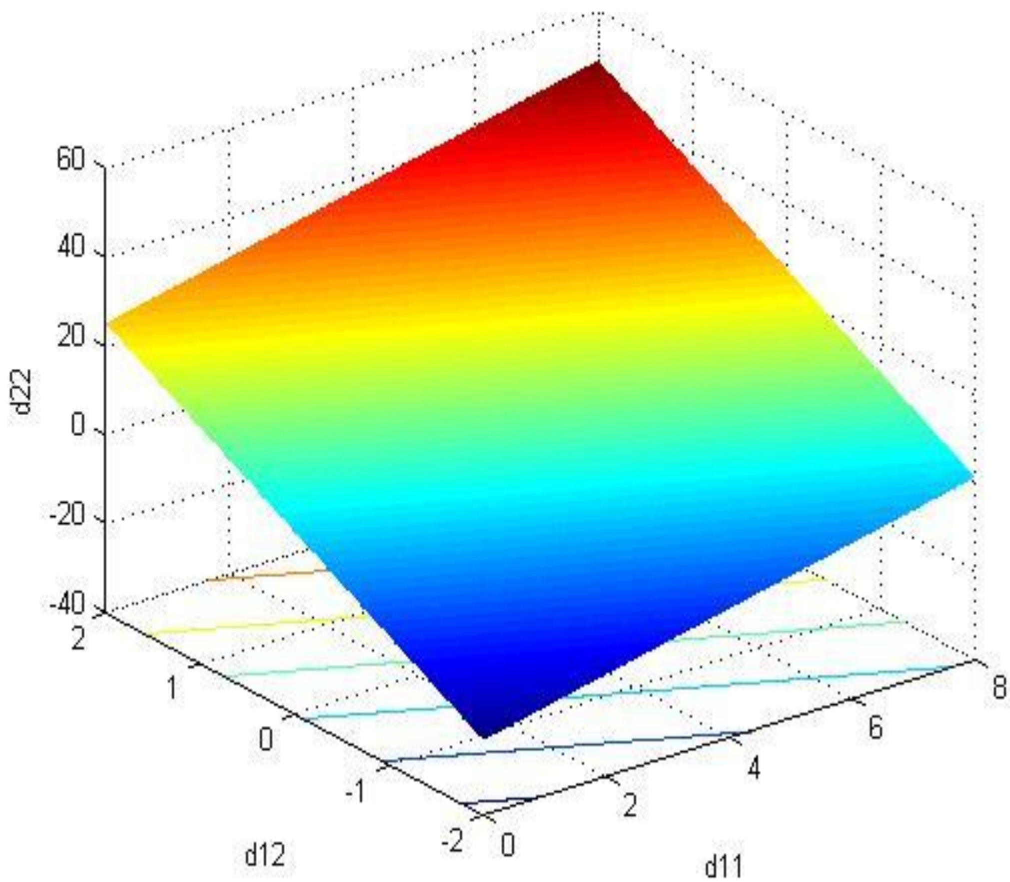


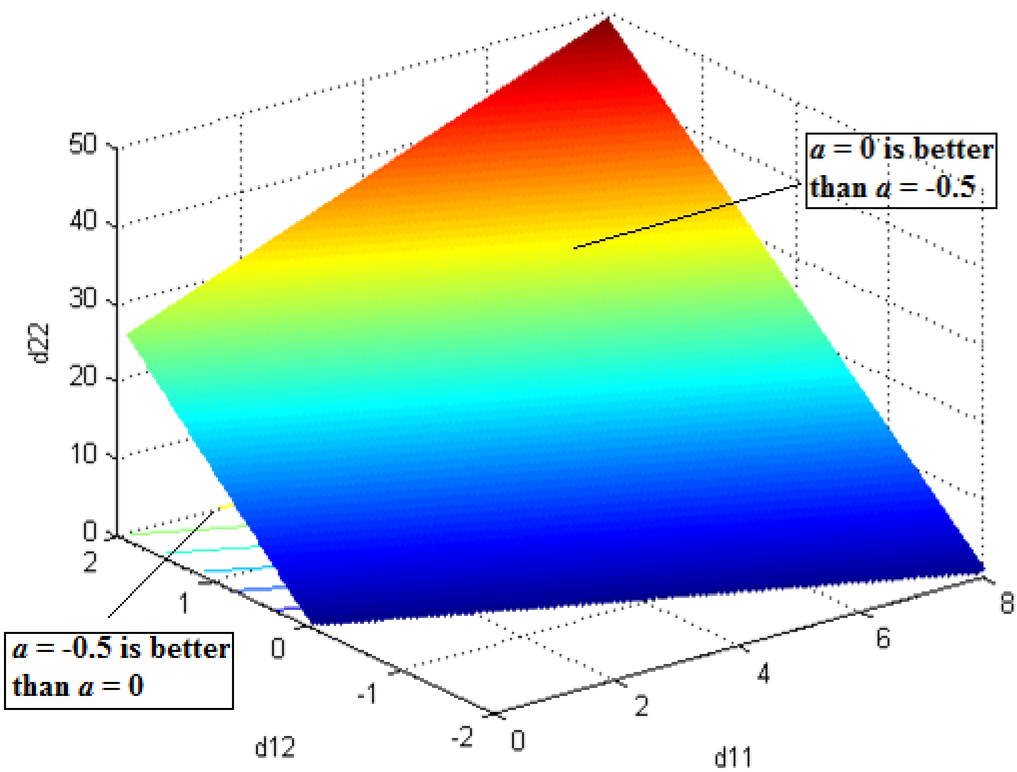
Figure 3.4: Plot of $d_{22} > 1 + 3d_{11} + \frac{6d_{12}}{0.5}$

for $0 < d_{11} < 8$ and $-2 < d_{12} < 2$



Note that the bottom part of the graph shows the contour line.

Figure 3.5: Comparison between cohort designs with $a = 0$ and $a = -0.5$

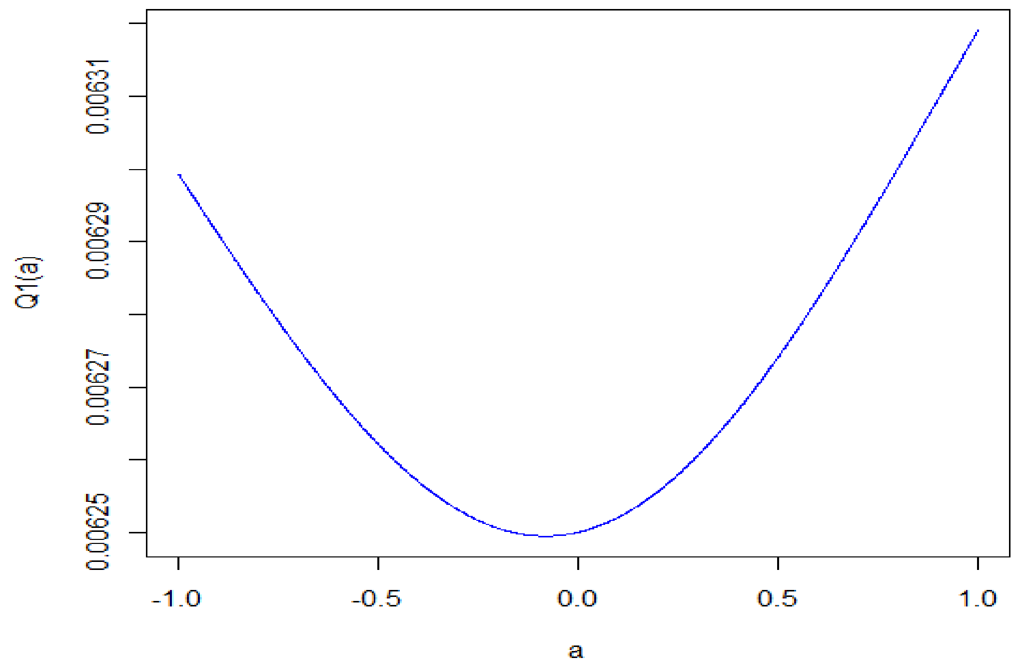


In Figure 3.5, all points $\{d_{11}, d_{12}, d_{22}\}$ satisfying $\left\{ d_{22} > 1 + 3d_{11} + \frac{6d_{12}}{0.5} \text{ and } d_{22} > \frac{d_{12}^2}{d_{11}} \right\}$ are in the region above the flat quadrangle surface. For the points within such region, design with $a = 0$ is *D*-optimal. Hence, the design with $a = 0$ is better over the design with $a = -0.5$. For all points on the flat surface, the designs with $a = 0$ and $a = -0.5$ are comparable. Moreover, for points beneath the surface, the design with $a = -0.5$ is better than the design with $a = 0$. Therefore, picking up any point above the flat surface generates a matrix **D**= $\begin{bmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{bmatrix}$. For this given **D**, the cohort design with equidistant time points, i.e. $a = 0$, is more efficient than the design with non-equidistant time points, i.e. $a = -0.5$. For example, for the point located above the surface with $d_{11} = 2$, $d_{12} = 1$, and $d_{22} = 20$, the design with $a = 0$ is better according to the *D*-optimality criterion.

3.3.5 *D*-optimal designs with $a \neq 0$

As we have discussed earlier, in longitudinal studies, designs with non-equidistant time points are more efficient in certain cases. The design with $a \neq 0$ is *D*-optimal when d_{11}, d_{12}, and d_{22} are under certain conditions so that $D_f^{(1)} < 0$. For instance, by part 3(*i*) of Theorem 3.3, $D_f^{(1)} < 0$ for d_{11}, d_{12}, and d_{22} satisfying $\max\left\{-\sqrt{d_{11}d_{22}},\ \frac{a}{6}(1+3d_{11}-d_{22})\right\} <$ $d_{12} \leq 0,$ where $-1 < a < 0$. Suppose we choose a matrix **D** satisfying such condition, i.e. **D** = $\begin{bmatrix} 15 & -1 \\ -1 & 10 \end{bmatrix}$. By Theorem 3.1, we know given this matrix **D**, choosing $a = 0$ is not *D*-optimal. To find the *D*-optimal design $\tau_D \in \mathrm{T}_{23(a)}$, we should maximize $Q_1(a)$ in (3.4). The numerical values of $Q_1(a)$ are calculated and plotted against the time point a ($-1 < a < 1$) in Figure 3.6. It can be observed that $Q_1(a)$ moves toward its maximum when a approaches to 1. If researchers can put a threshold on a ($-1 < a < 1$) such that $a \leq 0.8$ (Note: here 0.8 is chosen arbitrarily), then the design with $a = 0.8$ is *D*-optimal. Consequently, the *D*-optimal design is $\tau_D = \begin{Bmatrix} (-1, 0.8, 1) & (-1, 0.8, 1) \\ 0.5 & 0.5 \end{Bmatrix}$.

Figure 3.6: Plot of $Q_1(a)$ given $d_{11} = 15$, $d_{12} = -1$, and $d_{22} = 10$



3.4 *A*-optimal Designs

Now we are using *A*-optimality criterion to determine whether or not the 'best' allocation of repeated measurements is achieved by choosing equally spaced time points, i.e. (−1, 0, 1). Recall that the design $\tau_A \in \mathrm{T}_{23(a)}$ with the minimum value of $Q_2(a)$ is considered *A*-optimal. The exact expression of $Q_2(a)$ is given by

$$Q_2(a) = \frac{\begin{pmatrix} 6d_{11} + 6d_{22} + 2a^2 d_{11} \\ +2a^2 d_{22} + a^2 + 5 \end{pmatrix} \begin{pmatrix} 3d_{11} + 2d_{22} + a^2 d_{22} + 2ad_{12} + 1 \\ +(6 + 2a^2)(d_{11}d_{22} - d_{12}^2) \end{pmatrix}}{\begin{pmatrix} (6a^2 + 18)d_{11} + (2a^4 + 10a^2 + 12)d_{22} + 2a^2 + 6 \\ +(4a^4 + 24a^2 + 36)(d_{11}d_{22} - d_{12}^2) + (4a^3 + 12a)d_{12} \end{pmatrix}} = \frac{\mathrm{Num}_2(a)}{\mathrm{Den}_2(a)}. \quad (3.13)$$

It should be noted that $\text{Den}_2(a)$ is exactly the same as $\text{Num}_1(a)$ defined in (3.9).

3.4.1 Comparison of $Q_2(0)$ with $Q_2(a)$

For $a \neq 0$ and $-1 < a < 1$, the comparison of $Q_2(a)$ with $Q_2(0)$ has the ability to discriminate between the design with non-equidistant time points (-1, a, 1) and the design with equidistant time points (-1, 0, 1), with respect to A-optimality criterion. To check if designs with $a = 0$ are better than the designs with $a \neq 0$, we consider

$$Q_2(a) - Q_2(0) = \frac{\text{Num}_2(a)\text{Den}_2(0) - \text{Num}_2(0)\text{Den}_2(a)}{\text{Den}_2(a)\text{Den}_2(0)}, \qquad (3.14)$$

where

$$\begin{aligned}
&\text{Num}_2(a)\text{Den}_2(0) - \text{Num}_2(0)\text{Den}_2(a) \\
&\quad = -4a^2\left(3d_{11} + 2d_{22} + 2a^2 d_{11}d_{22} + a^2 d_{22} + 6(d_{11}d_{22} - d_{12}^2) + 1 + 2ad_{12} - 2a^2 d_{12}^2\right) \\
&\qquad \left(3d_{11} + 2d_{22} + 6(d_{11}d_{22} - d_{12}^2) + 1\right), \\
&\text{Den}_2(a)\text{Den}_2(0) = 12\left(3d_{11} + 2d_{22} + 2a^2 d_{11}d_{22} + a^2 d_{22} + 6(d_{11}d_{22} - d_{12}^2) + 1 + 2ad_{12} - 2a^2 d_{12}^2\right) \\
&\qquad (a^2+3)\left(3d_{11} + 2d_{22} + 6(d_{11}d_{22} - d_{12}^2) + 1\right).
\end{aligned}$$

Consequently, $Q_2(a) - Q_2(0)$ can be simplified to

$$Q_2(a) - Q_2(0) = \frac{1}{a^2+3} - \frac{1}{3}, \qquad (3.15)$$

which is independent of d_{11}, d_{22}, and d_{12}. Notice that (1) when $a = 0$, $Q_2(0) = Q_2(a)$ and obviously $Q_2(0) - Q_2(a) = 0$; (2) when $a \neq 0$ and $-1 < a < 1$, $Q_2(a) - Q_2(0) < 0$, which implies that design with $a = 0$ is not an A-optimal design for all d_{11}, d_{22}, and d_{12}.

Since choosing $a = 0$ is not the best allocation of time point in regard to A-optimality, now we consider two time points that are different from zero: a_1 and a_2 such that $-1 < a_1, a_2 < 1$. We have

$$Q_2(a_1) - Q_2(a_2) = \left(Q_2(a_1) - Q_2(0)\right) - \left(Q_2(a_2) - Q_2(0)\right)$$
$$= \frac{1}{a_1^2+3} - \frac{1}{a_2^2+3} = \frac{a_2^2 - a_1^2}{(a_1^2+3)(a_2^2+3)}. \qquad (3.16)$$

Theorem 3.7

We have

1. $Q_2(a_1) \le Q_2(a_2)$ when $a_1^2 \ge a_2^2$,
2. $Q_2(a_1) \ge Q_2(a_2)$ when $a_1^2 \le a_2^2$,

   where the equality holds if and only if $a_1 = a_2$.

Proof. Clear from (3.16).

By Theorem 3.7, we can draw the following conclusions: (1) the design with $a \neq 0$ is better than the design with $a = 0$ with respect to A-optimality; (2) for $a \neq 0$ and $-1 < a < 1$, the design $\tau_1 = \begin{Bmatrix} (-1, a, 1) & (-1, a, 1) \\ 0.5 & 0.5 \end{Bmatrix}$ and the design $\tau_2 = \begin{Bmatrix} (-1, -a, 1) & (-1, -a, 1) \\ 0.5 & 0.5 \end{Bmatrix}$ are indistinguishable under A-optimality; (3) If we could put a threshold on a such that $a^2 \le a_0$, where a_0 is an arbitrary value between 0 and 1 (Note: choosing a^2 very close to 1 should be avoided, as design class $T_{23(a)}$ involves three distinct time points), then the designs $\tau_A = \begin{Bmatrix} (-1, -\sqrt{a_0}, 1) & (-1, -\sqrt{a_0}, 1) \\ 0.5 & 0.5 \end{Bmatrix}$ and $\tau_A = \begin{Bmatrix} (-1, \sqrt{a_0}, 1) & (-1, \sqrt{a_0}, 1) \\ 0.5 & 0.5 \end{Bmatrix}$ are A-optimal for all d_{11}, d_{22}, and d_{12}.

Chapter 4

Comparison of Two Design Classes: $T_{23(a)}$ and $T_{23(-a,\, a)}$

4.1 Introduction

In chapter 3, we consider a special situation $T_{23(a)}$ for the design class T_{23} where both cohorts share the same time points $(-1, a, 1)$. In practice, for a longitudinal study in two cohorts with three repeated measurements, the time points may vary from cohort to cohort. In this chapter, we consider a design class $T_{23(-a,\, a)}$ within T_{23} by assuming time points $(-1, -a, 1)$ for the 1st cohort and time points $(-1, a, 1)$ for the 2nd cohort $(-1 < a < 1)$. Since longitudinal data are often correlated, AR(1) structure for the error correlation matrix $\boldsymbol{R}_i$ $(i = 1,\ 2)$ is considered. We present some general results with their applications in comparison of design classes $T_{23(a)}$ and $T_{23(-a,\, a)}$, with respect to D- and A-optimality criteria.

4.2 Design Class $T_{23(-a,\, a)}$

Consider cohort designs with m = 3 repeated measurements obtained from C = 2 cohorts with n subjects each. We define the class of cohort designs $T_{23(-a,a)}$ as

$$T_{23(-a,a)} = \left\{ \begin{matrix} (-1,-a,1) & (-1,a,1) \\ 0.5 & 0.5 \end{matrix} \middle| -1 < a < 1 \right\}, \tag{4.1}$$

and a cohort design $\tau = \begin{pmatrix} (-1,-a,1) & (-1,a,1) \\ 0.5 & 0.5 \end{pmatrix}$ for some specified values of a satisfying $-1 < a < 1,$ where 0.5 on the second row is the relative sample size indicating that the two cohorts are equally sized. It should be noted that the time point '$-a$' from the 1st cohort and the time point 'a' from the 2nd cohort are both located at the same distance of $|a|$ from the middle point '0', but in different directions.

4.3 Auto-Correlatoin Structure

For longitudinal data, each subject is observed on multiple occasions over time. The repeated measurements taken on the same subject are typically correlated. Because correlations among repeated measurements tend to decrease when measurements are taken further apart in time, one often assumes that the error correlation matrix $\boldsymbol{R}_i$ follows a first-order autoregressive correlation structure, i.e. AR(1). This structure has been used in many longitudinal studies, for example, Tekle, Tan, and Berger (2008), Winkens, Schouten, Breukelen, and Berger (2005). Other types of error correlation matrix and their uses were studies by Verbeke and Molenberghs (2000).

The auto-correlation between errors from two time points $t_{j(i)}$ and $t_{j'(i)}$ $(j, j' = 1,\ldots,n, j \neq j'$, $i = 1,2)$ in cohort i has the form $\rho^{|t_{j(i)}-t_{j'(i)}|}$, where ρ is a correlation parameter ($0 < \rho < 1$) and $|t_{j(i)} - t_{j'(i)}|$ refers to the absolute time separation between measurements j and j' made on subjects from the i^{th} cohort. Consequently, the correlations between the errors depend on how far apart they are in time. The correlation between the errors decreases as the time points lie farther apart. For the time points $t_i' = (t_{1(i)}, t_{2(i)}, t_{3(i)}),\ i = 1,\ 2,\ \boldsymbol{R}_i$ with AR(1) structure is given by

$$\boldsymbol{R}_i = \begin{bmatrix} 1 & \rho^{|t_{1(i)}-t_{2(i)}|} & \rho^{|t_{1(i)}-t_{3(i)}|} \\ \rho^{|t_{1(i)}-t_{2(i)}|} & 1 & \rho^{|t_{2(i)}-t_{3(i)}|} \\ \rho^{|t_{1(i)}-t_{3(i)}|} & \rho^{|t_{2(i)}-t_{3(i)}|} & 1 \end{bmatrix}. \qquad (4.2)$$

4.4 Comparison of $\mathrm{T}_{23(a)}$ and $\mathrm{T}_{23(-a,\, a)}$

Design class $\mathrm{T}_{23(a)}$: $\left\{ \begin{matrix} (-1, a, 1) & (-1, a, 1) \\ 0.5 & 0.5 \end{matrix} \middle| -1 < a < 1 \right\}.$

Design class $\mathrm{T}_{23(-a,a)}$: $\left\{ \begin{matrix} (-1, -a, 1) & (-1, a, 1) \\ 0.5 & 0.5 \end{matrix} \middle| -1 < a < 1 \right\}.$

We now compare the above two design classes under the linear mixed effects model stated in (2.3) with AR(1) correlated errors. The *D*- and *A*-optimality criteria are used in finding a better design class that gives precise estimation of the model parameters. As previously mentioned in Chapter 2, the variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ is

$$n\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \left[\boldsymbol{X}_1'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{X}_1 + \boldsymbol{X}_2'\boldsymbol{\Sigma}_2^{-1}\boldsymbol{X}_2 \right]^{-1}, \qquad (4.3)$$

where the variance-covariance matrix of $\boldsymbol{y}_{j(i)}$, $\boldsymbol{\Sigma}_i$ (i = 1, 2), is

$$\boldsymbol{\Sigma}_i = \sigma^2(\boldsymbol{Z}_i\mathbf{D}\boldsymbol{Z}_i' + \boldsymbol{R}_i). \qquad (4.4)$$

We note that $n\text{Var}(\hat{\boldsymbol{\beta}})$ in (4.3) depends on five model parameters: d_{11}, d_{12}, d_{22}, σ^2, and ρ and one design parameter a. We assume the following range of those parameters in the current chapter.

Table 4.1: The range of parameters

a	d_{11}	d_{22}	d_{12}	ρ	σ^2
(−1, 1)	> 0	> 0	0	(0, 1)	> 0

The matrices $\boldsymbol{X}_i$, $\boldsymbol{R}_i$, and $\boldsymbol{\Sigma}_i$ (i = 1, 2) play an important role in our analysis. Therefore, we list five important properties of these matrices below.

For design class $\text{T}_{23(-a,a)}$ under the model (2.3), we have

Property 1: $$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \boldsymbol{X}_1 \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \boldsymbol{X}_2, \qquad (4.5)$$

Property 2: $$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \boldsymbol{R}_2 \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = \boldsymbol{R}_1, \qquad (4.6)$$

Property 3: $$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \boldsymbol{\Sigma}_2^{-1} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = \left(\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \boldsymbol{\Sigma}_2 \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \right)^{-1}, \qquad (4.7)$$

Property 4: If $\mathbf{D} = \begin{bmatrix} d_{11} & 0 \\ 0 & d_{22} \end{bmatrix}$, then $$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \boldsymbol{\Sigma}_2 \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = \boldsymbol{\Sigma}_1, \qquad (4.8)$$

Property 5: If $\mathbf{D}=\begin{bmatrix} d_{11} & 0 \\ 0 & d_{22} \end{bmatrix}$, then $\boldsymbol{X}_2'\boldsymbol{\Sigma}_2^{-1}\boldsymbol{X}_2=\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\boldsymbol{X}_1'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{X}_1\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. (4.9)

4.4.1 *D*-optimality criterion

Between $T_{23(a)}$ and $T_{23(-a,\, a)}$, we are interested in identifying a design class with a smaller generalized variance of the estimators of $\boldsymbol{\beta}=(\beta_0,\beta_1)'$, $\det\{\mathrm{Var}(\hat{\boldsymbol{\beta}})\}$.

Theorem 4.1. If $d_{12}=0$, then $\det\left\{\mathrm{Var}(\hat{\boldsymbol{\beta}})_{T_{23(-a,a)}}\right\}\le\det\left\{\mathrm{Var}(\hat{\boldsymbol{\beta}})_{T_{23(a)}}\right\}$ for all d_{11}, d_{22}, σ^2, ρ and a.

Proof. For design class $T_{23(-a,a)}$, we denote

$$\boldsymbol{X}_1'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{X}_1=\begin{bmatrix} w_{11} & w_{12} \\ w_{12} & w_{22} \end{bmatrix}. \qquad (4.10)$$

By Property 5 in (4.9) we get

$$\boldsymbol{X}_2'\boldsymbol{\Sigma}_2^{-1}\boldsymbol{X}_2=\begin{bmatrix} w_{11} & -w_{12} \\ -w_{12} & w_{22} \end{bmatrix},$$

$$\left[n\mathrm{Var}(\hat{\boldsymbol{\beta}})_{T_{23(-a,a)}}\right]^{-1}=2\begin{bmatrix} w_{11} & 0 \\ 0 & w_{22} \end{bmatrix},$$

$$\det\left\{\left[2n\mathrm{Var}(\hat{\boldsymbol{\beta}})\right]_{T_{23(-a,a)}}^{-1}\right\}=w_{11}w_{22}. \qquad (4.11)$$

For design class $T_{23(a)}$, we have from (4.3) that

$$\boldsymbol{X}_1'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{X}_1=\boldsymbol{X}_2'\boldsymbol{\Sigma}_2^{-1}\boldsymbol{X}_2=\begin{bmatrix} w_{11} & -w_{12} \\ -w_{12} & w_{22} \end{bmatrix},$$

$$\left[n\mathrm{Var}(\hat{\boldsymbol{\beta}})_{\mathrm{T}_{23(a)}}\right]^{-1} = 2\begin{bmatrix} w_{11} & -w_{12} \\ -w_{12} & w_{22} \end{bmatrix},$$

$$\det\left\{\left[2n\mathrm{Var}(\hat{\boldsymbol{\beta}})\right]^{-1}_{\mathrm{T}_{23(a)}}\right\} = w_{11}w_{22} - w_{12}^2. \qquad (4.12)$$

From (4.11) and (4.12), we obtain

$$\det\left\{2n\mathrm{Var}(\hat{\boldsymbol{\beta}})_{\mathrm{T}_{23(-a,a)}}\right\} \le \det\left\{2n\mathrm{Var}(\hat{\boldsymbol{\beta}})_{\mathrm{T}_{23(a)}}\right\}.$$

The equality holds if and only if $w_{12} = 0$ and the rest is clear.

We note that $\mathrm{T}_{23(-a,a)}$ and $\mathrm{T}_{23(a)}$ are indistinguishable with respect to *D*-optimality criterion when $d_{12} = 0$ and $w_{12} = 0$. So we investigate when w_{12} is zero. We consider first a special situation where $a = 0$.

4.4.2 The Situation when *a* = 0

If we assume the additional time point is placed in the middle of the study period [−1, 1], i.e. $a = 0$, for both design classes $\mathrm{T}_{23(-a,a)}$ and $\mathrm{T}_{23(a)}$, then obviously

$$\mathrm{T}_{23(-a,a)} = \mathrm{T}_{23(a)} = \begin{Bmatrix} (-1,0,1) & (-1,0,1) \\ 0.5 & 0.5 \end{Bmatrix}. \quad (\text{Given } a = 0)$$

As a result, $\mathrm{T}_{23(a)}$ and $\mathrm{T}_{23(-a,a)}$ become two identical designs and they have the same *D*-efficiency. In other words, we have $\det\left\{\mathrm{Var}(\hat{\boldsymbol{\beta}})_{\mathrm{T}_{23(-a,a)}}\right\} = \det\left\{\mathrm{Var}(\hat{\boldsymbol{\beta}})_{\mathrm{T}_{23(a)}}\right\}$ when $a = 0$. The relationship between a, d_{12}, and w_{12} is presented in the following theorem.

Theorem 4.2. For design classes $\mathrm{T}_{23(a)}$ and $\mathrm{T}_{23(-a,\, a)}$, if $a = d_{12} = 0$, then $w_{12} = 0$.

Proof. If $d_{12} = 0$, $\mathbf{D}$ is a diagonal matrix such that $\mathbf{D} = \begin{bmatrix} d_{11} & 0 \\ 0 & d_{22} \end{bmatrix}$. When $a = 0$, we have

$$w_{12} = \begin{pmatrix} -1 & 0 & 1 \end{pmatrix} \mathbf{\Sigma}_1^{-1} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \qquad (4.12)$$

and $$\mathbf{\Sigma}_1 \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = \left[2d_{22} + (1-\rho^2) \right] \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}. \qquad (4.13)$$

From (4.12), we get $\begin{pmatrix} -1 & 0 & 1 \end{pmatrix} \mathbf{\Sigma}_1^{-1} = \begin{pmatrix} -1 & 0 & 1 \end{pmatrix} \left[2d_{22} + (1-\rho^2) \right]^{-1}$.

From (4.13), we have $$w_{12} = \begin{pmatrix} -1 & 0 & 1 \end{pmatrix} \mathbf{\Sigma}_1^{-1} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 & 0 & 1 \end{pmatrix} \left[2d_{22} + (1-\rho^2) \right]^{-1} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 0$$

and the result follows.

Notice that when $w_{12} = 0$, the matrices $\text{Var}(\hat{\boldsymbol{\beta}})$ for design classes $T_{23(-a,a)}$ and $T_{23(a)}$ are diagonal. Based on Theorem 4.2, we conclude that if both cohorts have equidistant time points $(-1, 0, 1)$ and the random interpret $b_{0j(i)}$ and the random slope $b_{1j(i)}$ are uncorrelated, then the estimated parameters of β_0 and β_1 are also uncorrelated.

4.4.3 *A*-optimality criterion

In this section, *A*-optimality criterion is applied in finding a better design class out of $T_{23(-a,a)}$ and $T_{23(a)}$. Recall that *A*-optimality minimizes sum of the diagonal elements of $\text{Var}(\hat{\boldsymbol{\beta}})$. We obtain the results below.

Theorem 4.3. If d_{12} = 0, then $\operatorname{tr}\left\{\operatorname{Var}(\hat{\boldsymbol{\beta}})_{\mathrm{T}_{23(-a,a)}}\right\} \le \operatorname{tr}\left\{\operatorname{Var}(\hat{\boldsymbol{\beta}})_{\mathrm{T}_{23(a)}}\right\}$ for all d_{11}, d_{22}, σ^2, ρ and a.

Proof. For design class $\mathrm{T}_{23(-a,a)}$, from (4.11) we have

$$2n\operatorname{Var}(\hat{\boldsymbol{\beta}})_{\mathrm{T}_{23(-a,a)}} = \begin{bmatrix} w_{11} & 0 \\ 0 & w_{22} \end{bmatrix}^{-1}.$$

For design class $\mathrm{T}_{23(a)}$, from (4.12) we have

$$2n\operatorname{Var}(\hat{\boldsymbol{\beta}})_{\mathrm{T}_{23(a)}} = \begin{bmatrix} w_{11} & -w_{12} \\ -w_{12} & w_{22} \end{bmatrix}^{-1}.$$

Hence, the trace of $2n\operatorname{Var}(\hat{\boldsymbol{\beta}})$ calculated for $\mathrm{T}_{23(-a,a)}$ and $\mathrm{T}_{23(a)}$ are:

$$\operatorname{tr}\left\{2n\operatorname{Var}(\hat{\boldsymbol{\beta}})_{\mathrm{T}_{23(-a,a)}}\right\} = \frac{w_{11}+w_{22}}{w_{11}w_{22}},\ \operatorname{tr}\left\{2n\operatorname{Var}(\hat{\boldsymbol{\beta}})_{\mathrm{T}_{23(a)}}\right\} = \frac{w_{11}+w_{22}}{w_{11}w_{22}-w_{12}^2}.$$

Since $w_{12}^2 \ge 0$, we have $\operatorname{tr}\left\{2n\operatorname{Var}(\hat{\boldsymbol{\beta}})_{\mathrm{T}_{23(-a,a)}}\right\} \le \operatorname{tr}\left\{2n\operatorname{Var}(\hat{\boldsymbol{\beta}})_{\mathrm{T}_{23(a)}}\right\}$ and the equality holds if and only if w_{12} = 0. The rest is clear.

Theorem 4.3 states that the design class $\mathrm{T}_{23(-a,a)}$ is as efficient as or more efficient than the design class $\mathrm{T}_{23(a)}$ in terms of *A*-optimality criterion with d_{12} = 0. As a result, given d_{12} is zero, the design class $\mathrm{T}_{23(-a,a)}$ is preferred over $\mathrm{T}_{23(a)}$ with respect to both *D*- and *A*-optimality criteria.

4.5 Numerical Study and Results

The findings in this chapter are illustrated by the example of longitudinal study on children's growth pattern in Section 2.4. The study period is five years and three measures of height are taken from each child. There two cohorts defined by birth year and we assume n subjects in each cohort. For model parameters, we assume $\mathbf{D}=\begin{bmatrix}5 & 0\\ 0 & 10\end{bmatrix}$, $\sigma^2=1$, and $\rho=0.5$.

$T_{23(-a,a)}$	$T_{23(a)}$
1st cohort: $(-1, -a, 1)$	1st cohort: $(-1, a, 1)$
2nd cohort: $(-1, a, 1)$	2nd cohort: $(-1, a, 1)$

$(-1 < a < 1)$

Suppose we have two sets of data: one set is collected under design class $T_{23(-a,a)}$, and the other is collected under $T_{23(a)}$ shown in the table above. To examine the *D*-efficiencies of design classes $T_{23(a)}$ and $T_{23(-a,a)}$ graphically, the numerical values of $\det\left\{2n\mathrm{Var}(\hat{\boldsymbol{\beta}})\right\}$ are calculated for both design classes and graphed against a $(-1 < a < 1)$ in Figure 4.1.

From Figure 4.1, we observe that

(1) At time point a = 0, two curves reach the same maximum and are indistinguishable. This phenomenon indicates that given the aforementioned model parameters, a = 0 is the worst choice for both $T_{23(a)}$ and $T_{23(-a,a)}$ under *D*-optimality.

Figure 4.1: $\det\left\{2n\mathrm{Var}(\hat{\boldsymbol{\beta}})\right\}$ for $T_{23(a)}$ and $T_{23(-a,a)}$ against a

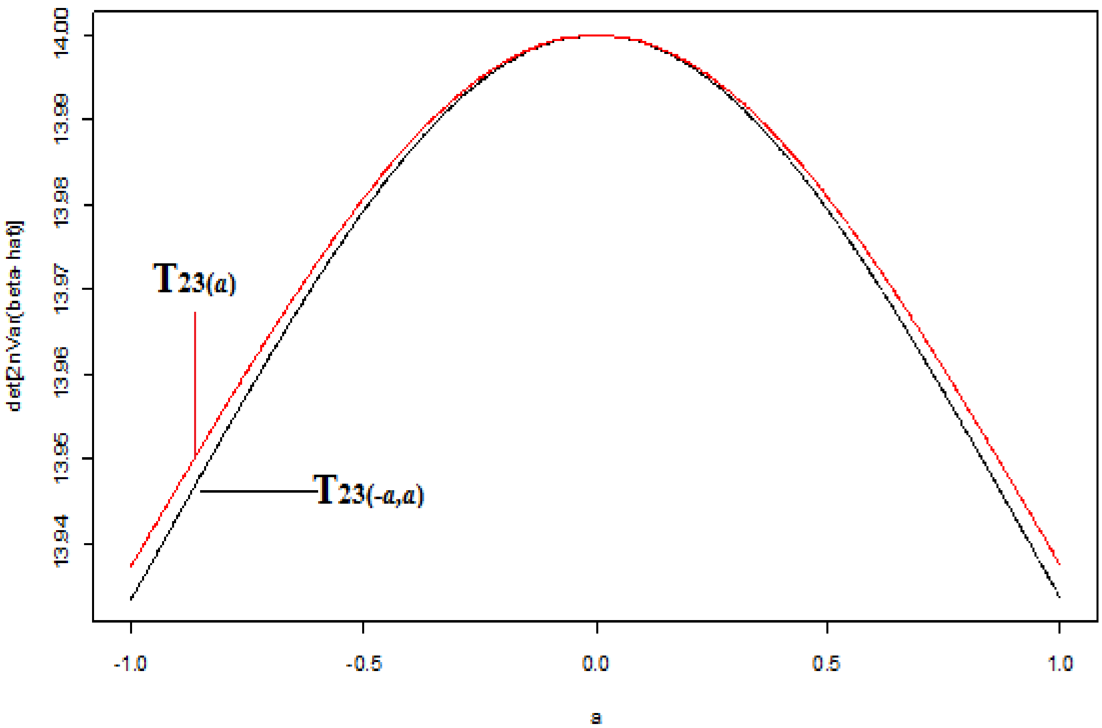


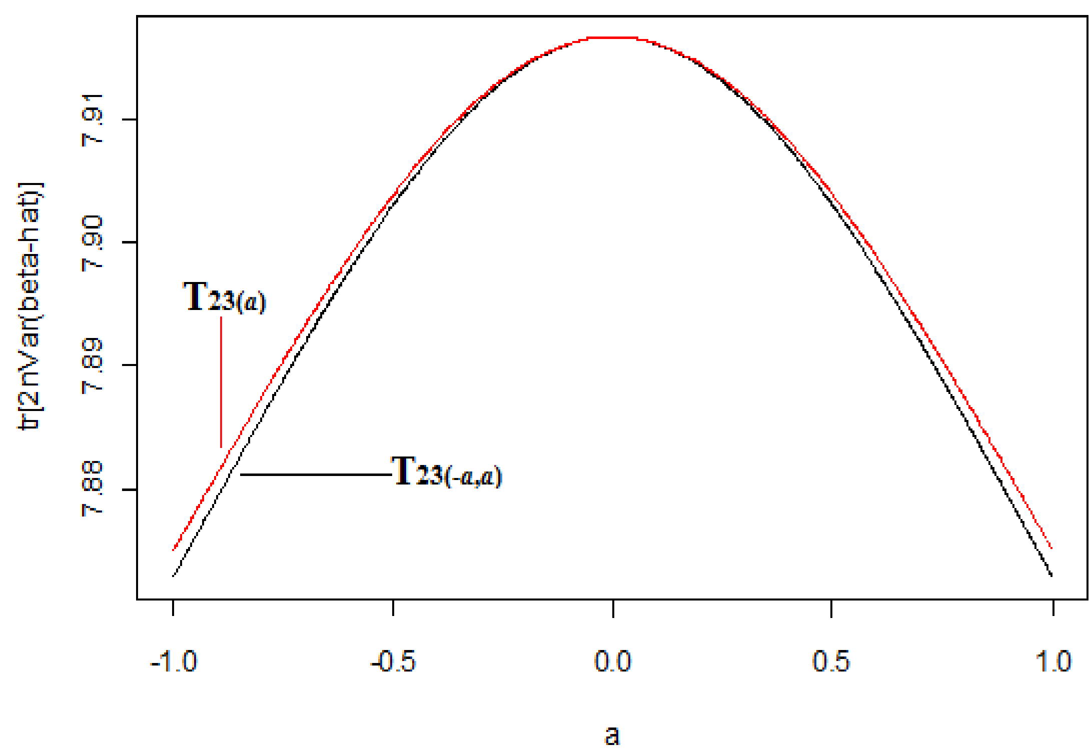
Table 4.2: The numerical value of $\det\left\{2n\mathrm{Var}(\hat{\boldsymbol{\beta}})\right\}$ for $T_{23(a)}$ and $T_{23(-a,a)}$

a	Value of $\det\left\{2n\mathrm{Var}(\hat{\boldsymbol{\beta}})\right\}$	
	$T_{23(a)}$	$T_{23(-a,a)}$
± 0.9	13.94685	13.94336
± 0.7	13.96490	13.96239
± 0.5	13.98077	13.97929
± 0.3	13.99272	13.99213
± 0.1	13.99917	13.99910
0.0	14.00000	14.00000

(2) As the value of $|a|$ increases, the two curves move further apart and become distinguishable. Moreover, the curve for $T_{23(a)}$ is always on top of the curve for $T_{23(-a,a)}$ when $a \neq 0$ and $-1 < a < 1$. Therefore, we conclude that design class $T_{23(-a,a)}$ is more efficient in this case. From Table 4.2, the difference in $\det\left\{2n\mathrm{Var}(\hat{\boldsymbol{\beta}})\right\}$ between these two designs is small due to the facts that the assumed **D** matrix has small numerical values for d_{11}, d_{12}, and d_{22}, and the time interval $[-1, 1]$ is short.

We now perform the analysis using A-optimality. We compute the numerical values of $\mathrm{tr}\left\{2n\mathrm{Var}(\hat{\boldsymbol{\beta}})\right\}$ for design classes $T_{23(a)}$ and $T_{23(-a,a)}$, respectively. After fixing model parameters, the function $\mathrm{tr}\left\{2n\mathrm{Var}(\hat{\boldsymbol{\beta}})\right\}$ depends on only one design parameter, namely a.

Figure 4.2: $\mathrm{tr}\left\{2n\mathrm{Var}(\hat{\boldsymbol{\beta}})\right\}$ for $T_{23(a)}$ and $T_{23(-a,a)}$ against a



The trace of $2n\text{Var}(\hat{\boldsymbol{\beta}})$ is plotted against a ($-1 < a < 1$) in Figure 4.2. It can be seen that the pattern of trace resembles that of the determinant. For $a \neq 0$ and $-1 < a < 1$, the A-efficiency of design class $T_{23(-a,a)}$ is slightly higher than that of design class $T_{23(a)}$. The design classes $T_{23(-a,a)}$ and $T_{23(a)}$ become identical when $a = 0$.

Table 4.3: The numerical value of $\text{tr}\left\{2n\text{Var}(\hat{\boldsymbol{\beta}})\right\}$ for $T_{23(a)}$ and $T_{23(-a,a)}$

a	Value of $\text{tr}\left\{2n\text{Var}(\hat{\boldsymbol{\beta}})\right\}$	
	$T_{23(a)}$	$T_{23(-a,a)}$
± 0.9	7.881234	7.879263
± 0.7	7.893266	7.891846
± 0.5	7.903846	7.903010
± 0.3	7.911812	7.911479
± 0.1	7.916113	7.916074
0.0	7.916667	7.916667

In conclusion, design class $T_{23(-a,a)}$ performs as well as or better than design class $T_{23(a)}$, if there is no correlation between the random effects. Therefore, when planning a longitudinal cohort design, if it is known before data collection that the random intercept and random slope are uncorrelated, and the error correlation has AR(1) structure, then design class $T_{23(-a,a)}$ should be implemented.

Chapter 5

D-Optimal Cohort Designs for $T_{23(a)}$ with $R_i \neq I$

5.1 Introduction

The covariance matrix of repeated measurements (2.5) for the linear mixed effects model is composed of two sources of variation, namely, the within-subject variation and the between-subject variation. The within-subject variation is determined by matrix **D**, whereas the between-subject variation depends on the structure of the error correlation matrix $\boldsymbol{R}_i$ (i = 1, 2). The form of error correlation matrix $\boldsymbol{R}_i$ is important in finding D-optimal cohort designs for linear mixed models. In this chapter, we consider a popular error correlation structure: compound symmetric (CS). The D-optimal cohort designs are obtained for design class $T_{23(a)}$ under the linear mixed effects model (2.3) with compound symmetric error correlation matrix. The comparison between cohort designs with $a = 0$ and $a \neq 0$ is also discussed.

5.2 Compound Symmetric (CS)

Historically, compound symmetric structure is one of the first structures used for the analysis of repeated measurements data. The correlation between errors from two time points $t_{j(i)}$ and $t_{j'(i)}$, ($j = 1,\dots,n$, $i = 1,2$) in the i^{th} cohort is ρ ($-1 \leq \rho \leq 1$). That is, there is a correlation between two separate measurements on the same subject, but it is assumed that the correlation is constant regardless of how far apart the measurements are. For time points $\boldsymbol{t}_i = (t_{1(i)}, t_{2(i)}, t_{3(i)})'$, matrix $\boldsymbol{R}_i$ with compound symmetric structure is

$$\boldsymbol{R}_1 = \boldsymbol{R}_2 = \boldsymbol{R} = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}, \qquad (5.1)$$

where ρ represents the correlation among the repeated measurements from the same subject. Note that because $\boldsymbol{R}$ in (5.1) must be positive-definite, i.e. $|\boldsymbol{R}| = 2\rho^3 - 3\rho^2 + 1 > 0$, the range of ρ is restricted to $-0.5 < \rho < 1$.

5.3 *D*-optimal Designs

The *D*-optimal designs for mixed effects models were studied in the literature. Atkins and Cheng (1999) investigated optimal designs for an random intercept quadratic polynomial model with independent errors. Bischoff (1993) studied *D*-optimal designs for linear models with correlated errors. In this section, we are interested in finding *D*-optimal cohort designs for design class

$$\mathrm{T}_{23(a)} = \left\{ \begin{matrix} (-1,a,1) & (-1,a,1) \\ 0.5 & 0.5 \end{matrix} \middle| -1 < a < 1 \right\}$$

corresponding to the random intercept and random slope model described in (2.3):

$$\boldsymbol{y}_{j(i)} = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{b}_{j(i)} + \boldsymbol{\varepsilon}_{j(i)},$$

where $\boldsymbol{y}_{j(i)} = (y_{1j(i)}, y_{2j(i)}, y_{3j(i)})'$ is the 3 × 1 vector of repeated measurements taken on the j^{th} subject (j = 1,…,n) at time points $\boldsymbol{t}_i = (t_{1(i)}, t_{2(i)}, t_{3(i)})'$ in the i^{th} cohort (i = 1, 2),

$$\mathbf{X}_i = \boldsymbol{Z}_i = \mathbf{X} = \boldsymbol{Z} = \begin{bmatrix} 1 & -1 \\ 1 & a \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \underline{\boldsymbol{j}} & \underline{\boldsymbol{a}} \end{bmatrix}, \text{ with } \underline{\boldsymbol{a}} = (-1, a, 1)' \text{ and } \underline{\boldsymbol{j}} = (1,1,1)'.$$

From (3.3), we compute $\det[2n\mathrm{Var}(\hat{\boldsymbol{\beta}})/\sigma^2]^{-1}$ and denote it as $Q_1^*(a)$ such that

$$Q_1^*(a) = \left(\underline{\boldsymbol{j}}'(\mathbf{ZDZ}' + \boldsymbol{R})^{-1}\underline{\boldsymbol{j}}\right)\left(\underline{\boldsymbol{a}}'(\mathbf{ZDZ}' + \boldsymbol{R})^{-1}\underline{\boldsymbol{a}}\right) - \left(\underline{\boldsymbol{j}}'(\mathbf{ZDZ}' + \boldsymbol{R})^{-1}\underline{\boldsymbol{a}}\right)^2, \qquad (5.2)$$

where $\boldsymbol{R}$ is defined in (5.1) and

$$(\mathbf{ZDZ}' + \boldsymbol{R})^{-1} = \begin{pmatrix} d_{11} - 2d_{12} + d_{22} + 1 & d_{11} - d_{12} + a(d_{12} - d_{22}) + \rho & d_{11} - d_{22} + \rho \\ d_{11} - d_{12} + a(d_{12} - d_{22}) + \rho & d_{11} + 2ad_{12} + a^2 d_{22} + 1 & d_{11} + d_{12} + a(d_{12} + d_{22}) + \rho \\ d_{11} - d_{22} + \rho & d_{11} + d_{12} + a(d_{12} + d_{22}) + \rho & d_{11} + 2d_{12} + d_{22} + 1 + \rho \end{pmatrix}.$$

In terms of *D*-optimality, the design τ_D with the minimum value of $\det[2n\mathrm{Var}(\hat{\boldsymbol{\beta}})/\sigma^2]$ which is equivalent to the maximum value of $Q_1^*(a)$, is considered *D*-optimal.

5.3.1 Equidistant time points

We identify D-optimal designs within the design class $\mathrm{T}_{23(a)}$ by maximizing $Q_1^*(a)$. In certain cases, $Q_1^*(a)$ is maximized at a = 0, which indicates that the design with equidistant time points (−1, 0, 1) is D-optimal. Those certain cases are presented in the following two tables.

Table 5.1: D-optimality region for non-positive a

a	d_{11}	d_{22}	d_{12}	ρ
(−1, 0]	> 0	> 0	$-\sqrt{d_{11}d_{22}} < d_{12} < \sqrt{d_{11}d_{22}}$ and $d_{12} < \frac{a}{6}(1+3d_{11}-d_{22}+2\rho)$	(−0.5, 1)

Table 5.2: D-optimality region for non-negative a

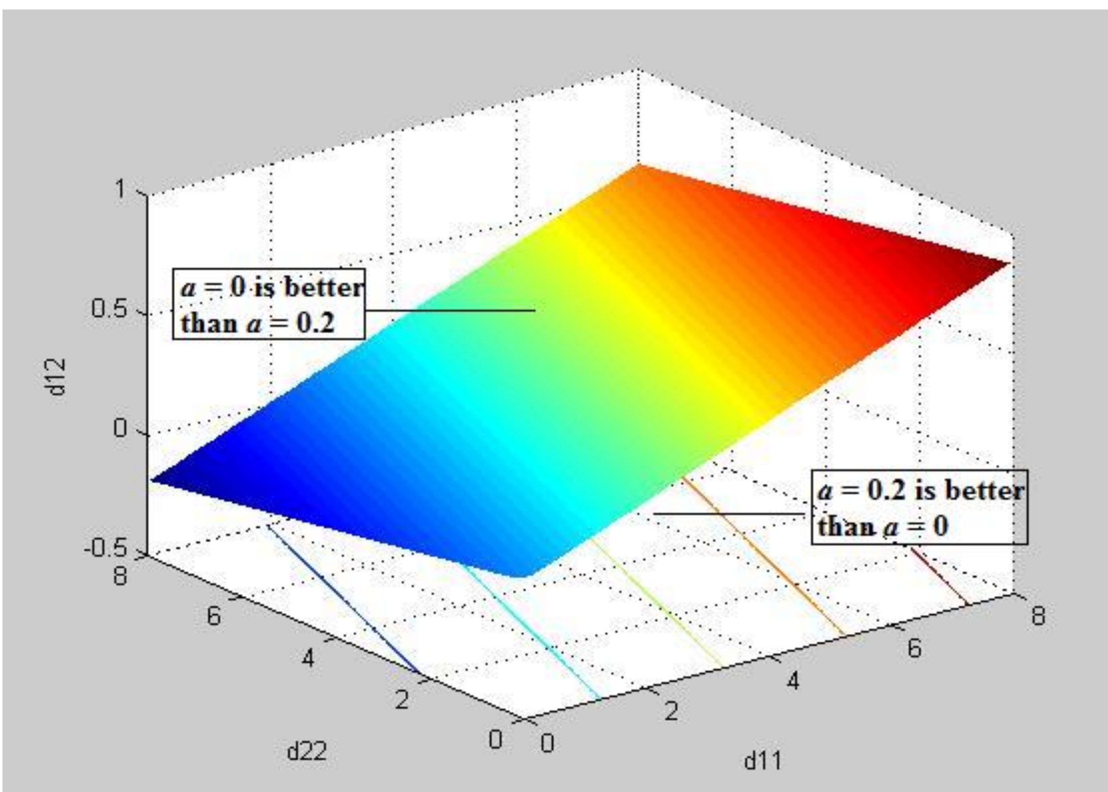
a	d_{11}	d_{22}	d_{12}	ρ
[0, 1)	> 0	> 0	$-\sqrt{d_{11}d_{22}} < d_{12} < \sqrt{d_{11}d_{22}}$ and $d_{12} > \frac{a}{6}(1+3d_{11}-d_{22}+2\rho)$	(−0.5, 1)

When we search for optimal solutions over $-1 < a \leq 0$, the design with $a = 0$ is the most efficient design, if matrix **D** and ρ satisfy the conditions shown in Table 5.1. On the other hand, when $0 \leq a < 1$, the design with $a = 0$ is D-optimal provided that matrix **D** and ρ are in the optimality region illustrated by Table 5.2. The above two D-optimality regions follow directly from Theorem 5.3 and 5.6 that we will discuss in next three sections.

Because '0' is the center of the study period, the comparison between cohort designs with $a = 0$ and $a \neq 0$ is in fact the comparison between two types of designs, namely the design with equidistant time points and the design with non-equidistant time points. As we have already mentioned, if matrix **D** and ρ fall into the optimality region listed in Table 5.1 or Table 5.2, then the design with equidistant time points (−1, 0, 1) is preferred.

For example, consider the cohort designs with $a = 0$ and $a = 0.2$, respectively. For the linear mixed effects model in (2.3) with serial correlation $\rho = 0.6$, we evaluate the efficiency of those two cohort designs graphically by the 3-D plot in Figure 5.1.

Figure 5.1: Comparison between cohort designs with $a = 0$ and $a = 0.2$



It should be noted that X-axis represents d_{11} ranged from 0 to 8, Y-axis represents d_{22} ranged from 0 to 8, and the vertical axis is d_{12}. Based on Theorem 5.3 in Section 5.3.4, the region above the flat surface is defined by two inequalities $-\sqrt{d_{11}d_{22}} < d_{12} < \sqrt{d_{11}d_{22}}$

and $d_{12} > \frac{0.2}{6}(2.2 + 3d_{11} - d_{22})$.

For any point $\{d_{11}, d_{12}, d_{22}\}$ in this region, D-efficiency of the design with $a = 0$ is higher than that of the design with $a = 0.2$. In this case, the design with equidistant time points is favored. Moreover, for points located on the flat surface in Figure 5.1, the designs with $a = 0$ and $a = 0.2$ are indistinguishable with respect to D-optimality.

5.3.2 Non-equidistant time points

Sometimes, however, the cohort designs with $a \neq 0$ yield more efficient estimators of model parameters. Now we present the situations where the design with non-equidistant time points is preferred as follows.

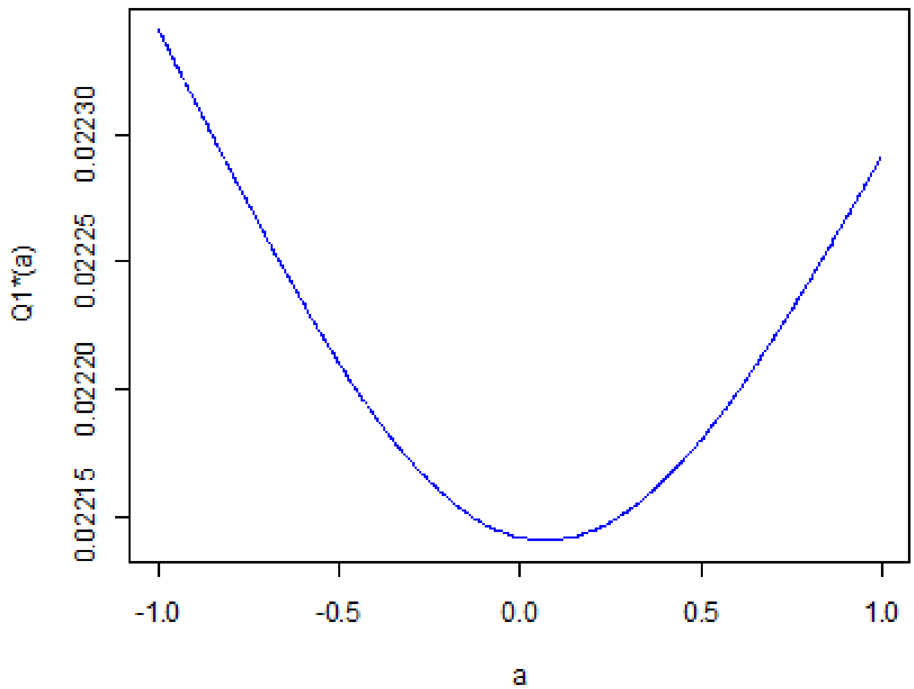
When $a \in (-1, 0): d_{11} > 0, d_{22} > 0, -\sqrt{d_{11}d_{22}} < d_{12} < \sqrt{d_{11}d_{22}}$ and $d_{12} > \frac{a}{6}(1 + 3d_{11} - d_{22} + 2\rho)$.

When $a \in (0, 1): d_{11} > 0, d_{22} > 0, -\sqrt{d_{11}d_{22}} < d_{12} < \sqrt{d_{11}d_{22}}$ and $d_{12} < \frac{a}{6}(1 + 3d_{11} - d_{22} + 2\rho)$.

As an example, consider the comparison between cohort designs with $a = 0$ and $a = 0.2$ again. In Figure 5.1, for all points $\{d_{11}, d_{12}, d_{22}\}$ under the flat surface, the design with $a = 0.2$ is more efficient than the design with $a = 0$. For instance, consider a point such that $d_{11} = 8$, $d_{12} = 0.5$, and $d_{22} = 5$. Since this points falls beneath the surface, we know $a = 0$ is not D-optimal. In fact, given $\mathbf{D} = \begin{bmatrix} 8 & 0.5 \\ 0.5 & 5 \end{bmatrix}$, the D-optimal design $\tau_D \in \mathrm{T}_{23(a)}$ can be computed by maximizing det $[2n\mathrm{Var}(\hat{\boldsymbol{\beta}})/\sigma^2]^{-1}$, i.e. $Q_1^*(a)$, with respect to a. The numerical values of $Q_1^*(a)$ are calculated and graphed against a ($0 < a < 1$) in Figure 5.2.

We observe that $Q_1^*(a)$ approaches its maximum when a gets closer to -1. If one can put a threshold on the lower bound of a such that $-0.8 \le a$, then the design with $a = -0.8$ is D-optimal. In other words, the D-optimal design is $\tau_D = \begin{Bmatrix} (-1, -0.8, 1) & (-1, -0.8, 1) \\ 0.5 & 0.5 \end{Bmatrix}$.

Figure 5.2: Plot of $Q_1^*(a)$ given $d_{11} = 8$, $d_{12} = 0.5$, $d_{22} = 5$, and $\rho = 0.6$



5.3.3 Comparison of $Q_1^*(0)$ with $Q_1^*(a)$

The evaluation of designs with equidistant and non-equidistant time points in terms of D-optimality leads to the comparison of $Q_1^*(0)$ with $Q_1^*(a)$ for $a \neq 0$ and $-1 < a < 1$.

The function in (5.2) can be rewritten as

$$Q_1^*(a) = \frac{2(a^2+3)}{\begin{pmatrix} 3d_{11} + 2d_{22} + (6+2a^2)(d_{11}d_{22} - d_{12}^2) + a^2 d_{22} + 2ad_{12} + 1 \\ +\rho(a^2 d_{22} + 1 - 3d_{11} + 4d_{22} - 2\rho - 2ad_{12}) \end{pmatrix}} = \frac{\text{Num}_1^*(a)}{\text{Den}_1^*(a)}, \qquad (5.3)$$

where $\text{Num}_1^*(a) > 0$ and $\text{Den}_1^*(a) > 0$ for all a ($-1 < a < 1$). In addition, $Q_1^*(a)$ is a function of four model parameters d_{11}, d_{12}, d_{22}, ρ and one design parameter a. In our analysis, we assume the following conditions on those five parameters.

Table 5.3: The parameter conditions

a	d_{11}	d_{22}	d_{12}	ρ
$(-1, 1)$	> 0	> 0	$\left(-\sqrt{d_{11}d_{22}}, \sqrt{d_{11}d_{22}}\right)$	$(-0.5, 1)$

As we mentioned earlier in Chapter 3, we assume the additional measurement cannot be taken at the beginning ($a \neq -1$) or at the end ($a \neq 1$) of the study. Here, both d_{11} and d_{22} are greater than zero, because they are variances for the random intercept and random slope, respectively. To make sure that matrix **D** is positive-definite, i.e. $|\mathbf{D}| = d_{11}d_{22} - d_{12}^2 > 0$, we restrict the absolute value of d_{12} to be less than $\sqrt{d_{11}d_{22}}$.

We consider the difference between $Q_1^*(0)$ and $Q_1^*(a)$

$$Q_1^*(0) - Q_1^*(a) = \frac{\text{Num}_1^*(0)}{\text{Den}_1^*(0)} - \frac{\text{Num}_1^*(a)}{\text{Den}_1^*(a)} = \frac{f_{(1)}}{\text{Den}_1^*(0)\text{Den}_1^*(a)}, \qquad (5.4)$$

where

$$f_{(1)} = \text{Num}_1^*(0)\text{Den}_1^*(a) - \text{Num}_1^*(a)\text{Den}_1^*(0).$$

Since $\text{Den}_1^*(0) > 0$, $\text{Den}_1^*(a) > 0$, we have

$$Q_1^*(0) - Q_1^*(a) > 0 \iff f_{(1)} > 0,$$

where $f_{(1)}$ is further written as

$$f_{(1)} = 2a(\rho - 1)(a - 6d_{12} + 3ad_{11} - ad_{22} + 2a\rho). \qquad (5.5)$$

Denote

$g^{(0)} = 2a,$

$g^{(1)} = \rho - 1 \leq 0,$

$$g^{(2)} = a - 6d_{12} + 3ad_{11} - ad_{22} + 2a\rho = a(1 + 3d_{11} - d_{22} + 2\rho) - 6d_{12}. \qquad (5.6)$$

Then, $f_{(1)}$ can be rewritten as the product of $g^{(0)}$, $g^{(1)}$, and $g^{(2)}$ such that

$$f_{(1)} = g^{(0)} g^{(1)} g^{(2)}. \qquad (5.7)$$

It should be pointed out that when $a = 0$, $g^{(0)} = 0$ and $f_{(1)} = 0$. Thus, the designs with a = 0 and $a \neq 0$ are indistinguishable with respect to *D*-optimality criterion. When $f_{(1)} > 0$, the cohort design with a = 0 is *D*-optimal. However, when $f_{(1)} < 0$, the cohort design with $a \neq 0$ is *D*-optimal. Since the comparison of $Q_1^*(0)$ with $Q_1^*(a = 0)$ is unnecessary, we perform our analysis on $f_{(1)}$ for $a \neq 0$ and $-1 < a < 1$ through the three cases in Table 5.4.

Table 5.4: Analysis of $f_{(1)}$ in three cases

$g^{(2)} > 0$	$g^{(2)} < 0$	$g^{(2)} = 0$
I	II	III

5.3.4 The Situation with $-1 < a < 0$

We evaluate $Q_1^*(0) - Q_1^*(a)$ over the region $-1 < a < 0$ by analyzing the sign of $f_{(1)}$ under the aforementioned three cases. Table 5.5 shows the signs of $g^{(0)}$, $g^{(1)}$, $g^{(2)}$, and $f_{(1)}$. Notice that the function $g^{(2)}$ is important in determining the sign of $f_{(1)}$.

Table 5.5: Analysis of $f_{(1)}$ for $-1 < a < 0$

Case	$g^{(0)}$	$g^{(1)}$	$g^{(2)}$	$f_{(1)}$
I	< 0	< 0	> 0	> 0
II	< 0	< 0	< 0	< 0
III	< 0	< 0	$= 0$	$= 0$

We present results comparing $Q_1^*(0)$ with $Q_1^*(a)$ in the next three theorems.

Theorem 5.1

When $-1 < a < 0$, we have

1. $f_{(1)} \geq 0$ for all a, d_{11}, d_{22}, d_{12}, and ρ satisfying $g^{(2)} \geq 0$,
2. $f_{(1)} < 0$ for all a, d_{11}, d_{22}, d_{12}, and ρ satisfying $g^{(2)} < 0$.

Proof.

1. First note that $g^{(0)} > 0$ and $g^{(1)} > 0$. Suppose a, d_{11}, d_{22}, d_{12}, and ρ satisfying $g^{(2)} \geq 0$, it is clear from (5.7) that $f_{(1)} = g^{(0)} g^{(1)} g^{(2)} \geq 0$, and the result follows.
2. Suppose a, d_{11}, d_{22}, d_{12}, and ρ satisfying $g^{(2)} < 0$, we know $g^{(0)} > 0$ and $g^{(1)} > 0$. Then it is clear from (5.7), $f_{(1)} = g^{(0)} g^{(1)} g^{(2)} < 0$, and the result follows.

Theorem 5.2

When $-1 < a < 0$,

1. The condition all a, d_{11}, d_{22}, d_{12}, and ρ satisfying $g^{(2)} \geq 0$ implies that d_{12} must satisfy

$$-\sqrt{d_{11}d_{22}} < d_{12} \leq \min\left\{\sqrt{d_{11}d_{22}}, \frac{a(1+3d_{11}-d_{22}+2\rho)}{6}\right\} \text{for all } a, d_{11}, d_{22}, d_{12}, \text{ and } \rho .$$

2. The condition all a, d_{11}, d_{22}, d_{12}, and ρ satisfying $g^{(2)} < 0$ implies that d_{12} must satisfy

$$\max\left\{-\sqrt{d_{11}d_{22}}, \frac{a(1+3d_{11}-d_{22}+2\rho)}{6}\right\} < d_{12} < \sqrt{d_{11}d_{22}} \text{ for all } a, d_{11}, d_{22}, d_{12}, \text{ and } \rho .$$

Proof.

1. Suppose a, d_{11}, d_{22}, d_{12}, and ρ satisfying $g^{(2)} \geq 0,$ then from (5.6),

$$g^{(2)} = a(1+3d_{11}-d_{22}+2\rho)-6d_{12} \geq 0, \text{ which implies } d_{12} \leq \frac{a(1+3d_{11}-d_{22}+2\rho)}{6}.$$

Since $-\sqrt{d_{11}d_{22}} < d_{12} < \sqrt{d_{11}d_{22}},$ we have

$$-\sqrt{d_{11}d_{22}} < d_{12} \leq \min\left\{\sqrt{d_{11}d_{22}}, \frac{a(1+3d_{11}-d_{22}+2\rho)}{6}\right\} \text{ and the result follows.}$$

2. Suppose a, d_{11}, d_{22}, d_{12}, and ρ satisfying $g^{(2)} < 0,$ then from (5.6),

$$g^{(2)} = a(1+3d_{11}-d_{22}+2\rho)-6d_{12} < 0, \text{ which implies } d_{12} > \frac{a(1+3d_{11}-d_{22}+2\rho)}{6}.$$

Since $-\sqrt{d_{11}d_{22}} < d_{12} < \sqrt{d_{11}d_{22}}$, we have

$\max\left\{-\sqrt{d_{11}d_{22}}, \frac{a(1+3d_{11}-d_{22}+2\rho)}{6}\right\} < d_{12} < \sqrt{d_{11}d_{22}}$ and the result follows.

Theorem 5.3

When $-1 < a < 0$, we have

1. $f_{(1)} \geq 0$ for all a, d_{11}, d_{22}, d_{12}, and ρ satisfying

$$-\sqrt{d_{11}d_{22}} < d_{12} \leq \min\left\{\sqrt{d_{11}d_{22}}, \frac{a(1+3d_{11}-d_{22}+2\rho)}{6}\right\},$$

2. $f_{(1)} < 0$ for all a, d_{11}, d_{22}, d_{12}, and ρ satisfying

$$\max\left\{-\sqrt{d_{11}d_{22}}, \frac{a(1+3d_{11}-d_{22}+2\rho)}{6}\right\} < d_{12} < \sqrt{d_{11}d_{22}}.$$

Proof. Clear from Theorem 5.1 and Theorem 5.2.

5.3.5 The Situation with 0 < *a* < 1

Now we analyze $f_{(1)}$ via the three cases shown in Table 5.4 for $0 < a < 1$. The sign of $f_{(1)}$ depends on three functions, namely $g^{(0)}$, $g^{(1)}$, and $g^{(2)}$. It can be seen that $f_{(1)}$ is negative for case I, positive for case II, and zero for case III.

Table 5.6: Analysis of $f_{(1)}$ for $0 < a < 1$

Case	$g^{(0)}$	$g^{(1)}$	$g^{(2)}$	$f_{(1)}$
I	> 0	< 0	> 0	< 0
II	> 0	< 0	< 0	> 0
III	> 0	< 0	$= 0$	$= 0$

Theorem 5.4

When $0 < a < 1$, we have

1. $f_{(1)} \geq 0$ for all a, d_{11}, d_{22}, d_{12}, and ρ satisfying $g^{(2)} \leq 0$,

2. $f_{(1)} < 0$ for all a, d_{11}, d_{22}, d_{12}, and ρ satisfying $g^{(2)} > 0$.

Proof. Similar to proofs of Theorem 5.1.

Theorem 5.5

When $0 < a < 1$,

1. The condition all a, d_{11}, d_{22}, d_{12}, and ρ satisfying $g^{(2)} \leq 0$ implies that d_{12} must satisfy

   $$\max\left\{-\sqrt{d_{11}d_{22}}, \frac{a(1+3d_{11}-d_{22}+2\rho)}{6}\right\} \leq d_{12} < \sqrt{d_{11}d_{22}} \text{ for all } a,\ d_{11},\ d_{22},\ d_{12}, \text{ and } \rho.$$

2. The condition all a, d_{11}, d_{22}, d_{12}, and ρ satisfying $g^{(2)} > 0$ implies that d_{12} must satisfy

   $$-\sqrt{d_{11}d_{22}} < d_{12} < \min\left\{\sqrt{d_{11}d_{22}}, \frac{a(1+3d_{11}-d_{22}+2\rho)}{6}\right\} \text{ for all } a,\ d_{11},\ d_{22},\ d_{12}, \text{ and } \rho.$$

Proof. Similar to proofs of Theorem 5.2.

Theorem 5.6

When $0 < a < 1$, we have

1. $f_{(1)} \geq 0$ for all a, d_{11}, d_{22}, d_{12}, and ρ satisfying

   $$\max\left\{-\sqrt{d_{11}d_{22}}, \frac{a(1+3d_{11}-d_{22}+2\rho)}{6}\right\} \leq d_{12} < \sqrt{d_{11}d_{22}},$$

2. $f_{(1)} < 0$ for all a, d_{11}, d_{22}, d_{12}, and ρ satisfying

$$-\sqrt{d_{11}d_{22}} < d_{12} < \min\left\{\sqrt{d_{11}d_{22}}, \frac{a(1+3d_{11}-d_{22}+2\rho)}{6}\right\}.$$

Proof. Clear from Theorem 5.4 and Theorem 5.5.

Theorems 5.1 − 5.6 state the general results of the comparison between $Q_1^*(0)$ and $Q_1^*(a)$ for $a \neq 0$ and $-1 < a < 1$. These theorems provide us with analytical solutions to the comparison between cohort designs with equidistant time points and non-equidistant time points. To be specific, for a, d_{11}, d_{12}, d_{22}, ρ satisfying certain conditions so that $f_{(1)} > 0$, we find that designs with $a = 0$ is preferred. On the other hand, for a, d_{11}, d_{12}, d_{22}, ρ satisfying certain conditions so that $f_{(1)} < 0$, optimal allocation of time points is achieved by choosing $a \neq 0$.

Chapter 6

Optimal Cohort Designs for T_{23} with $\boldsymbol{R_i}$ = I and D = I

6.1 Introduction

Finding optimal cohort designs analytically for longitudinal studies when $a_1 \neq a_2$ is very difficult because of the complexity of $\mathrm{Var}(\hat{\boldsymbol{\beta}})$. In fact, in the literature [4], [16], [27-30], all optimal cohort designs have been computed numerically. In this chapter, we introduce and describe our approach for deriving *D*-, *A*-, and *E*-optimal cohort designs analytically for the design class T_{23} in (2.10) under the linear mixed effects model in (2.3) with the covariance of random effects **D** = **I** and uncorrelated errors, i.e. $\boldsymbol{R}_{\mathrm{i}}$ = **I**. Finally, the findings of optimal designs are summarized.

6.2 Optimal Cohort Designs

The class of cohort designs T_{23} is given by

$$T_{23} = \left\{ \begin{matrix} (-1, a_1, 1) & (-1, a_2, 1) \\ 0.5 & 0.5 \end{matrix} \middle| -1 < a_1 \le a_2 < 1 \right\}.$$

As we have mentioned earlier, a longitudinal cohort study models the responses as a function of time for groups (cohorts) of subjects. The precision of estimates of model parameters is measured by their covariance matrix.

Consider the linear mixed effects model described in (2.3) with $\mathbf{D} = \mathbf{I}$ and $\boldsymbol{R}_i = \mathbf{I}$

$$\boldsymbol{y}_{j(i)} = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{b}_{j(i)} + \boldsymbol{\varepsilon}_{j(i)}, \qquad i = 1, 2,\ j = 1, \dots, n,$$

$$\text{where } \boldsymbol{X}_i = \boldsymbol{Z}_i = \begin{pmatrix} 1 & 1 \\ 1 & a_i \\ 1 & -1 \end{pmatrix}.$$

The covariance matrix can be expressed as

$$\frac{n\mathrm{Var}(\hat{\boldsymbol{\beta}})}{\sigma^2} = \left[\boldsymbol{X}_1' \boldsymbol{\Sigma}_1^{-1} \boldsymbol{X}_1 + \boldsymbol{X}_2' \boldsymbol{\Sigma}_2^{-1} \boldsymbol{X}_2 \right]^{-1} = [\mathrm{V}(a_1) + \mathrm{V}(a_2)]^{-1}$$

$$= \begin{bmatrix} \dfrac{2a_1^2 + 9}{12 + 3a_1^2} + \dfrac{2a_2^2 + 9}{12 + 3a_2^2} & \dfrac{a_1}{12 + 3a_1^2} + \dfrac{a_2}{12 + 3a_2^2} \\ \dfrac{a_1}{12 + 3a_1^2} + \dfrac{a_2}{12 + 3a_2^2} & \dfrac{3a_1^2 + 8}{12 + 3a_1^2} + \dfrac{3a_2^2 + 8}{12 + 3a_2^2} \end{bmatrix}^{-1}, \qquad (6.1)$$

which is a function of two design parameters, namely a_1 and a_2. Here, we divide by σ^2 in order to consider an optimality criterion that is scale free, and n represents the sample size of each cohort. In the following sections, three optimality criteria will be used in finding the optimal allocation of time points for design class T_{23}. First, we will consider these three optimality criteria separately, and then we will describe how to apply the combination of these three criteria to identity the optimal designs.

6.2.1 *D*-optimality

A *D*-optimal cohort design is the one among all possible cohort designs in T_{23} for which $\det\left\{n\mathrm{Var}(\hat{\boldsymbol{\beta}})/\sigma^2\right\}$ is minimized. To simplify the computation procedure, we consider $\det[V(a_1) + V(a_2)]$ and denote it by $DV(a_1, a_2)$

$$DV(a_1,a_2)=\left(\frac{2a_1^2+9}{12+3a_1^2}+\frac{2a_2^2+9}{12+3a_2^2}\right)\left(\frac{3a_1^2+8}{12+3a_1^2}+\frac{3a_2^2+8}{12+3a_2^2}\right)-\left(\frac{a_1}{12+3a_1^2}+\frac{a_2}{12+3a_2^2}\right)^2.$$

We write

$$DV(a_1, a_2) = DV_1(a_1, a_2) - DV_2(a_1, a_2), \qquad (6.2)$$

where

$$DV_1(a_1, a_2) = \left(\frac{2a_1^2+9}{12+3a_1^2}+\frac{2a_2^2+9}{12+3a_2^2}\right)\left(\frac{3a_1^2+8}{12+3a_1^2}+\frac{3a_2^2+8}{12+3a_2^2}\right), \qquad (6.3)$$

$$DV_2(a_1, a_2) = \left(\frac{a_1}{12+3a_1^2}+\frac{a_2}{12+3a_2^2}\right)^2. \qquad (6.4)$$

Property 1. For $-1<a_1\le a_2<1$, $DV_2(a_1, a_2)$ attains its minimum when $a_1 = -a_2$.

Proof. $DV_2(a_1,a_2)=\left(\frac{a_1}{12+3a_1^2}+\frac{a_2}{12+3a_2^2}\right)^2=\frac{9\left[(4+a_1a_2)(a_1+a_2)\right]^2}{\left(12+3a_1^2\right)^2\left(12+3a_2^2\right)^2}$. For $-1<a_1\le a_2<1$,

$\left(12+3a_1^2\right)^2\left(12+3a_2^2\right)^2 > 0$, $9\left[(4+a_1a_2)(a_1+a_2)\right]^2 \geq 0$, and $4+a_1a_2>0$. Therefore, the minimum value of $DV_2(a_1, a_2)$ is zero, and this is attained when $a_1+a_2=0 \Leftrightarrow a_1=-a_2$, and the rest is clear.

Property 2. For $-1<a_1\leq a_2<1$, $\min\{DV_2(a_1, a_2)\} = 0$.

Proof. Clear from Property 1.

Now we denote

$$DV_3(a) = \frac{2a^2+9}{12+3a^2},\ DV_4(a) = \frac{-a^2+1}{12+3a^2}. \qquad (6.5)$$

It follows that

$$DV_3(a_1, a_2) = DV_3(a_1) + DV_3(a_2) = \frac{2a_1^2+9}{12+3a_1^2}+\frac{2a_2^2+9}{12+3a_2^2}, \qquad (6.6)$$

$$DV_4(a_1, a_2) = DV_4(a_1) + DV_4(a_2) = \frac{-a_1^2+1}{12+3a_1^2}+\frac{-a_2^2+1}{12+3a_2^2}. \qquad (6.7)$$

Property 3. $DV_1(a_1, a_2) = DV_3(a_1, a_2)\,[DV_3(a_1, a_2) - DV_4(a_1, a_2)]$.

Proof. Clear from (6.3), (6.6), and (6.7).

Property 4. For $-1<a_1\leq a_2<1$, $DV_3(a_1, a_2) > 0$ and $\lim_{a_1^2\to 1, a_2^2\to 1} DV_4(a_1, a_2) = 0$.

Proof. From (6.6), $DV_3(a_1, a_2) = \dfrac{51(a_1^2+a_2^2)+12a_1^2a_2^2+216}{(12+3a_1^2)(12+3a_2^2)}$.

For $-1 < a_1 \le a_2 < 1$, $a_1^2 \ge 0$ and $a_2^2 \ge 0$, so both the numerator and the denominator are positive and hence $DV_3(a_1, a_2) > 0$. $\lim_{a_1^2 \to 1, a_2^2 \to 1} DV_4(a_1, a_2) = \lim_{a_1^2 \to 1} DV_4(a_1) + \lim_{a_2^2 \to 1} DV_4(a_2)$. It is clear that $\lim_{a_1^2 \to 1} -a_1^2 + 1 = 0$ and $\lim_{a_1^2 \to 1} 12 + 3a_1^2 = 15$. Therefore, $\lim_{a_1^2 \to 1} DV_4(a_1) = 0$. Equivalently, we can show that $\lim_{a_2^2 \to 1} DV_4(a_2) = 0$. The rest is clear.

Property 5. For $-1 < a_1 \le a_2 < 1$, $\max\{DV_1(a_1, a_2)\} \to \left(\frac{22}{15}\right)^2$, as $a_1^2 \to 1$ and $a_2^2 \to 1$.

Proof. We have $DV_3(a_1, a_2) = \frac{22}{15} + \frac{3}{15} DV_4(a_1, a_2)$. From Property 3,

$$
\begin{aligned}
DV_1(a_1, a_2) &= DV_3(a_1, a_2)\left[DV_3(a_1, a_2) - DV_4(a_1, a_2)\right] \\
&= \left(\frac{22}{15}\right)^2 - \frac{22 \times 9}{15^2} DV_4(a_1, a_2) - \frac{3 \times 12}{15^2}\left(DV_4(a_1, a_2)\right)^2.
\end{aligned}
$$

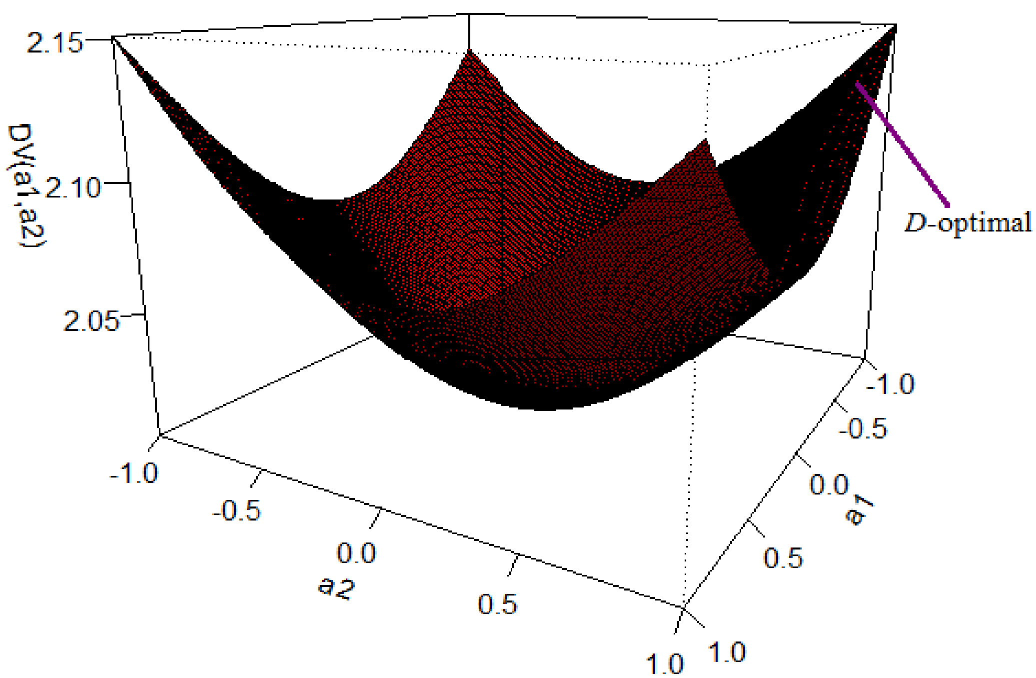
From Property 4, $\lim_{a_1^2 \to 1, a_2^2 \to 1} DV_4(a_1, a_2) = 0$. Consequently, $\lim_{a_1^2 \to 1, a_2^2 \to 1} DV_1(a_1, a_2) = \left(\frac{22}{15}\right)^2$.

The rest is clear.

According to *D*-optimality criterion, we can obtain the *D*-optimal cohort design by maximizing the function $DV(a_1, a_2)$ in (6.2). Based on Property 1 to 5, we know $\min\{DV_2(a_1, a_2)\} = 0$, when $a_1 = -a_2$. If users could put a threshold on a_i ($i = 1, 2$) such that $a_i^2 \le a_0$, where a_0 is an arbitrary value between 0 and 1, then the value of $DV_1(a_1, a_2)$ is maximized at $a_1^2 = a_2^2 = a_0$. It can be checked that for $-1 < a_1 \le a_2 < 1$, the time points (a_1, a_2) that maximize $DV_1(a_1, a_2)$ and minimize $DV_2(a_1, a_2)$ simultaneously are

$\left(-\sqrt{a_0},\sqrt{a_0}\right)$, which are presented graphically in Figure 6.1. Therefore, the D-optimal cohort design for design class T_{23} is $\tau_D = \begin{Bmatrix} (-1,-\sqrt{a_0},1) & (-1,\sqrt{a_0}1) \\ 0.5 & 0.5 \end{Bmatrix}$, given $\mathbf{D} = \mathbf{I}$, $\boldsymbol{R}_i = \mathbf{I}$.

Figure 6.1: The plot of DV(a_1, a_2) for $-1 < a_1 < 1$ and $-1 < a_2 < 1$



6.2.2 *A*-optimality

The A-optimality criterion is based on the sum of the diagonal elements of $n\text{Var}(\hat{\boldsymbol{\beta}})/\sigma^2$ in (6.1).

We have

$$\left[\frac{n\text{Var}(\hat{\boldsymbol{\beta}})}{\sigma^2}\right]^{-1} = \begin{bmatrix} \alpha & \gamma \\ \gamma & \beta \end{bmatrix}, \qquad (6.8)$$

where

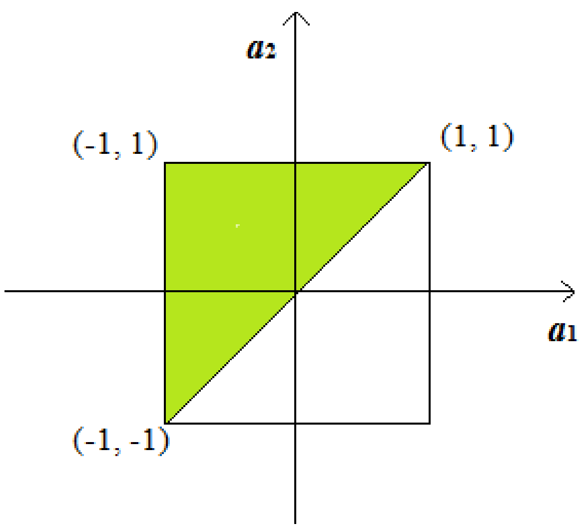
$$\alpha = \frac{2a_1^2+9}{12+3a_1^2} + \frac{2a_2^2+9}{12+3a_2^2},\ \beta = \frac{3a_1^2+8}{12+3a_1^2} + \frac{3a_2^2+8}{12+3a_2^2},\ \text{and}\ \gamma = \frac{a_1}{12+3a_1^2} + \frac{a_2}{12+3a_2^2}. \qquad (6.9)$$

It can be checked that

$$\operatorname{tr}\left\{\frac{n\operatorname{Var}(\hat{\boldsymbol{\beta}})}{\sigma^2}\right\} = \frac{\alpha+\beta}{\alpha\beta-\gamma^2} = \frac{3(10a_1^2a_2^2+37a_1^2+37a_2^2+136)}{24a_1^2a_2^2+85a_1^2+85a_2^2-2a_1a_2+288}. \qquad (6.10)$$

In order to find the *A*-optimal design, we search for time points (a_1, a_2) that minimize equation (6.10) over the region of $-1 < a_1 \le a_2 < 1$ indicated in Figure 6.2.

Figure 6.2: Region of $-1 < a_1 \le a_2 < 1$



Property 6. Denote $\operatorname{tr}\left\{\frac{n\operatorname{Var}(\hat{\boldsymbol{\beta}})}{\sigma^2}\right\}$ by $f(a_1, a_2)$. We have $\min\{f(a_1,a_2)\} \to 1.364$, as $a_1 \to -1$ and $a_2 \to 1$.

Proof. By solving $\frac{\partial f(a_1, a_2)}{a_1} = 0$ and $\frac{\partial f(a_1, a_2)}{a_2} = 0$, we find a_1 = a_2 = 0, which is the critical point inside the above region. At this critical point, $f(0,0)$ = 1.417. We then examine the minimum on the boundary. We observe that $\lim_{a_1 \to -1, a_1 \to 1} f(a_1, a_2)$ = 1.364, $\lim_{a_1 \to 1, a_1 \to 1} f(a_1, a_2)$ = 1.375, and $\lim_{a_1 \to -1, a_1 \to -1} f(a_1, a_2)$ = 1.375. Therefore, the minimum value of $f(a_1, a_2)$ approaches 1.364 when $a_1 \to -1$ and $a_2 \to 1$. The rest is clear.

Figure 6.3: The plot of tr{ $n\text{Var}(\hat{\boldsymbol{\beta}})/\sigma^2$ } for $-1 < a_1 < 1$ and $-1 < a_2 < 1$

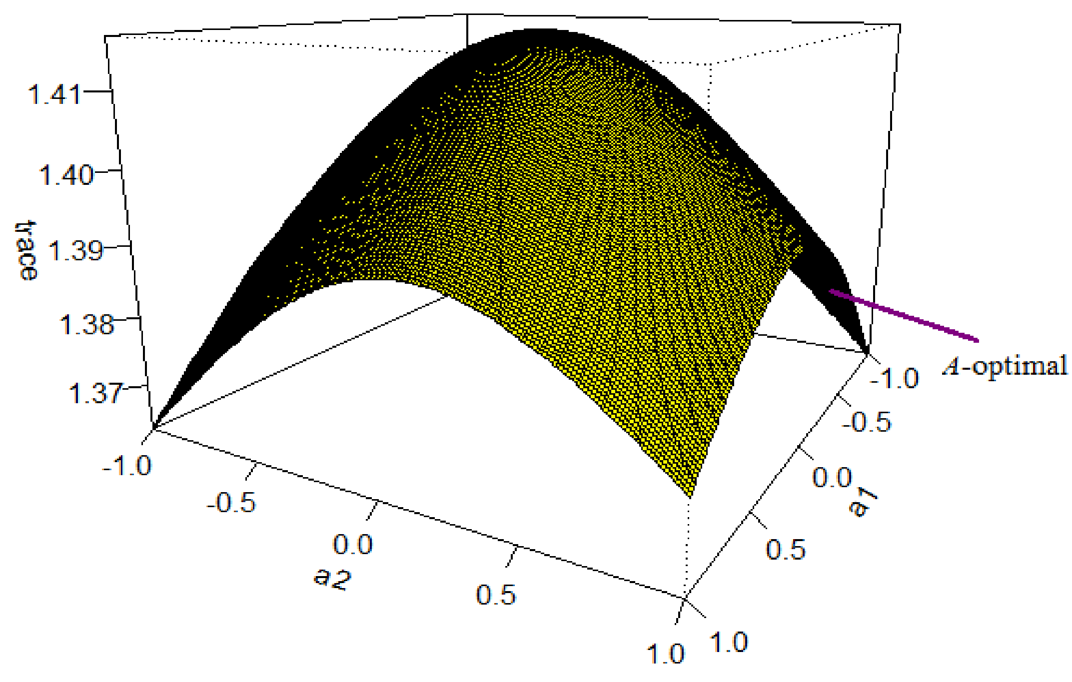


Figure 6.3 presents the values of tr{ $n\text{Var}(\hat{\boldsymbol{\beta}})/\sigma^2$ }, where X-axis is a_1 ($-1 < a_1 < 1$) and Y-axis is a_2 ($-1 < a_2 < 1$). Notice that tr{ $n\text{Var}(\hat{\boldsymbol{\beta}})/\sigma^2$ } is getting close to the minimum as $a_1 \to -1$ and $a_2 \to 1$.

As we have discussed earlier, if one can put a threshold on a_i (i = 1, 2) such that $a_i^2 \le a_0$, ($0 < a_0 < 1$), then tr{ $n\mathrm{Var}(\hat{\boldsymbol{\beta}})/\sigma^2$ } is minimized at the time points $\left(-\sqrt{a_0}, \sqrt{a_0}\right)$ based on Property 6 and Figure 6.3. Consequently, with respect to A-optimality criterion, the optimal cohort design for design class T_{23} is $\tau_A = \begin{Bmatrix} (-1,-\sqrt{a_0},1) & (-1,\sqrt{a_0}1) \\ 0.5 & 0.5 \end{Bmatrix}$ as well, given $\mathbf{D} = \mathbf{I}$, and $\boldsymbol{R}_\mathrm{i} = \mathbf{I}$.

6.2.3 *E*-optimality

Another popular optimality criterion is E-optimality, which minimizes the maximum eigenvalue (characteristic root) of the covariance matrix $n\mathrm{Var}(\hat{\boldsymbol{\beta}})/\sigma^2$ in (6.1). This is equal to minimizing the variance of the least well-estimated contrast $\alpha\boldsymbol{\beta}$, subject to $\alpha'\alpha = 1$. Hence, E in the name of the criterion stands for extreme.

It can be shown that the maximum characteristic root for $n\mathrm{Var}(\hat{\boldsymbol{\beta}})/\sigma^2$ is $1/\lambda_2$, where

$$\lambda_2 = \frac{(\alpha+\beta)-\sqrt{(\alpha-\beta)^2+4\gamma^2}}{2}$$
$$= \frac{1}{2}\left(\frac{5a_1^2+17}{12+3a_1^2}+\frac{5a_2^2+17}{12+3a_2^2}-\sqrt{\left(\frac{1-a_1^2}{12+3a_1^2}+\frac{1-a_2^2}{12+3a_2^2}\right)^2+\left(\frac{a_1}{12+3a_1^2}+\frac{a_2}{12+3a_2^2}\right)^2}\right).$$

Define

$$\mathrm{DV}_5(a_1, a_2) = \mathrm{DV}_5(a_1) + \mathrm{DV}_5(a_2) = \frac{5a_1^2+17}{12+3a_1^2}+\frac{5a_2^2+17}{12+3a_2^2}, \qquad (6.11)$$

$$\mathrm{DV}_6(a_1, a_2) = \sqrt{\left(\frac{1-a_1^2}{12+3a_1^2}+\frac{1-a_2^2}{12+3a_2^2}\right)^2+\left(\frac{a_1}{12+3a_1^2}+\frac{a_2}{12+3a_2^2}\right)^2}. \qquad (6.12)$$

Combining (6.10) and (6.11), λ_2 can be expressed as

$$\lambda_2 = \frac{1}{2}[\ \mathrm{DV}_5(a_1, a_2) - \mathrm{DV}_6(a_1, a_2)]. \qquad (6.13)$$

The design $\tau_E \in \mathrm{T}_{23}$ with the maximum value of λ_2 in (6.13) is *E*-optimal.

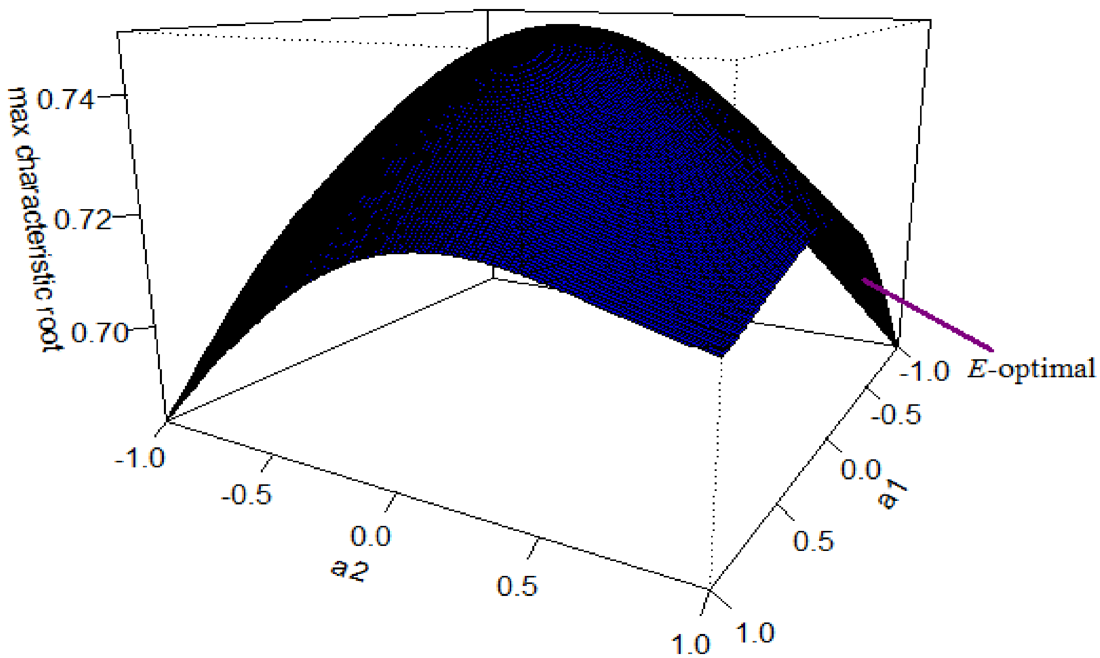
Property 7. For $-1 < a_1 \le a_2 < 1$, $\max\{\mathrm{DV}_5(a_1, a_2)\} \to \frac{44}{15}$ as $a_1^2 \to 1$ and $a_2^2 \to 1$.

Proof. For $\mathrm{DV}_5(a_1) = \frac{5a_1^2+17}{12+3a_1^2}$, set $\frac{\partial \mathrm{DV}_5(a_1)}{\partial a_1} = \frac{18a_1}{\left(12+3a_1^2\right)^2} = 0$, we have $a_1 = 0$. When $a_1 \in (-1,\ 0)$, $\frac{\partial \mathrm{DV}_5(a_1)}{\partial a_1}$ is negative and when $a_1 \in (0,\ 1)$, $\frac{\partial \mathrm{DV}_5(a_1)}{\partial a_1}$ is positive. Since $\frac{\partial \mathrm{DV}_5(a_1)}{\partial a_1}$ changes from negative to positive at 0, $\min\{\mathrm{DV}_5(a_1)\} = \frac{17}{12}$ at $a_1 = 0$, and $\max\{\mathrm{DV}_5(a_1)\} \to \frac{22}{15}$ when $a_1^2 \to 1$. Equivalently, we can prove $\max\{\mathrm{DV}_5(a_2)\} \to \frac{22}{15}$ as $a_2^2 \to 1$. The rest is clear.

Property 8. For $-1 < a_1 \le a_2 < 1$, $\min\{\mathrm{DV}_6(a_1, a_2)\} \to 0$, as $a_1 \to -1$ and $a_2 \to 1$.

Proof. Based on (6.12), when $a_1 = -a_2$, $\mathrm{DV}_6(a_1, a_2)$ is simplified to $\frac{1-a_1^2}{12+3a_1^2}+\frac{1-a_2^2}{12+3a_2^2} >$ 0. Clearly, $\lim_{a_1 \to -1, a_1 \to 1} \mathrm{DV}_6(a_1, a_2) = 0$ and the rest is clear.

Figure 6.4: The plot of max characteristic root ($1/\lambda_2$)
for $-1 < a_1 < 1$ and $-1 < a_2 < 1$



In Figure 6.4, we observe that the value of $1/\lambda_2$ becomes smaller as a_1 approaches -1 and a_2 approaches 1. To compute *E*-optimal cohort designs analytically, we apply Property 7 and 8 in determining the time points (a_1, a_2) that lead to the maximum value of λ_2 in (6.13). For $-1 < a_1 \le a_2 < 1$, we notice that (a_1, a_2) that maximize $DV_5(a_1, a_2)$ and minimize $DV_6(a_1, a_2)$ concurrently are: $a_1 = -\sqrt{a_0}$ and $a_2 = \sqrt{a_0}$, where a_0 ($0 < a_0 < 1$) is the limit on a_i (i = 1, 2) such that $a_i^2 \le a_0$. As a result, the *E*-optimal cohort design in design class T_{23} is $\tau_E = \begin{Bmatrix} (-1, -\sqrt{a_0}, 1) & (-1, \sqrt{a_0} 1) \\ 0.5 & 0.5 \end{Bmatrix}$, given $\mathbf{D} = \mathbf{I}$, $\boldsymbol{R}_i = \mathbf{I}$.

6.2.4 Combination

In this section, we would like to consider the situations where *D*-, *A*- and *E*-optimality criteria are used at the same time in identifying an optimal cohort design in design class T_{23} with $\mathbf{D} = \mathbf{I}$, $\boldsymbol{R}_i = \mathbf{I}$.

The DV(a_1, a_2) in (6.2) has two eigenvalues, namely

$$\lambda_1 = \frac{(\alpha+\beta)+\sqrt{(\alpha-\beta)^2+4\gamma^2}}{2} \text{ and } \lambda_2 = \frac{(\alpha+\beta)-\sqrt{(\alpha-\beta)^2+4\gamma^2}}{2},$$

where α, β, and γ are defined in (6.9).

Instead of deriving optimal designs with respect to determinant, maximum characteristic root, and trace of the covariance matrix separately, we can find an optimal cohort design which takes into account of these three criteria simultaneously by considering the next three conditions:

Condition 1: $\max\{\lambda_1\lambda_2\} = \max\{\alpha\beta-\gamma^2\}$

Condition 2: $\min\left\{\frac{1}{\lambda_1}+\frac{1}{\lambda_2}\right\} = \min\left\{\frac{\alpha+\beta}{\alpha\beta-\gamma^2}\right\}$,

Condition 3: $\min\{\max(\lambda_1, \lambda_2)\} = \min\left\{\frac{(\alpha+\beta)+\sqrt{(\alpha-\beta)^2+4\gamma^2}}{2}\right\}$.

These three conditions can be satisfied simultaneously if

$$\gamma^2 = 0, \text{ and } \alpha = \beta. \qquad (6.14)$$

Substituting $\gamma^2 = 0$ and $\beta = \alpha$ into *Condition* 1 to 3, we get max{ $\lambda_1\lambda_2$ } = 2α, min $\left\{\frac{1}{\lambda_1}+\frac{1}{\lambda_2}\right\} = \frac{2}{\alpha}$, and min{max(λ_1, λ_2)} = α .

Property 9. For $-1 < a_1 \le a_2 < 1$, if $\gamma^2 = 0$, then $a_1 = -a_2$.

Proof. Clear from (6.9).

Property 10. For $-1 < a_1 \le a_2 < 1$, $\lim_{a_1^2 \to 1, a_2^2 \to 1} (\alpha - \beta) = 0$.

Proof. Clear from (6.9).

Our objective is to identify a cohort design that has a maximum precision for the parameter estimators in the model (2.3) with $\mathbf{D} = \mathbf{I}$, $\boldsymbol{R}_i = \mathbf{I}$. We set a threshold on a_i (i = 1, 2) such that $a_i^2 \le a_0$ ($0 < a_0 < 1$). According to Property 10, we have $\alpha \cong \beta$, when $a_i^2 = a_0$. By Property 9, $\gamma^2 = 0$ when $a_1 = -\sqrt{a_0}$ and $a_2 = \sqrt{a_0}$.

Therefore, under *D*-, *A*-, and *E*-optimality criteria, the optimal design for T_{23} = $\left\{ \begin{matrix} (-1, a_1, 1) & (-1, a_2, 1) \\ 0.5 & 0.5 \end{matrix} \middle| -1 < a_1 \le a_2 < 1 \right\}$ is

$$\tau^* = \left\{ \begin{matrix} (-1, -\sqrt{a_0}, 1) & (-1, \sqrt{a_0} 1) \\ 0.5 & 0.5 \end{matrix} \right\}. \qquad (6.15)$$

In other words, for designs of two equally sized cohorts with three time points, given $\mathbf{D} = \mathbf{I}$, $\boldsymbol{R}_{\mathrm{i}} = \mathbf{I}$, the optimal design is selected by placing the 2nd time point close to the beginnig of the study period for the 1st cohort, and placing the 2nd time point close to the end of the study period for the 2nd cohort.

In fact, the optimal design τ^* (6.15) is a special case of the designs with non-equidistant time points discussed in Chapter 3. When $\mathbf{D} = \mathbf{I}$, $d_{12} = 0$, by Theorems 3.1 and 3.4 we know $\mathrm{D}_f^{(1)} < 0$, and thus given $\mathbf{D} = \mathbf{I}$, $\boldsymbol{R}_{\mathrm{i}} = \mathbf{I}$, the D-optimal design for T_{23} is the design with non-equidistant time points.

Chapter 7
Restricted Maximum Likelihood Estimation of Variance Components

7.1 Introduction

In a linear mixed effects model, there are two types of parameters: the fixed-effect parameters $\boldsymbol{\beta}$, and the variance component parameters in $\mathbf{D}$, $\boldsymbol{R}_i$ (i =1, 2), and σ^2. Maximum likelihood (ML) and restricted maximum likelihood (REML) estimation are methods commonly used to estimate these parameters.

The maximum likelihood (ML) estimation is a method of obtaining estimates of unknown parameters by optimizing a likelihood function. The maximum likelihood estimates of the parameters are the values of the parameters that maximize the likelihood function. The restricted maximum likelihood (REML) estimates the variance parameters by maximizing the likelihood function of the transformed observations $\mathbf{A}\underline{\boldsymbol{y}}$, where **A** is a (6*n* − 2) × 6n matrix such that $\mathrm{E}(\mathbf{A}\underline{\boldsymbol{y}}) = 0$.

In this chapter, we first discuss the theory of REML estimation of variance components in linear mixed effects models. We then introduce our method of finding the REML estimate using three criterion functions: function Δ, the log-likelihood l^*, and function P. We also present three numerical examples to illustrate our proposed method.

7.2 Restricted Maximum Likelihood Estimation

We have discussed optimal cohort designs in previous chapters and demonstrated that such optimal designs depend on the variance component parameters in **D**, $\boldsymbol{R}_i$ (i =1, 2), and σ^2 . To estimate these unknown variance components, the restricted maximum likelihood (REML) estimation procedure is now illustrated for the mixed effects model below.

7.2.1 Model and design class

We consider the linear mixed effects model described in (2.3)

$$\boldsymbol{y}_{j(i)} = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_{j(i)} + \boldsymbol{\varepsilon}_{j(i)},$$

where $\boldsymbol{y}_{j(i)} = (y_{1j(i)}, y_{2j(i)}, y_{3j(i)})'$ is the vector of repeated measurements for subject j in the i^{th} cohort (i = 1, 2, j = 1,…,n), $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ are fixed unknown parameters, $\boldsymbol{b}_{j(i)}$ is the vector of random regression coefficients of subject j within cohort i, with mean **0** and a covariance matrix $\sigma^2 \mathbf{D} = \sigma^2 \begin{bmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{bmatrix}$. The 3 × 1 random error vector $\boldsymbol{\varepsilon}_{j(i)}$ has mean **0** and a covariance matrix $\sigma^2 \boldsymbol{R}_i$, where σ^2 is the common variance for error components and for simplicity we assume $\boldsymbol{R}_i = \mathbf{I}$ in this chapter.

We consider the design class $\mathrm{T}_{23(a)}$ which is given by

$$\mathrm{T}_{23(a)} = \left\{ \begin{matrix} (-1, a, 1) & (-1, a, 1) \\ 0.5 & 0.5 \end{matrix} \middle| -1 < a < 1 \right\}.$$

We have

$$\mathbf{X}_i = \boldsymbol{Z}_i = \begin{bmatrix} 1 & -1 \\ 1 & a \\ 1 & 1 \end{bmatrix}. \qquad (7.1)$$

As we mentioned in Chapter 2, we denote the vector of all observations $\underline{\boldsymbol{y}} = (\boldsymbol{y}'_{1(1)}, ..., \boldsymbol{y}'_{n(1)}, \boldsymbol{y}'_{1(2)}, ..., \boldsymbol{y}'_{n(2)})'$. Its expectation and covariance are

$$\mathrm{E}(\underline{\boldsymbol{y}}) = \begin{bmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \mathbf{X}\boldsymbol{\beta}, \qquad \mathrm{Var}(\underline{\boldsymbol{y}}) = \sigma^2 \left[\begin{array}{ccc|ccc} \mathbf{V}_1 & & 0 & 0 & \cdots & 0 \\ & \ddots & & \vdots & \ddots & \vdots \\ 0 & & \mathbf{V}_1 & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & \mathbf{V}_2 & & 0 \\ \vdots & \ddots & \vdots & & \ddots & \\ 0 & \cdots & 0 & 0 & & \mathbf{V}_2 \end{array} \right] = \sigma^2 \mathbf{V}, \qquad (7.2)$$

where $\mathbf{V}_i = \boldsymbol{Z}_i \mathbf{D} \boldsymbol{Z}'_i + \mathbf{I}$ (i = 1, 2).

7.2.2 REML estimators

As we discussed earlier, the restricted maximum likelihood (REML) estimators for the variance components are obtained by maximizing the likelihood function of the transformed observations $\mathbf{A}\underline{\boldsymbol{y}}$, where $\mathbf{A}$ is a $(6n-2) \times 6\text{n}$ matrix such that

$$\text{rank}(\mathbf{A}) = 6n - 2, \quad \text{E}(\mathbf{A}\underline{\boldsymbol{y}}) = 0. \qquad (7.3)$$

A possible matrix **A** that satisfy the above condition is

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & -1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 & 0 & -1 \\ a-1 & 2 & -(a+1) & a-1 & 2 & -(a+1) & \cdots & -(a+1) & a-1 & 2 & -(a+1) \end{pmatrix}$$
$$= \begin{pmatrix} \mathbf{I}_3 & -\mathbf{I}_3 & \underset{3\times3}{\mathbf{0}} & \cdots & \underset{3\times3}{\mathbf{0}} & \underset{3\times3}{\mathbf{0}} \\ \underset{3\times3}{\mathbf{0}} & \mathbf{I}_3 & -\mathbf{I}_3 & \cdots & \underset{3\times3}{\mathbf{0}} & \underset{3\times3}{\mathbf{0}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \underset{3\times3}{\mathbf{0}} & \underset{3\times3}{\mathbf{0}} & \underset{3\times3}{\mathbf{0}} & \cdots & \mathbf{I}_3 & -\mathbf{I}_3 \\ \underline{\boldsymbol{a}}_1' & \underline{\boldsymbol{a}}_1' & \underline{\boldsymbol{a}}_1' & \cdots & \underline{\boldsymbol{a}}_1' & \underline{\boldsymbol{a}}_1' \end{pmatrix},$$

where $\underline{\boldsymbol{a}}_1' = \left(a-1 \quad 2 \quad -(a+1)\right)$.

It should be pointed out that the choice of **A** is not unique. An alternative transforming matrix may be obtained as **B** = **TA**, where **T** is any $(6n - 2) \times (6n - 2)$ matrix with rank(**T**) $= 6n - 2$. Moreover, the REML estimators, in fact, do not depend on the choice of **A**.

We use the vector $\boldsymbol{\theta}$ to denote all parameters contained in matrix **D**. In this case,

$$\boldsymbol{\theta} = \begin{bmatrix} d_{11} \\ d_{12} \\ d_{22} \end{bmatrix}. \qquad (7.4)$$

We write $\underline{\boldsymbol{t}} = \mathbf{A}\underline{\boldsymbol{y}}$. The $\underline{\boldsymbol{t}}$ is distributed $\mathrm{N}\left(\mathbf{0},\ \sigma^2\mathbf{AVA'}\right)$ with mean known and thus free of the fixed effects. The likelihood function is given by

$$\begin{aligned} L(\sigma^2, \boldsymbol{\theta}) &= \frac{1}{(2\pi)^{3n-1} \,|\, \sigma^2\mathbf{AVA'} \,|^{1/2}} \exp\left\{-\frac{1}{2}\underline{\boldsymbol{t}}'(\sigma^2\mathbf{AVA'})^{-1}\underline{\boldsymbol{t}}\right\} \\ &= (2\pi)^{1-3n}(\sigma^2)^{1-3n} \,|\, \mathbf{AVA'} \,|^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}\underline{\boldsymbol{t}}'(\mathbf{AVA'})^{-1}\underline{\boldsymbol{t}}\right\}. \end{aligned}$$

The corresponding log-likelihood function expressed in terms of $\underline{\boldsymbol{y}}$ is given by

$$l(\sigma^2, \boldsymbol{\theta}) = (1-3n)\log 2\pi + (1-3n)\log\sigma^2 - \frac{1}{2}\log|\,\mathbf{AVA'}\,| - \frac{1}{2\sigma^2}\underline{\boldsymbol{y}}'\mathbf{A}'(\mathbf{AVA'})^{-1}\mathbf{A}\underline{\boldsymbol{y}}.$$

After eliminating terms not involving σ^2 and $\boldsymbol{\theta}$, we have

$$l^* = (1-3n)\log\sigma^2 - \frac{1}{2}\log|\,\mathbf{AVA'}\,| - \frac{1}{2\sigma^2}\underline{\boldsymbol{y}}'\mathbf{A}'(\mathbf{AVA'})^{-1}\mathbf{A}\underline{\boldsymbol{y}}. \qquad (7.5)$$

As we mentioned, the REML estimators are obtained by maximizing the log-likelihood function l^* in (7.5). We need the estimation equations for maximizing l^*. The estimation equations can be found by taking the partial derivative of l^* with respect to unknown variance parameters: σ^2 and $\boldsymbol{\theta} = (d_{11}, d_{12}, d_{22})'$.

Differentiating the log-likelihood l^*, we obtain

$$\frac{\partial l^*}{\partial \sigma^2} = \frac{1-3n}{\sigma^2} + \frac{1}{2\sigma^4}\underline{\boldsymbol{y}}'\mathbf{A}'(\mathbf{AVA}')^{-1}\mathbf{A}\underline{\boldsymbol{y}},$$

$$\frac{\partial l^*}{\partial \boldsymbol{\theta}_i} = -\frac{1}{2}\operatorname{tr}\left[\mathbf{A}'(\mathbf{AVA}')^{-1}\mathbf{A}\frac{\partial \mathbf{V}}{\partial \boldsymbol{\theta}_i}\right] + \frac{1}{2\sigma^2}\underline{\boldsymbol{y}}'\mathbf{A}'(\mathbf{AVA}')^{-1}\mathbf{A}\frac{\partial \mathbf{V}}{\partial \boldsymbol{\theta}_i}\mathbf{A}'(\mathbf{AVA}')^{-1}\mathbf{A}\underline{\boldsymbol{y}}, \qquad (7.6)$$

where $\boldsymbol{\theta}_i$ (i = 1, 2, 3) is the i^{th} component of vector $\boldsymbol{\theta}$ in (7.4).

Equating (7.6) to zero gives the REML estimators:

$$\hat{\sigma}^2 = \frac{\underline{\boldsymbol{y}}'\mathbf{A}'(\mathbf{AVA}')^{-1}\mathbf{A}\underline{\boldsymbol{y}}}{6n-2}, \qquad (7.7)$$

$$\operatorname{tr}\left[\mathbf{A}'(\mathbf{AVA}')^{-1}\mathbf{A}\frac{\partial \mathbf{V}}{\partial \boldsymbol{\theta}_i}\right] = \frac{1}{\sigma^2}\underline{\boldsymbol{y}}'\mathbf{A}'(\mathbf{AVA}')^{-1}\mathbf{A}\frac{\partial \mathbf{V}}{\partial \boldsymbol{\theta}_i}\mathbf{A}'(\mathbf{AVA}')^{-1}\mathbf{A}\underline{\boldsymbol{y}}. \qquad (7.8)$$

The REML estimators are defined as the maximizers of the log-likelihood function l^* in (7.5). Clearly, the REML equation (7.8) has no closed form solutions and has to be solved numerically. The most common iterative algorithms used for solving nonlinear equations are the expectation-maximization (EM) algorithm, the Newton-Raphson (N-R) algorithm, and the Fisher scoring algorithm.

The EM algorithm is an iterative method which alternates between performing an expectation (E) step, which computes the expectation of the log-likelihood evaluated using the current estimate for the parameters, and a maximization (M) step, which computes parameters maximizing the expected log-likelihood found on the E step. These parameter-estimates are then used to determine the distribution of the latent variables in the next E step. General descriptions of the EM algorithm can be found in Dempster et al. (1977) and Laird et al. (1987). EM algorithm is often used to maximize complicated likelihood functions or to find good starting values of the parameters. The EM approach is currently used by the procedures in software R and Stata.

The Newton-Raphson (N-R) algorithm minimizes an objective function defined as −2 times the log-likelihood function l^* in (7.5). It does so by computing the Jacobian linearization of the objective function around an initial guess point, and using this linearization to move closer to the nearest zero. Analytical formulas for N-R algorithm are given in Lindstrom and Bates (1988). The N-R algorithm and its variations are the most commonly used algorithms in REML estimation of linear mixed effects models. In fact, the N-R algorithm is the iterative method used by SAS PROC MIXED procedure.

The Fisher scoring algorithm can be considered as a modification of the N-R algorithm. The only difference is that Fisher scoring uses the expected Hessian matrix (square matrix of second-order partial derivatives of l^* with respect to unknown variance parameters), whereas the N-R algorithm uses the observed one. The softwares SAS and SPSS use the Fisher scoring.

7.2.3 Our iterative method

We now present our iterative method for i = 1, 2, 3.

From (7.8), we write

$$L_{(i)} = \operatorname{tr}\left[\mathbf{A}'(\mathbf{AVA}')^{-1}\mathbf{A}\frac{\partial \mathbf{V}}{\partial \boldsymbol{\theta}_i}\right], \quad R_{(i)} = \frac{1}{\sigma^2}\underline{\boldsymbol{y}}'\mathbf{A}'(\mathbf{AVA}')^{-1}\mathbf{A}\frac{\partial \mathbf{V}}{\partial \boldsymbol{\theta}_i}\mathbf{A}'(\mathbf{AVA}')^{-1}\mathbf{A}\underline{\boldsymbol{y}}.$$

In addition to the log-likelihood function l^* in (7.5) as a criterion function, we define other two criterion functions for finding the accurate final solution of REML estimators based on the above concept. The criterion functions are

$$\Delta = |L_{(1)} - R_{(1)}| + |L_{(2)} - R_{(2)}| + |L_{(3)} - R_{(3)}|, \qquad (7.9)$$

$$\mathrm{P} = |\sigma^2 - \hat{\sigma}^2| + |d_{22} - \hat{d}_{22}| + |d_{11} - \hat{d}_{11}| + |d_{12} - \hat{d}_{12}|. \qquad (7.10)$$

We want to maximize the log-likelihood function l^* while minimize the criterion functions Δ and P.

Our iterative method requires the following steps:

(i) Assign initial values to the variance parameters d_{11}, d_{22}, and d_{12}.

(ii) Estimate σ^2 by solving (7.7).

(iii) Use the initial values of d_{11}, d_{22}, and d_{12} and updated $\hat{\sigma}^2$ value from step (ii) to calculate new estimates: $\hat{d}_{11}$, $\hat{d}_{22}$, and $\hat{d}_{12}$ that make $|L_{(i)} - R_{(i)}|$ (i = 1, 2, 3) close to zero.

(iv) Repetition of (ii) and (iii), ending at (ii), is continued until Δ is small, l^* is large, and two consecutive Δ values as well as two consecutive l^* values simultaneously become less than or equal to 10^{-6}, that is,

$$|\Delta_s - \Delta_{s-1}| \le 10^{-6} \text{ and } |l_s^* - l_{s-1}^*| \le 10^{-6},$$

where Δ_s and Δ_{s-1} are the values of criterion function Δ in (7.9) at the s and (s − 1) stages of iteration respectively. Similarly, l_s^* and l_s^* are the values of l^* in (7.5) at the s and (s − 1) stages of iteration respectively.

We obtain the most accurate final solution of REML estimators when the numerical value of Δ is minimum, the numerical value of l^* in (7.5) is maximum, and the numerical value of P is relatively small. The smaller numerical value of Δ, or the larger value of l^*, or the smaller numerical value of P achieves the better accuracy on the final solutions of REML estimators. The criterion function P is proposed in (7.10) because we want our REML estimators be close to the true variance parameter values.

7.3 Numerical Examples

We now illustrate our proposed iterative method in details.

7.3.1 Estimation of variance components for *n* = 1

First, we consider the situation that each cohort consists of a single subject. For the three observations on the subject in the i^{th} (i = 1, 2) cohort, the model is

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i,$$

which is a special case of the model (2.3) with j = 1.

For design class $T_{23(a)}$, we have

$$\underline{\boldsymbol{y}} = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} = \begin{bmatrix} y_{1(1)} \\ y_{2(1)} \\ y_{3(1)} \\ y_{1(2)} \\ y_{2(2)} \\ y_{3(2)} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & a \\ 1 & 1 \\ 1 & -1 \\ 1 & a \\ 1 & 1 \end{bmatrix}, \mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & 0 \\ 0 & \mathbf{V}_1 \end{bmatrix} \text{ where } \mathbf{V}_1 = \boldsymbol{X}_1\mathbf{D}\boldsymbol{X}_1' + \mathbf{I}. \qquad (7.11)$$

As we have discussed earlier, to estimate the variance components without dealing with the fixed effects, we apply a transformation to the data through a matrix **A** that satisfies the condition in (7.3) with $n = 1$. A feasible matrix **A** in this case is

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \\ a-1 & 2 & -(a+1) & a-1 & 2 & -(a+1) \end{bmatrix} = \begin{bmatrix} \mathbf{I} & -\mathbf{I} \\ \underline{\boldsymbol{a}}_1' & \underline{\boldsymbol{a}}_1' \end{bmatrix},$$

where $\underline{\boldsymbol{a}}_1' = \left(a-1 \quad 2 \quad -(a+1)\right)$.

We denote

$$\begin{aligned}
&\underline{\boldsymbol{j}}' = \left(1 \quad 1 \quad 1\right), \\
&\underline{\boldsymbol{a}}_2' = \left(-1 \quad a \quad 1\right), \\
&\mathrm{K} = \left(\begin{array}{c} \left((2+a^2)(d_{11}d_{22} - d_{12}^2) + d_{11}\right)\left(\underline{\boldsymbol{j}}'(\boldsymbol{y}_1 - \boldsymbol{y}_2)\right)^2 + \left(3(d_{11}d_{22} - d_{12}^2) + d_{22}\right)\left(\underline{\boldsymbol{a}}_2'(\boldsymbol{y}_1 - \boldsymbol{y}_2)\right)^2 \\ + 2\left(-a(d_{11}d_{22} - d_{12}^2) + d_{12}\right)\left(\underline{\boldsymbol{j}}'(\boldsymbol{y}_1 - \boldsymbol{y}_2)\right)\left(\underline{\boldsymbol{a}}_2'(\boldsymbol{y}_1 - \boldsymbol{y}_2)\right) \end{array} \right), \\
&\mathrm{T} = (\boldsymbol{y}_1 - \boldsymbol{y}_2)'(\boldsymbol{y}_1 - \boldsymbol{y}_2) + \frac{\left(\underline{\boldsymbol{a}}_1'(\boldsymbol{y}_1 + \boldsymbol{y}_2)\right)^2}{\underline{\boldsymbol{a}}_1'\underline{\boldsymbol{a}}_1}. \qquad (7.12)
\end{aligned}$$

It can be checked that the log-likelihood function l^* ignoring the part not dependent on the variance parameters is

$$l^* = -2\log\sigma^2 - \frac{\mathrm{T}}{4\sigma^2} - \frac{1}{2}\log\left|\mathbf{V}_1\right| + \frac{\mathrm{K}}{4\sigma^2\left|\mathbf{V}_1\right|}. \qquad (7.13)$$

Differentiating l^* in (7.13) with respect to σ^2 and equating it to zero gives

$$\hat{\sigma}^2 = \frac{1}{8}\left(\mathrm{T} - \frac{\hat{\mathrm{K}}}{\left|\hat{\mathbf{V}}_1\right|} \right).$$

The REML estimation functions regarding the variance parameters d_{11}, d_{22}, and d_{12} respectively are

$$\begin{bmatrix} \left((2+a^2)\hat{d}_{22}+1\right)\left(\underline{\boldsymbol{j}}'(\boldsymbol{y}_1-\boldsymbol{y}_2)\right)^2 + 3\hat{d}_{22}\left(\underline{\boldsymbol{a}}_2'(\boldsymbol{y}_1-\boldsymbol{y}_2)\right)^2 \\ -2a\hat{d}_{22}\left(\underline{\boldsymbol{j}}'(\boldsymbol{y}_1-\boldsymbol{y}_2)\right)\left(\underline{\boldsymbol{a}}_2'(\boldsymbol{y}_1-\boldsymbol{y}_2)\right) \end{bmatrix} = \left[3+(6+2a^2)\hat{d}_{22}\right]\left(-6\hat{\sigma}^2+\mathrm{T}\right),$$

$$\begin{bmatrix} (2+a^2)\hat{d}_{11}\left(\underline{\boldsymbol{j}}'(\boldsymbol{y}_1-\boldsymbol{y}_2)\right)^2 + (3\hat{d}_{11}+1)\left(\underline{\boldsymbol{a}}_2'(\boldsymbol{y}_1-\boldsymbol{y}_2)\right)^2 \\ -2a\hat{d}_{11}\left(\underline{\boldsymbol{j}}'(\boldsymbol{y}_1-\boldsymbol{y}_2)\right)\left(\underline{\boldsymbol{a}}_2'(\boldsymbol{y}_1-\boldsymbol{y}_2)\right) \end{bmatrix} = \left[(a^2+2)+(6+2a^2)\hat{d}_{11}\right]\left(-6\hat{\sigma}^2+\mathrm{T}\right),$$

$$\begin{bmatrix} -(2+a^2)\hat{d}_{12}\left(\underline{\boldsymbol{j}}'(\boldsymbol{y}_1-\boldsymbol{y}_2)\right)^2 - 3\hat{d}_{12}\left(\underline{\boldsymbol{a}}_2'(\boldsymbol{y}_1-\boldsymbol{y}_2)\right)^2 \\ +(2a\hat{d}_{12}+1)\left(\underline{\boldsymbol{j}}'(\boldsymbol{y}_1-\boldsymbol{y}_2)\right)\left(\underline{\boldsymbol{a}}_2'(\boldsymbol{y}_1-\boldsymbol{y}_2)\right) \end{bmatrix} = \left[a-(6+2a^2)\hat{d}_{12}\right]\left(-6\hat{\sigma}^2+\mathrm{T}\right). \quad (7.14)$$

The exact expression for solutions of (7.14) does not exist since the equations are non-linear in variables. We denote the left-hand side of equations in (7.14) as $L_{(1)}$, $L_{(2)}$, and $L_{(3)}$, respectively, and expressions on the right-hand side of (7.14) are $R_{(1)}$, $R_{(2)}$, $R_{(3)}$, respectively.

To further demonstrate our iterative method, we simulate a data set with n = 1 subject per cohort assuming the following parameter values:

$$\sigma^2 = 0.0005,\ \mathbf{D} = \begin{bmatrix} d_{11}=0.05 & d_{12}=0.0005 \\ d_{12}=0.0005 & d_{22}=0.10 \end{bmatrix},\ \boldsymbol{R}_i = \mathbf{I},\ \beta_0 = 1, \text{ and } \beta_1 = 5. \quad (7.15)$$

Table 7.1: The simulated data with $n = 1$

Cohort (i)	Subject (j)	Observations		
1	1	−3.9925	1.0388	5.9497
2	1	−3.9826	1.0250	6.0084

The simulated data set is shown in Table 7.1. We use the Optimization Tool in MATLAB with ‘fmincon’ function and ‘active set’ algorithm to perform our iterative procedure. We present the numerical values of REML estimators obtained by our method as well as by SAS in Table 7.2.

Table 7.2: Solution of REML estimators with values of Δ, l^*, and P for $n = 1$

Parameter	σ^2	d_{11}	d_{22}	d_{12}	Δ	l^*	P
True	0.0005	0.05	0.10	0.0005	20.2157	11.5953	-
Our Estimate	0.0008	0.0465	0.1120	0.0560	0.0097	14.6862	0.0713
SAS Estimate	0.0008	0.0465	0.0835	0.0621	0.0098	14.6848	0.0819

We can see that our estimates of variance parameters are fairly close to the true values. The REML estimates of σ^2 and d_{11} of our method are identical up to at least four decimal places compared with SAS results. In addition, our REML estimates lead to smaller numerical value of Δ, larger numerical value of l^*, and smaller numerical value of P. Therefore, our iterative method is showing favorable results for this simulated data.

7.3.2 Estimation of variance components for *n* = 2

In this section, demonstration of our iterative procedure is given in terms of a simulated data set with *n* = 2 subjects per cohort shown in Table 7.3. This data set is generated using the parameter values in (7.15) under the cohort design τ from the design class $T_{23(a)}$ with *a* = 0:

$$\tau = \begin{Bmatrix} (-1,0,1) & (-1,0,1) \\ 0.5 & 0.5 \end{Bmatrix}.$$

Notice that the data in Table 7.1 are used as the observations for the first subject (*j* =1) in the i^{th} (*i* = 1, 2) cohort.

Table 7.3: The simulated data with *n* = 2

Cohort (*i*)	Subject (*j*)	Observations		
1	1	−3.9925	1.0388	5.9497
1	2	−4.0413	1.0672	6.0261
2	1	−3.9826	1.0250	6.0084
2	2	−4.0045	1.0230	6.0126

For the three measurements taken on the j^{th} subject in cohort *i*, the assumed model is

$$\boldsymbol{y}_{j(i)} = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_{j(i)} + \boldsymbol{\varepsilon}_{j(i)}, \qquad i = 1,2,\ j = 1,2,$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ are the fixed parameters, $\boldsymbol{b}_{j(i)} = (b_{0j(i)}, b_{1j(i)})'$ are random parameters, $\boldsymbol{\varepsilon}_{j(i)} = (\varepsilon_{1j(i)}, \varepsilon_{2j(i)}, \varepsilon_{3j(i)})'$ is the random error vector with mean **0** and the covariance $\sigma^2\mathbf{I}$, $\boldsymbol{b}_{j(i)}$ and $\boldsymbol{\varepsilon}_{j(i)}$ are independent. The design matrices $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are given by

$$\boldsymbol{X}_i = \boldsymbol{Z}_i = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

The vector of all observations is denoted by $\underline{\boldsymbol{y}} = (\boldsymbol{y}_{1(1)}, \boldsymbol{y}_{2(1)}, \boldsymbol{y}_{1(2)}, \boldsymbol{y}_{2(2)})'$ with its covariance $\text{Var}(\underline{\boldsymbol{y}}) = \sigma^2\mathbf{V} = \sigma^2 diag(\mathbf{V}_1)$ where

$$\mathbf{V}_1 = \boldsymbol{X}_1\mathbf{D}\boldsymbol{X}_1' + \mathbf{I} = \begin{bmatrix} d_{11} - 2d_{12} + d_{22} + 1 & d_{11} - d_{12} & d_{11} - d_{22} \\ d_{11} - d_{12} & d_{11} + 1 & d_{11} + d_{12} \\ d_{11} - d_{22} & d_{11} + d_{12} & d_{11} + 2d_{12} + d_{22} + 1 \end{bmatrix}. \qquad (7.16)$$

For finding the REML estimators, matrix **A** that satisfies the condition in (7.3) with n = 2 is given by

$$\underset{10\times 12}{\mathbf{A}} = \begin{bmatrix} \mathbf{I}_3 & -\mathbf{I}_3 & \underset{3\times 3}{\mathbf{0}} & \underset{3\times 3}{\mathbf{0}} \\ \underset{3\times 3}{\mathbf{0}} & \mathbf{I}_3 & -\mathbf{I}_3 & \underset{3\times 3}{\mathbf{0}} \\ \underset{3\times 3}{\mathbf{0}} & \underset{3\times 3}{\mathbf{0}} & \mathbf{I}_3 & -\mathbf{I}_3 \\ \underline{\boldsymbol{a}}_1' & \underline{\boldsymbol{a}}_1' & \underline{\boldsymbol{a}}_1' & \underline{\boldsymbol{a}}_1' \end{bmatrix},$$

where $\underline{\boldsymbol{a}}_1' = \begin{pmatrix} -1 & 2 & -1 \end{pmatrix}$.

Define the transformed data as $\underline{\boldsymbol{y}}^* = \mathbf{A}\underline{\boldsymbol{y}} = (\underline{\boldsymbol{y}}_1^*, \underline{\boldsymbol{y}}_2^*, \underline{\boldsymbol{y}}_3^*, \underline{\boldsymbol{y}}_4^*)'$, then the l^* can be rewritten as

$$l^* = -5\log\sigma^2 - \frac{3}{2}\log|\mathbf{V}_1| - \frac{\text{E}}{2\sigma^2}, \qquad (7.17)$$

where

$$\text{E} = \frac{\underline{\boldsymbol{y}}_4^{*2}}{24} + \frac{3}{4}\underline{\boldsymbol{y}}_1^{*\prime}\mathbf{V}_1^{-1}\underline{\boldsymbol{y}}_1^* + \underline{\boldsymbol{y}}_2^{*\prime}\mathbf{V}_1^{-1}\underline{\boldsymbol{y}}_2^* + \frac{3}{4}\underline{\boldsymbol{y}}_3^{*\prime}\mathbf{V}_1^{-1}\underline{\boldsymbol{y}}_3^* + \underline{\boldsymbol{y}}_1^{*\prime}\mathbf{V}_1^{-1}\underline{\boldsymbol{y}}_2^* + \underline{\boldsymbol{y}}_2^{*\prime}\mathbf{V}_1^{-1}\underline{\boldsymbol{y}}_3^* + \frac{1}{2}\underline{\boldsymbol{y}}_1^{*\prime}\mathbf{V}_1^{-1}\underline{\boldsymbol{y}}_3^*,$$

$$|\mathbf{V}_1| = 3d_{11} + 2d_{22} - 6d_{12}^2 + 6d_{11}d_{22} + 1. \qquad (7.18)$$

Differentiating l^* in (7.17) with respect to σ^2 and equating it to zero, we have

$$\hat{\sigma}^2 = \frac{\mathrm{E}}{5}.$$

Now we take the partial derivative of l^* with respect to variance parameters: d_{11}, d_{22}, and d_{12} respectively and set those REML equations equal to zero, we obtain

$$\left(\frac{3}{4}\underline{\boldsymbol{y}}_1^{*\prime}\mathrm{W}\underline{\boldsymbol{y}}_1^* + \underline{\boldsymbol{y}}_2^{*\prime}\mathrm{W}\underline{\boldsymbol{y}}_2^* + \frac{3}{4}\underline{\boldsymbol{y}}_3^{*\prime}\mathrm{W}\underline{\boldsymbol{y}}_3^* + \underline{\boldsymbol{y}}_1^{*\prime}\mathrm{W}\underline{\boldsymbol{y}}_2^* + \underline{\boldsymbol{y}}_2^{*\prime}\mathrm{W}\underline{\boldsymbol{y}}_3^* + \frac{1}{2}\underline{\boldsymbol{y}}_1^{*\prime}\mathrm{W}\underline{\boldsymbol{y}}_3^*\right) = (9+18d_{22})\sigma^2\left|\mathbf{V}_1\right|,$$
$$\left(\frac{3}{4}\underline{\boldsymbol{y}}_1^{*\prime}\mathrm{G}\underline{\boldsymbol{y}}_1^* + \underline{\boldsymbol{y}}_2^{*\prime}\mathrm{G}\underline{\boldsymbol{y}}_2^* + \frac{3}{4}\underline{\boldsymbol{y}}_3^{*\prime}\mathrm{G}\underline{\boldsymbol{y}}_3^* + \underline{\boldsymbol{y}}_1^{*\prime}\mathrm{G}\underline{\boldsymbol{y}}_2^* + \underline{\boldsymbol{y}}_2^{*\prime}\mathrm{G}\underline{\boldsymbol{y}}_3^* + \frac{1}{2}\underline{\boldsymbol{y}}_1^{*\prime}\mathrm{G}\underline{\boldsymbol{y}}_3^*\right) = (6+18d_{11})\sigma^2\left|\mathbf{V}_1\right|,$$
$$\left(\frac{3}{4}\underline{\boldsymbol{y}}_1^{*\prime}\mathrm{F}\underline{\boldsymbol{y}}_1^* + \underline{\boldsymbol{y}}_2^{*\prime}\mathrm{F}\underline{\boldsymbol{y}}_2^* + \frac{3}{4}\underline{\boldsymbol{y}}_3^{*\prime}\mathrm{F}\underline{\boldsymbol{y}}_3^* + \underline{\boldsymbol{y}}_1^{*\prime}\mathrm{F}\underline{\boldsymbol{y}}_2^* + \underline{\boldsymbol{y}}_2^{*\prime}\mathrm{F}\underline{\boldsymbol{y}}_3^* + \frac{1}{2}\underline{\boldsymbol{y}}_1^{*\prime}\mathrm{F}\underline{\boldsymbol{y}}_3^*\right) = -36d_{12}\sigma^2\left|\mathbf{V}_1\right|, \qquad (7.19)$$

where

$$\mathrm{W} = \begin{bmatrix} w_1 & w_1 & w_1 \\ w_2 & w_2 & w_2 \\ w_3 & w_3 & w_3 \end{bmatrix},$$

$$w_1 = 6d_{12} + 4d_{22} + 6d_{12}^2 + 4d_{22}^2 + 9d_{11}d_{12} + 6d_{12}d_{22} + 1,$$
$$w_2 = 4d_{22} + 6d_{12}^2 + 4d_{22}^2 + 1,$$
$$w_3 = -6d_{12} + 4d_{22} + 6d_{12}^2 + 4d_{22}^2 - 9d_{11}d_{12} - 6d_{12}d_{22} + 1,$$

$$\mathrm{G} = \begin{bmatrix} g_{11} & 0 & -g_{11} \\ g_{21} & 0 & -g_{21} \\ g_{31} & 0 & -g_{31} \end{bmatrix}, \qquad (7.20)$$

$$g_{11} = 6d_{11} + 4d_{12} + 6d_{11}d_{12} + 4d_{12}d_{22} + 9d_{11}^2 + 6d_{12}^2 + 1,$$
$$g_{21} = 2d_{12}(3d_{11} + 2d_{22} + 2),$$
$$g_{31} = -6d_{11} + 4d_{12} + 6d_{11}d_{12} + 4d_{12}d_{22} - 9d_{11}^2 - 6d_{12}^2 - 1,$$

$$F = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}, \qquad (7.21)$$

$$\begin{aligned}
f_{11} &= -9d_{11}^2 - 15d_{11}d_{22} - 6d_{11} - 12d_{12}^2 - 10d_{12}d_{22} - 10d_{12} - 4d_{22}^2 - 4d_{22} - 2,\\
f_{12} &= -6d_{11} - 4d_{12} - 6d_{11}d_{12} - 4d_{12}d_{22} - 9d_{11}^2 - 6d_{12}^2 - 1,\\
f_{13} &= (d_{12} - 3d_{11} + 2d_{22})(3d_{11} + 2d_{22} + 2),\\
f_{21} &= -4d_{12} - 4d_{22} - 6d_{11}d_{12} - 4d_{12}d_{22} - 6d_{12}^2 - 4d_{22}^2 - 1,\\
f_{22} &= -2d_{12}(3d_{11} + 2d_{22} + 2),\\
f_{23} &= -4d_{12} + 4d_{22} - 6d_{11}d_{12} - 4d_{12}d_{22} + 6d_{12}^2 + 4d_{22}^2 + 1,\\
f_{31} &= (d_{12} + 3d_{11} - 2d_{22})(3d_{11} + 2d_{22} + 2),\\
f_{32} &= 6d_{11} - 4d_{12} - 6d_{11}d_{12} - 4d_{12}d_{22} + 9d_{11}^2 + 6d_{12}^2 + 1,\\
f_{33} &= 9d_{11}^2 - 15d_{11}d_{22} + 6d_{11} + 12d_{12}^2 - 10d_{12}d_{22} - 10d_{12} + 4d_{22}^2 + 4d_{22} + 2. \qquad (7.22)
\end{aligned}$$

The close form expressions of REML estimators are impossible to obtain. We have to compute such REML estimators numerically by iterative methods. Let $L_{(1)}$, $L_{(2)}$, and $L_{(3)}$ denote the left-hand side of equations in (7.19). Define the right-hand side of equations in (7.19) as $R_{(1)}$, $R_{(2)}$, and $R_{(3)}$, respectively. A criterion function Δ is defined as $\sum_{i=1}^{3} | L_{(i)} - R_{(i)} |$.

Table 7.4: Solution of REML estimators with values of Δ, l^*, and P for n = 2

Parameter	σ^2	d_{11}	d_{22}	d_{12}	Δ	l^*	P
True	0.0005	0.05	0.10	0.0005	0.0466	27.0389	--
Our Estimate	0.0009	0.0198	0.2580	0.0720	0.0446	29.0532	0.2600
SAS Estimate	0.0010	0.0198	0.2586	0.0720	0.0461	29.0636	0.2607

For the data presented in Table 7.3, using the criterion functions l^* in (7.17) and Δ in (7.9), we have obtained the REML estimates of the variance parameters shown in Table 7.4. The estimation procedures are performed using the Optimization Toolbox in MATLAB with ‘fmincon’ as the solver, ‘active set’ as the search algorithm. The estimates generated by SAS using PROC MIXED are also presented in Table 7.4.

It can be seen that, with P = 0.26, our estimates of variance components are relatively close to the true values. The REML estimates found by our method are identical up to at least three decimal places compared with SAS results. Furthermore, our REML estimates perform better than SAS estimates with respect to the criterion P. However, SAS estimates are slightly better than ours in terms of Δ and l^* criteria. In the next section, we will take one real data set to illustrate our described method.

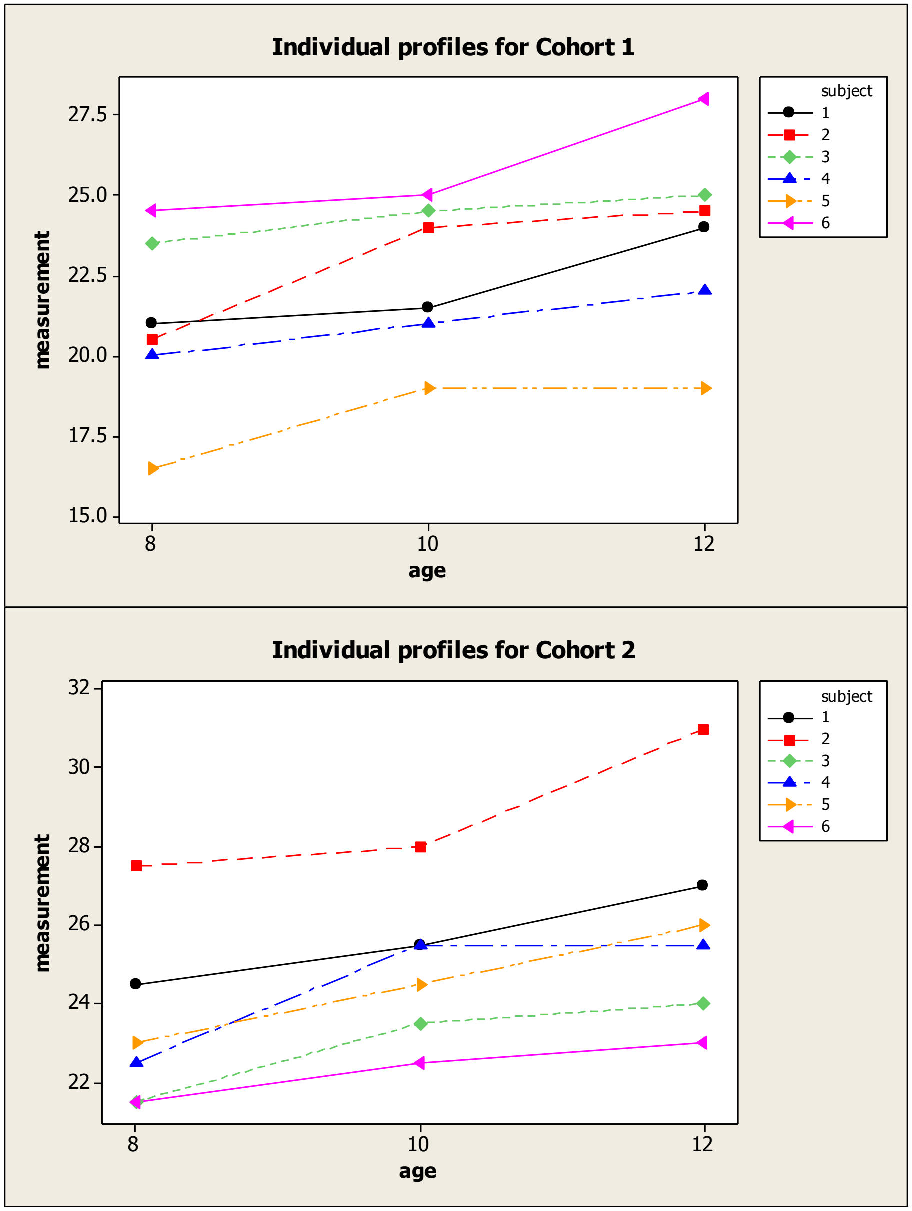
7.3.3 Estimation of variance components for a real data set

Consider, as an example, a longitudinal study on facial growth made on 11 girls and 16 boys (Potthoff and Roy, 1964 [20]). For each subject, the distance from the center of the pituitary to the maxillary fissure was recorded at age 8, 10, 12, and 14 years. Here, cohort is defined by gender. The 1st cohort consists of girls and the 2nd cohort consists of boys. For illustration purpose, we randomly select 6 girls and 6 boys. For any chosen subject, the measurements observed at age 8, 10, and 12 are used. The facial growth data are shown in Table 7.5.

Table 7.5: The facial growth data with n = 6

Cohort (i)	Subject (j)	Observations		
1	1	21.0	21.5	24.0
1	2	20.5	24.0	24.5
1	3	23.5	24.5	25.0
1	4	20.0	21.0	22.0
1	5	16.5	19.0	19.0
1	6	24.5	25.0	28.0
2	1	24.5	25.5	27.0
2	2	27.5	28.0	31.0
2	3	21.5	23.5	24.0
2	4	22.5	25.5	25.5
2	5	23.0	24.5	26.0
2	6	21.5	22.5	23.0

Figure 7.1: Plots of individual profiles for Cohort 1 and Cohort 2



In Figure 7.1, the individual facial growth profiles of six girls and six boys are displayed. The plots seem to indicate that the linear mixed effects model in (2.3) would summarize both the individual development of facial growth in those subjects and their average profile as well. For the three measurements taken on the j^{th} subject in cohort i, the model is given by

$$\boldsymbol{y}_{j(i)} = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_{j(i)} + \boldsymbol{\varepsilon}_{j(i)}, \qquad i = 1, 2,\ j = 1, \ldots 6,$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ is the vector of fixed intercept and fixed slope. Since the subjects differ in baseline facial measurement and in facial growth speed, $\boldsymbol{b}_{j(i)} = (b_{oj(i)}, b_{1j(i)})'$ is the vector of random intercept and slope with mean **0** and a covariance matrix $\sigma^2\,\mathbf{D}$ = $\sigma^2 \begin{bmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{bmatrix}$. The 3 × 1 random error vector $\boldsymbol{\varepsilon}_{j(i)}$ has mean **0** and a covariance $\sigma^2\mathbf{I}$, where σ^2 is the common variance for error components. The design matrices $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are

$$\boldsymbol{X}_i = \boldsymbol{Z}_i = \begin{bmatrix} 1 & t_{1(i)} \\ 1 & t_{2(i)} \\ 1 & t_{3(i)} \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

Note that for simplicity the time interval [8, 12] is rescaled to [−1, 1] by applying a linear transformation:

$$\frac{x - \frac{8+12}{2}}{\frac{12-8}{2}} = \frac{x-10}{2}, \text{ where } x \text{ is the age of the subject } (8 \le x \le 12).$$

The logarithm of the restricted likelihood function expressed in terms of all observations $\underline{\boldsymbol{y}} = (\boldsymbol{y}'_{1(1)}, \ldots, \boldsymbol{y}'_{6(1)},\ \boldsymbol{y}'_{1(2)}, \ldots, \boldsymbol{y}'_{6(2)})'$ is simplified to

$$l^* = -17\log\sigma^2 - \frac{11}{2}\log\left|\mathbf{V}_1\right| - \frac{1}{2\sigma^2}\underline{\boldsymbol{y}}'\mathbf{A}'(\mathbf{AVA}')^{-1}\mathbf{A}\underline{\boldsymbol{y}},$$

where $|\mathbf{V}_1|$ represents the determinant of $\mathbf{V}_1$ and it is defined in (7.18), the 36 × 36 matrix $\mathbf{V} = diag(\mathbf{V}_1)$. A matrix $\mathbf{A}$ satisfying rank($\mathbf{A}$) = 34 and $\mathbf{AX} = 0$ is

$$\underset{34\times 36}{\mathbf{A}} = \begin{pmatrix} \mathbf{I}_3 & -\mathbf{I}_3 & \underset{3\times3}{\mathbf{0}} & \cdots & \underset{3\times3}{\mathbf{0}} & \underset{3\times3}{\mathbf{0}} \\ \underset{3\times3}{\mathbf{0}} & \mathbf{I}_3 & -\mathbf{I}_3 & \cdots & \underset{3\times3}{\mathbf{0}} & \underset{3\times3}{\mathbf{0}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \underset{3\times3}{\mathbf{0}} & \underset{3\times3}{\mathbf{0}} & \underset{3\times3}{\mathbf{0}} & \cdots & \mathbf{I}_3 & -\mathbf{I}_3 \\ \underline{\boldsymbol{a}}_1' & \underline{\boldsymbol{a}}_1' & \underline{\boldsymbol{a}}_1' & \cdots & \underline{\boldsymbol{a}}_1' & \underline{\boldsymbol{a}}_1' \end{pmatrix}$$

with $\underline{\boldsymbol{a}}_1' = \begin{pmatrix} -1 & 2 & -1 \end{pmatrix}$.

Using the estimation equations in (7.7) and (7.8), the REML estimates are computed via our iterative method and are displayed in Table 7.6 with values of Δ and l^*. SAS results are also presented for comparison.

Table 7.6: Solution of REML estimators with values of Δ and l^* for facial growth data

Parameter	σ^2	d_{11}	d_{22}	d_{12}	Δ	l^*
Our Estimate	0.4430	15.9736	0.0474	0.8704	0.0021	59.2016
SAS Estimate	0.4430	15.9736	0.0474	0.8704	0.0021	59.2016

It should be noticed that the criterion function P cannot be applied to this observed facial growth data, because the true variance component values are unknown. The REML estimates found by our method are identical up to at least four decimal places compared with SAS results. We conclude that our proposed method is equally good as the SAS PROC MIXED procedure.

Chapter 8
Conclusions

In this dissertation, we address the design problem for longitudinal cohort studies and propose an iterative method for computing the REML estimators of the variance components in the linear mixed effects models. The mixed longitudinal designs with two fully overlapping cohorts are considered. The *D*-, *A*-, and *E*-optimality criteria are used in finding the optimal allocation of time points to maximize the information for the estimation of the fixed parameters in the model.

We find optimal cohort designs for the design class $T_{23(a)}$ with correlation matrix $\boldsymbol{R_i} = \mathbf{I}$ (i = 1, 2) and a general matrix **D** under the *D*- and *A*-optimality criteria. The *D*-optimal design for $T_{23(a)}$ with $\boldsymbol{R}_i = \mathbf{I}$ and a general **D** is the design with equidistant time points (−1, 0, 1) for both cohorts, if $\mathrm{D}_f^{(1)} > 0$. We compute optimal cohort designs analytically for the design class T_{23} with covariance matrix $\mathbf{D} = \mathbf{I}$ and error correlation matrix $\boldsymbol{R_i} = \mathbf{I}$. The *D*-, *A*-, and *E*- optimal design for the design class T_{23} with $\boldsymbol{R}_i = \mathbf{I}$ and $\mathbf{D} = \mathbf{I}$ is the design with time points $(-1, -\sqrt{a_0}, 1)$ for the 1st cohort and $(-1, \sqrt{a_0} 1)$ for the 2nd cohort, where $a_i^2 \le a_0$ $(0 < a_0 < 1)$, $i = 1, 2$. We compare between cohort designs with equidistant and non-equidistant time points. We have learned that when the covariance of the random effects satisfies certain conditions, the design with equidistant time points is preferred. For instance, the *D*-optimal design for $T_{23(a)}$ with $\boldsymbol{R}_i = \mathbf{CS}$ and a general $\mathbf{D}$ is the design with equidistant time points (−1, 0, 1) for both cohorts, if $-\sqrt{d_{11}d_{22}} < d_{12} \le \min\left\{\sqrt{d_{11}d_{22}}, \frac{a(1+3d_{11}-d_{22}+2\rho)}{6}\right\}$. However, under certain cases, for example, the second case in Theorem 5.4, the design with non-equidistant time points is better. In addition, we present general results with their applications in comparison of design classes $T_{23(a)}$ and $T_{23(-a,\, a)}$, with respect to *D*- and *A*-optimality criteria. The design class $T_{23(-a,\, a)}$ is preferred over $T_{23(a)}$ with respect to both *D*- and *A*-optimality criteria, if $\mathbf{D}$ is a diagonal matrix.

The Restricted Maximum likelihood (REML) estimators for the variance components are also obtained using three criterion functions l^*, Δ, and P. We have compared our estimates with the SAS estimates for two simulated data sets and one observed facial growth data set. Our estimates are comparable with the SAS estimates with respect to l^*, Δ, and P.

Bibliography

[1] Atkinson, A.C. and Donev, A.N. (1992). *Optimum Experimental Design*. Oxford: Science Publications.

[2] Bandera, E.V., Freudenheim, J.L., Marshall, J.R., Zielenzy, M., Priore, R.L., Brasure, J., Baptiste, M., Graham, S. (1997). Diet and alcohol consumption and lung cancer risk in the New York state Cohort (United States). *Cancer Causes and Control* 8, 828–840.

[3] Biedermann, S. and Dette, H. (2005). Numerical construction of parameter maximin *D*-optimal designs for binary response models. *South African Statistical Journal*, 39 (2), 127-161.

[4] 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.

[5] Calle, E.E., Rodriguez, C., Jacobs, E.J., Almon, M.L., Chao, A., McCullough, M.L., and Feigelson, H.S. (2002). The American Cancer Society Cancer Prevention Study II Nutrition Cohort. *Cancer* 94, 500–511.

[6] Corbeil, R.R. and Searle, S.R. (1975). Restricted maximum likelihood (REML) estimation of variance components in the mixed model. *Technometrics* 18, No. 1, 31-38.

[7] Dette, H., Lopez, I.M., Rodriguez, I.M.O. and Pepelyshev, A. (2004). Maximin efficient design of experiment for exponential regression models. *Journal of Statistical Planning and Inference* 136, 4397-4418.

[8] Diggle, P. J., Liang, K. Y. & Zeger, S. L. (1994), *Analysis of Longitudinal Data*. Oxford University Press.

[9] Dik, M.G., Deeg, D.J.H., Bouter, L.M., Corder, E.H., Kok, A., and Jonker, C. (2000). Stroke and apolipoprotein E ε4 are independent risk factors for cognitive decline: a population-based study. *Stroke* 31, 2431–2436.

[10] Hartley, H.O. and Rao, J.N.K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika* 54, 93-108.

[11] Kiefer, J. and Wolfowitz, J. (1959). Optimum Designs in Regression Problems. *Annals Math. and Statistics.*, Vol 30, 271-294

[12] King, J. and Wong, W.K. (2000). Minimax *D*-optimal designs for the logistic model. *Biometrics* 56 (4), 1263-1267.

[13] Longford, N.T. (1993). *Random Coefficient Models*. New York: Oxford University Press.

[14] Mathew, T. and Sinha, B.K. (2001). Optimal designs for binary data under logistic regression. *Journal of Statistical Planning and Inference* 93, 295-307.

[15] Meyer, K. (1985). Maximum likelihood estimation of variance components for a multivariate mixed model with equal design matrices. *Biometrics* 41 (1), 153-165.

[16] Moerbeek, M. (2005). Robust properties of *A*-, *D*-, and *E*-optimal designs for polynomial growth models with autocorrelated errors. *Computational Statistics and Data Analysis* 48, 765-778.

[17] Ortega-Azurdy, S.A., Tan, F.E.S., and Berger, M.P.F. (2007). The effect of dropout on the efficiency of *D*-optimal designs of linear mixed models. *Statistics in Medicine*, DOI: 10.1002/sim.3108 (in press).

[18] Ouwens, M.J.N.M., Tan, F.E.S. and Berger, M.P.F. (2002). Maximin *D*-optimal designs for longitudinal mixed-effects models. *Biometrics* 58, 735-741.

[19] Pinheiro, J.C., and Bates, D.M. (2000). *Mixed Effects Models in S and S-Plus*. New York: Springer-Verlag.

[20] Potthof, R.F., and Roy, S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, 51, 313-326.

[21] Pocock, S.J. (1983). *Clinical trials: A practical approach*. Chichester: John Wiley & Sons.

[22] Pukelsheim, F. (1993). *Optimal Design of Experiments*. New York: Wiley.

[23] Rao, C.R. (1973). *Linear Statistical Inference and Its Applications*. New York: Wiley.

[24] Sebastiani, P. and Settimi, R. (1997). A note on D-optimal designs for a logistic regression model. *Journal of Statistical Planning and Inference* 59, 396-402.

[25] Silvey, S. D. (1980). *Optimal Design*. New York: Chapman and Hall.

[26] Snijders, T.A.B. and Bosker, R.J. (1999). *Multilevel Analysis*. London: SAGE.

[27] 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.

[28] Tekle, F.B., Tan, F.E.S. and Berger, M.P.F. (2008). *D*-optimal cohort designs for linear mixed-effects models. *Statistics in Medicine* 27, 2586-2600.

[29] Tekle, F.B., Tan, F.E.S. and Berger, M.P.F. (2011). Too many cohorts and repeated measurements are a waste of resources. *Journal of Clinical Epidemiology*, DOI: 10.1016/j.jclinepi.2010.11.023 (in press).

[30] Tekle, F.B., Tan, F.E.S. and Berger, M.P.F. (2008). Maximin *D*-optimal designs for binary longitudinal responses. *Computational Statistics and Data Analysis* 52 (12), 5253-5262.

[31] Verbeke, G., and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer.

[32] Vickers, A.J. (2003) How many repeated measures in repeated measures designs? Statistical issues for comparative trials. *BMC Medical Research Methodology* 2003, 3:22

[33] Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. Transactions of the American Mathematical Society 54, 426-482.

[34] Willett, J.B., Singer, J.D., & Martin, N.C. (1998). The design and analysis of longitudinal studies of development and psychopathology in context: Statistical models and methodological recommendations. *Development and Psychopathology*, 10, 395-426.

[35] Winkens, B., Schouten, H. J.A., van Breukelen, G. J. P., and Berger, M. P. F. (2007). Optimal designs for clinical trials with second-order polynomial treatment effects. *Statistical Methods in Medical Research*, 16, 523 - 537.

[36] Winkens, B., Schouten, H.J.A., Breukelen, G.J.P, and Berger, M.P.F. (2005). Optimal time-points in clinical trials with linearly divergent treatment effects. *Statistics in Medicine* 24, 3743-3756.

[37] Woolson, R.F., Leeper, J.D., and Clark, W.R. (1978). Analysis of incomplete data from longitudinal and mixed longitudinal studies. *Journal of the Royal Statistical Society*, 141, 242-252.

[38] Zacks, S. (1971). *The Theory of Statistical Inference*. New York: Wiley.