

Bayesian optimal designs of binary repeated measurements

Citation for published version (APA):

Abebe, H. T. (2014). *Bayesian optimal designs of binary repeated measurements*. [Doctoral Thesis, Maastricht University]. Boekenplan. <https://doi.org/10.26481/dis.20140129ha>

Document status and date:

Published: 01/01/2014

DOI:

[10.26481/dis.20140129ha](https://doi.org/10.26481/dis.20140129ha)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

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

[Link to publication](#)

General rights

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

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

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

www.umlib.nl/taverne-license

Take down policy

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

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Bayesian Optimal Designs of Binary Repeated Measurements

Haftom Temesgen Abebe

Bayesian Optimal Designs of Binary Repeated Measurements

Dissertation

to obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus, Prof. dr. L.L.G. Soete
in accordance with the decision of the Board of Deans,
to be defended in public on
Wednesday 29 January 2014 at 16.00 hours

by

Haftom Temesgen Abebe

Supervisor

Prof. dr. M.P.F. Berger

Co-supervisors

Dr. E.S. Tan

Dr. G.J.P. van Breukelen

Assessment committee

Prof. dr. M.H. Prins (chair)

Dr. M.J.J.M. Candel

Prof. dr. P. Goos (University of Antwerp)

Prof. dr. H.J.A. Hoijtink (University of Utrecht)

Prof. dr. M.P. Zeegers

ISBN: 978 90 8666 320 0

Printing: Boekenplan | Maastricht



© Haftom Temesgen Abebe, Maastricht 2014

The studies presented in this thesis were supported by the School for Public Health and Primary Care (CAPHRI), Maastricht University.

In loving memory of my Grandmother, Alemash Hagos.

To my parents, Temesgen Abebe and Beriha Yigzaw, and my adorable wife, Desta Abay.

Contents

1 Introduction	2
1.1 An example of a time-structured study with a binary outcome.....	3
1.2 The logistic mixed effects model.....	5
1.2.1 First order penalized quasi-likelihood (PQL1)	6
1.2.2 Extension of generalized estimating equations (GEE)	7
1.3 Bayesian optimal designs.....	7
1.3.1 Bayesian D-optimality criterion.....	8
1.4 Cost function and relative efficiency.....	9
1.5 The outline of the thesis.....	10
2 On the choice of a prior for Bayesian D-optimal designs for the logistic regression model with a single predictor	13
2.1 Introduction.....	15
2.2 Bayesian D-optimality criterion.....	16
2.2.1 The logistic regression design problem.....	18
2.2.2 Relative efficiency.....	19
2.2.3 Maximin Bayesian procedure	19
2.2.4 Algorithm.....	20
2.3 Numerical study and Results.....	20
2.3.1 Hyperparameter values.....	20
2.3.2 Results.....	21
2.4 Conclusion.....	28
3 Bayesian D-optimal designs for the two parameter logistic mixed effect model	31
3.1 Introduction.....	33

3.2 Logistic mixed effects models.....	34
3.2.1 Variance-covariance matrix of the parameter estimators.....	34
3.3 Optimal designs and relative efficiency.....	35
3.3.1 Bayesian D-optimal designs.....	35
3.3.2 Implementation.....	36
3.3.3 Cost consideration and relative efficiency.....	37
3.3.4 Maximin Bayesian procedure.....	38
3.4 Numerical study and Results.....	39
3.4.1 Models and Hyperparameter values.....	39
3.4.2 Results.....	40
3.5 Conclusions.....	48
4 Bayesian design for dichotomous repeated measurements with autocorrelation	51
4.1 Introduction.....	53
4.2 Motivating Example.....	54
4.3 Model and Variance-covariance matrix of the parameter estimators.....	55
4.4 Optimal designs and relative efficiency.....	57
4.4.1 Bayesian D-optimality criterion.....	57
4.4.2 Cost function and relative efficiency.....	58
4.5 Numerical approach.....	59
4.5.1 Method of Optimization.....	59
4.5.2 Priors and covariance structures.....	60
4.6 Results.....	61
4.7 Application.....	70
4.8 Summary and Conclusion.....	71
5 Robustness of Bayesian D-optimal design for the logistic mixed model against misspecification of autocorrelation	75
5.1 Introduction.....	77
5.2 Model specification.....	78
5.3 Estimation methods.....	79
5.3.1 First order penalized quasi-likelihood (PQL1).....	79
5.3.2 Extension of generalized estimating equations (GEE).....	80
5.4 Optimal designs and relative efficiency.....	81
5.4.1 Bayesian D-optimality criterion.....	81
5.4.2 Cost function and relative efficiency.....	82
5.5 Numerical approach.....	83
5.5.1 Method of Optimization.....	83

5.5.2 Priors and covariance structures.....	84
5.6 Results.....	85
5.7 Summary and Conclusion.....	94
6 A computer program for Bayesian D-optimal binary repeated measurements designs	97
6.1 Introduction.....	99
6.2 Model specification and estimation methods	100
6.2.1 The logistic mixed effects model	100
6.2.2 Estimation methods.....	100
6.3 Optimal designs and relative efficiency.....	102
6.3.1 Bayesian D-optimality criterion.....	102
6.3.2 Cost function and relative efficiency.....	103
6.3.3 Method of Optimization.....	104
6.4 BODMixed_Logistic program.....	105
6.4.1 Description and use of the program.....	105
6.5 Example: The respiratory infection study.....	110
6.6 Conclusions.....	114
7 Summary and recommendations	115
7.1 Summary.....	116
7.2 Overall conclusions and recommendations.....	118
7.3 Ideas for future research.....	118
8 Samenvatting en aanbevelingen	121
8.1 Samenvatting.....	122
8.2 Globale conclusies en aanbevelingen.....	124
References	125
Acknowledgements	131
About the author	133

CHAPTER 1

Introduction

1 Introduction

This thesis deals with Bayesian optimal designs for binary responses, mainly for repeated measurements over time. The basic idea underlying optimal design is that statistical inference about model parameters of interest, for instance a time effect or a treatment effect, can be improved by optimally selecting levels of the independent variables, such as the time points of measurements in a longitudinal study or the dosages of a drug in dose-effect studies. These levels are chosen such as to minimize the variability of the parameter estimators by minimizing a real-valued target function of the variance-covariance matrix of the parameter estimators, which is usually referred to as an optimality criterion (see, e.g., Silvey, 1980).

In optimal design theory, many different optimality criteria, such as the D-, A-, G-, and V-optimality criteria have been developed to measure the performance of a design (Atkinson et al., 2007; Berger and Wong, 2009). Each of these criteria serves a different goal. In general, an optimality criterion can be characterized as an estimation criterion or a prediction criterion. The D- and A-optimality criteria are estimation criteria, as they concern the minimization of the variance of the parameter estimators, while the G- and V-optimality criteria are concerned with accurate response predictions. Restricting optimal designs to estimation there is a variety of criteria that leads to different designs (Chaloner and Verdinelli, 1995). This thesis focuses on the D-optimality criterion, which is the most commonly used one in practice and performs well in terms of other criteria (Atkinson et al., 2007; Goos, 2002). The D-optimality criterion seeks to minimize the determinant of the variance-covariance matrix of the parameter estimators and it is not affected by linear transformation of the scale of independent variables. Such property does not hold for other optimality criteria.

In case of non-linear models like the logistic regression model, the variance-covariance matrix of the parameter estimators depend on the unknown parameters. Thus, in order to find and to implement the optimal design, the model parameter values should be known in advance. However, the parameter values are always unknown as the design is planned to obtain data for estimating them. The traditional solution to this problem is to design the study to be optimal for a best guess of the parameter values, which leads to a locally optimal design (Chernoff, 1953; Abdelbasit and Plackett, 1983). However, such a locally optimal design may not be efficient when the true parameter values differ from those best guesses, that is, the design may not be robust to parameter misspecification. Thus, locally optimal design faces the serious problem of how to take into account the uncertainty about the parameter values.

To overcome the problem of local optimality, three methods are proposed in the literature namely, the sequential procedure that uses information from previous runs to update estimates of the model parameters and thereby also to update the design which is called a sequential optimal design (see, e.g., Abdelbasit and Plackett, 1983; Wu, 1985; Berger and Wong, 2009), the maximin or minimax approach, which selects the design that minimizes the maximum possible variance of the estimators, or equivalently, that maximizes their minimum efficiency over a user selected region of the parameter values (see, e.g., Dette, 1997; Ouwens et al., 2002; Tekle et al., 2008; Tan, 2011), and Bayesian methods, described by Chaloner and

Larntz (1989) and Chaloner and Verdinelli (1995), among others. Although the sequential design procedure is simple, it may take more time and resources than a non-sequential one, and may not be feasible due to restrictions on the length of the research period. A potential objection to the maximin approach is that the maximin design may occur at the edges of the parameter space, and if some points in the parameter space have a very low probability then their importance (weight) in the design criterion may be over-emphasised by the maximin criterion (Atkinson et al., 2007).

The Bayesian approach takes the uncertainty of the parameter values of the statistical model into account by using a prior distribution on the unknown parameters rather than single-value guesses. This will give more flexibility to a researcher. The Bayesian optimal design method allows for a subjective construction of a prior distribution that captures all existing information about the uncertainty of the parameters values. Since the choice of a prior distribution is a problem in practice, we will discuss this in some more detail in one of the chapters of this thesis.

The Bayesian optimal design literature outside the normal linear model is mostly restricted to binary response models (see, e.g., Khuri et al., 2006), and the majority of the Bayesian literature on design issues does not discuss or justify the choice of a prior distribution. Moreover, to our knowledge no work has been done on Bayesian optimal designs for longitudinal studies with a binary response under cost constraints. In this research, we try to fill in these gaps. More specifically, we will mainly focus on Bayesian D-optimal designs for mixed effects logistic models with repeated measures under cost constraints. The mixed effects logistic regression model is very useful in modelling longitudinal binary outcomes and is often applied in biomedical research.

1.1 An example of a time-structured study with a binary outcome

The Indonesian Children's Health study is a prospective longitudinal study of preschool children aged up to 6 years and was conducted in 6 rural villages in the Purwakarta district, West Java, Indonesia between 1977 and 1978 (Sommer et al., 1984; Diggle et al., 1994). Over 3000 preschool children were re-examined medically for up to six consecutive quarters to assess whether they suffered from respiratory or diarrhoeal infection and xerophthalmia, an ocular manifestation of vitamin A deficiency.

In many repeated measurement studies, the individual trend over time is assumed to be characterized by a first- or second-order polynomial function. Although higher-order polynomials are also possible, they are less prominent in the literature. A second-order (quadratic) polynomial is more flexible than a first-order (linear) one and can handle monotonic as well as non-monotonic time effects within a restricted time space. Figure 1.1 shows some patterns that can be fitted by a quadratic growth curve. Interindividual differences in the precise pattern can be included as random effects in the model. For example, Diggle et al. (1994) focused on the question of whether vitamin A deficient children are at increased risk of respiratory infection, and found a curvilinear pattern such as Figure 1.1

(a) and fitted a random intercept logistic model with a quadratic time effect for the longitudinal data on respiratory infection of 275 Indonesian children who were examined quarterly up to six visits.

In planning a longitudinal study such as in the Indonesian Children's health study, researchers face the problem of choosing an adequate number and timing of the repeated measurements. Usually, these choices are based on non-statistical criteria. Using optimal design theory, this thesis addresses the choice of the number and timing of measurements, taking into account the costs for obtained the data. Researchers usually have a limited budget for a study and the number of repeated measurements per subject usually needs to be restricted. Therefore, it is reasonable to take the cost of measuring into account when a design is optimized, and find the optimal number of measurements and the allocations of these measurements over time for the available fixed budget. Although optimal designs with cost constraints are found in the optimal design literature, the current thesis will address the problem of Bayesian D-optimal designs for longitudinal studies for a binary response with cost constraints, and also address the problem of a choice of the prior distribution for the Bayesian designs. This topic has not yet been fully studied before. In the next sub-sections, the logistic mixed effects model and the Bayesian optimal design criterion will be presented, as well as suitable cost function for longitudinal studies with a relative efficiency criterion that takes costs into account when comparing different designs.

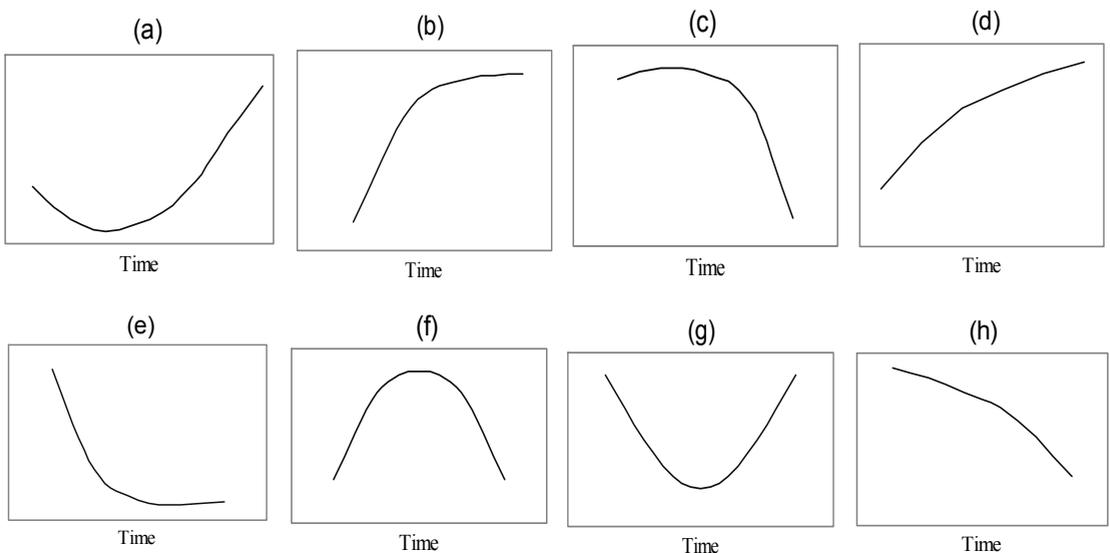


Figure 1.1 Examples of second-order polynomial functions over time.

1.2 The logistic mixed effects model

Let the $q \times 1$ vector $\mathbf{y}_i = (y_{i1}, \dots, y_{iq})'$ be binary responses y_{ij} of subject i at q time points, $i = 1, 2, \dots, N$ and $j = 1, \dots, q$. It is assumed that all subjects have measurements at the same time points, and that, conditional on the subject-specific random effect vector \mathbf{b}_i , the binary responses y_{ij} of \mathbf{y}_i are assumed to be Bernoulli distributed with probability of success $p(y_{ij} = 1 | \mathbf{b}_i)$. These probabilities are related to the fixed and random effects via the logit link function. The corresponding logistic mixed effect model is given by:

$$\text{logit}(p(y_{ij} = 1 | \mathbf{b}_i)) = \log \left(\frac{p(y_{ij} = 1 | \mathbf{b}_i)}{1 - p(y_{ij} = 1 | \mathbf{b}_i)} \right) = \mathbf{x}_j' \boldsymbol{\beta} + \mathbf{z}_j' \mathbf{b}_i, \quad (1.1)$$

where the $p \times 1$ vector \mathbf{x}_j is the design vector of the explanatory variables at the j th measurement for subject i , $\boldsymbol{\beta}$ is the corresponding $p \times 1$ vector of fixed polynomial time effects, and \mathbf{z}_j is the $r \times 1$ design vector for the random effects that is usually a subset of vector \mathbf{x}_j . The vector \mathbf{b}_i is the corresponding $r \times 1$ vector of random effects, which is assumed to have a multivariate normal distribution with mean zero and covariance matrix \mathbf{D} . For instance, if a quadratic ($p = 3$) time effect is assumed, then the design vector is $\mathbf{x}_j' = (1 \ t_j \ t_j^2)$ and $\boldsymbol{\beta} = (\beta_0 \ \beta_1 \ \beta_2)'$, where t_j is the time point of the j th measurement, and β_0 , β_1 and β_2 are the fixed effects. Suppose that a random intercept and random linear slope are assumed. Then the design vector is $\mathbf{z}_j' = (1 \ t_j)$ and $\mathbf{b}_i = (b_{0i} \ b_{1i})'$, where b_{0i} and b_{1i} are the corresponding random (subject-specific) deviations from these fixed effects. Then, according to model (1), the log-odds of a positive response for subject i at time t_j is given by:

$$\text{logit}(p(y_{ij} = 1 | \mathbf{b}_i)) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_j + \beta_2 t_j^2. \quad (1.2)$$

Due to the random effects in model (1.1) and (1.2), the log-likelihood cannot be written down in closed form (see, e.g., Molenberghs and Verbeke, 2005). Hence, either numerical methods or approximations to the log-likelihood must be used. Numerical methods require large computational resources and more importantly they require full knowledge of the data, making them computationally inconvenient for optimal design procedures. To overcome this problem, approximation methods are employed. Optimal designs obtained by approximation methods were studied by Moerbeek et al. (2001), Moerbeek and Maas (2005), Ouwens et al. (2006), Woods and van de Ven (2006), Tekle et al. (2008), Niaparast (2009) and Niaparast and Schwabe (2013). There is a large statistical literature on various approximation methods, but here, for the purpose of obtaining optimal designs, we will focus on the two most frequently used ones, which are implemented in commercially available software packages: Penalized quasi-likelihood (PQL) (Breslow and Clayton 1993; Jang and Lim 2009) and an

extended version of generalized estimating equations (GEE) (Molenberghs and Verbeke 2005). Details of these methods will be given in the next sub-sections.

1.2.1 First order penalized quasi-likelihood (PQL1)

We use an approximate variance-covariance matrix of the parameters $\boldsymbol{\beta}$ based on first order penalized quasi-likelihood (PQL1) estimation. See Breslow and Clayton (1993) and Jang and Lim (2009) for details. PQL1 performs well in terms of point estimation since it produces the smallest mean squared error (MSE) and the bias of the estimators decrease as the sample size increases (Breslow and Clayton 1993; Moerbeek et al. 2003; Jang and Lim 2009). The variance-covariance matrix of the estimator $\hat{\boldsymbol{\beta}}$ for the logistic mixed effects models (1) is approximated in PQL1 by:

$$\text{var}(\hat{\boldsymbol{\beta}}) \approx (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}, \quad (1.3)$$

where \mathbf{X} is the $Nq \times p$ design matrix formed by stacking $\{\mathbf{x}_j'\}$ for N subjects and q time points, and the $Nq \times Nq$ block-diagonal matrix \mathbf{V} has N blocks of $q \times q$ variance-covariance matrices given by:

$$\mathbf{v}_i \approx \mathbf{w}_i^{-1/2} \mathbf{R} \mathbf{w}_i^{-1/2} + \mathbf{Z} \mathbf{D} \mathbf{Z}', \quad (1.4)$$

where \mathbf{R} is the $q \times q$ correlation matrix that captures autocorrelation of the responses over time, \mathbf{Z} is the $q \times r$ design matrix with rows \mathbf{z}_j' , $j = 1, \dots, q$, the $r \times r$ matrix \mathbf{D} is the variance-covariance matrix of the random effects, and \mathbf{w}_i^{-1} is the diagonal matrix of the conditional variances of the transformed responses given the random effects \mathbf{b}_i , which is equal to the inverse of the diagonal matrix of the conditional variances of the untransformed responses given the random effects \mathbf{b}_i (see for detail Moerbeek et al. 2001; Molenberghs and Verbeke 2005 page 270). Note that, under conditional independence \mathbf{R} is an identity matrix and equation (1.4) becomes

$$\mathbf{v}_i \approx \mathbf{w}_i^{-1} + \mathbf{Z} \mathbf{D} \mathbf{Z}'. \quad (1.5)$$

The diagonal matrix \mathbf{w}_i is given by:

$$\mathbf{w}_i = \text{diag}(w_{i1}^{b_i}, \dots, w_{iq}^{b_i}), \quad (1.6)$$

where $w_{ij}^{b_i} = p(y_{ij}=1|\mathbf{b}_i)(1-p(y_{ij}=1|\mathbf{b}_i))$, for $i = 1, \dots, N$ and $j = 1, \dots, q$. Since the random effects are unknown in the design stage, we will generate \mathbf{b}_i from a multivariate normal distribution with mean zero and variance-covariance \mathbf{D} .

1.2.2 Extension of generalized estimating equations (GEE)

Another approximation method is the extended GEE. The variance-covariance matrix of the estimator $\hat{\beta}$ for the logistic mixed effects model (1.1), based on the extension of the GEE approach (Zeger et al., 1988; Molenberghs and Verbeke, 2005), is approximated by:

$$\text{var}(\hat{\beta}) \approx \left(\sum_{i=1}^N \frac{\partial \mathbf{P}_i'}{\partial \beta} \mathbf{u}_i^{-1} \frac{\partial \mathbf{P}_i}{\partial \beta'} \right)^{-1}, \quad (1.7)$$

where $\mathbf{P}_i = (p(y_{i1}|\mathbf{b}_i), \dots, p(y_{iq}|\mathbf{b}_i))'$, $\frac{\partial \mathbf{P}_i'}{\partial \beta} = \mathbf{X}_i' \mathbf{W}_i$, the $p \times q$ matrix \mathbf{X}_i with rows $\{\mathbf{x}_j'\}$ for $j = 1, \dots, q$ and the working variance-covariance matrix \mathbf{u}_i of the responses is given by:

$$\mathbf{u}_i \approx \mathbf{w}_i^{1/2} \mathbf{R} \mathbf{w}_i^{1/2} + \mathbf{w}_i \mathbf{Z} \mathbf{D} \mathbf{Z}' \mathbf{w}_i', \quad (1.8)$$

where \mathbf{w}_i is the diagonal matrix of the conditional variances of responses given the random effects $\mathbf{b}_i = 0$ and is given by:

$$\mathbf{w}_i = \text{diag}(w_{i1}^{b_i=0}, \dots, w_{iq}^{b_i=0}), \quad (1.9)$$

where $w_{ij}^{b_i=0} = p(y_{ij}=1|\mathbf{b}_i=0)(1-p(y_{ij}=1|\mathbf{b}_i=0))$, for $i = 1, \dots, N$ and $j = 1, \dots, q$ (see, Molenberghs and Verbeke, 2005 page 443).

Under conditional independence \mathbf{R} vanishes and equation (1.8) reduces to

$$\mathbf{u}_i \approx \mathbf{w}_i + \mathbf{w}_i \mathbf{Z} \mathbf{D} \mathbf{Z}' \mathbf{w}_i'. \quad (1.10)$$

In the present research, a first order auto regressive (AR1) correlation structure is considered for the correlation matrix \mathbf{R} . The (j, l) element of the correlation matrix \mathbf{R} is then equal to $\rho^{|t_j - t_l|}$, where $l = 1, \dots, q$, and so ρ is the autocorrelation coefficient between two responses at a time distance of one, that is, $\rho = \text{Corr}(y_{ij}, y_{il})$ for which $|t_j - t_l| = 1$. This autocorrelation structure implies that repeated measurements closer in time are more highly correlated and that the correlation decreases as the distance between the time points increases.

1.3 Bayesian optimal designs

Bayesian methods are widely used in many statistical procedures for data analysis. The Bayesian optimal design problem has also been actively studied in scientific research by Chaloner and Lantz (1989), Chaloner and Verdinelli (1995) and Atkinson et al. (2007), among others. The Bayesian optimal design method requires that the researcher comes up with a prior distribution that captures all the existing information for the model parameters.

By incorporating historical knowledge about the model parameters of interest into the prior, Bayesian design methods can help to improve the design of new studies. In practice, there is usually some information available at the beginning of a scientific study. Some may come from similar previous studies, a pilot study or from different experts in the field. Usually the source of the information is not unique and several similar studies may provide different information. The information may take on different forms, such as information on the model parameters or only as a certain function of the model parameters. The first step in the Bayesian approach is to capture all prior information in the form of a probability distribution for the model parameters of interest. This distribution is referred to as the prior distribution. The Bayesian approach then resolves the dependency of optimal design on unknown parameters by averaging the proposed design criterion over the prior distribution. Bayesian optimal designs can be different for different design criteria. In the next section, the Bayesian D-optimality criterion, which is the most often considered in the literature, will be presented.

1.3.1 Bayesian D-optimality criterion

To introduce the notation for the optimality criterion, suppose a design be denoted by ξ , where ξ is an element of the design space Ξ . The variance-covariance matrix of the maximum likelihood estimator $\hat{\beta}$ for a model with unknown parameters β and design ξ is denoted by $\text{Var}(\hat{\beta}, \xi)$. Optimal design minimizes a real-valued function of $\text{Var}(\hat{\beta}, \xi)$, for instance the determinant. This is usually difficult when the model is nonlinear or when a nonlinear function of the coefficients of a linear model is of interest, since in such cases, the $\text{Var}(\hat{\beta}, \xi)$ depends on the unknown parameters β . The Bayesian approach solves this by using a prior distribution for the unknown parameters and minimizing a real-valued function of the posterior variance-covariance matrix of β . However, computations involving the exact posterior distribution are often intractable for nonlinear models, and so asymptotic approximations may be used (see, e.g., Berger, 1985 page 224; Chaloner and Larntz, 1989 and Khuri et al., 2006). The most common form of a normal approximation to the posterior distribution of β is

$$\beta \mid y, \xi \sim N(\hat{\beta}, \text{Var}(\hat{\beta}, \xi)). \quad (1.11)$$

This approximation assumes that the sample size and/or the prior variance is large.

The D-optimal design ξ^* is the design among all possible designs $\xi \in \Xi$ for which the determinant of the variance-covariance matrix of $\hat{\beta}$ is minimized or equivalently, for which the determinant of the inverse of the variance-covariance matrix of $\hat{\beta}$ is maximized (Atkinson et al., 2007). Using this approximation (1.11), many authors (Chaloner and Larntz, 1989; Chaloner and Verdinelli, 1995; Dette, 1996; Sebastiani and Settini, 1998; Khuri et al., 2006;

Atkinson et al., 2007; Braess and Dette, 2007) define the Bayesian D-optimality criterion as follows

$$\phi_D(\xi) = \int_{\beta} \log \det[\text{Var}(\hat{\beta}, \xi)]^{-1} \pi(\beta) d\beta, \quad (1.12)$$

where $\pi(\beta)$ is the prior distribution for β . The design which maximizes (1.12) is the Bayesian D-optimal design. Note that (1.12) extends the non-Bayesian D-criterion by integrating its natural logarithm over the prior distribution, thereby giving more weight to more likely parameter values and less weight to improbable values.

There are several different forms of the Bayesian D-optimality criterion (see Atkinson et al., 2007; Berger and Wong, 2009), but the design criterion (1.12) is the most natural one for non-linear models for three reasons: Firstly, this criterion arises naturally from the expected gain of Shannon information (Lindley, 1956; Chaloner and Verdinelli, 1995; Atkinson et al., 2007). Secondly, concavity is always guaranteed for this criterion function, so that a global equivalence theorem will hold (Firth and Hinde, 1997). Thirdly, its numerical results are the most satisfactory (see, Atkinson et al., 2007 page 297).

In this thesis, we will adopt the Bayesian D-optimal criterion (1.12) for longitudinal and cross-sectional studies with binary responses, with the focus being on time-structured repeated measures of a binary outcome. The purpose is to identify and propose an optimal design in terms of the number and allocation of time points with constraints on the study costs.

1.4 Cost function and relative efficiency

The precision of estimating the parameters increases by taking more measurements per subject. However, the addition of subjects and measurements per subject will increase the costs of the study. In practice, the number of subjects and the number of time points at which measurements are taken per subject are often limited by budget constraints. Thus, it is reasonable to take into account the costs of a longitudinal study when designs are compared with each other. There are two main cost components in repeated measurements: the costs for the recruitment of subjects and the costs of the repeated measurements once a subject is recruited. Let the cost of recruiting a subject be C_1 , and the cost of obtaining a repeated measurement per subject be C_2 . Then the total costs (budget) of a longitudinal study with q repeated measurements for N subjects, excluding all costs that do not depend on N or q , can be defined by the linear cost function:

$$C = C_1N + C_2Nq = NC_2(k + q), \quad (1.13)$$

where $k = C_1/C_2$ is the ratio of the costs of adding a new subject to the costs of an additional measurement per subject. So the total costs of a longitudinal study depend on the number of repeated measurements as well as the number of subjects. Thus, to compare fairly, we will

consider designs with the same total costs C . This means that the designs that are compared can differ in terms of the number of subjects N and the number of the repeated measurements q .

The relative efficiency (RE) can be used to compare designs. Let $\phi_D(\xi_q|\pi)$ denote the value of design criterion (1.12) for the optimal design ξ_q with q time points with prior distribution π for the fixed effects β . Then the relative efficiency of an arbitrary design ξ_s with s time points, relative to the optimal design ξ_q is defined as:

$$\text{RE}(\xi_s; \xi_q|\pi) = \frac{k+q}{k+s} \left[\exp \left\{ \frac{\phi_D(\xi_s|\pi) - \phi_D(\xi_q|\pi)}{p} \right\} \right], \quad (1.14)$$

where $(q,s) \geq p$ and p is the number of unknown fixed parameters. If the value of this relative efficiency is close to unity, then the design ξ_s is about equally efficient as the optimal design ξ_q for a given prior π . The inverse of this relative efficiency determines the required number of subjects N for a design ξ_s to have the same efficiency as the optimal design ξ_q . Note that this relative efficiency (1.14) is an extension of the relative efficiency given by Chaloner and Larntz (1989). When the ratio $\frac{k+q}{k+s}$ is one, it reduces to their relative efficiency. This ratio follows from (1.14) and takes into account that the design with the smaller number of repeated measurements can have more subjects for the same budget.

1.5 The outline of the thesis

The **Chapters 2-6** in this thesis may be read as self-contained articles and some text in one chapter may therefore be repeated in other chapters to maintain clarity. The last chapter (**Chapter 7**) lists overall conclusions and a summary of this dissertation, and ideas for future work. A brief introduction to each chapter is given below.

Chapter 2. This chapter addresses the problem of a choice of the prior distribution for Bayesian D-optimal designs in the context of fixed effects logistic regression with a single, continuous explanatory variable. We investigate what prior distribution leads to a Bayesian optimal design which remains efficient for other choices of the prior distribution. We compute numerically Bayesian D-optimal designs for various hyperparameter values (i.e., parameter values for the prior distribution) of a normal and a uniform prior for the fixed logistic model and then determine which hyperparameter values lead to the most robust and efficient design for the logistic model. We also compare the normal and uniform prior distributions with respect to each other, and evaluate the performance of Bayesian D-optimal designs relative to the locally D-optimal designs. Moreover, the relative efficiencies of designs with equidistant equally weighted design points relative to the Bayesian D-optimal designs are computed.

Chapter 3. This chapter builds upon chapter 2 by deriving the Bayesian D-optimal number and allocation of time points for the logistic mixed effects model with a linear time effect in the absence of autocorrelation, under cost constraints and using priors based on the results from chapter 2. The Bayesian optimal number and timing (allocation) of time points for a given subject-to-measurement cost ratio is obtained for different prior means and also for different covariance structures of the random effects. Maximin Bayesian D-optimal designs are proposed to overcome the problem of dependence of Bayesian designs on the choice of prior distribution. To show the consequences of taking a prior with a small variance on the design choice and its robustness, we compute locally D-optimal designs and maximin locally D-optimal designs for different covariance structures under different subject-to-measurement cost ratios, and then we compare these with the maximin Bayesian D-optimal design in terms of minimum relative efficiency over all priors. Moreover, to see how efficient equidistant design points are compared to the optimal designs, we compute the relative efficiencies of equidistant time points versus the Bayesian D-optimal designs for different subject-to-measurement cost ratios.

Chapter 4. This chapter extends the work in the previous chapter in two ways. First, a linear growth (first-order polynomial) function may be too restrictive because the time-structured pattern may not be strictly linear. A quadratic growth function is more flexible than a linear one and can handle monotonic as well as nonmonotonic time effects within a restricted time space. Secondly, repeated measurements often show autocorrelation. Accordingly, this chapter studies Bayesian D-optimal designs for the logistic mixed effects model with quadratic time effects and autocorrelations between the time-structured measurements. The effect of the covariance structure of the random effects on the Bayesian D-optimal designs is also studied.

Chapter 5. Robustness of Bayesian D-optimal designs for the logistic mixed effects model against misspecification of autocorrelation is considered in this chapter. Since in general, researchers do not know the size of the autocorrelation parameter in the design stage, it would be helpful to know whether designs are robust against incorrect specification of the autocorrelation among the errors in the model. The Bayesian D-optimal designs for different numbers of time points, different values of the autocorrelation, i.e., $\rho \in [0, 0.9]$, and different covariance structures under various prior distributions are computed numerically by optimizing a scalar function of the approximate variance-covariance matrix based on first order penalized quasi likelihood (PQL1) and an extended version of generalized estimating equations (GEE). We first compare PQL1 with extended GEE in terms of the resulting Bayesian D-optimal designs for a given number of time points, prior distribution, covariance structure of the random effects, and autocorrelation. We also compare the two methods in terms of the Bayesian optimal number of time points for a given subject-to-measurement cost ratio. Subsequently, the effect of different values of the autocorrelation on the Bayesian D-optimal designs and the robustness of the latter against incorrect specification of the autocorrelation are studied.

Chapter 6. A new user friendly computer program is introduced and described in this chapter. A number of papers have been published on optimal designs for longitudinal studies with binary responses. However, to our knowledge there is no computer program that helps researchers to plan such studies in an optimal way using the Bayesian approach. The new program helps researchers to identify the optimal number and allocations of time points under cost constraints for linear, quadratic and cubic growth models with a user-specified covariance structure for the random effects. Moreover, it helps to assess the effect of autocorrelation on the optimal number and allocation of time points.

CHAPTER 2

On the choice of a prior for Bayesian D-optimal designs for the logistic regression model with a single predictor

This chapter will be published in *Communications in Statistics-Simulation and Computation* (article in press) with co-authors Frans E.S. Tan, Gerard J.P. Van Breukelen, Jan Serroyen and Martijn P.F. Berger

Abstract

The Bayesian design approach accounts for uncertainty of the parameter values on which optimal design depends, but Bayesian designs themselves depend on the choice of a prior distribution for the parameter values. This chapter investigates Bayesian D-optimal designs for two-parameter logistic models, using numerical search. We show three things: 1) a prior with large variance leads to a design which remains highly efficient under other priors, 2) uniform and normal priors lead to equally efficient designs, and 3) designs with four or five equidistant equally weighted design points are highly efficient relative to the Bayesian D-optimal designs.

Keywords: Bayesian D-optimal designs; logistic regression model; maximin Bayesian D-optimal design; locally D-optimal designs; relative efficiency.

2.1 Introduction

Logistic models are used increasingly in a wide variety of biomedical, social sciences and marketing studies. For example, clinical trials investigate the effect of a treatment on recovery from a disease. In an educational study one may estimate the effect of a new teaching method on the pass rate for an exam. The precision of estimating the unknown parameters of a logistic model depends to a large extent on the design used in the experiment (Khuri et al., 2006). A researcher designing an experiment has to select the levels of the independent variables and allocate the subjects to the different levels of the independent variables. For example, in designing a dose-response study, we need to choose the dosages to be included and to decide on the sample size per dosage. Optimal design theory can help in making these decisions. An optimal design maximizes a real-valued function of the Fisher information matrix, which is usually referred to as an optimality criterion (see, e.g., Silvey, 1980). However, the optimal design depends on the unknown parameter values in non-linear models like the logistic regression model. A common approach is to use a best guess of the parameter values, which leads to locally optimal designs (Chernoff, 1953; Abdelbasit and Plackett, 1983). A design which is optimal for a best guess, however, may not be efficient for other parameter values. So the ability to find locally optimal designs still faces the serious problem of how to take into account the uncertainty about the parameter values.

Three approaches have been suggested to deal with this uncertainty in the parameter values: A sequential procedure that uses information from previous runs (Wu, 1985; among others), a maximin approach which maximizes the minimal efficiency (Müller, 1995; Dette, 1997; Dette and Sahm, 1997; Müller and Pazman, 1998; King and Wong, 2000; Imhof, 2001; Ouwens et al., 2002; Tekle et al., 2008), and a Bayesian approach that assumes a prior distribution for the parameters (Chaloner, 1989; Chaloner and Larntz, 1989; Chaloner and Verdinelli, 1995; Dette, 1996; Han and Chaloner, 2004; Dror and Steinberg, 2006; Woods et al., 2006; Atkinson et al., 2007; Gotwalt et al. 2009; Woods and van de Ven, 2011). Construction of an optimal design using the sequential approach may be quite time consuming and costly, and therefore is not always feasible (Chaloner and Larntz, 1989), although two-stage design (Dragalin et al., 2008) may sometimes be feasible. A potential objection to the maximin approach is that the maximin design may occur at the edges of the parameter space (Atkinson et al., 2007 page 258).

In the present chapter, we study Bayesian optimal design, which allows a subjective construction of a prior distribution that captures all existing information about the uncertainty of the parameters values. To construct a prior distribution, researchers need evidence from past studies on the topic of interest in which the statistical model of interest was applied. However, researchers may differ in the studies that they are aware of, and also in the studies that they prefer to draw upon, and this will lead to differences between researchers with respect to their choice of a prior. Thus, the choice of the prior distribution is subjective, since there is no standard way available to quantify the prior information. The resulting Bayesian optimal design can vary substantially when different prior distributions are used. So it is

important to check the sensitivity of a Bayesian optimal design to the choice of a prior distribution before the design is implemented. This sensitivity check consists of the following two questions: 1. How can we choose a suitable prior distribution for the parameters of the model. For example, should we use a uniform or a normal prior? 2. How do we choose appropriate hyperparameter values, i.e., parameter values for the prior distribution that lead to the most robust designs. The Bayesian design literature outside the normal linear model is vastly restricted to binary response models (Chaloner and Larntz, 1989; Khuri et al., 2006; Atkinson et al., 2007; among others), and the majority of the Bayesian literature on design issues for logistic models, does not discuss or justify the choice of a prior distribution (Chaloner and Larntz, 1989; Ford et al., 1992; Khuri et al., 2006).

The aim of this study is to investigate what prior distribution leads to a Bayesian optimal design which is robust, i.e., which remains highly efficient for other choices of the prior distribution. We will first investigate numerically Bayesian D-optimal designs for various hyperparameter values for a normal prior and a uniform prior, and determine which hyperparameter values lead to the most robust and efficient design for the logistic model. We will also compare the normal and uniform prior distributions with respect to each other, and evaluate the performance of Bayesian D-optimal designs with respect to the locally D-optimal designs for the logistic model. Finally, we will compare the relative efficiencies of designs with equidistant equally weighted design points to the Bayesian D-optimal designs.

In the next section we will introduce the Bayesian D-optimality criterion, and formulate the relative efficiency as a measure of performance of a design as in Chaloner and Larntz (1989). The numerical study and its results are presented in section 2.3, followed by a conclusion in section 2.4.

2.2 Bayesian D-optimality criterion

To introduce the notation for the criterion, suppose that an experiment is to be designed by choosing k values of the design variable x and at each of the k values of x_i an independent observation y_i will be available. Let the data be denoted by $y^T = (y_1, \dots, y_k)$ and θ be the p -vector of unknown parameters of interest. Denote the contribution to the likelihood function by a single observation y_i made at x_i by $P(y_i|\theta, x_i)$. The Fisher information matrix for θ , given an observation y_i at design point x_i , is

$$I(x_i, \theta) = -E \left[\left(\frac{\partial \ln(p(y_i|x_i, \theta))}{\partial \theta} \right) \left(\frac{\partial \ln(p(y_i|x_i, \theta))}{\partial \theta} \right)^T \right], \quad (2.1)$$

(Sebastiani and Settimi, 1998; Retout and Mentré, 2003). An important consideration in optimal design theory is to focus on the normalized information matrix, by expanding the definition of a design to include a probability design measure ξ in the design space Ξ . ξ is a

probability measure with a finite number of design points. Any approximate (continuous) design ξ can be represented in the form

$$\xi = \left\{ \begin{array}{cccc} x_1 & x_2 & \dots & x_k \\ w_1 & w_2 & \dots & w_k \end{array} \right\}, \quad (2.2)$$

where x_1, x_2, \dots, x_k are design points, and w_1, w_2, \dots, w_k are the associated weights or proportions of subjects at the k distinct design points, respectively, with $\sum_{i=1}^k w_i = 1$ and $0 \leq w_i \leq 1$, for all i 's. The expected normalized information matrix for a design ξ supported with weight w_i at x_i is defined by

$$M(\xi, \theta) = \sum_{i=1}^k w_i I(x_i, \theta). \quad (2.3)$$

(see Sebastiani and Settimi, 1998; Braess and Dette, 2007; Atkinson et al., 2007). The information matrix (2.3) is a function of the unknown parameter vector θ and is inversely proportional to the asymptotic variance-covariance matrix of the parameter estimators for θ . The goal of optimal experimental design is the maximization of a function of (2.3). However, for non-linear models, the information matrix (2.3) depends on the unknown θ . We consider the Bayesian approach, where a prior distribution is assumed for θ , which is then incorporated into an appropriate design criterion. There are several such optimality criteria, such as D-, A- or G- optimality. Here we restrict ourselves to the D-criterion for two reasons: Firstly, the D-optimal design is not affected by linear transformation of the scale of independent variables. This means that if we rescale the design interval, we can deduce the D-optimal design directly from the one constructed over the original design interval. Such property does not hold for other optimality criteria. Secondly, D-optimality leads to minimization of the volume of the confidence ellipsoid for the parameters.

Most of the Bayesian optimality criteria functions are based on normal approximations to the posterior distribution of the parameter vector θ , as computations involving the exact posterior distribution are intractable, (Khuri et al., 2006). Several such approximations to the posterior are available (Berger, 1985 page 224) and involve either the observed or expected Fisher information matrix. The most common form of such a normal approximation states that under regularity conditions, the posterior distribution of θ is approximately multivariate normal with mean equal to the maximum likelihood estimate, $\hat{\theta}$, and variance-covariance matrix which is equal to the inverse of the observed Fisher information matrix. This approximation assumes that the sample size is large, so that the additional effect of the prior information about the parameter vector θ on the variance-covariance matrix is negligible. Using this approximation, many authors (Chaloner and Larntz, 1989; among others) define the Bayesian D-optimality criterion as a function of the non-singular information matrix (2.3), by analogy with the non-Bayesian theory of optimal design, as follows:

$$\phi_D(\xi|\pi) = E_\theta \log |M(\xi, \theta)| = \int_\theta \log |M(\xi, \theta)| \pi(\theta) d(\theta), \quad (2.4)$$

where $\pi(\theta)$ is a prior distribution for θ . The design which maximizes (2.4) is the Bayesian D-optimal design. There are several alternative forms of the Bayesian D-optimal criterion (see Atkinson et al., 2007; Berger and Wong, 2009). The design criterion (2.4) is the most natural Bayesian formulation of D-optimality criterion for non-linear models, because: a) this criterion arises naturally from the expected gain of Shannon information (Lindley, 1956; Chaloner and Verdinelli, 1995; Atkinson et al., 2007); b) for this criterion function, concavity is always guaranteed so that a global equivalence theorem will hold (Firth and Hinde, 1997); and c) its numerical results are the most satisfactory (Atkinson et al., 2007). In general, maximization of functions of the form (2.4) has to be done numerically for most of the commonly used models and prior distributions. The integration can be done numerically by sampling from the prior distribution (Atkinson et al., 2007). In this chapter, we concentrate on the logistic model since this is a very often applied model, but the theory and methods apply to any non-linear model.

2.2.1 The logistic regression design problem

Suppose that a Bernoulli response y_i is observed at a value x_i of the explanatory variable. The probability of success $p(y_i = 1)$ according to the logistic regression model with two parameters is

$$p(y_i = 1|x_i, \theta) = \frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)}, \quad (2.5)$$

where $\theta^T = (\alpha, \beta)$ is unknown, $x \in \mathfrak{R}$, and $-\alpha/\beta$ is the value of x at which the probability of success is 0.5 and β is the slope in the logit scale. The purpose is to find approximate (continuous) optimal designs. The log-likelihood function of observation i for (2.5) is given by:

$$\ln P(y_i|x_i, \theta) = y_i(\alpha + \beta x_i) - \ln(1 + \exp(\alpha + \beta x_i)). \quad (2.6)$$

Then using equation (1), the Fisher information matrix for θ is

$$I(x_i, \theta) = \frac{\exp(\alpha + \beta x_i)}{(1 + \exp(\alpha + \beta x_i))^2} \begin{bmatrix} 1 & x_i \\ x_i & x_i^2 \end{bmatrix}. \quad (2.7)$$

By substituting equation (2.7) in equation (2.3) we obtain the expected normalized information matrix of the logistic model for a design ξ . In order to find such designs for the logistic model (2.5) we use the Bayesian D-optimality criterion function in (2.4).

2.2.2 Relative Efficiency

To compare designs, we use their relative efficiencies. Several authors used the ratio of the two criteria values to compute their relative efficiencies. See for example, Yu et al. (2008), and Biedermann et al. (2011). In this chapter we use the formulation as given by Chaloner and Larntz (1989), and Song and Wong (1998), because it has a nice interpretation in terms of sample size. Suppose that design ξ^* is the optimal design constructed with design criterion $\phi_D(\xi^*|\pi)$. The relative efficiency of an arbitrary design ξ relative to the optimal design ξ^* is then defined as:

$$RE(\xi|\xi^*(\pi); \pi) = \exp\left\{\frac{\phi_D(\xi|\pi) - \phi_D(\xi^*|\pi)}{p}\right\}, \quad (2.8)$$

where p is the number of unknown parameters. This enables us to explore how good design ξ is compared to the optimal design ξ^* . If the value of this relative efficiency is close to unity, the design ξ is highly efficient. The inverse of the relative efficiency (2.8) determines the required sample size for a design ξ to have the same efficiency as the optimal design ξ^* . For instance, twice as many observations are needed for a design ξ with $RE(\xi|\xi^*(\pi); \pi) = 0.5$ to be as efficient as the optimal design ξ^* . Note that Chaloner and Larntz (1989) gave the inverse of (2.8), so that their relative efficiency (RE) becomes greater than one.

2.2.3 Maximin Bayesian procedure

The Bayesian optimal design can vary substantially when different prior distributions are used. For some examples, see Chaloner and Larntz (1989) and Khuri et al. (2006). Some of the designs may have a very low relative efficiency (2.8) compared to others. To find a prior distribution that leads to a highly efficient design under other prior distributions, we propose a maximin Bayesian procedure, which is based on the relative efficiency (2.8) and consists of four steps. Let the priors under consideration be π_1, \dots, π_q , and ξ_1, \dots, ξ_q be their respective Bayesian D-optimal designs.

1. Compute ξ_1, \dots, ξ_q .
2. Compute $RE(\xi_m|\xi_l; \pi_l)$ for $1 \leq m, l \leq q$ using (2.8).
3. For each m find $\min_l RE(\xi_m|\xi_l; \pi_l)$.
4. Choose m such that $\min_l RE(\xi_m|\xi_l; \pi_l)$ is maximized. This maximum relative efficiency value is referred to as the maximin value (MMV) and the corresponding design ξ_m is the maximin Bayesian D-optimal design. This design performs best in terms of ‘the worst case’.

So the MMV is defined as:

$$\text{MMV} = \max_{m=1,2,\dots,q} \left\{ \min_{l=1,2,\dots,q} \text{RE}(\xi_m | \xi_l; \pi_l) \right\} \quad (2.9)$$

2.2.4 Algorithms

The optimal designs were found using both numerical optimization and a step called consolidation. The program uses the FMINCON functions of MATLAB version 7.10.0499 (R2010a). This function performs constrained non-linear optimization. The program requires a number of starting points and so we repeat the optimization, starting with 2 design points with equal weights as initial values and increasing the number of design points steadily. We then choose the design which optimizes the criterion (2.4) on the smallest number of design points. After searching for optimal designs in this way, the global optimality of the design is then checked by evaluating the directional derivative over possible values of x (user defined) in the design space. Consolidation was proposed by Atkinson et al. (2007). There are two parts to the consolidation step:

1. Design points with small (< 0.0001) weights are dropped.
2. Nearby (< 0.1) design points are replaced by their average with weight equal to the sum of the weights of the averaged points.

A pseudo Monte Carlo method is used to sample from the priors to estimate the design criterion in equation (2.4), because it is simple to implement in practice. However, there are several sampling schemes alternatives to pseudo Monte Carlo sampling which are fast and use smaller samples, such as the quadrature method (Gotwalt et al., 2009) and the use of Latin Hypercube sampling (Woods and van de Ven, 2011). A large sample size is required to obtain a good integral approximation of the design criterion (2.4). Because of computation time, we let the sample size vary from 10^5 for the priors with the smallest variance to 5×10^5 for the priors with the largest variance. The computing time varied between 1 and 6 hours (computer processor: Intel Core i7-920, 2.67 GHz).

2.3 Numerical study and Results

2.3.1 Hyperparameter values

We considered uniform and normal prior distributions of α and β with bounded design regions of x for the numerical search of Bayesian D-optimal designs with criterion (2.4). Without loss of generality, we code x as $[-1, 1]$. It was assumed that α and β were independent and that there is some prior knowledge on the location of α and β . We choose five different values for each of the means μ_α and μ_β , i.e., $\mu_\alpha = [0, 0.5, 1, 1.5, 2]$ and $\mu_\beta = [1, 1.5, 2, 2.5, 3]$ and three variances, i.e., $\sigma^2 = [1, 5, 10]$. All 25 combinations of the mean

values were used to generate 25 different prior distributions per σ^2 . Thus in total there were 75 normal prior distributions and 75 uniform prior distributions. For example, the uniform prior distribution of $\mu_\alpha = 0$ and $\mu_\beta = 1$ with $\sigma^2 = 1, 5$ and 10 are $\{-1.73, 1.73\}$, $[-0.73, 2.73]$, $\{-3.87, 3.87\}$, $[-2.87, 4.87]$ and $\{-5.48, 5.48\}$, $[-4.48, 6.48]$, respectively.

2.3.2 Results

2.3.2.1 Bayesian D-optimal designs

Bayesian D-optimal designs were computed for each of the 75 prior distributions of α and β of the logistic model (2.5) for the uniform distributions and the normal distributions separately. The global optimality of the design was then checked for all the Bayesian D-optimal designs by evaluating the directional derivative over possible values of x . To save space we only present the Bayesian D-optimal designs for some choices of the prior distributions of the parameters. The Bayesian D-optimal designs for the uniform priors were similar to those of the normal priors. Hence, the results shown in Figure 2.1 are only for the uniform prior distributions. Figure 2.1 shows the Bayesian D-optimal allocations of design points for different choices of uniform prior distributions of α and β with $\sigma^2 = 1, 5$ and 10 . The plots in the three columns of Figure 2.1 correspond to prior means $\mu_\alpha = 0$, $\mu_\alpha = 1$ and $\mu_\alpha = 2$. Each set of row plots corresponds to $\sigma^2 = 1$, $\sigma^2 = 5$ and $\sigma^2 = 10$, respectively. In each plot, the vertical axis shows the prior mean values of β , and the horizontal axis shows the locations of the design points, and separate marker sizes are given for the different weights categories of the design points.

It can be observed from Figure 2.1 that the Bayesian optimal design varies substantially when different prior distributions are used. There are several interesting features of this plot. First of all, the number of Bayesian D-optimal design points increases with σ^2 , from 2 or 3 design points for $\sigma^2 = 1$ to 4 or 5 design points for $\sigma^2 = 5$, and to 5 or 6 design points for $\sigma^2 = 10$. Chaloner and Larntz (1989) and Khuri et al. (2006) have found a similar effect for the logistic model. Secondly, the Bayesian D-optimal designs are symmetric around $x = 0$ when $\mu_\alpha = 0$ and approximately also when σ^2 is large relative to μ_α . Atkinson and Haines (1996) have found a similar result for locally D-optimal designs for the logistic model with two explanatory variables. Third, the Bayesian optimal location of design points includes the two end points of the design region when μ_β , the mean value of the regression parameter, is small, and some of the design points move away from the end points of the design region as μ_β increases for $\sigma^2 = 1$. Atkinson and Haines (1996) and Chipman and Welch (1996) have found a similar phenomenon in their comparison of D-optimal design for generalized linear models. They show that the optimal design points for the logistic model ‘move in’ from the edge of the design region. Recently, Tekle et al. (2008) found a similar result for the logistic mixed-effects models using locally D-optimal designs.

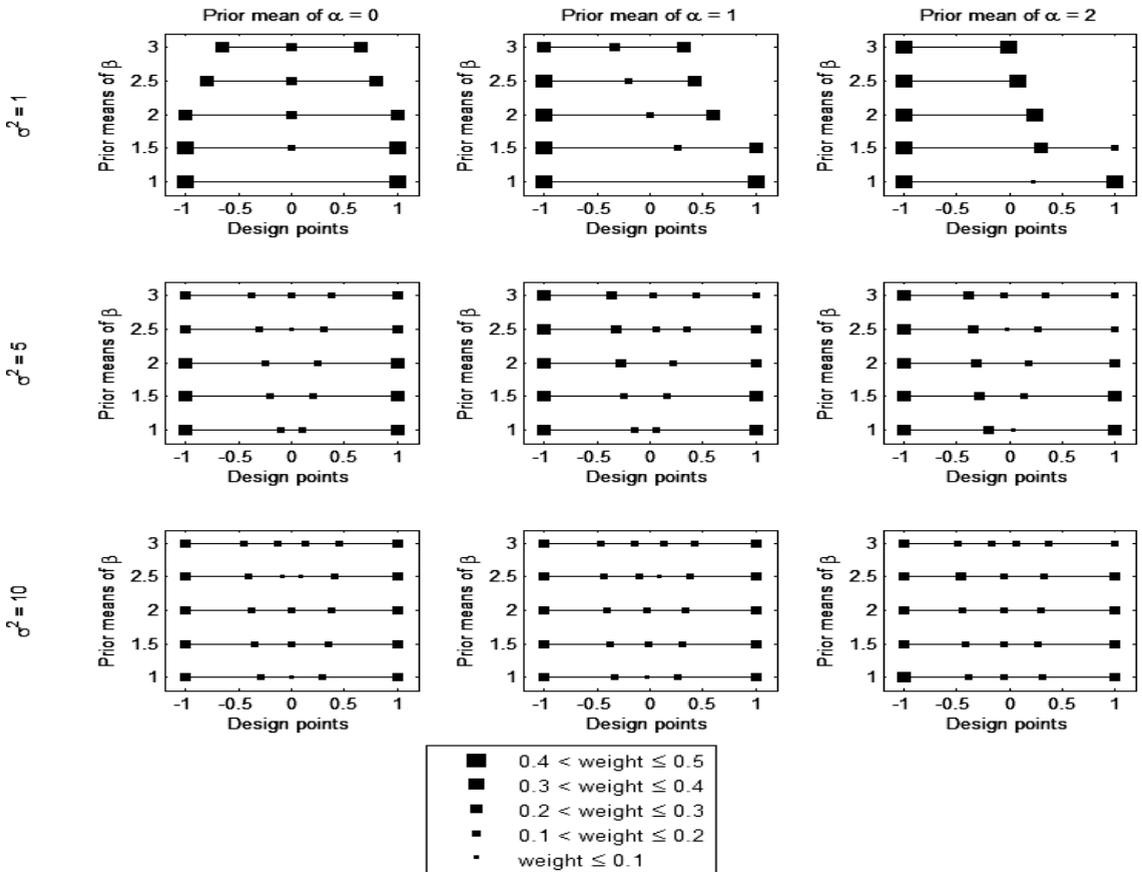


Figure 2.1 Bayesian D-optimal design points for independent uniform priors for the logistic model.

2.3.2.2 Choice of prior distributions

Bayesian D-optimal designs depend on the prior distributions of the parameters. Therefore, we want to determine what prior distribution leads to a Bayesian D-optimal design which remains highly efficient under other prior distributions, i.e., we look for a Bayesian optimal design which is robust against a change of the prior distribution. This can be done using the maximin Bayesian procedure. The plots in Figure 2.2 show the minimum relative efficiencies of the 75 Bayesian D-optimal designs (one design per prior with α and β) over all the 75 normal prior distributions of α and β with a separate plot per $\sigma^2 = 1, 5$ and 10 . In each plot, the horizontal axis represents the prior mean of α , and the vertical axis corresponds to the minimum relative efficiencies, and a separate line is given for each prior mean of β . It can be

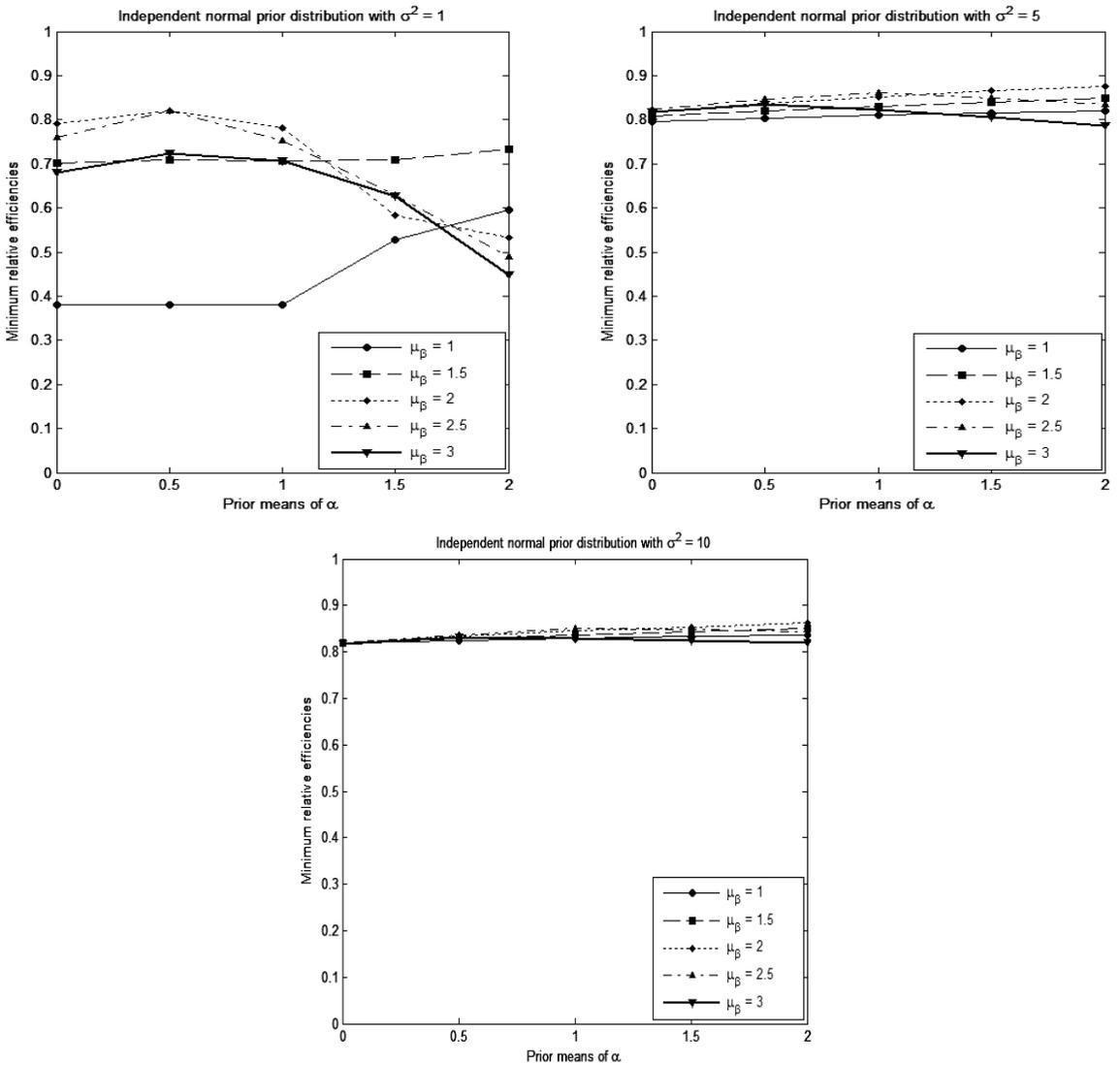


Figure 2.2 Minimum relative efficiency plots of Bayesian D-optimal designs for independent normal prior distributions.

seen from the plots that the maximin Bayesian D-optimal design is obtained by the prior with $\mu_\alpha = 2$, $\mu_\beta = 2$ and $\sigma^2 = 5$, and that it has a maximin value MMV of 0.8763. The prior with the same μ_α and μ_β but with $\sigma^2 = 10$ is also robust, with a minimum relative efficiency of

0.8615. In fact, all priors with $\sigma^2 = 5$ or 10 give a minimum relative efficiency above 0.80. In contrast, the Bayesian D-optimal designs for priors with $\sigma^2 = 1$ vary in minimum relative efficiency from 0.38 to 0.81, with the lowest efficiency occurring for small μ_α and μ_β . The plots for all uniform prior distributions gave very similar results to those of the normal priors and are therefore not shown here.

Figure 2.2 shows the counterintuitive result that the maximin Bayesian D-optimal design is not obtained by the prior with $\mu_\alpha = 1$ and $\mu_\beta = 2$ as expected, although the Bayesian D-optimal design for that prior is nearly as efficient as the maximin Bayesian D-optimal design, with minimum relative efficiency of 0.8523 for $\sigma^2 = 5$ and 0.8507 for $\sigma^2 = 10$. This result might perhaps be due to the effects of varying σ^2 . Let us therefore look at varying the prior means while keeping the σ^2 constant, by computing the relative efficiencies within each class of 25 priors with the same σ^2 . Table 2.1 shows the range and mean of minimum relative efficiencies computed over the 25 priors with the same σ^2 and likewise over all 75 priors for every value of the σ^2 . As expected the maximin Bayesian D-optimal design within the 25 priors with the same σ^2 is obtained by the prior $\mu_\alpha = 1$ and $\mu_\beta = 2$, with maximin values MMV of 0.8612, 0.9750 and 0.9907 for $\sigma^2 = 1, 5$ and 10, respectively.

Two other logical results are first, that the range of minimum relative efficiencies is smaller within the 25 priors per σ^2 than over the 75 priors, and decreases as the σ^2 increases, and second, that the mean of minimum relative efficiencies is larger within the 25 priors per σ^2 than over the 75 priors, and increases as the σ^2 increases. Very similar results were again obtained for uniform prior distributions (the results are not included here). Finally, note that the minimum relative efficiency within the 25 priors is very high (>0.95) for $\sigma^2 = 10$. This suggests that if a researcher has no information on μ_α and μ_β , and uses the maximin Bayesian D-optimal design with relatively large σ^2 , then the loss in efficiency incurred by a change of prior means is at most 5%, which can be compensated by sampling about 5% more subjects.

In conclusion, if a researcher has no information on the mean values of a prior distribution, choosing a design based on a prior with relatively large σ^2 will lead to Bayesian designs which remain highly efficient under other choices of prior means.

Table 2.1 Range (first row) and mean (second row) of minimum relative efficiencies of normal priors per σ^2 for the logistic model.

σ^2	Range and mean of minimum relative efficiencies	
	Within the 25 normal priors with the same σ^2	Over all 75 priors
1	(0.5279, 0.8612)	(0.3809, 0.8194)
	0.7179	0.7016
5	(0.9162, 0.9750)	(0.7854, 0.8763)
	0.9406	0.8221
10	(0.9588, 0.9907)	(0.8171, 0.8615)
	0.9755	0.8333

2.3.2.3 Comparison between normal and uniform prior distributions

The results in the previous section show that a Bayesian D-optimal design based on a prior distribution with a large σ^2 is relatively robust, that is, highly efficient for other priors as well. The next question is which distribution, uniform or normal prior for α and β leads to the most robust design. To answer this question, we computed the relative efficiency (RE) of the Bayesian D-optimal design for the uniform prior relative to the normal prior and vice versa, for every pair of μ_α and μ_β , and $\sigma^2 = 5$ (the results are not included here). The relative efficiencies of the two distributions with respect to each other were very similar and very high, always above 0.997. This suggests that the loss in efficiency due to switching from a normal to a uniform prior or vice versa, is at most 0.30%. Thus, choosing a uniform or normal prior does not seem to matter in terms of efficiency.

2.3.2.4 Performance of Bayesian optimal design relative to locally optimal designs

Robustness against a change of μ_α and μ_β by taking a large σ^2 comes at the price of some local efficiency loss. Therefore, we now study the performance of Bayesian D-optimal designs relative to locally D-optimal designs, which can be seen as Bayesian D-optimal designs for priors with zero variance. We assumed different true values for $\alpha = 0, 1$ and 2 , and for $\beta = 1, 2$ and 3 , respectively and $\sigma^2 = 5$. For each of the nine combinations, we evaluated the local relative efficiencies of the Bayesian D-optimal designs (with $\sigma^2 = 5$) with respect to the locally D-optimal designs (with $\sigma^2 = 0$). The local relative efficiencies are

shown in Figure 2.3 for different choices of normal prior distributions of α and β of the logistic model (2.5). The plots in the three columns of Figure 2.3 correspond to $\beta \sim N(1, 5)$, $\beta \sim N(2, 5)$ and $\beta \sim N(3, 5)$. Each set of row plots corresponds to $\alpha \sim N(0, 5)$, $\alpha \sim N(1, 5)$ and $\alpha \sim N(2, 5)$, respectively. In each plot, the horizontal axis represents the true value of α and the vertical axis corresponds to the local relative efficiencies and a separate line is given for each true value of β .

As can be seen from Figure 2.3, the local relative efficiency tends to decrease as the means of the prior distributions depart further from the true values of α and β , but is always above 0.70. If $\mu_\alpha = \text{true } \alpha$ and $\mu_\beta = \text{true } \beta$, then, the local relative efficiency is above 0.83 in all cases. We also computed the local relative efficiencies of Bayesian D-optimal designs with $\sigma^2 = 1$ with respect to the locally D-optimal designs for each of the 9 combinations of true values of α and β (the plots are not included here) and again the local relative efficiencies were largest when $\mu_\alpha = \text{true } \alpha$ and $\mu_\beta = \text{true } \beta$, that is, the local relative efficiency was above 0.94 in these cases, against values becoming as low as 0.48 if the means were not equal to the true values. In summary, if we use the Bayesian D-optimal designs instead of the locally D-optimal designs with prior means equal to the true values, then the local loss in efficiency is at most 6% for $\sigma^2 = 1$, against 17% for $\sigma^2 = 5$. So robustness against a change of prior mean by taking a large σ^2 has the drawback of a lower local efficiency compared with a prior with the same mean but smaller variance. Finally, we computed the local relative efficiencies of the Bayesian D-optimal designs for uniform prior distributions of α and β with $\sigma^2 = 5$ compared to the locally D-optimal designs, and found very similar results as for the normal priors.

To conclude, if the means of the prior distributions are far away from the true values of α and β , Bayesian D-optimal designs based on a prior with relatively large σ^2 are much more efficient than Bayesian D-optimal designs based on a prior with relatively small σ^2 . In contrast, if the means of the prior distributions are close to the true values of α and β , Bayesian D-optimal designs based on a prior with relatively small σ^2 are somewhat more efficient.

Thus, if a researcher has some information on the means of prior distribution, he or she should choose a design based on prior distributions with relatively small σ^2 , or use the locally optimal design. However, if there is little or no information, such designs can lead to a substantial efficiency loss and a design based on a prior distribution with relatively large σ^2 should be used.

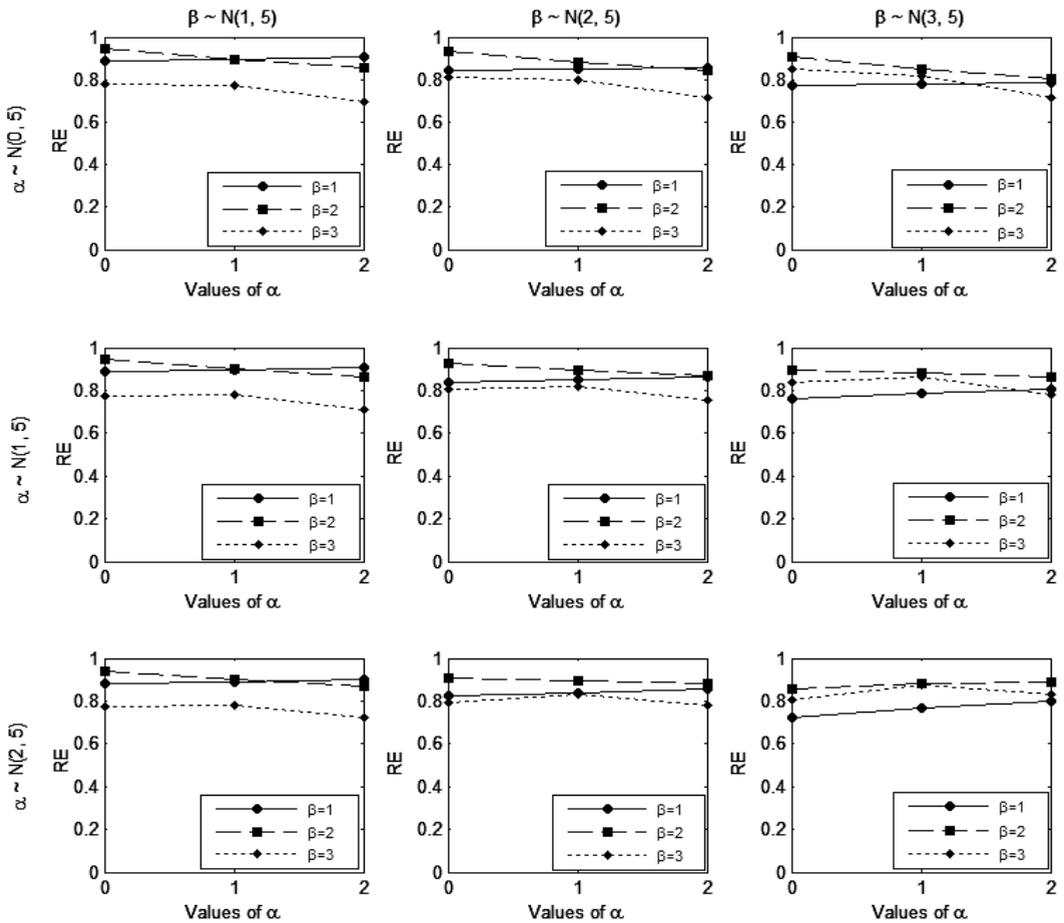


Figure 2.3 Relative efficiency plots of Bayesian D-optimal designs versus locally D-optimal designs for various true values of α and β .

2.3.2.5 Equidistant equally weighted design points relative to Bayesian optimal design

The results in the previous sections show that choosing a design based on a prior with a relatively large σ^2 will lead to a highly efficient design under other prior distributions. Finally, let us have a look at the efficiencies of equidistant equally weighted design points, with the first and last point on the boundaries of the design space, against the Bayesian D-optimal designs for each of the 25 priors for $\sigma^2 = 1, 5$ and 10 . Table 2.2 shows the minimum relative efficiencies of 2, 3, 4, 5 and 6 equidistant equally weighted design points computed over the 25 normal priors with the same σ^2 . It can be seen that 4 and 5 equidistant equally weighted design points are very efficient, especially for large σ^2 , and the loss in efficiency is

then smaller than 8%. A very similar result was obtained for uniform prior distributions (the results are not shown here). In general, choosing a design with 4 or 5 equidistant equally weighted design points seems to be a good choice in terms of robustness.

Table 2.2 The minimum relative efficiencies of equidistant equally weighted design points (with the first and last point on the boundaries of the design space) over all the 25 normal priors per σ^2 .

Equidistant equal weighted design points	Minimum relative efficiencies		
	$\sigma^2 = 1$	$\sigma^2 = 5$	$\sigma^2 = 10$
2	0.5279	0.4856	0.3809
3	0.8168	0.8939	0.8600
4	0.8156	0.9419	0.9547
5	0.8075	0.9287	0.9631
6	0.6566	0.7667	0.7931

2.4 Conclusion

Literature on Bayesian optimal design generally does not discuss or justify the choice of a prior distribution, even though it is known that the choice of the prior distribution can be highly subjective and can affect the design substantively. The main purpose of this chapter was to investigate how researchers could choose a prior distribution which leads to a Bayesian optimal design that remains highly efficient under other prior distributions.

Our results show that if we have little information about the mean values of the prior distribution, a prior distribution with relatively large variance will lead to a Bayesian design which remains highly efficient under other prior distributions, i.e., is robust against a change of prior distribution. In this case Bayesian D-optimal designs based on a prior distribution with relatively small variances may lead to very inefficient designs. We also investigated which prior distribution, uniform or normal leads to the most efficient design and found that both distributions are generally equally efficient with respect to each other. Finally, we studied the efficiencies of equidistant equally weighted design points against the Bayesian D-optimal designs and found that 4 and 5 equidistant equally weighted design points are highly efficient. The prior distributions used in this chapter, normal and uniform for the parameters of the logistic model, may not always be appropriate but they serve to illustrate our technique.

The methodology that we have described in this chapter can be applied to other non-linear design problems as well.

In conclusion, if a researcher has little information about the mean values of the prior distribution, he or she either should choose a Bayesian design based on a prior distribution with relatively large variance or should choose 4 or 5 equidistant equally weighted design points, because these design points are more or less robust against a change of the means of the prior distributions. In contrast, if the researcher uses a prior distribution with relatively small variance, the efficiency loss incurred by changing the means of the prior can be substantial. On the other hand, if reasonable information is available about the mean values of the prior distribution, designs based on a prior distribution with relatively large variance may lead to a loss of efficiency. Thus it is important that researchers choose their priors very carefully before the design is implemented. It is re-assuring that the choice between normal and uniform prior appears to have no consequences.

CHAPTER 3

Bayesian D-optimal designs for the two parameter logistic mixed effects model

This chapter is published in *Computational Statistics and Data Analysis* (2014), 71: 1066-1076, with co-authors Frans E.S. Tan, Gerard J.P. Van Breukelen and Martijn P.F. Berger

Abstract

Bayesian optimal designs for binary longitudinal responses analyzed with mixed logistic regression describing a linear time effect are considered. In order to find the optimal number and allocations of time points, for different priors, cost constraints and covariance structures of the random effects, a scalar function of the approximate information matrix based on the first order penalized quasi likelihood (PQL1) is optimized. To overcome the problem of dependence of Bayesian designs on the choice of prior distributions, maximin Bayesian D-optimal designs are proposed. The results show that the optimal number of time points depends on the subject-to-measurement cost ratio and increases with the cost ratio. Furthermore, maximin Bayesian D-optimal designs are highly efficient and robust under changes in priors. Locally D-optimal designs are also investigated and maximin locally D-optimal designs are found to have much lower minimum relative efficiency and fewer time points than maximin Bayesian D-optimal designs. When comparing the efficiencies of designs with equidistant time points with the Bayesian D-optimal designs, it was found that three or four equidistant time points are advisable for small cost ratios and five or six equidistant time points for large cost ratios.

Keywords: Bayesian D-optimal designs; logistic mixed effects model; maximin Bayesian D-optimal design; maximin locally D-optimal designs; minimum relative efficiency.

3.1 Introduction

Over the past few decades the design of longitudinal studies for dichotomous data has not received much attention from statisticians. To a large extent the precision of estimators of the unknown parameters of the statistical model depends on the design used in the experiment (Khuri et al., 2006). A researcher designing a longitudinal study has to decide on the number and allocation of time points. Optimal design theory can help in making this decision. However, optimal design depends on the unknown parameter values in non-linear models like the logistic regression model. Much effort has been devoted to this problem. A common approach is to design an experiment to be efficient for a best guess of the parameter values, which leads to locally optimal designs (see, e.g., Chernoff, 1953; Abdelbasit and Plackett, 1983). Such a design, however, may not be efficient for other parameter values. Thus, the ability to find locally optimal designs still faces the serious problem of how to take into account the uncertainty about the parameter values.

To handle the problem of local optimality, three methods are proposed in the literature namely, the sequential procedure (see, e.g., Abdelbasit and Plackett, 1983; Wu, 1985), the minimax or maximin approach (see, e.g., Dette, 1997; Ouwens et al., 2002; Tekle et al., 2008; Tan, 2011), and the use of Bayesian methods, described by Chaloner and Larntz (1989), Chaloner and Verdinelli (1995), Atkinson et al. (2007), Tommasi and López-Fidalgo (2010), and Drovandi et al. (2013). Due to restrictions on the length of the research period, the sequential approximation of an optimal design may not be feasible. A potential objection to the maximin approach is that the maximin design may occur at the edges of the parameter space and, if some points in the parameter space have a very low probability, then their weight may be over-emphasised (Atkinson et al., 2007).

In the present chapter, we will use the Bayesian approach, which is a more natural approach that formally accounts for the prior uncertainty of the parameter values. The Bayesian design literature outside the normal linear model is mostly restricted to binary response models (see, e.g., Khuri et al., 2006). Optimal designs for fixed effects models for binary data, especially logistic models have been investigated in the literature. Optimal designs were discussed by Bischoff (1993), Mentré et al. (1997), Atkins and Cheng (1999), Tan and Berger (1999), Berger and Tan (2004) and Berger and Wong (2009) for the linear random effects model. Recently, Ouwens et al. (2006) and Tekle et al. (2008) have extended the work on optimal designs for logistic models with random effects using maximin D-optimal designs. However, to the authors' knowledge no work has been done on Bayesian optimal designs for longitudinal studies with a binary response. In this chapter, we will investigate numerically Bayesian D-optimal designs for the two parameter logistic mixed effects model with repeated measures of a binary response.

The chapter is organized as follows. The next section introduces the logistic mixed effects model and the variance-covariance matrix of parameter estimators. In Section 3.3 the Bayesian D-optimality criterion and the corresponding algorithm to construct these designs are given as well as a discussion of a linear cost function and the relative efficiency as a

measure for the comparison of designs. In section 3.4 a numerical study and its results are presented, followed by the conclusions in section 3.5.

3.2 Logistic mixed effects models

Let the $q \times 1$ vector $\mathbf{y}_i = (y_{i1}, \dots, y_{iq})'$ be the binary responses y_{ij} (either 0 or 1) of subject i at q time points, $i = 1, 2, \dots, N$ and $j = 1, \dots, q$. It is assumed that, conditional on the subject-specific random effect vector \mathbf{b}_i , the binary responses y_{ij} are independently Bernoulli distributed with probability of success $p(y_{ij} = 1 | \mathbf{b}_i)$. These probabilities are related to the fixed and random-effects via the logit link. The corresponding logistic mixed effects model is given by:

$$\text{logit}(p(y_{ij} = 1 | \mathbf{b}_i)) = \log\left(\frac{p(y_{ij} = 1 | \mathbf{b}_i)}{1 - p(y_{ij} = 1 | \mathbf{b}_i)}\right) = \mathbf{x}_j' \boldsymbol{\beta} + \mathbf{z}_j' \mathbf{b}_i, \quad (3.1)$$

where the $p \times 1$ vector \mathbf{x}_j is the design vector of the explanatory variables at the j th measurement for subject i , $\boldsymbol{\beta}$ is the corresponding $p \times 1$ vector of unknown fixed effects, and \mathbf{z}_j is the $r \times 1$ design vector for the random effects that is usually a subset of vector \mathbf{x}_j , \mathbf{b}_i is the corresponding $r \times 1$ vector of unknown random effects, which is assumed to have a multivariate normal distribution with mean zero and covariance matrix \mathbf{D} .

To simplify, let us assume that all subjects have measurements at the same time points, but this can be relaxed. In this chapter, we consider a linear ($p = 2$) time effect, i.e., the design vector is $\mathbf{x}_j' = (1 \ t_j)$ and $\boldsymbol{\beta} = (\beta_0 \ \beta_1)'$, where t_j is the time point of the j th measurement, and β_0 and β_1 are the fixed intercept and slope effects. Suppose that for the linear time effect, a random intercept and random slope are assumed. The design vector is then $\mathbf{x}_j' = \mathbf{z}_j' = (1 \ t_j)$ and $\mathbf{b}_i = (b_{0i} \ b_{1i})'$, where b_{0i} and b_{1i} are the corresponding random (subject-specific) deviations from these fixed effects. Thus, according to model (3.1), the log-odds of a positive response for subject i at time t_j is given by:

$$\text{logit}(p(y_{ij} = 1 | \mathbf{b}_i)) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_j. \quad (3.2)$$

The objective of this chapter is to determine the optimal number of time points q and the optimal allocation of these time points t_j ($j = 1, \dots, q$), where ‘optimal’ means that the parameters $\boldsymbol{\beta}$ are estimated as efficiently as possible.

3.2.1 Variance-covariance matrix of the parameter estimators

The log-likelihood cannot be written down in closed form due to the random effects in model (3.1). Hence, either numerical methods or approximations to the log-likelihood must be used.

Numerical methods require large computational resources and more importantly they require full knowledge of the data (Moerbeek et al., 2003; Han and Chaloner, 2004; Tekle et al., 2008), making them computationally inconvenient for optimal design procedures. Therefore, we will use an approximate information matrix based on a first order penalized quasi-likelihood (PQL1). See Breslow and Clayton (1993), and Jang and Lim (2009) for details. PQL1 performs best in terms of point estimation since it produces the smallest mean squared error (MSE) and the bias of the estimators decreases as the sample size increases (Breslow and Clayton, 1993; Moerbeek et al., 2003; Jang and Lim, 2009). The asymptotic variance-covariance matrix of $\hat{\boldsymbol{\beta}}$, which is the inverse of the information matrix of the parameters, is approximated in PQL1 by:

$$\text{var}(\hat{\boldsymbol{\beta}}) \approx (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}, \quad (3.3)$$

where $\hat{\boldsymbol{\beta}}$ is the estimator of $\boldsymbol{\beta}$ for the logistic mixed-effect model (3.1), \mathbf{X} is the $Nq \times p$ design matrix formed by stacking $\{\mathbf{x}_j'\}$ for N subjects, and the $Nq \times Nq$ block-diagonal matrix \mathbf{V} has N blocks of $q \times q$ variance-covariance matrices given by:

$$\mathbf{v}_i \approx \mathbf{w}_i^{-1} + \mathbf{Z}\mathbf{D}\mathbf{Z}', \quad (3.4)$$

where \mathbf{Z} is the $q \times r$ design matrix with rows \mathbf{z}_j' , $j = 1, \dots, q$, the $r \times r$ matrix \mathbf{D} is the variance-covariance matrix of the random effects, and \mathbf{w}_i^{-1} is the diagonal matrix of the conditional variances of the transformed responses given the random effects \mathbf{b}_i , which is equal to the inverse of the diagonal matrix of the conditional variances of the untransformed responses given the random effects \mathbf{b}_i (Breslow and Clayton, 1993; Moerbeek et al., 2001; Jang and Lim, 2009). The diagonal matrix of the conditional variances of the untransformed responses given the random effects \mathbf{b}_i , is given by:

$$\mathbf{w}_i = \text{diag}(w_{i1}^{b_i}, \dots, w_{iq}^{b_i}), \quad (3.5)$$

where $w_{ij}^{b_i} = \text{var}(y_{ij} | \mathbf{b}_i)$, for $i = 1, \dots, N$, $j = 1, \dots, q$. Since the random effects are unknown in the design stage, we will generate \mathbf{b}_i from a multivariate normal distribution with mean zero and variance-covariance \mathbf{D} .

3.3 Optimal designs and relative efficiency

3.3.1 Bayesian D-optimal designs

Let the design space with $q \times 1$ time vectors be denoted by Ξ , i.e. $\Xi = \{(t_1, \dots, t_q) : t_j \in [a, b]\}$. A design ξ_q is an element of the design space Ξ if it has q time points at which measurements

are taken for all N subjects. The amount of information in a design is measured by the information matrix, which is approximately equal to the inverse of the asymptotic variance-covariance matrix of the parameter estimators in (3.3). The D-optimal design ξ_q^* is the design among all possible designs ξ_q with q time points for which the determinant of the approximate information matrix ($\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$) is maximized or equivalently, the determinant of the variance-covariance matrix is minimized (Atkinson et al., 2007 page 151). The approximate information matrix depends on the unknown vector of regression parameters $\boldsymbol{\beta}$. The Bayesian approach resolves this dependence on unknown parameters by taking the expectation of the log of the determinant of the approximate information matrix over a prior distribution for the regression parameters $\boldsymbol{\beta}$, and then maximizing that expectation as a function of the design. Thus, the Bayesian D-optimality criterion as a function of the variance-covariance matrix (3.3), by analogy with the non-Bayesian optimal design, is defined as follows:

$$\phi_D(\xi|\pi) = E_{\boldsymbol{\beta}} \log |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| = \int \log |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| \pi(\boldsymbol{\beta}) d(\boldsymbol{\beta}), \quad (3.6)$$

where $\pi(\boldsymbol{\beta})$ is a prior distribution for $\boldsymbol{\beta}$. The design which maximizes (3.6) is the Bayesian D-optimal design. In this chapter, we will use this criterion because D-optimality minimizes the volume of the asymptotic confidence ellipsoid of the parameters and it does not depend on the coding used for the endpoints of the chosen time interval $[a, b]$. If the coding is transformed linearly, for instance into $[0, 1]$, then the D-optimal design for the new time interval is obtained by applying the same linear transformation to the D-optimal time points for the old interval (See Ouwens et al., 2006).

3.3.2 Implementation

The Bayesian optimal designs for logistic mixed effects models were obtained numerically by optimizing the criterion in (3.6). This involved two steps: 1) computation of criterion (3.6) for a given design, and 2) finding an optimal design by maximization of the criterion value. Details of this procedure will now be given.

3.3.2.1 Sampling parameter values from priors to compute the criterion

To construct Bayesian optimal designs for continuous prior distributions, all candidate designs must be evaluated in terms of their criterion values as defined by (3.6). This in turn requires a complicated integration over the prior distribution, which can not be easily done analytically. Numerical approximations can be done by sampling parameter values from the prior distribution and then replacing the integral in (3.6) with a summation over the sample (Atkinson et al., 2007; Chaloner and Verdinelli, 1995). Estimating (3.6), using the traditional sampling (pseudo Monte Carol) method requires very large samples from the prior to

minimize sampling error in the design choice. In this chapter, we use an Adaptive Rejection Metropolis Sampling (ARMS) algorithm (Gilks and Wild, 1992; Gilks, et al., 1995), which is a more efficient sampling algorithm that requires a smaller sample to obtain a good approximation of the design criterion (3.6).

3.3.2.2 Optimization algorithm for finding an optimal design

A program using Matlab's function FMINCON was implemented in MATLAB version 7.10.0499 (R2010a). The program performs constrained non-linear optimization and starts with an initial design ξ_0 . Without loss of generality, we coded the time interval as $[-1, 1]$, and we used equally spaced time points as initial designs. The following global search algorithm was used to obtain the Bayesian D-optimal designs for a given prior distribution of the parameters:

1. Take samples from the prior distributions of the parameters using ARMS.
2. Compute the Bayesian D-optimal design using $q = p$ equidistant time points as initial design, where p is the number of unknown fixed parameters of the model.
3. Increase the number of time points q by one and perform step 2 again to find the Bayesian optimal design (allocation) for the new value of q . Repeat step 2 and 3 until the maximum number of time point q (user specified) is reached.
4. Afterwards, select the optimal number of time points q for the Bayesian D-optimal design by computing the relative efficiencies of designs with different numbers of time points, for various cost constraints (to be discussed in the next sub-section).

3.3.3 Cost consideration and relative efficiency

Adding more measurements per subject obviously provides additional information and increases the precision of estimating the unknown parameters. However, in practice, the number of subjects and number of time points at which measurements are taken per subject are often limited by budget constraints. The addition of subjects and measurements per subject will increase the costs of the study. Therefore, it is reasonable to take into account the costs of measuring when the designs are compared. There are two main components of the costs of obtaining measurements for subjects. These are the costs for recruitment of a subject and the costs of the measurements once a subject is recruited. Let the cost per subject be C_1 and the cost per measurement be C_2 . Then the total costs (budget) of a longitudinal study with q time points for N subjects, excluding overhead cost can be represented by the linear cost function:

$$C = C_1N + C_2Nq = NC_2(k + q), \quad (3.7)$$

where $k = C_1/C_2$ is the ratio of the costs of adding a new subject to the costs of an additional observation per subject. This cost ratio will be varied from 0 through 1 to large values, to study its effect on the optimal design.

Relative efficiency can be used to compare different designs. To fairly compare different designs, we will consider designs with the same total costs. This means that the designs can differ in terms of the number of subjects N and the number and timing of the measurements q . We first computed the Bayesian D-optimal designs using N fixed and then corrected for costs and varying q and N as follows. Let $\phi_D(\xi_q^*|\pi)$ denote the value of design criterion (3.6) for the optimal design ξ_q^* with q time points. Then the relative efficiency (RE) of an arbitrary design ξ_s with s time points relative to the optimal design ξ_q^* is defined as:

$$\text{RE}(\xi_s; \xi_q^*(\pi)|\pi) = \frac{k+q}{k+s} \left[\exp \left\{ \frac{\phi_D(\xi_s|\pi) - \phi_D(\xi_q^*|\pi)}{p} \right\} \right], \quad (3.8)$$

where $q, s \geq p$. If the value of this relative efficiency is close to unity, the design ξ_s is equally efficient as the optimal design ξ_q^* . Note that this relative efficiency (3.8) is an extension of the relative efficiency given by Chaloner and Larntz (1989), and when the ratio $\frac{k+q}{k+s}$ is one, it reduces to the relative efficiency in Chaloner and Larntz (1989).

3.3.4 Maximin Bayesian procedure

While Bayesian optimal design solves the problem of local optimality, i.e. takes into account the uncertainty about the unknown parameters β , it faces a new problem, that of dependence on the choice of the prior distribution for β . To find a Bayesian D-optimal design that is highly efficient under various prior distributions, we propose a maximin Bayesian procedure. This is based on the relative efficiency (3.8) and involves 4 steps:

1. Compute the Bayesian D-optimal design ξ_l^* , for $l = 1, 2, \dots, R$, where R is the number of different prior distributions.
2. Compute the relative efficiencies $\{\text{RE}(\xi_m^*; \xi_l^*(\pi_l)|\pi_l)\}$, for $l, m = 1, 2, \dots, R\}$ for each pair of designs ξ_m^*, ξ_l^* using (3.8).
3. Select for each design ξ_m^* its smallest relative efficiency $\{\min_{l=1,2,\dots,R} \text{RE}(\xi_m^*; \xi_l^*(\pi_l)|\pi_l)\}$.
4. Select from the whole set of R smallest relative efficiencies the largest relative efficiency value. This largest relative efficiency value is referred to as the maximin value (MMV) and the corresponding design is the maximin Bayesian D-optimal design for these R different prior distributions. This Bayesian design is the most robust for the whole set of prior distributions.

So the MMV is defined as:

$$\text{MMV} = \max_{m=1,2,\dots,R} \left\{ \min_{l=1,2,\dots,R} \text{RE}(\xi_m^*; \xi_l^*(\pi_l)|\pi_l) \right\}. \quad (3.9)$$

3.4 Numerical study and Results

3.4.1 Models and Hyperparameter values

A number of different logistic models and prior distributions of the fixed parameters were studied. We considered the logistic mixed models with linear ($p = 2$) time effect and three different covariance structures \mathbf{D} (to be discussed below). We assumed an independent bivariate normal prior for the fixed effects β_0 and β_1 , and we also considered three different prior means (location parameters) for β_0 and β_1 , i.e., $\mu = [0, 1, 2]$, giving nine different priors and a range from 0.12 to 0.98 for the response probability in (3.1). We took a large prior variance, $\sigma^2 = 5$ for both fixed effects, as Abebe et al. (2014a) have found $\sigma^2 = 5$ to give designs which are robust under a change of the location parameters in the context of fixed logistic regression of cross-sectional data. We only consider positive location parameters for the prior because negative values for both location parameters gives the same designs as positive values for both. This follows from the fact that the information, which is a monotonic function of $p(y_{ij} = 1 | \mathbf{b}_i)(1 - p(y_{ij} = 1 | \mathbf{b}_i))$ remains the same. Having one negative location parameter and the other one positive needs reflected design points to obtain the same information.

The following three random effects covariance structures \mathbf{D} were considered:

- $\mathbf{D}_1 = 0.5$, i.e., a random intercept model with variance of the random intercept 0.5.

- $\mathbf{D}_2 = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$, i.e., a random intercept and random slope model with variances $d_{11} = 0.5$

for the intercept and $d_{22} = 0.5$ for the slope, and no slope-intercept covariance.

- $\mathbf{D}_3 = \begin{bmatrix} 0.5 & 0.25 \\ 0.25 & 0.5 \end{bmatrix}$, i.e., the same as \mathbf{D}_2 except that there is now a slope-intercept covariance of 0.25, giving a correlation of 0.5.

As described in section 3.2.1, the PQL1 method needs the random effects \mathbf{b}_i in order to calculate the diagonal matrix of the conditional variances of the responses (3.5), which in turn is needed to compute the \mathbf{V} -matrix in the efficiency criterion (3.6). Therefore, we sampled the values of \mathbf{b}_i from a multivariate normal distribution with mean zero and covariance matrix \mathbf{D} . Note that this must not be confused with the sampling described in section 3.3.2.1, which concerns the sampling of the fixed effects from their prior distribution to approximate the integral in (3.6). Here, we are concerned with the sampling of random effects from their distribution to estimate the determinant in (3.6).

To determine what sample size N from the random effects distribution is needed to give a stable estimator of the determinant in (3.6), we ran simulations with $N = [25, 50, 75, 100, 125, 150, 175, 200]$ from a multivariate normal with covariance structure \mathbf{D}_2 , generating 300 runs per value of N and computing the determinant per run. Figure 3.1 presents a box plot of these 300 determinants for every N , showing that the determinant becomes stable (i.e. constant over

runs) if N is at least 100. We repeated the simulation study with the covariance structures \mathbf{D}_1 and \mathbf{D}_3 and found very similar results. Therefore, we used $N = 100$ to estimate the determinant in (3.6) as part of the search for Bayesian D-optimal designs.

We sampled 30,000 parameter values of fixed effects from their priors to estimate the design criterion in (3.6). We restricted the number of time points q to a maximum of seven measures, which covers the number of time points in most longitudinal studies because of computation time. Different cost ratios $k = 0, 0.5, 1, 1.5, 2, 2.5$ and 5 were considered. This means when $k = 5$, the costs of recruiting a new subject is five times as expensive as the cost of one measurement. Cost ratios above 5 did not further change the design due to the constraint of at most seven time points. We performed all computations in MATLAB R2010a (32bit) using a computer with a 2.67 GHz Intel processor and 6 GB RAM. The total computing time to generate all the Bayesian optimal time points $2 \leq q \leq 7$ is about 72 hours per prior per covariance structure. A written MATLAB code is available and can be obtained from the first author upon request.

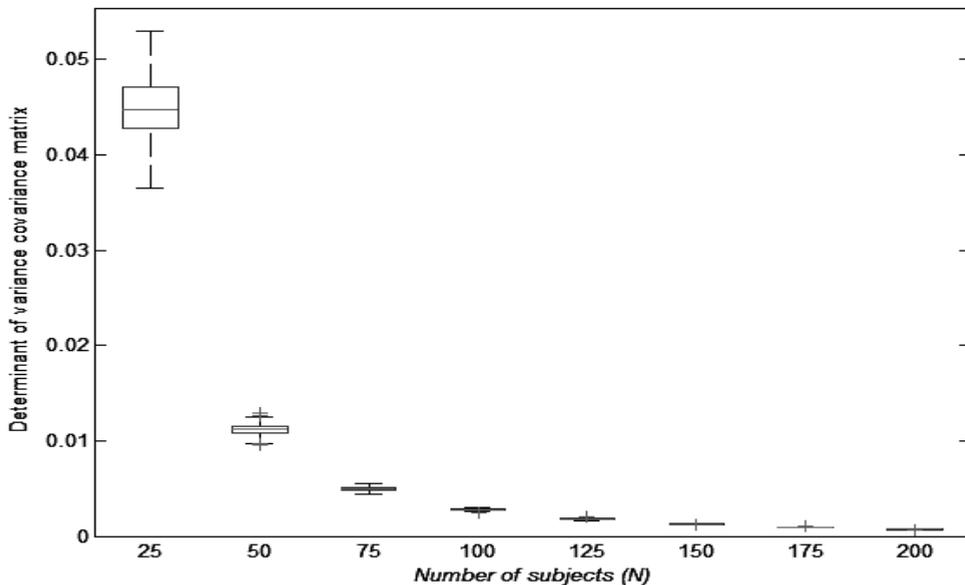


Figure 3.1 Box plots that show the variability of the determinant of the variance covariance matrix of parameter estimates from 300 different simulations per value of N .

3.4.2 Results

3.4.2.1 Bayesian D-optimal designs

Firstly, the Bayesian D-optimal design was computed for each number of time points $q \leq 7$ and for each of the nine different prior distributions and each of the three covariance structures. Thereafter, the Bayesian D-optimal designs for different q were compared with

each other, per prior per covariance structure and per cost ratio k . To save space, we only present the Bayesian optimal allocations of time points for $q = 2, 3$ and 4 and covariance structure D_2 . Figure 3.2 shows the Bayesian D-optimal allocations of time points for the nine priors of β for the logistic mixed effects model with linear time effect. The plots in the three columns of Figure 3.2 correspond to the prior distribution of β_0 . Each set of row plots corresponds to the number of time points $q = 2, 3$ and 4, respectively. In each plot, the vertical axis shows the prior mean values of β_1 and the horizontal axis shows the locations of the optimal time points. It can be observed that the Bayesian D-optimal design points depend more on the prior for β_1 than on the prior for β_0 . The Bayesian optimal location of design points includes the two end points of the design region when μ_{β_1} is small and for $q = 2$ some of the design points move away from the end points of the design region as μ_{β_1} increases. The Bayesian D-optimal time points are symmetric around $t = 0$ when $\mu_{\beta_0} = 0$ and become less symmetric as μ_{β_0} moves away from zero. This follows from the fact that the information, which is a monotonic function of $p(y_{ij}=1|\mathbf{b}_i)(1-p(y_{ij}=1|\mathbf{b}_i))$ is the same at time point t as at time point $-t$, if $\mu_{\beta_0} = 0$. Abebe et al. (2014a) found a similar phenomenon for a fixed effects logistic model using Bayesian D-optimal designs.

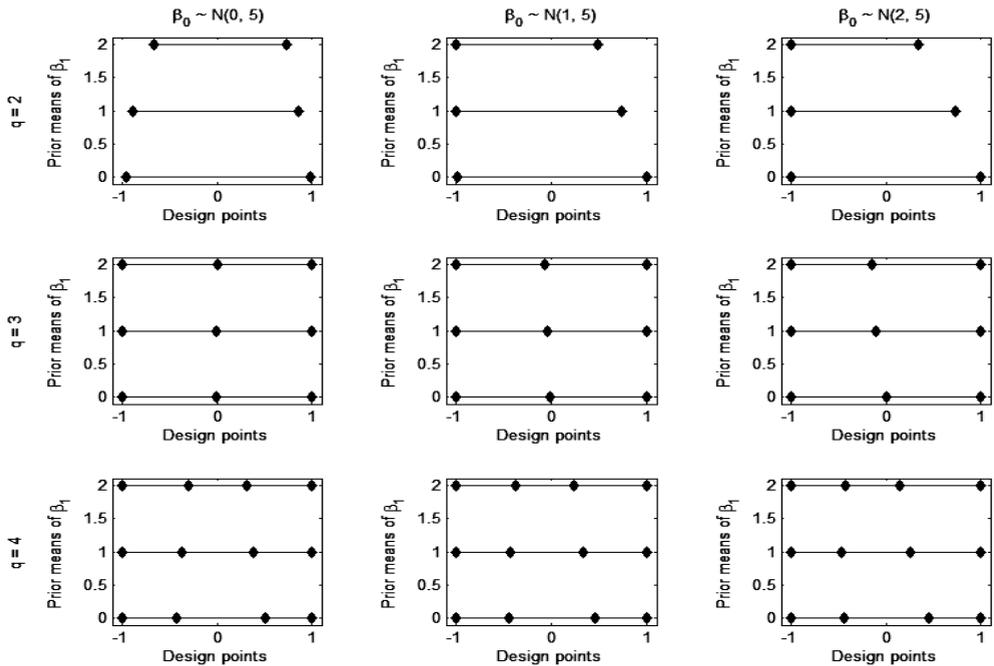


Figure 3.2 Bayesian D-optimal allocations of time points for the logistic mixed effects model with linear time effects and covariance structure D_2 .

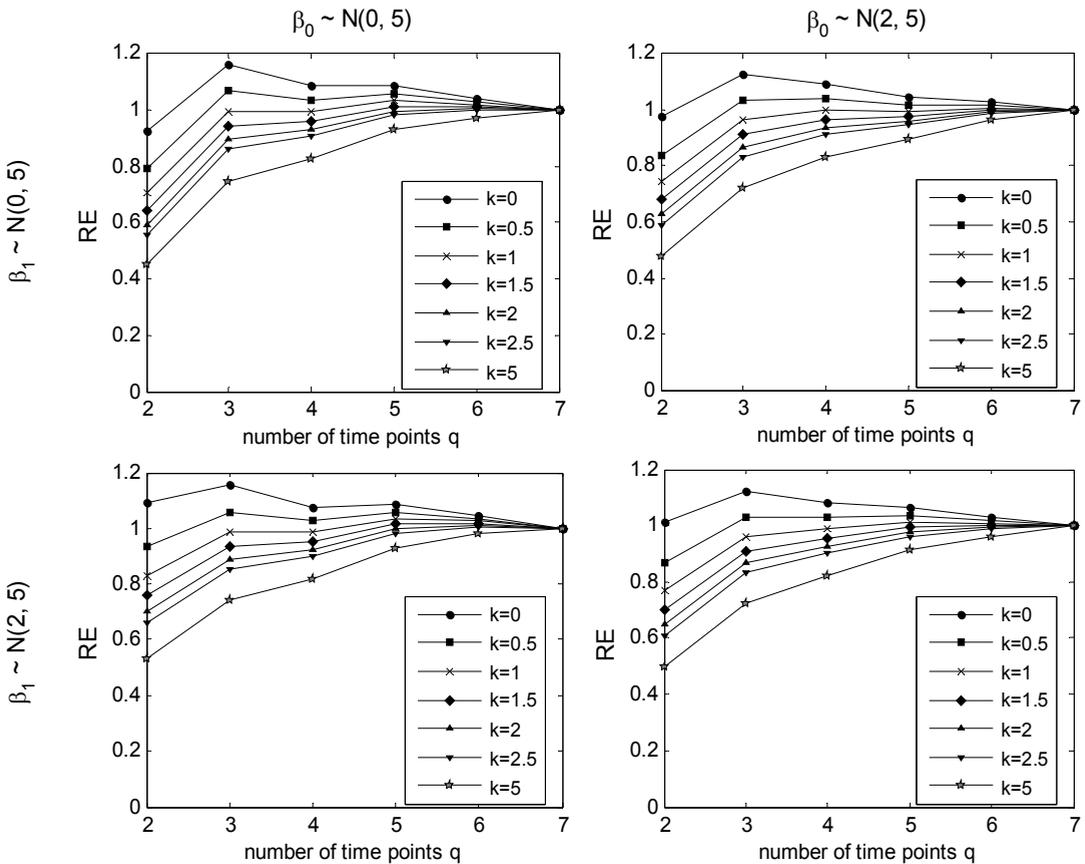


Figure 3.3 Relative efficiency of Bayesian optimal designs of q time points compared to a Bayesian optimal design with $q=7$ time points for different cost ratios k for the logistic mixed effects model with linear time effects and covariance structure \mathbf{D}_2 .

3.4.2.2 Selecting the optimal number of time points for a given cost ratio k

To find the optimal number of time points q for the Bayesian D-optimal design for a given prior distribution and cost ratio k , we computed the relative efficiencies of designs with $q < 7$ time points compared to a design with $q = 7$ time points within a given prior, \mathbf{D} and cost ratio k . Because the relative efficiency plots for different priors have a similar pattern, we will only present the relative efficiencies for the extreme prior means. Figure 3.3 shows the relative efficiency plots of $q = 2, 3, \dots, 7$ time points compared to a design with $q = 7$ time points for covariance structure \mathbf{D}_2 and a given cost ratio. To save space, we will not show the relative efficiency plots for the other two covariance structures. In Figure 3.3, the prior distribution of β_0 varies across the columns and the prior distribution of β_1 varies across the rows. In each

plot, the relative efficiencies are given at the vertical axis and the number of time points at the horizontal axis, and a separate line is given for the different cost ratios k .

In each curve (cost ratio) within each plot (prior), the optimal number of time points is the number on the horizontal axis at which the relative efficiency is highest. This gives one Bayesian D-optimal design per prior per cost ratio k . It can be seen from these plots that the choice of the most efficient number of time points strongly depends on the cost ratio and to a lesser extent on the prior distribution of the parameters. In all cases, the optimal number of time points is three if the cost ratio is $k = 0$. The optimal number of time points is still three for cost ratio $k = 0.5$ except for four or five when $\mu_{\beta_1} = 2$. However, three time points are still highly efficient, i.e., above 0.98. The optimal number of time points is five or six for cost ratio $k = 1$. Six time points is optimal for cost ratio $k = 2$. In general, for a cost ratio $k > 2$, the relative efficiency increases with the number of time points (q).

3.4.2.3 Maximin Bayesian D-optimal designs

Although Bayesian optimal design theory solves the problem of local optimality by taking into account the uncertainty about the unknown parameter β , it now faces a new problem that the Bayesian D-optimal designs depend on the choice of the prior distributions for β . Therefore, we want to determine a Bayesian D-optimal design which is robust under various prior distributions for a given cost ratio k and covariance structure \mathbf{D} . This can be done by using the maximin Bayesian procedure. Table 3.1 shows the range of minimum relative efficiencies computed over the nine prior distributions of β_0 and β_1 for a given \mathbf{D} and cost ratio k . The maximin value (MMV) for each cost ratio k is the upper bound of each range. For all the cost ratios k this MMV is very high, that is, above 0.98 for all \mathbf{D} . The range of minimum relative efficiencies is smallest for cost ratio $k = 0$ in all \mathbf{D} . Table 3.2 shows the maximin Bayesian D-optimal designs for each cost ratio k and covariance structure \mathbf{D} . It can be seen that, for $k = 0$, the maximin Bayesian D-optimal designs has three time points for all \mathbf{D} , and for $k = 1$, it is six time points for all \mathbf{D} . The number of time points for the maximin Bayesian D-optimal design for $k = 2.5$ or 5 is always seven for all \mathbf{D} . For other cost ratios, there are some differences between the three covariance structures with respect to the number of time points for the maximin Bayesian D-optimal design, but these should not be taken too seriously. For example, for $k = 0.5$, the maximin Bayesian D-optimal design has three time points for \mathbf{D}_2 against four for \mathbf{D}_3 , but three time points is also highly efficient in \mathbf{D}_3 , with a minimum relative efficiency of 0.9753. A similar result holds for $k = 1.5$ and 2 , where taking six instead of seven time points gives a minimum relative efficiency above 0.98 for \mathbf{D}_3 . In Table 3.2, can be seen that the maximin Bayesian D-optimal designs have only three or four *distinct* time points except in one case where it is five, i.e., for \mathbf{D}_2 and $k = 5$. In general, the allocation and number of time points of the maximin Bayesian D-optimal designs depend on the cost ratio k and the optimal number of time points of the Bayesian designs increases with k .

Table 3.1 Range of minimum relative efficiencies of the nine Bayesian D-optimal designs versus each other, per k and per \mathbf{D} for the logistic mixed model with linear time effect.

cost ratio k	Range of minimum relative efficiencies		
	\mathbf{D}_1	\mathbf{D}_2	\mathbf{D}_3
$k = 0$	(0.9888, 0.9967)	(0.9884, 0.9970)	(0.9901, 0.9971)
$k = 0.5$	(0.9414, 0.9799)	(0.9444, 0.9895)	(0.9585, 0.9769)
$k = 1$	(0.9430, 0.9896)	(0.9543, 0.9857)	(0.9585, 0.9886)
$k = 1.5$	(0.9396, 0.9848)	(0.9509, 0.9963)	(0.9781, 0.9940)
$k = 2$	(0.9464, 0.9884)	(0.9815, 0.9899)	(0.9781, 0.9941)
$k = 2.5$	(0.9464, 0.9884)	(0.9748, 0.9880)	(0.9781, 0.9941)
$k = 5$	(0.9464, 0.9884)	(0.9748, 0.9895)	(0.9781, 0.9941)

Table 3.2 Maximin Bayesian optimal design per cost ratio k and per covariance structure \mathbf{D} .

cost ratio k	Maximin Bayesian D-optimal designs								
	\mathbf{D}_1			\mathbf{D}_2			\mathbf{D}_3		
$k = 0$	-1.0000	-0.0674	1.0000	-1.0000	-0.0674	1.0000	-1.0000	-0.0843	1.0000
$k = 0.5$	-1.0000	-1.0000	-0.3676	0.2427	1.0000	1.0000	-1.0000	-0.0405	1.0000
$k = 1$	-1.0000	-1.0000	-0.2621	0.2758	1.0000	1.0000	-1.0000	-1.0000	-0.2228
$k = 1.5$	-1.0000	-1.0000	-0.2576	0.1978	1.0000	1.0000	-1.0000	-1.0000	-0.2645
$k = 2$	-1.0000	-1.0000	-1.0000	-0.1636	0.2983	1.0000	1.0000	-1.0000	-1.0000
$k = 2.5$	-1.0000	-1.0000	-1.0000	-0.1636	0.2983	1.0000	1.0000	-1.0000	-1.0000
$k = 5$	-1.0000	-1.0000	-1.0000	-0.1636	0.2983	1.0000	1.0000	-1.0000	-1.0000

3.4.2.4 Maximin locally D-optimal designs

To show the effect of taking a prior with small variance on the design choice and the robustness of such designs against a change of the location of parameters for the prior, we computed locally D-optimal designs and then maximin locally D-optimal designs for a given cost ratio k and covariance structure \mathbf{D} , by assuming different values for $\beta_0 = 0, 1$ and 2 , and for $\beta_1 = 0, 1$ and 2 , respectively. Note that a locally optimal design is the extreme case of a

prior variance being equal to zero. For each of the nine combinations of β_0 and β_1 , and each of the three covariance structures \mathbf{D} , the locally D-optimal designs with number of time points $q \leq 7$ were computed. We selected the optimal q , by computing the local relative efficiencies $\{RE(\xi_q^*; \xi_7^*(\boldsymbol{\beta})|\boldsymbol{\beta})\}$ of designs with $q < 7$ time points compared to a design with $q = 7$ time points within each pair of β_0 and β_1 , \mathbf{D} and k . Finally, we compared the nine locally D-optimal designs per \mathbf{D} and k in terms of their minimum local relative efficiency over the nine pairs of β_0 and β_1 , analogously to the procedure in the preceding section. Table 3.3 shows the range of those nine minimum local relative efficiencies for each covariance structure \mathbf{D} and each cost ratio k . The MMV for each cost ratio k is the upper bound of each range. For all the cost ratios k this MMV is high, that is, above 0.83 in \mathbf{D}_1 and \mathbf{D}_3 , and above 0.89 in \mathbf{D}_2 . In general, the MMV increases as the cost ratio k increases.

The maximin locally D-optimal design has two distinct time points only, and the number of time points q varies from two for $k = 0$ to seven for $k = 5$ (the results are not included here). In contrast, the maximin Bayesian D-optimal design (preceding section) has three or more distinct time points and the number of time points q varied from three for $k = 0$ to seven for $k = 5$.

In general, for a given cost ratio the number of time points at which measurements are taken is larger for the maximin Bayesian D-optimal design than for the maximin locally D-optimal design. The uncertainty in the prior information causes this, i.e., a prior with a larger variance yields a larger number of design points. It was observed empirically by many authors that increasing the uncertainty in the prior information increases the number of points in the Bayesian optimal designs for a fixed effects logistic model (see, e.g., Chaloner and Larntz, 1989; Khuri et al., 2006; Abebe et al., 2014a).

Table 3.3 Range of minimum relative efficiencies of the nine locally D-optimal designs versus each other, per k and per \mathbf{D} for the logistic mixed model with linear time effect.

cost ratio k	Range of minimum relative efficiencies		
	\mathbf{D}_1	\mathbf{D}_2	\mathbf{D}_3
$k = 0$	(0.6375, 0.8317)	(0.6780, 0.8997)	(0.6552, 0.8610)
$k = 0.5$	(0.6349, 0.8503)	(0.6780, 0.8997)	(0.6836, 0.8642)
$k = 1$	(0.6409, 0.8566)	(0.6985, 0.8995)	(0.6836, 0.8642)
$k = 1.5$	(0.6428, 0.8734)	(0.7093, 0.9028)	(0.6987, 0.8699)
$k = 2$	(0.6278, 0.8884)	(0.7033, 0.9082)	(0.7060, 0.8643)
$k = 2.5$	(0.6301, 0.8972)	(0.6953, 0.9211)	(0.7060, 0.8579)
$k = 5$	(0.6332, 0.9086)	(0.6873, 0.9365)	(0.7060, 0.8702)

3.4.2.5 Equidistant design points relative to Bayesian D-optimal designs

The results in the previous section show that choosing a design using the maximin Bayesian procedure gives a minimum relative efficiency of at least 0.98 and that using a maximin locally optimal design gives a minimum relative efficiency of at least 0.83. However, in practice, repeated measures are often planned at equidistant time points, which raises the question how efficient such designs are. Therefore, we computed and plotted the relative efficiencies $\{ \text{RE}(\xi_{eq}; \xi_B^*(\pi) | \pi) \}$ for 2, 3, ..., 7 equidistant design points, with the first and last point on the boundaries of the design space, with respect to the Bayesian D-optimal design for a given prior, \mathbf{D} and cost ratio k . Note that the optimal number of time points q for the Bayesian design follows from Figure 3.3 for a given prior and cost ratio for \mathbf{D}_2 and this q for the Bayesian design remains fixed as we vary the number of equidistant time points from 2 to 7. Since the relative efficiency plots of equidistant versus Bayesian optimal design were very similar for all considered priors and the three different covariance structures \mathbf{D} , Figure 3.4 shows the results for extreme prior mean values and covariance structure \mathbf{D}_2 only. It can be seen from these plots that for cost ratio $k = 0$ and 0.5, three equidistant design points is highly efficient, that is, above 0.98, because the Bayesian optimal time points are very close to equidistant. The relative efficiencies of four equidistant design points for cost ratio $k = 1$ are above 0.94. In general, three or four equidistant design points are highly efficient for all priors compared to Bayesian optimal designs for small cost ratio $k \leq 1$, having a relative efficiency above 0.90. Five or six equidistant design points are highly efficient for all cost ratios k , having a relative efficiency above 0.81.

Table 3.4 shows the number of maximin equidistant design points and the range of minimum relative efficiencies of 2, 3, 4, 5, 6 and 7 equidistant design points for a given \mathbf{D} and cost ratio k , where the minimum relative efficiency of each equidistant design is obtained over all nine priors. The upper bound of each range is the MMV for the maximin equidistant design. For example, for $k = 0$ and in \mathbf{D}_1 , the number '3' in the cell of Table 3.4 represents the number of the maximin equidistant design points with MMV 0.9890. As can be seen from Table 3.4, the maximin equidistant design is highly efficient compared to the Bayesian optimal design. However, the MMV decreases as the cost ratio k increases because the number of coinciding time points of the Bayesian optimal designs increases with the cost ratio k . Thus, the Bayesian optimal time points differ more from equidistance as the cost ratio increases. Furthermore, the number of maximin equidistant design points is the same in all \mathbf{D} for a given k .

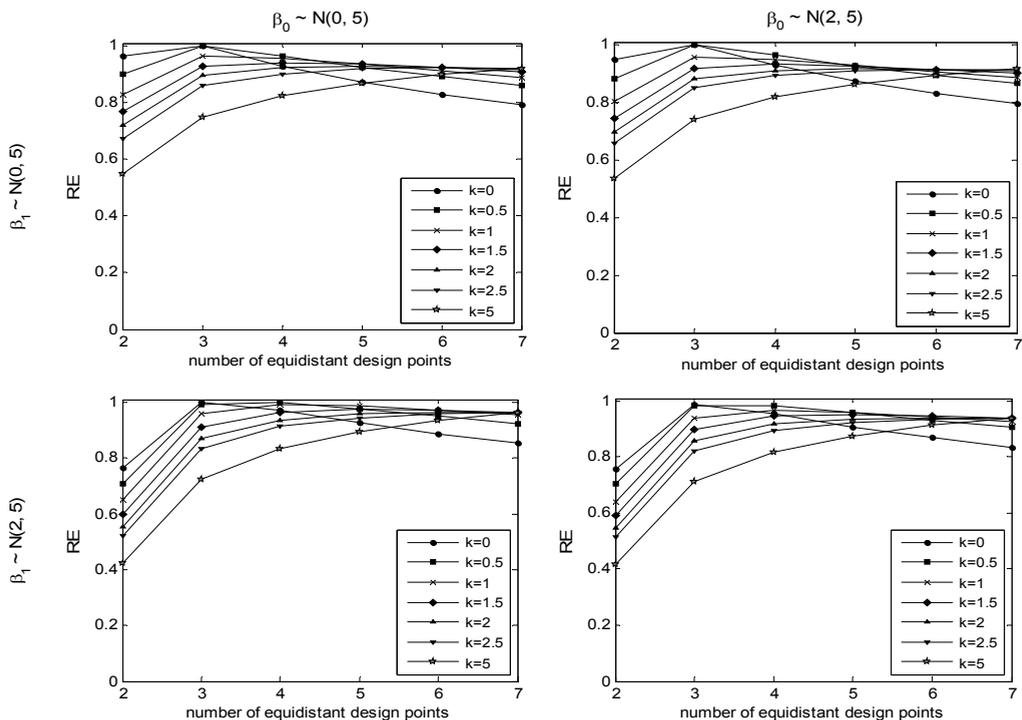


Figure 3.4 Relative efficiency of equidistant design points relative to Bayesian optimal design for different cost ratios k for the logistic mixed effects model with linear time effects and covariance structure D_2 .

Table 3.4 Number of maximin equidistant design points (and range of minimum relative efficiencies of 2, 3, 4, 5, 6 and 7 equidistant design points relative to the Bayesian optimal design, over all nine priors) for each D and cost ratio k .

cost ratio k	Number of maximin equidistant design points and range of min REs		
	D_1	D_2	D_3
$k = 0$	3 (0.7363, 0.9890)	3 (0.7583, 0.9895)	3 (0.7672, 0.9875)
$k = 0.5$	3 (0.6640, 0.9555)	3 (0.7042, 0.9845)	3 (0.7058, 0.9687)
$k = 1$	4 (0.6003, 0.9179)	4 (0.6402, 0.9491)	4 (0.6417, 0.9464)
$k = 1.5$	4 (0.5487, 0.8866)	4 (0.5903, 0.9320)	4 (0.5880, 0.9172)
$k = 2$	5 (0.5083, 0.8235)	5 (0.5473, 0.9174)	5 (0.5490, 0.9008)
$k = 2.5$	6 (0.4717, 0.8150)	6 (0.5135, 0.9152)	6 (0.5151, 0.8928)
$k = 5$	7 (0.3873, 0.8778)	7 (0.4170, 0.9165)	7 (0.4182, 0.8967)

Note: The relative efficiencies of equidistant design points are computed relative to Bayesian optimal designs with prior variance, $\sigma^2 = 5$.

3.5 Conclusions

This chapter addressed the problem of choosing the optimal number and allocation of time points for the logistic mixed effects model for longitudinal studies with a binary response and a linear time effect from a Bayesian perspective. A subject-to-measurement cost ratio was taken into account in the design optimization, because the number of subjects and number of time points at which measurements are taken per subject are often limited by budget constraints, and the cost per subject may be quite different from the cost per measurement. Firstly, the Bayesian D-optimal allocation of time points was presented for $q = 2, 3, \dots, 7$ time points and three different covariance structures. Following this, the Bayesian D-optimal designs for different q 's were compared with each other in terms of their relative efficiencies under the various prior distributions, for each covariance structure and different cost ratios. Maximin Bayesian designs were obtained to overcome the problem of dependency of Bayesian designs on the choice of the prior distribution. We also computed locally D-optimal designs and maximin locally D-optimal designs for the three different covariance structures under different cost ratios to show the consequences of taking a prior with a small variance on the design choice and its robustness. Finally, the relative efficiencies of equidistant time points versus the Bayesian D-optimal designs were also studied for different cost ratios.

The findings for the Bayesian D-optimal designs show that the optimal number of time points depends on the cost ratio. In particular, the Bayesian D-optimal designs with three time points are highly efficient when the cost ratio is relatively small, i.e., when measurements are very expensive compared to the costs of sampling subjects. The optimal number of time points of the Bayesian designs increases with the cost ratio. The maximin Bayesian D-optimal design has three or four distinct time points depending on the cost ratio, and the other time points coincide at either of the two design boundaries. If repeated measures at the same time point are not feasible in practice, then researchers should measure repeatedly at nearby time points. Further, the Bayesian D-optimal designs for various priors are highly efficient relative to each other, at least if the prior variance is large (see Table 3.1). Accordingly, the maximin Bayesian D-optimal designs obtained by maximizing the minimum relative efficiency over all priors, have a very high efficiency if the prior variance is large. Abebe et al. (2014a) have found a similar result for the fixed effects logistic model.

Locally D-optimal designs, which are special cases of Bayesian D-optimal designs with a prior variance of zero, always have two distinct time points, and the number of (coinciding) time points increases with the cost ratio. The maximin locally D-optimal design gives a much lower relative efficiency than the maximin Bayesian D-optimal design. The number of time points at which measurements are taken is larger for the maximin Bayesian optimal design than for the maximin locally optimal design.

Finally, we showed that for small cost ratios, three or four equidistant design points are highly efficient compared to the Bayesian optimal designs, while for large cost ratios, five or six equidistant design points are highly efficient.

In conclusion, the optimal number of time points increases with the cost ratio. Maximin Bayesian optimal designs have a high minimum relative efficiency and are robust against a change of priors. Finally, if there are limited computational resources, our recommendation is that researchers should choose three or four equidistant design points when the cost of a repeated measurement is the same or more than the cost of sampling a subject, and choose five or six equidistant design points when the cost of sampling a subject is more than the cost of a repeated measurement. These equidistant designs are highly efficient relative to the Bayesian optimal design for all priors considered and thus quite robust.

The results of this chapter assumed that all subjects are available for the study during the study period and that there is no dropout. Further research is needed to study the effect of dropout on optimal design issues. Another important aspect for further research is the extension of the results for the purpose of comparing two or more groups over time.

Acknowledgements

We wish to thank Tekle (2008) for his permission to use part of his computer programme (POLS-Logistic).

CHAPTER 4

Bayesian design for dichotomous repeated measurements with autocorrelation

This chapter will be published in *Statistical Methods in Medical Research* (article in press) with co-authors Frans E.S. Tan, Gerard J.P. Van Breukelen and Martijn P.F. Berger

Abstract

In medicine and health sciences binary outcomes are often measured repeatedly to study their change over time. A problem for such studies is that designs with an optimal efficiency for some parameter values may not be efficient for other values. To handle this problem, we propose Bayesian designs which formally account for the uncertainty in the parameter values for a mixed logistic model which allows quadratic changes over time. Bayesian D-optimal allocations of time points are computed for different priors, costs, covariance structures and values of the autocorrelation. Our results show that the optimal number of time points increases with the subject-to-measurement cost ratio, and that neither the optimal number of time points nor the optimal allocations of time points appear to depend strongly on the prior, the covariance structure, or on the size of the autocorrelation. It also appears that for subject-to-measurement cost ratios up to five, four equidistant time points, and for larger cost ratios, five or six equidistant time points are highly efficient. Our results are compared with the actual design of a respiratory infection study in Indonesia and it is shown that, selection of a Bayesian optimal design will increase efficiency, especially for small cost ratios.

Keywords: Bayesian D-optimal designs; logistic mixed-effects model; cost ratio; relative efficiency; number of time points, autocorrelation.

4.1 Introduction

Longitudinal studies are carried out in many fields of science to study the change of a particular outcome variable over time. In a clinical trial, for example, patients are studied over a period of time and classified in terms of presence or absence of a clinical symptom. In smoking prevention studies, pupils in primary and secondary school may be followed up to study the prevalence of smoking as a function of age. Researchers of longitudinal studies face the problem of choosing the optimal allocation and selection of the time points at which measurements are taken. The number of time points is usually determined by non-statistical criteria, and the time points are frequently chosen to be equally spaced. Some examples are, the Indonesian Children's Health study, whereby more than 3000 preschool children were examined medically for up to six consecutive quarters to assess the presence of a respiratory infection (Sommer et al., 1984; Diggle et al., 1994), and the French EPIDOS study, where 7575 women aged 75 years or older were annually visited for 6 years to study the risk factors for hip fracture (Dargent-Molina et al., 1996; Carrière and Bouyer, 2002). However, these design choices may be inefficient because the precision of the model parameter estimates of interest, such as the rate of change, depends on the study design; in particular on the number and allocation of the time points of measurement.

Longitudinal data are often modelled using a polynomial model. Although a linear growth or first-order polynomial function is often assumed in longitudinal studies, this may be too restrictive because the pattern may not be strictly linear. A quadratic growth function (second-order polynomial) is more flexible and can also handle monotonic trends as well as u-shaped trends. Interindividual differences in the precise pattern can be included as random effects in the model. For example, Diggle et al. (1994) found a curvilinear pattern and fitted a logistic mixed-effects model with a quadratic time effect for the longitudinal data on respiratory infection in 275 Indonesian children who were examined quarterly up to six visits.

The purpose of this chapter is to find the optimal number and allocation of distinct time points for longitudinal studies with a binary response and quadratic time effects, using optimal design theory. In the case of non-linear models optimal designs depend on the unknown parameter values. Much effort has been devoted to this problem. A common approach is to design an experiment for an initial estimate of the parameters, which leads to locally optimal designs (see, e.g., Chernoff, 1953). However, a design which is optimal for an initial estimate of the parameters, may not be optimal for other parameter values and if the initial estimate of the parameters is poor, the design's performance may be far from optimal. Therefore, locally optimal designs face the serious problem of how to take the uncertainty about the parameter values into account. Three methods have been proposed in the literature to overcome this problem, namely (a) the sequential procedure, which uses information from previous runs, (e.g. as in group sequential design), (b) the maximin approach, which maximizes the minimal efficiency over the parameter space, that is, the range of plausible parameter values, and (c) the Bayesian approach, which assumes a prior distribution for the unknown parameter values (see, e.g., Abdelbasit and Plackett, 1983; Dette, 1997; Chaloner

and Larntz, 1989). Due to restrictions on the total length of the research period, the sequential approach may not always be feasible especially for a longitudinal study. A potential objection to the maximin approach is that the maximin design may occur at the edges of the parameter space and if some points in the parameter space have a very low probability, then their weight may be over-emphasized (Atkinson et al., 2007 page 258). In this chapter, we will use the Bayesian approach, which allows inclusion of a prior distribution that captures all existing information about the uncertainty of the parameter values.

A number of papers on the optimal allocation and selection of design points have been published. For the fixed effects logistic model, this has been investigated in literature by Abdelbasit and Plackett (1983); Chaloner and Larntz (1989) and Abebe et al. (2014a), among others. For the linear random effects model optimal designs are discussed by Tan and Berger (1999), Berger and Tan (2004) and Berger and Wong (2009), among others. Ouwens et al. (2006) and Tekle et al. (2008) have extended the work on optimal designs for logistic mixed-effects models using maximin D-optimal designs. However, there is no work, to the authors' knowledge, on Bayesian optimal designs for longitudinal studies with a binary response in the presence of autocorrelation.

In the present chapter, we will numerically investigate Bayesian optimal designs for the logistic mixed-effects model with quadratic time effects and autocorrelations between the repeated measurements. The results will give information about the optimal number and allocation of time points for different subject-to-measurement cost ratios and different priors. Moreover, the effects of autocorrelation and the covariance structure of the random effects on the Bayesian optimal designs will be studied. Finally, we will address the question whether or not equidistant time points are highly efficient compared to Bayesian D-optimal designs, and we will apply the results to a published longitudinal study on respiratory infection (Sommer et al., 1984; Diggle et al., 1994).

This chapter is organized as follows. In section 4.2, the motivating case study is introduced, while the quadratic time effect logistic mixed model and the variance-covariance matrix of parameter estimators are described in section 4.3. The Bayesian D-optimality criterion and the relative efficiency as a measure for the comparison of designs are given in section 4.4. In section 4.5, we will describe our numerical approach. Thereafter, results are presented in section 4.6. The numerical results are illustrated by the motivating example in section 4.7. Finally, a brief summary and conclusion are provided in section 4.8.

4.2 Motivating Example

The results of the simulation study in this chapter will be applied to a published longitudinal study with dichotomous data. The Indonesian Children's Health study, a prospective longitudinal study of preschool children aged up to 6 years was conducted in 6 rural villages of the Purwakarta district, West Java, Indonesia between 1977 and 1978 (Sommer et al., 1984; Diggle et al., 1994). More than 3000 preschool children were re-examined medically

for up to six consecutive quarters to assess whether they suffered from respiratory or diarrhoeal infection and xerophthalmia, an ocular manifestation of vitamin A deficiency. Diggle et al. (1994) focused on the question of whether vitamin A deficient children are at increased risk of respiratory infection, and analysed the evolution of risk of respiratory infection on 275 preschool children who were examined for up to 6 consecutive quarters. The model considered by Diggle et al. (1994) was a logistic mixed-effects model with a quadratic time effect. The quadratic time effect was tested and found statistically significant.

In planning a longitudinal study such as in the example above, the researcher faces the problem of choosing the best number and timing of the repeated measurements. Usually, these choices are based on non-statistical criteria, and the choices may be statistically inefficient, that is, lead to imprecise parameter estimates. Using optimal design theory, this chapter addresses the choice of number and timing of measurements, taking into account costs for measuring. After the numerical studies we will get back to the Indonesian study to see how efficient the actual design was.

4.3 Model and variance-covariance matrix of parameter estimators

Let the $q \times 1$ vector $\mathbf{y}_i = (y_{i1}, \dots, y_{iq})'$ be the binary responses y_{ij} of subject i at q time points, $i = 1, 2, \dots, N$ and $j = 1, \dots, q$. It is assumed that all subjects have measurements at the same time points, and that, conditional on the subject-specific random effect vector \mathbf{b}_i , the binary responses y_{ij} of \mathbf{y}_i are Bernoulli distributed with probability of success $p(y_{ij} = 1 | \mathbf{b}_i)$. These probabilities are related to the fixed and random effects via the logit link and the corresponding logistic mixed-effects model is given by:

$$\text{logit}(p(y_{ij} = 1 | \mathbf{b}_i)) = \log\left(\frac{p(y_{ij} = 1 | \mathbf{b}_i)}{1 - p(y_{ij} = 1 | \mathbf{b}_i)}\right) = \mathbf{x}_j' \boldsymbol{\beta} + \mathbf{z}_j' \mathbf{b}_i, \quad (4.1)$$

where the $p \times 1$ vector \mathbf{x}_j is the design vector of the explanatory variables at the j th measurement for subject i , $\boldsymbol{\beta}$ is the corresponding $p \times 1$ vector of fixed effects, and \mathbf{z}_j is the $r \times 1$ design vector for the random effects that is usually a subset of vector \mathbf{x}_j . The vector \mathbf{b}_i is the corresponding $r \times 1$ vector of random effects, which is assumed to have a multivariate normal distribution with mean zero and covariance matrix \mathbf{D} .

A quadratic ($p = 3$) time effect is considered, i.e., the design vector is $\mathbf{x}_j' = (1 \ t_j \ t_j^2)$ and $\boldsymbol{\beta} = (\beta_0 \ \beta_1 \ \beta_2)'$, where t_j is the time point of the j th measurement, and β_0 , β_1 and β_2 are the fixed effects. Suppose that the intercept and linear time effect both vary randomly between subjects. Then the design vector is $\mathbf{z}_j' = (1 \ t_j)$ and $\mathbf{b}_i = (b_{0i} \ b_{1i})'$, where b_{0i} and b_{1i}

are the corresponding random (subject-specific) deviations from these fixed effects. Thus, according to model (4.1), the log-odds of a positive response for subject i at time t_j is given by:

$$\text{logit}(p(y_{ij}=1|\mathbf{b}_i)) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_j + \beta_2 t_j^2. \quad (4.2)$$

Due to the random effects in models (4.1) and (4.2), the log-likelihood cannot be written down in closed form. Hence, either numerical methods or approximations to the log-likelihood must be used. Numerical methods are computationally intensive and require full knowledge of the data (Moerbeek et al., 2003; Han and Chaloner, 2004), which makes them computational inconvenient for optimal design. Therefore, for the purpose of optimal design, we will use an approximate information matrix, which is the inverse of the asymptotic variance-covariance matrix of the parameter estimators based on an extension of the generalized estimating equation (GEE) approach (Zeger et al., 1988; Molenberghs and Verbeke, 2005). The variance-covariance matrix of the estimator $\hat{\boldsymbol{\beta}}$ of the fixed effects parameter vector $\boldsymbol{\beta}$ for the logistic mixed-effects model (4.1) with autocorrelation, based on the extension of the GEE approach is approximated by:

$$\text{var}(\hat{\boldsymbol{\beta}}) \approx \left(\sum_{i=1}^N \frac{\partial \mathbf{P}_i'}{\partial \boldsymbol{\beta}} \mathbf{v}_i^{-1} \frac{\partial \mathbf{P}_i}{\partial \boldsymbol{\beta}'} \right)^{-1}, \quad (4.3)$$

where $\hat{\boldsymbol{\beta}}$ is the estimator of $\boldsymbol{\beta}$ for model (4.1), $\mathbf{P}_i = (p(y_{i1}|\mathbf{b}_i), \dots, p(y_{iq}|\mathbf{b}_i))'$, and the working variance-covariance matrix \mathbf{v}_i of the responses is given by:

$$\mathbf{v}_i \approx \boldsymbol{\omega}_i^{1/2} \mathbf{R} \boldsymbol{\omega}_i^{1/2} + \mathbf{w}_i \mathbf{Z} \mathbf{D} \mathbf{Z}' \mathbf{w}_i', \quad (4.4)$$

where \mathbf{R} is the $q \times q$ residual correlation matrix, that captures autocorrelation of the responses over time, \mathbf{Z} is the $q \times r$ design matrix with rows \mathbf{z}_j' , $j = 1, \dots, q$ and \mathbf{w}_i is the diagonal matrix of the conditional variances of responses given the random effects $\mathbf{b}_i = 0$. The residuals are the discrepancy between observed and expected outcomes given the random effects $\mathbf{b}_i = 0$. The diagonal matrix of the conditional variances $\boldsymbol{\omega}_i$ of responses given the random effects $\mathbf{b}_i = 0$ is given by:

$$\boldsymbol{\omega}_i = \text{diag}(w_{i1}^{b_i=0}, \dots, w_{iq}^{b_i=0}), \quad (4.5)$$

where $w_{ij}^{b_i=0} = p(y_{ij}=1|\mathbf{b}_i=0)(1 - p(y_{ij}=1|\mathbf{b}_i=0))$, for $i = 1, \dots, N$, $j = 1, \dots, q$ (see Molenberghs and Verbeke, 2005 chapter 24 for details).

In the present chapter, a first order autoregressive (AR1) correlation structure is considered for the residual correlation matrix \mathbf{R} , i.e., the (j, l) element of the residual

correlation matrix R is equal to $\rho^{|t_j - t_l|}$, where $l = 1, \dots, q$, and so ρ is the autocorrelation coefficient between two residuals for which $|t_j - t_l| = 1$. This autocorrelation structure implies that repeated measurements closer in time are more highly correlated and that the correlation decreases as the distance between the time points increases.

4.4 Optimal design and relative efficiency

4.4.1 Bayesian D-optimality criterion

To introduce the notation for the optimality criterion, let \mathcal{E} be the design space containing all designs with $q \times 1$ time vectors, i.e.,

$$\mathcal{E} = \left\{ (t_1, t_2, \dots, t_q)' : t_j \in [a, b] \right\}, \quad (4.6)$$

where t_1, t_2, \dots, t_q are the q time points, and $q \geq p$ with p being the number of fixed parameters of the model. The time interval $[a, b]$ is assumed to be fixed by substantive constraints within the field of application. A design ξ_q is an element of the design space \mathcal{E} if it has q time points within the time interval $[a, b]$ at which measurements are taken for all N subjects. The amount of information in a design is measured by the information matrix, which is approximately equal to the inverse of the asymptotic variance-covariance matrix of the parameter estimators in (4.3). The goal of optimal experimental design is to minimize a function of (4.3). An optimal design maximizes a real-valued function of the information matrix, which is usually known as optimality criterion (see, e.g., Silvey, 1980). There are several optimality criteria proposed in the literature, such as the D-, A- or G-optimality criteria. In this chapter, we will use the D-optimality criterion for two reasons: 1. It minimizes the volume of the asymptotic confidence ellipsoid of the parameters. 2. It does not depend on the coding used for the endpoints of the chosen time interval $[a, b]$. This means that if the coding for the time interval is linearly transformed, a D-optimal design for the new time interval is obtained by applying the same linear transformation to the D-optimal design for the old interval (See Ouwens et al., 2006 for proof). The D-optimal design ξ_q^* is the design among all possible designs ξ_q with q time points for which the determinant of the approximated variance-covariance matrix (4.3) is minimized or equivalently, for which the determinant of the approximated information matrix is maximized (Atkinson et al., 2007 page 151). The approximation to the asymptotic variance-covariance matrix (4.3) depends on the parameter vector β . We consider the Bayesian approach, where a prior distribution is assumed for the regression parameters β , which is then incorporated into an appropriate design criterion. The Bayesian D-optimality criterion as a function of variance-covariance matrix (4.3), by analogy with the non-Bayesian optimal design, is defined as follows:

$$\phi_D(\xi|\pi) = E_{\beta} \log \det \left(\sum_{i=1}^N \frac{\partial \mathbf{P}_i'}{\partial \beta} \mathbf{v}_i^{-1} \frac{\partial \mathbf{P}_i}{\partial \beta'} \right) = \int_{\beta} \log \det \left(\sum_{i=1}^N \frac{\partial \mathbf{P}_i'}{\partial \beta} \mathbf{v}_i^{-1} \frac{\partial \mathbf{P}_i}{\partial \beta'} \right) \pi(\beta) d(\beta), \quad (4.7)$$

where $\pi(\beta)$ is the prior distribution for β . The design which maximizes (4.7) is the Bayesian D-optimal design.

4.4.2 Cost function and relative efficiency

The precision of estimating the parameters increases by taking more measurements per subject. However, the addition of subjects and measurements per subject will increase the costs of the study. In practice, the number of subjects and the number of time points at which measurements are taken per subject are often limited by budget constraints. Thus, it is reasonable to take into account the costs of a longitudinal study when designs are compared with each other. There are two main cost components of obtaining measurements for subjects. These are the costs for the recruitment of subjects and the costs of the measurements once a subject is recruited. Let the cost of recruiting a subject be C_1 and the cost of one measurement per subject be C_2 . Then the total costs of a longitudinal study with q time points for N subjects, excluding overhead cost can be defined by the linear cost function:

$$C = C_1 N + C_2 N q = N C_2 (k + q), \quad (4.8)$$

where $k = C_1/C_2$ is the ratio of the costs of adding a new subject to the costs of an additional measurement per subject. We will vary this cost ratio from $k = 0$ to $k = 50$, and study its effect on the optimality of the Bayesian design.

To compare different designs, we will use their relative efficiencies and to do so fairly we will consider designs with the same total costs C . This means that the designs can differ in terms of the number of subjects N and the number and timing of the measurements q . First we compute the Bayesian D-optimal designs using N fixed, and then we correct for costs and varying q and N as follows.

Let $\phi_D(\xi_q^*|\pi)$ denote the value of design criterion (4.7) for the optimal design ξ_q^* with q time points. Then the relative efficiency (RE) of an arbitrary design ξ_s with s time points relative to the optimal design ξ_q^* is defined as:

$$\text{RE}(\xi_s; \xi_q^*(\pi)|\pi) = \frac{k+q}{k+s} \left[\exp \left\{ \frac{\phi_D(\xi_s|\pi) - \phi_D(\xi_q^*|\pi)}{p} \right\} \right], \quad (4.9)$$

where $q, s \geq p$ and π is the prior distribution for the fixed effects. If the value of this relative efficiency is close to unity, then the design ξ_s is about equally efficient as the optimal design ξ_q^* for a given prior π . The inverse of this relative efficiency determines the required number

of subjects N for a design ξ_s to have the same efficiency as the optimal design ξ_q^* . Note that this relative efficiency (4.9) is an extension of the relative efficiency given by Chaloner and Larntz (1989). When the ratio $\frac{k+q}{k+s}$ is one, it reduces to their relative efficiency. This ratio follows from (4.8) and takes into account that the design with the smaller number of time points can have more subjects for the same budget.

4.5 Numerical approach

4.5.1 Method of Optimization

The Bayesian D-optimal designs for the logistic mixed-effect model with quadratic ($p = 3$) time effects were found numerically by maximization of the criterion value (4.7) among all candidate designs for a given prior distribution of the parameters. Details of this will be given in the next sub-sections.

4.5.1.1 Sampling parameter values from priors to compute the criterion

To construct Bayesian designs for continuous prior distributions, we first need to evaluate the integral of the criterion over the prior distribution and then compute all candidate designs in terms of their criterion values (4.7). However, evaluation of the integration over the prior distribution is very complicated and can not easily be done analytically. A numerical approximation of the integral is necessary. Numerical approximations can be done by sampling parameter values from the prior distribution and then replacing the integral in (4.7) with a summation over the sample (Atkinson et al., 2007; Chaloner and Verdinelli, 1995). Estimating (4.7) using the traditional sampling (pseudo Monte Carlo) method requires very large samples from the prior to reduce the sample-to-sample variability to the point where different samples do not lead to different design choices. Thus, this approach is costly in terms of computing time. In this chapter, we use an Adaptive Rejection Metropolis Sampling (ARMS) algorithm (Gilks and Wild, 1992; Gilks et al., 1995), which is a more efficient sampling algorithm that requires a smaller sample to obtain a good approximation of the design criterion (4.7).

We performed all computations in MATLAB R2010a (32bit) using a computer with a 2.67 GHz Intel processor and 6 GB RAM. We sampled 30,000 parameter values of fixed effects vector from their priors using the ARMS to estimate the design criterion in (4.7). The total computing time to generate the Bayesian optimal time points for all values of q that we consider is about 2 hours and 30 minutes per prior per covariance structure and per autocorrelation. Using the traditional sampling, the total computing time is roughly about 2.5 times as long.

4.5.1.2 Optimization algorithm for finding an optimal design

To find candidate designs and in particular the optimal design, the program uses the FMINCON function of MATLAB version 7.10.0499 (R2010a). This function performs constrained non-linear optimization and requires an initial design ξ_0 . Without loss of generality, we coded the time interval as $[-1, 1]$, and we used equally spaced time points as initial designs. The following global search algorithm was used to find the Bayesian D-optimal designs for a given prior distribution of the parameters:

1. Take samples from the prior distribution of the parameters using ARMS.
2. Compute the Bayesian D-optimal allocation of q time points, using $q = p$ equidistant time points as initial design, where p is the number of fixed parameters of the model.
3. Increase the number of time points q by one and perform step 2 again to find the Bayesian optimal design (allocation) for the new value of q . Repeat step 2 and 3 until the maximum number of time point q (user specified) is reached.
4. Thereafter, select the optimal number of time points q for the Bayesian D-optimal design by computing the relative efficiencies of designs with different number of time points against each other for a given subject-to-measurement cost ratio. This will be described in the next section.

4.5.2 Priors and covariance structures

For the logistic mixed model with quadratic ($p = 3$) time effects, a number of different prior distributions of the fixed parameters and three different covariance structures \mathbf{D} were considered. We assumed an independent trivariate normal prior for the fixed effects β_0 , β_1 and β_2 . We considered three different prior mean values for each of the fixed effects, i.e., $\mu = [0, 1, 2]$, and a large prior variance $\sigma^2 = 5$ for each of the fixed effects. All the 27 combinations of the prior mean values were used to generate 27 different prior distributions. We used $N = 100$ subjects to estimate the determinant in (4.7) as part of the search for Bayesian D-optimal designs, because Abebe et al. (2014b) found by performing a simulation study with N between 25 and 200 that the estimated determinant of the variance-covariance matrix of parameter estimates becomes stable when N is at least 100. Since that result was obtained for the case of a mixed-effects logistic model without autocorrelation, we verified it by simulations for the present mixed-effects model with autocorrelation.

The following three random effects covariance structures \mathbf{D} were considered:

- $\mathbf{D}_1 = 0.5$, i.e., a random intercept model with variance of the random intercept 0.5.
- $\mathbf{D}_2 = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$, i.e., a random intercept and random linear slope model with variances $d_{11} = 0.5$ for the intercept and $d_{22} = 0.5$ for the slope, and no slope-intercept covariance.

- $\mathbf{D}_3 = \begin{bmatrix} 0.5 & 0.25 \\ 0.25 & 0.5 \end{bmatrix}$, i.e., the same as \mathbf{D}_2 except that there is now a slope-intercept covariance of 0.25, giving a correlation of 0.5.

As published work already showed that having more than ten measurements is a waste of resources (Winkens et al., 2007; Tekle et al., 2011), we restricted the number of time points q to a maximum of ten measurements. Since the optimal combination of N and q depends on the relative costs of subjects and measurements, the following different cost ratios $k = 0, 0.5, 1, 2, 5, 10, 20$ and 50 were considered. This means that, for instance when $k = 10$, the costs of recruiting a new subject is ten times as expensive as the cost of one measurement. As a specification of ρ is needed to compute the working variance-covariance of the responses in (4.4) and time-structured data are likely to be positively correlated, we considered values of the autocorrelation parameter ρ in the range $[0.1, 0.9]$.

4.6 Results

4.6.1 Effect of autocorrelation on Bayesian optimal designs

We computed the Bayesian D-optimal designs for number of time points $3 \leq q \leq 10$, for each of the 27 different prior distributions, and each of the three covariance structures, for different values of the autocorrelation, i.e., $\rho = 0.1, 0.2, \dots, 0.9$. We verified that the results (Bayesian D-optimal designs) are not sensitive to the choice of the initial design points by implementing different starting points for a given number of time points q and prior in step 2 of the algorithm in section 4.5.1.2. The different starting points lead to the same optimal allocations of design points. To save space we will only present the Bayesian D-optimal allocations of time points for some choices of q . Further, since the pattern of the plots was the same for all priors, we will only present Bayesian D- design points for the prior mean $[2, 2, 2]$ for different values of the autocorrelation for the logistic mixed-effects model with quadratic time effects (Figure 4.1). Each column corresponds to a number of time points q , and each row corresponds to a covariance structure \mathbf{D} . Each plot shows the optimal design points (x-axis) as a function of the autocorrelation ρ (y-axis). It can be seen from these plots that the Bayesian D-optimal designs are fairly independent of ρ under all q , except for $q = 3$, where the optimal design points shift a bit for large value of ρ . Further, the allocations of time points are essentially the same for all three covariance structures \mathbf{D} of the random parameters.

From Figure 4.1 it is observed that the Bayesian D-optimal allocations of time points are hardly affected by the size of the autocorrelation. To check this we computed the relative efficiencies of the Bayesian D-optimal designs for $\rho = 0.1, 0.2, \dots, 0.9$ compared to a Bayesian optimal design for $\rho = 0.1$. In Figure 4.2, the vertical axis shows the relative efficiencies for the covariance structure \mathbf{D}_2 , and the horizontal axis shows the autocorrelations (ρ). Separate lines are given for the different number of time points q . It can be seen that the

relative efficiencies are all very high, i.e., above 0.955 for all q and all ρ . Together, Figure 4.1 and Figure 4.2 indicate that the Bayesian optimal designs are hardly affected by the value of the autocorrelations at least within the range ρ from 0.1 to 0.9. Thus, in the remaining part of this chapter, we will only present results for autocorrelation $\rho = 0.5$.

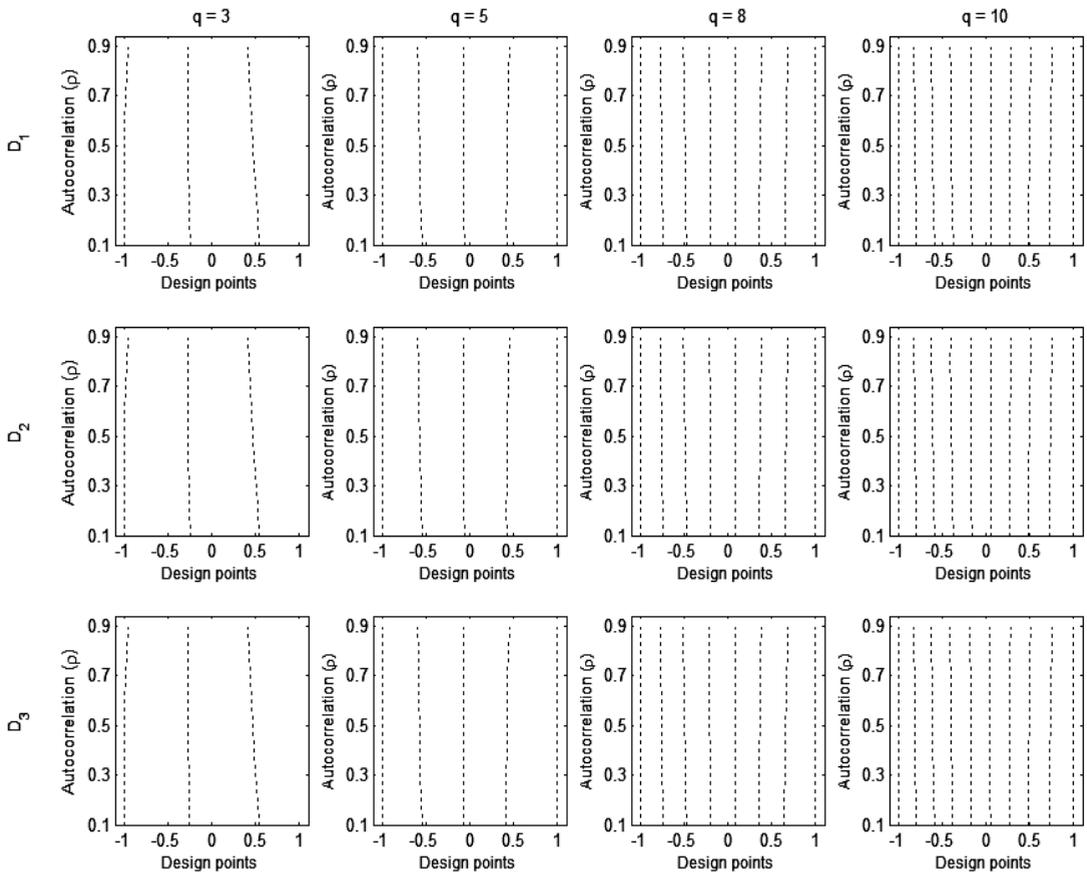


Figure 4.1 Bayesian D-optimal allocations of time points for the logistic mixed effects model with quadratic time effects for different values of autocorrelation ρ and for the prior mean $[2, 2, 2]$.

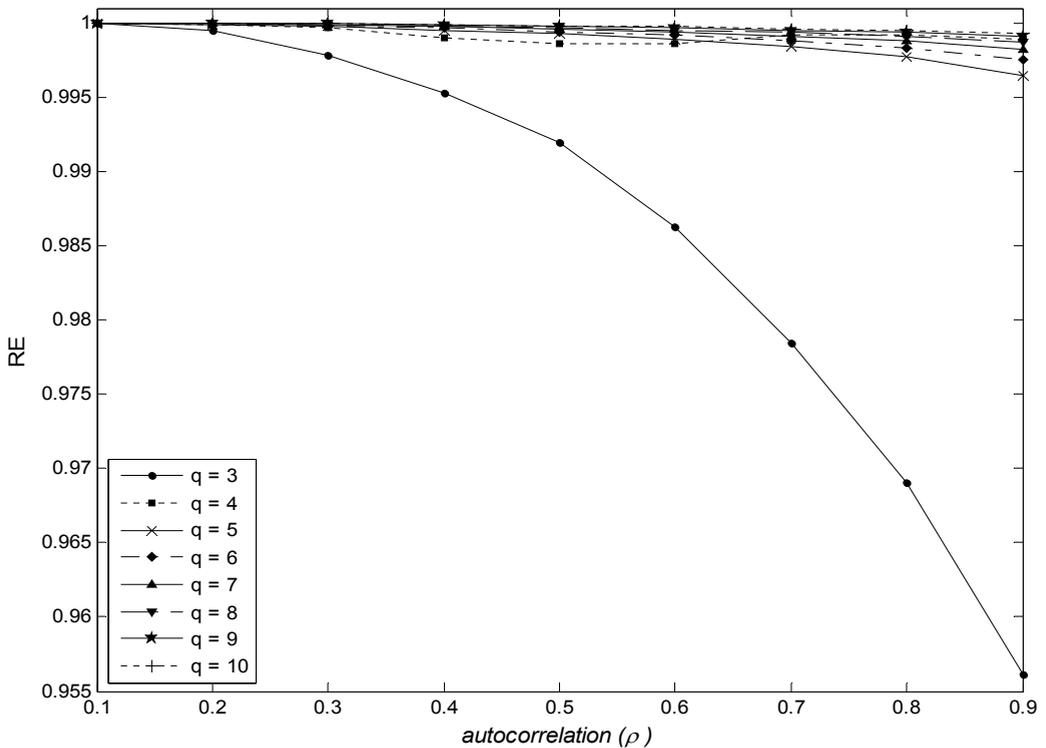


Figure 4.2 RE of the Bayesian optimal designs for $\rho \leq 0.9$ compared to the Bayesian optimal design for $\rho = 0.1$ for the covariance structure \mathbf{D}_2 .

4.6.2 Effect of prior and covariance structure on Bayesian optimal designs

Figure 4.1 already shows that for one prior, the Bayesian D-optimal allocations of time points are essentially the same for all three covariance structures \mathbf{D} of the random parameters. However, we also considered other covariance structures \mathbf{D} of the random parameters with a variance of five or ten and even the fixed effects case, i.e., with no random effects, and also a covariance structure \mathbf{D} which include a random quadratic slope. Here we always found very similar Bayesian D-optimal designs. Thus, the Bayesian D-optimal designs appear to be hardly affected by the choice of a covariance structure \mathbf{D} for the random parameters. Hence, we will only discuss the results for covariance structure \mathbf{D}_2 (random intercept and random slope without covariance). Further, to save space we will only present the Bayesian optimal allocations of time points for $q = 3, 4$ and 5 . Figure 4.3 shows these allocations for each of the 27 prior distributions of the parameters β . Each column in Figure 4.3 corresponds to a prior distribution of β_0 . Each row corresponds to a number of time points q . Within each plot, the vertical axis shows the prior mean of β_2 and the horizontal axis shows the locations of the

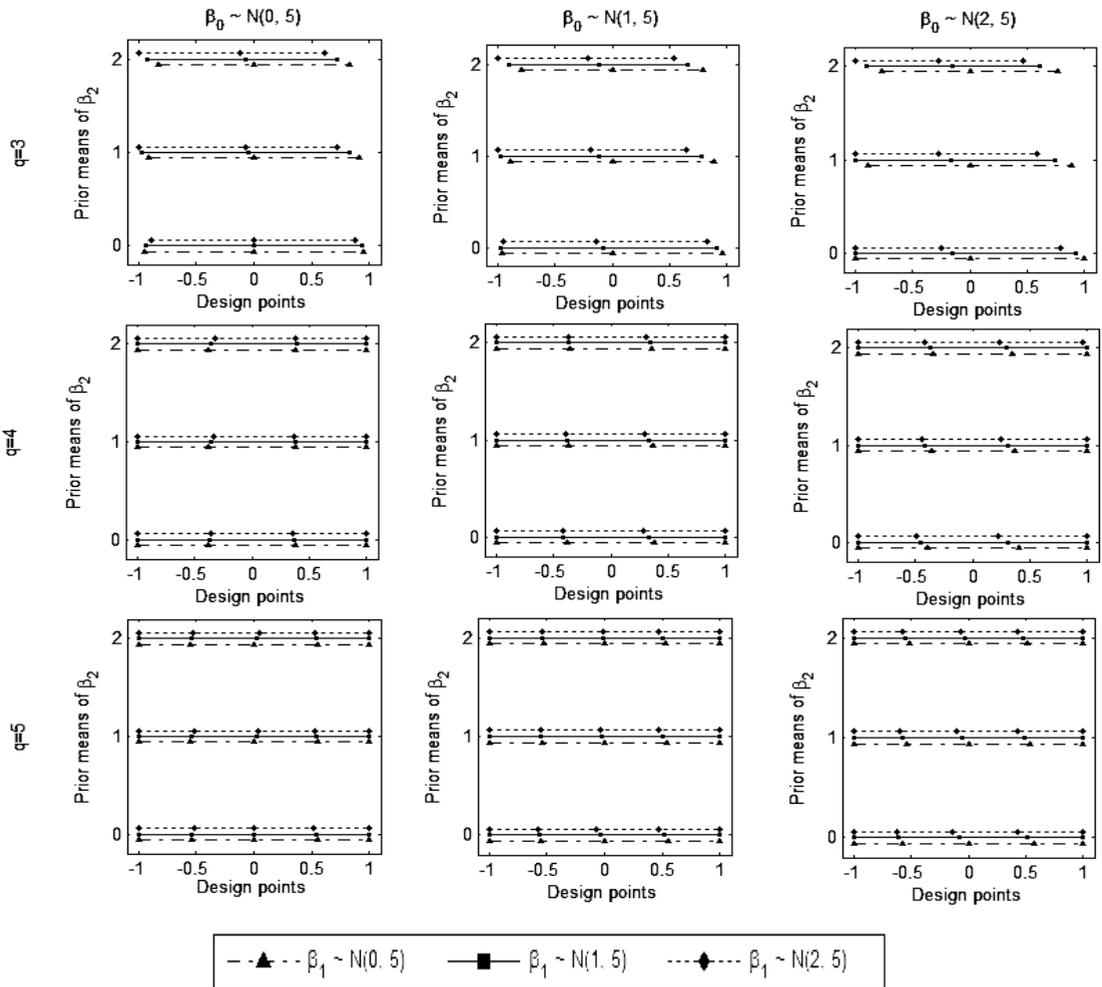


Figure 4.3 Bayesian D-optimal allocations of time points for the logistic mixed effects model with quadratic time effects with covariance structure D_2 .

optimal design points, and a separate line is given for each prior mean of β_1 . It can be observed from Figure 4.3 that the Bayesian D-optimal designs vary between the priors for $q = 3$, whereas, for $q = 4$ and 5 , the Bayesian optimal designs are similar under different prior distributions.

The Bayesian D-optimal time points are always symmetric around $t = 0$ when the prior mean of β_1 is zero, and also symmetric when both the prior mean of β_0 and β_2 are zero. The Bayesian D-optimal design becomes asymmetric as the prior means move away from zero.

This follows from the fact that the information, which is a function of $p(y_{ij}=1|\mathbf{b}_i)(1-p(y_{ij}=1|\mathbf{b}_i))$ is the same at time point t as at time point $-t$, if either the prior mean of β_1 is zero or both the prior mean of β_0 and β_2 are zero. Abebe et al. (2014a) have found a similar effect for Bayesian D-optimal designs for the fixed logistic model. The Bayesian optimal location of design points includes the two end points of the time interval, that is, $t = -1$ and $t = +1$, for every number of measurements q , except for $q = 3$, where the first or the last design point moves away from the end points of the time interval as the prior mean of β_1 or the prior mean of β_2 moves away from zero. Tekle et al. (2008) found a similar phenomenon for maximin designs of logistic mixed-effects models. They show that the optimal design points ‘move away’ from the edge of the design region as the value of the parameter increases.

To summarize, the Bayesian optimal allocation of time points is hardly affected by the choices of prior means and by the choices of covariance structures that we have considered and differs only slightly from equidistance for all q , except for $q = 3$.

4.6.3 Sensitivity of locally optimal design to the location of the parameter values

Figure 4.3 shows that the Bayesian optimal allocation of time points is robust against a change of prior means with a large prior variance. To study the effect of parameter values on the locally optimal design, i.e., Bayesian optimal design with prior variance zero, we computed locally D-optimal designs, by assuming different values for $\beta_0 = 0, 1$ and 2 , $\beta_1 = 0, 1$ and 2 and $\beta_2 = 0, 1$ and 2 , respectively. For each of the 27 combinations of β_0 , β_1 and β_2 , and each of the three covariance structures \mathbf{D} , and for autocorrelation $\rho = 0.5$, the locally D-optimal designs for $q = 3, 4, \dots, 10$ time points were computed. We then computed the relative efficiencies (RE) of the 27 locally optimal designs versus each other for a given number of time points q and \mathbf{D} . Finally, we compared the 27 locally D-optimal designs per q and \mathbf{D} in terms of their minimum relative efficiencies over the 27 combinations of β_0 , β_1 and β_2 . To save space we will only present the results for the time points $q = 3, 4, \dots, 7$. Table 4.1 shows the range of these 27 minimum relative efficiencies for each covariance structure \mathbf{D} and each q . Here it can be seen that the minimum relative efficiencies tend to increase with the number of time points q , from 0.55 for $q = 3$ to values near 1 for $q \geq 6$. The range of the minimum relative efficiencies is the same for all \mathbf{D} for a given q , and this range is large for $q \leq 5$. If we misspecify the location of the regression parameter values, the maximum loss in efficiency is about 45% for $q = 3$ and about 40% for $q = 4$. For larger number of time points ($q > 4$), the loss in efficiency is smaller than 9%.

To compare, Table 4.2 shows the range of minimum relative efficiencies computed over the 27 Bayesian D-optimal designs with prior variance of 5 for β_0 , β_1 and β_2 and a given q

and \mathbf{D} . It can be seen that the minimum relative efficiencies are all very high, above 0.96 for $q > 3$ and all \mathbf{D} , and that the minimum relative efficiencies increase with the number of time points q , from 0.69 for $q = 3$ to values near 1 for time points $q \geq 4$. Further, a comparison of Table 4.1 and Table 4.2 shows that the minimum relative efficiencies are much higher in Table 4.2 than in Table 4.1 for $q < 6$ and close to 1 for $q > 3$. This indicates that Bayesian D-optimal designs are robust against a change of priors, at least given a large prior variance. In contrast, the locally optimal designs (Bayesian designs with a prior variance zero) are not robust against a change of the location of the regression parameter values, unless $q > 4$.

Table 4.1 Range of minimum relative efficiencies of the 27 locally D-optimal designs versus each other, per q and per \mathbf{D} for the logistic mixed model with quadratic time effects.

Number of time points q	Range of minimum relative efficiencies		
	\mathbf{D}_1	\mathbf{D}_2	\mathbf{D}_3
$q = 3$	(0.5544, 0.7836)	(0.5641, 0.7867)	(0.5687, 0.7862)
$q = 4$	(0.6042, 0.8759)	(0.6125, 0.8759)	(0.6156, 0.8750)
$q = 5$	(0.9111, 0.9833)	(0.9167, 0.9850)	(0.8818, 0.9726)
$q = 6$	(0.9665, 0.9947)	(0.9673, 0.9945)	(0.9678, 0.9945)
$q = 7$	(0.9768, 0.9963)	(0.9771, 0.9960)	(0.9777, 0.9960)

Table 4.2 Range of minimum relative efficiencies of the 27 Bayesian D-optimal designs versus each other, per q and per \mathbf{D} for the logistic mixed model with quadratic time effects.

Number of time points q	Range of minimum relative efficiencies		
	\mathbf{D}_1	\mathbf{D}_2	\mathbf{D}_3
$q = 3$	(0.6911, 0.8909)	(0.6944, 0.8928)	(0.6964, 0.8922)
$q = 4$	(0.9686, 0.9809)	(0.9689, 0.9811)	(0.9675, 0.9816)
$q = 5$	(0.9848, 0.9924)	(0.9863, 0.9925)	(0.9852, 0.9926)
$q = 6$	(0.9901, 0.9957)	(0.9903, 0.9958)	(0.9905, 0.9959)
$q = 7$	(0.9936, 0.9970)	(0.9937, 0.9971)	(0.9938, 0.9972)

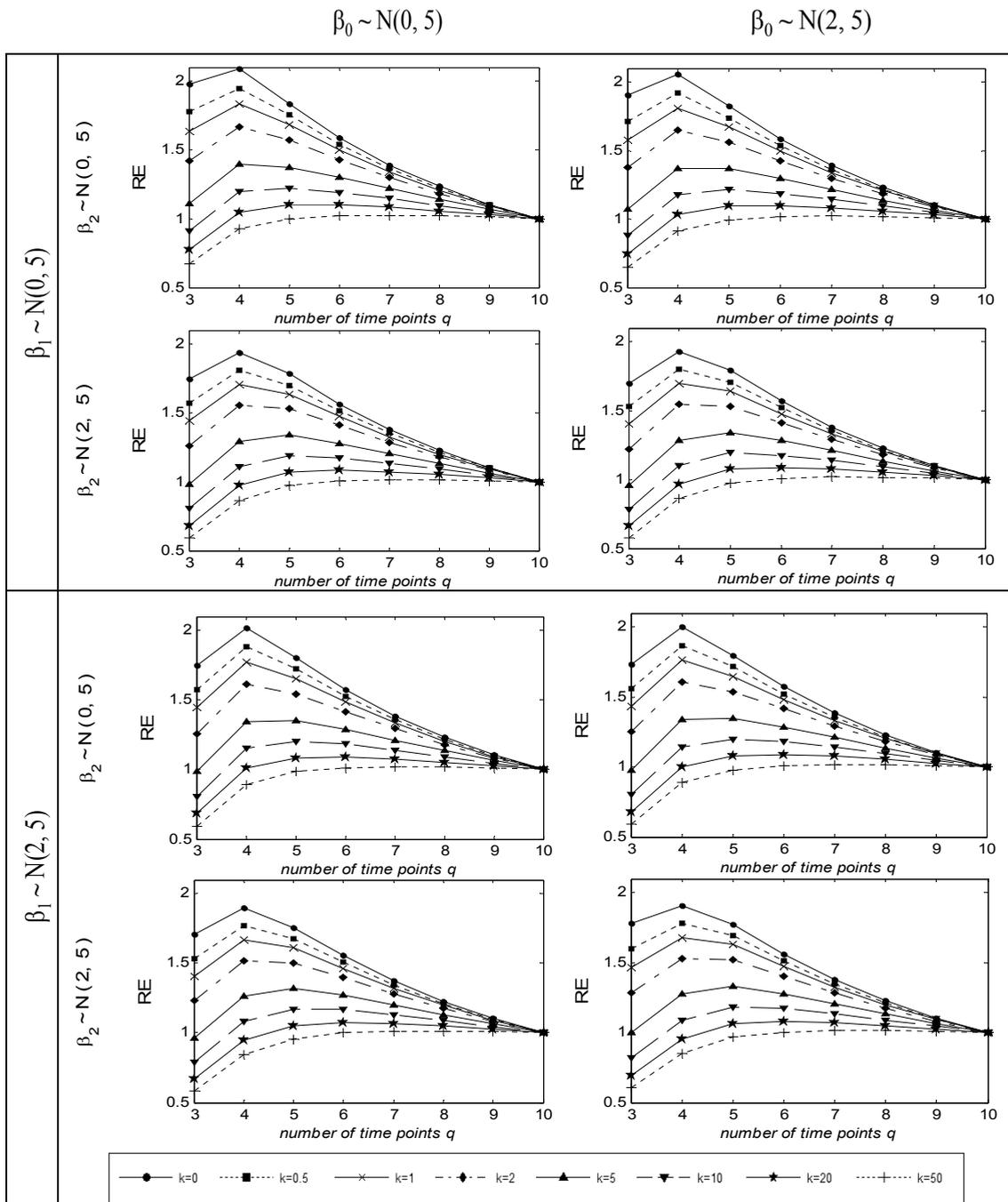


Figure 4.4 Relative efficiency of q time points compared with $q = 10$ time points for different cost ratios k for the logistic mixed model with quadratic time effects with covariance structure D_2 .

4.6.4 Selection of the optimal number of time points for a given cost ratio k

To find the optimal number of time points for the Bayesian D-optimal design with a given prior distribution and cost ratio k , we computed the relative efficiencies of Bayesian designs with $3 \leq q < 10$ time points compared to a Bayesian design with $q = 10$ time points for a given prior, \mathbf{D} and cost ratio k . The efficiency plots for models with the different covariance structures \mathbf{D} of the random parameters had a very similar pattern. Therefore, we will only present the relative efficiency plots for the random intercept and slope logistic model with covariance matrix \mathbf{D}_2 and $\rho = 0.5$. Further, since the efficiency plots for different priors also had a similar pattern, we will only present the relative efficiencies for extreme prior mean values. Figure 4.4 shows the relative efficiency plots for Bayesian designs with $q = 3, 4, \dots, 10$ time points compared to a Bayesian design with $q = 10$ time points for covariance structure \mathbf{D}_2 and for various cost ratios k . In the figure, the prior of β_0 varies across the columns and the priors of β_1 and β_2 vary across the rows. In each plot, the relative efficiencies are given on the vertical axis and the number of time points on the horizontal axis, with a separate line per cost ratio k . In each curve (cost ratio) within each plot (prior), the optimal number of time points is the number on the horizontal axis for which the relative efficiency is highest. This gives one Bayesian D-optimal design per prior and per cost ratio k . It can be seen from these plots that the optimal number of time points increases with an increasing cost ratio k . As the cost ratio k increases from $k = 0$ to $k = 10$, the optimal number of time points moves from $q = 4$ to $q = 5$, and for $k > 10$ the optimal number moves from $q = 5$ through $q = 6$ to $q = 7$ for $k = 50$. In summary, the Bayesian optimal number of time points q is the same for all considered priors and covariance structures, and only depends on the subject-to-measurement cost ratio k , ranging from $q = 4$ for $k = 0$ to $q = 7$ for $k = 50$.

4.6.5 Designs with equidistant time points relative to Bayesian optimal designs

The results in the previous sections show that for the logistic mixed model with quadratic time effects, the Bayesian optimal allocation of time points is only weakly affected by the choice of prior by choice of the covariance structure and by size of the autocorrelation, and that the Bayesian optimal allocation differs slightly from equidistance. The optimal number of time points of the Bayesian design increases with an increasing subject-to-measurement cost ratio. In most studies, the number and allocation of the time points are usually determined by non-statistical criteria, and the time points are frequently chosen to be equally spaced, which raises the question how efficient such designs are. Therefore, we computed and plotted the relative efficiencies $\{RE(\xi_{eq}; \xi^*(\pi)|\pi)\}$ for 3, 4, ..., 10 equidistant design points, with the first and last point on the boundaries of the design space, relative to the Bayesian D-optimal design for a given prior, covariance structure and cost ratio k . Since the pattern of the plots

was very similar for all considered priors and all covariance structures, we will only present results for extreme prior mean values and covariance structure \mathbf{D}_2 in Figure 4.5.

In each plot in Figure 4.5, the vertical axis shows the relative efficiencies of equidistance versus the Bayesian D-optimal design, and the horizontal axis shows the number of equidistant time points, and a separate line is given per cost ratio k . It can be seen that for cost ratios $k \leq 2$, four equidistant design points are highly efficient relative to the Bayesian optimal design, and that the relative efficiency decreases as the number of equidistant design points increases. For cost ratio $k = 5$, five equidistant design points is the most efficient of all equidistant designs, but four equidistant design points is still highly efficient. For cost ratio $k = 10$ and $k = 20$, five and six equidistant design points are highly efficient, respectively. For $k = 50$, seven or eight equidistant design points are highly efficient. In general, the relative efficiency is highest when the number of equidistant designs points is equal to the optimal number of time points for the Bayesian design and this relative efficiency is always above 0.96. Moreover, for a cost ratio $k \leq 5$, four equidistant design points are highly efficient, always above 0.95, and for a cost ratio $k > 5$, five or six equidistant design points are highly efficient, always above 0.95 for all the priors and covariance structures considered.

4.7 Application

4.7.1 The respiratory infection study

Let us now return to the example of the respiratory infection study in Indonesian children, introduced in section 4.2, to compare its actual design with its 6 equidistant time points with the Bayesian D-optimal designs under different priors, covariance structures \mathbf{D} and cost ratios k . Equation (4.9) was used to compute the relative efficiency of the actual design compared to the Bayesian optimal design for a given prior, covariance structure \mathbf{D} , and cost ratio k . Note that the optimal number of time points of the Bayesian optimal design depends on the cost ratio k , varying from $q = 4$ for $k \leq 2$ to $q = 7$ for $k = 50$ (see Figure 4.4). Table 4.3 shows the range of relative efficiencies of the actual design of the Indonesian respiratory infection study computed over all 27 priors for a given covariance structure \mathbf{D} and cost ratio k . In line with our results it can be seen that the efficiency of the actual design relative to the Bayesian design increases with the cost ratio k , from 0.76 to values near 1 for cost ratios $k \geq 10$. The range of the relative efficiencies is the same for all \mathbf{D} for a given cost ratio k , and this range is large for cost ratios $k \leq 2$, because the optimal Bayesian designs vary more between the priors when the number of time points q is small. In other words, Bayesian optimal designs become more similar for different priors as the number of time points q increases. For small cost ratios k , if we use the actual design with its 6 equidistant time points instead of the 4 time point Bayesian design, the loss in efficiency is at most 24%, which can be compensated by sampling about 31% more children. For large cost ratios ($k \geq 10$), if we use the 6 equidistant time points instead of the Bayesian design, the loss in efficiency is at most about 3%. This can be compensated by sampling about 3% more children.

Table 4.3 Range of relative efficiencies of the actual design of the Indonesian respiratory infection study relative to the Bayesian D-optimal designs over all 27 priors for each D and cost ratio k for quadratic time effects.

Cost ratio k	Range of relative efficiencies		
	D_1	D_2	D_3
$k = 0$	(0.7619, 0.8214)	(0.7602, 0.8187)	(0.7603, 0.8187)
$k = 0.5$	(0.7912, 0.8530)	(0.7894, 0.8502)	(0.7895, 0.8502)
$k = 1$	(0.8163, 0.8801)	(0.8145, 0.8772)	(0.8165, 0.8771)
$k = 2$	(0.8571, 0.9241)	(0.8552, 0.9211)	(0.8553, 0.9210)
$k = 5$	(0.9350, 0.9640)	(0.9329, 0.9631)	(0.9331, 0.9631)
$k = 10$	(0.9740, 0.9929)	(0.9734, 0.9931)	(0.9734, 0.9931)
$k = 20$	(0.9929, 0.9998)	(0.9931, 0.9999)	(0.9927, 0.9998)
$k = 50$	(0.9820, 0.9943)	(0.9833, 0.9949)	(0.9835, 0.9949)

4.8 Summary and Conclusion

The present chapter addresses the problem of the choice of the optimal number and allocation of time points for the logistic mixed-effects model with quadratic time effects for binary repeated measures with autocorrelation. Since budget constraints limit the number of subjects and number of time points at which measurements can be taken, and the cost per subject may be quite different from the cost per measurement, a subject-to-measurement cost ratio was considered to compare designs with each other. First the Bayesian D-optimal designs for $q = 3, 4, \dots, 10$ time points were computed for autocorrelations $\rho = 0.1, 0.2, \dots, 0.9$ and three different covariance structures under the various different prior distributions. Then, the effect of autocorrelation, prior and covariance structure of the random effects on the Bayesian D-optimal designs were studied for each q . The Bayesian D-optimal designs for different q 's were compared with each other in terms of their relative efficiencies for each prior distribution, for each covariance structure, per subject-to-measurement cost ratio. Finally, the relative efficiencies of equidistant time points versus the Bayesian D-optimal designs were studied for different subject-to-measurement cost ratios.

With respect to the Bayesian D-optimal designs, the results show that the optimal *allocation* of a given number time points q is hardly affected by the size of the autocorrelation at least within the range $[0.1, 0.9]$ (see Figures 4.1 and 4.2). Moreover, the Bayesian optimal designs with autocorrelations are also hardly affected by the choice of a covariance structure for the random effects. Contrary to the case of a zero autocorrelation, where the Bayesian D-optimal design is indeed affected by the choice of a covariance structure (Abebe et al.,

2014b). Furthermore, in the presence of an autocorrelation, the Bayesian optimal allocation of time points is hardly affected by the prior for all q , except for $q = 3$. The Bayesian optimal design points include the two endpoints of the time interval, that is, $t = -1$ and $t = +1$, for all number of repeated measurements q , except for $q = 3$. For $q = 3$ the first or the last design point moves away from the end points of the time interval as the prior means of the regression parameters increase.

The Bayesian optimal *number* of time points depends on the subject-to-measurement cost ratio, and four or five time points are highly efficient for cost ratios less than or equal to ten ($k \leq 10$), whereas, for cost ratios greater than ten ($k > 10$), six or seven time points are most efficient. Moreover, the Bayesian optimal number of time points is the same for all priors and all covariance structures. Finally, for a given cost ratio, the relative efficiencies of designs with equidistant time points relative to Bayesian optimal designs are highest when the number of equidistant time points are equal to the Bayesian optimal number of time points. This is because the Bayesian optimal allocation of time points are always approximately equidistant, except for $q = 3$ which was not optimal for any cost ratio anyway. The loss of efficiency by choosing equidistant time points instead of the Bayesian optimal designs remains less than 5%.

In conclusion, neither the optimal number nor the optimal allocation of time points appears to depend strongly on the prior, the covariance structure, or on the value of the autocorrelation. The optimal number of time points of the Bayesian D-optimal designs depends on the subject-to-measurement cost ratio and increases as the cost ratio increases. If there are limited computational resources, our recommendation is that researchers can use equidistant time points instead of Bayesian D-optimal designs, and they only need to consider the optimal number of time points in relation to the subject-to-measurement cost ratio. When the subject-to-measurement cost ratio is less than or equal to five, four equidistant time points are highly efficient, while, for a cost ratio greater than five, five or six equidistant time points are highly efficient. These equidistant designs are highly efficient relative to the Bayesian D-optimal designs for all priors and covariance structures considered and thus quite robust. Moreover, the precise allocation of the time points appears to be less crucial in general as long as researchers use the optimal number of time points for their cost ratio, and keep distance between successive measurements.

This research focuses on Bayesian optimal designs for longitudinal studies with a binary outcome under budget constraints. For continuous outcomes the effects of subject-to-measurement cost ratios on optimal designs have been studied in the literature (Bloch, 1986; Galbraith and Marschner, 2002; Winkens et al., 2007; Tekle et al., 2011). Winkens et al. (2007) and Tekle et al. (2011) have used locally D-optimal designs to determine the optimal number of time points for a given subject-to-measurement cost ratio, whereas Bloch (1986) and Galbraith and Marschner (2002) used statistical power as criterion. Furthermore, the results of this chapter assumed that all subjects are available over the total study period and that there is no dropout. However, in longitudinal studies missing data due to dropout are common and it is difficult at the design stage to predict the level of the dropout. The impact of

anticipated dropout on power as well as ways to allow dropout in linear mixed models at the design stage were reported by Galbraith and Marschner (2002) and Verbeke and Lesaffre (1999). For continuous outcomes, optimal designs that do take dropout into account have been studied by Imhof et al. (2004) and Azurduy et al. (2008). They showed that under certain conditions, D-optimal designs are robust against dropout and the loss in efficiency is moderate. Further research is needed to see whether their results will still hold in the Bayesian context and for binary outcomes. Other important issues for future work are Bayesian design for models using splines and for models comparing two or more groups over time.

CHAPTER 5

Robustness of Bayesian D-optimal design for the logistic mixed model against misspecification of autocorrelation

This chapter is under review with co-authors Frans E.S. Tan, Gerard J.P. Van Breukelen and Martijn P.F. Berger

Abstract

In medicine and health sciences mixed effects models are often used to study time-structured data. Optimal designs for such studies have been shown useful to improve the precision of the estimators of the parameters. However, optimal designs for such studies are often derived under the assumption of a zero autocorrelation between the errors, especially for binary data. Ignoring or misspecifying the autocorrelation in the design stage can result in loss of efficiency. This chapter addresses robustness of Bayesian D-optimal designs for the logistic mixed effects model for longitudinal data with a linear or quadratic time effect against incorrect specification of the autocorrelation. To find the Bayesian D-optimal allocations of time points for different values of the autocorrelation, under different priors for the fixed effects and different covariance structures of the random effects, a scalar function of the approximate variance-covariance matrix of the fixed effects is optimized. Two approximations are compared, one based on a first order penalized quasi likelihood (PQL1), and one based on an extended version of the generalized estimating equations (GEE). The results show that Bayesian D-optimal allocations of time points are robust against misspecification of the autocorrelation and are approximately equally spaced. Moreover, PQL1 and extended GEE give essentially the same Bayesian D-optimal allocation of time points for a given subject-to-measurement cost ratio. Furthermore, Bayesian optimal designs are hardly affected either by the choice of a covariance structure or by the choice of a prior distribution.

Keywords: Bayesian D-optimal designs; logistic mixed effects model; subject-to-measurement cost ratio; robustness; autocorrelation.

5.1 Introduction

Optimal designs for longitudinal studies have been shown useful to improve the precision of the estimators of unknown parameters of a statistical model. For the linear random effects model optimal designs were studied by Tan and Berger (1999) and Berger and Tan (2004), among others. Ouwens et al. (2006) and Tekle et al. (2008) considered optimal designs for the logistic mixed effect model using a maximin criterion. Sinha and Xu (2011) studied the performance of locally D-optimal sequential designs for the generalized linear mixed model. Niaparast (2009) and Niaparast and Schwabe (2013) derived optimal designs for a Poisson regression model with random effects using locally D-optimal designs. Han and Chaloner (2004) studied Bayesian optimal designs for nonlinear mixed effects models. Recently, Abebe et al. (2014b) have studied Bayesian D-optimal designs for a logistic mixed effects model.

Although optimal designs in the framework of linear and nonlinear mixed effects models have been studied extensively in the literature, very few studies have been done on optimal designs for longitudinal studies with a binary response in the presence of autocorrelated errors in the model. While autocorrelated errors are frequently assumed in data analysis, they are largely ignored in optimal design research. However, ignoring or misspecifying the autocorrelation among the errors in the design stage can result in loss of efficiency. Since researchers in general do not know the size of the autocorrelation parameter in the design stage, it would be helpful to know whether designs are robust against incorrect specification of the autocorrelation among the errors in the model. The present chapter addresses this robustness question for longitudinal studies with a binary response.

The generalized linear mixed model (GLMM) is the most frequently used model for the analysis of longitudinal binary data. Complications in the design process arise from the random effects. Due to the random effects in GLMM, neither the maximum likelihood estimator nor the Fisher information matrix can be written in closed form. Hence, either numerical methods or approximations to the likelihood must be used. Numerical methods require large computational resources and more importantly they require full knowledge of the data (Moerbeek et al. 2003; Han and Chaloner 2004), making them computationally inconvenient for optimal design procedures. To overcome this problem, approximation methods are employed. Optimal designs obtained by approximation methods were discussed by Moerbeek et al. (2001), Moerbeek and Maas (2005), Ouwens et al. (2006), Woods and van de Ven (2006), Tekle et al. (2008), Niaparast (2009) and Niaparast and Schwabe (2013). There is a large statistical literature on various approximation methods, but here, for the purpose of obtaining optimal designs, we will focus on the two most frequently used ones, which are implemented in commercially available software packages: Penalized quasi-likelihood (PQL) (Breslow and Clayton 1993; Jang and Lim 2009) and an extended version of generalized estimating equations (GEE) (Molenberghs and Verbeke 2005).

In the optimal design literature a number of optimality criteria have been developed, which optimize a scalar function of the information matrix. In this study, we will restrict ourselves to the most often used criterion, namely the D-optimality criterion, which does not depend on

the scale of the predictor. See Atkinson et al. (2007) for details. The information matrix for nonlinear models depends on the unknown parameters and so the optimal design will also depend upon the value of these parameters. Therefore, to find optimal designs we must have knowledge about the parameters in advance. A simple approach to this problem is to design an experiment to be efficient for a best guess of the parameter values, which leads to locally optimal designs (Chernoff 1953). However, such designs may not be efficient for all other parameter values.

In this chapter, the Bayesian approach will be applied to overcome the local optimality problem by using a prior distribution for the unknown parameter values. Research has been done on Bayesian designs for binary response models (see, e.g., Khuri et al. 2006) but, to our knowledge no research has been done yet on Bayesian optimal designs for binary data with autocorrelated errors. Therefore, the aim of this study is to investigate how robust Bayesian D-optimal designs are against misspecification of the autocorrelation in logistic mixed effects models with repeated measurements of a binary response. We will first compare numerically Bayesian D-optimal designs based on PQL with extended GEE in terms of the resulting Bayesian optimal designs for various subject-to-measurement cost ratios and priors. Subsequently, we will study the effect of different values of the autocorrelation on Bayesian optimal designs and the robustness of the latter against incorrect specification of the autocorrelation.

This chapter is organized as follows. In section 5.2, we briefly describe a quadratic time effect logistic mixed model. The variance-covariance matrix of parameter estimators will be described in section 5.3. The Bayesian D-optimality criterion and the corresponding relative efficiency are given in section 5.4. In section 5.5, we will describe our numerical approach. Thereafter, results are presented in section 5.6. Finally, a brief summary and conclusion is provided in section 5.7.

5.2 Model specification

Let $\mathbf{y}_i = (y_{i1}, \dots, y_{iq})'$ be the $q \times 1$ vector of binary responses y_{ij} of subject i at time points $j = 1, \dots, q$, for each subject $i = 1, 2, \dots, N$, and it is assumed that all subjects have measurements at the same time points. These responses y_{ij} are either 0 or 1 and obtained at values x_j of an explanatory variable x for subject i . Moreover, given the subject-specific random effect vector \mathbf{b}_i , the binary responses y_{ij} of \mathbf{y}_i are assumed to be Bernoulli distributed with probability of success $p(y_{ij} = 1 | \mathbf{b}_i)$. These probabilities are related to the fixed and random effects via the logit link function. The corresponding logistic mixed effects model is given by:

$$\text{logit}(p(y_{ij} = 1 | \mathbf{b}_i)) = \log\left(\frac{p(y_{ij} = 1 | \mathbf{b}_i)}{1 - p(y_{ij} = 1 | \mathbf{b}_i)}\right) = \mathbf{x}_j' \boldsymbol{\beta} + \mathbf{z}_j' \mathbf{b}_i, \quad (5.1)$$

where the $p \times 1$ vector \mathbf{x}_j is the design vector of the explanatory variables at the j th measurement for subject i , $\boldsymbol{\beta}$ is the corresponding $p \times 1$ vector of fixed effects, and \mathbf{z}_j is the $r \times 1$ design vector for the random effects that is usually a subset of vector \mathbf{x}_j . The vector \mathbf{b}_i is the corresponding $r \times 1$ vector of random effects, which is assumed to have a multivariate normal distribution with mean zero and covariance matrix \mathbf{D} .

We assume a quadratic ($p = 3$) time effect, i.e., the design vector is $\mathbf{x}_j' = (1 \ t_j \ t_j^2)$ and $\boldsymbol{\beta} = (\beta_0 \ \beta_1 \ \beta_2)'$, where t_j is the time point of the j th measurement, and β_0 , β_1 and β_2 are the fixed effects. Suppose that we allow the intercept and linear slope to vary randomly between subjects. Then the design vector is $\mathbf{z}_j' = (1 \ t_j)$ and $\mathbf{b}_i = (b_{0i} \ b_{1i})'$, where b_{0i} and b_{1i} are the corresponding random (subject-specific) deviations from the fixed effects. Thus, according to model (5.1), the log-odds of a positive response for subject i at time t_j is given by:

$$\text{logit}(p(y_{ij} = 1 | \mathbf{b}_i)) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_j + \beta_2 t_j^2. \quad (5.2)$$

5.3 Estimation methods

As mentioned in the introduction, for the purpose of optimal design, we will use an approximate variance-covariance matrix of the parameter estimators based on either of two methods: PQL (Breslow and Clayton 1993; Jang and Lim 2009) and an extension of GEE (Zeger et al. 1988; Molenberghs and Verbeke 2005). Details of these methods will be given in the next sub-sections.

5.3.1 First order penalized quasi-likelihood (PQL1)

We use an approximate variance-covariance matrix based on a first order penalized quasi-likelihood (PQL1). See Breslow and Clayton (1993) and Jang and Lim (2009) for details. PQL1 performs well in terms of point estimation since it produces the smallest mean squared error (MSE) and the bias of the estimators decrease as the sample size increases (Breslow and Clayton 1993; Moerbeek et al. 2003; Jang and Lim 2009). The variance-covariance matrix of the estimator $\hat{\boldsymbol{\beta}}$ of the parameter $\boldsymbol{\beta}$ for the logistic mixed effects models (5.1) is approximated in PQL1 by:

$$\text{var}(\hat{\boldsymbol{\beta}}) \approx (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}, \quad (5.3)$$

where \mathbf{X} is the $Nq \times p$ design matrix formed by stacking $\{\mathbf{x}_j'\}$ for N subjects and q time points, and the $Nq \times Nq$ block-diagonal matrix \mathbf{V} has N blocks of $q \times q$ variance-covariance matrices given by:

$$\mathbf{v}_i \approx \mathbf{w}_i^{-1/2} \mathbf{R}(\rho) \mathbf{w}_i^{1/2} + \mathbf{Z} \mathbf{D} \mathbf{Z}' . \quad (5.4)$$

The $q \times q$ matrix $\mathbf{R}(\rho)$ is the residual correlation matrix, \mathbf{Z} is the $q \times r$ design matrix with rows \mathbf{z}_j' , $j = 1, \dots, q$, the $r \times r$ matrix \mathbf{D} is the variance-covariance matrix of the random effects, and \mathbf{w}_i^{-1} is the diagonal matrix of the conditional variances of the transformed responses given the random effects \mathbf{b}_i , which is equal to the inverse of the diagonal matrix of the conditional variances of the untransformed responses given the random effects \mathbf{b}_i (See for detail Moerbeek et al. 2001; Molenberghs and Verbeke 2005 page 270). Note that, under conditional independence $\mathbf{R}(\rho)$ is an identity matrix and equation (5.4) becomes

$$\mathbf{v}_i \approx \mathbf{w}_i^{-1} + \mathbf{Z} \mathbf{D} \mathbf{Z}' . \quad (5.5)$$

The diagonal matrix of the conditional variances of the untransformed responses given the random effects \mathbf{b}_i , is given by:

$$\mathbf{w}_i = \text{diag}(w_{i1}^{b_i}, \dots, w_{iq}^{b_i}), \quad (5.6)$$

where $w_{ij}^{b_i} = \text{var}(y_{ij} | \mathbf{b}_i)$, for $i = 1, \dots, N$, $j = 1, \dots, q$. Since the random effects are unknown in the design stage, we will generate \mathbf{b}_i from a multivariate normal distribution with mean zero and variance-covariance \mathbf{D} .

5.3.2 Extension of generalized estimating equations (GEE)

The GEE approach has been extended by Zeger et al. (1988) and Molenberghs and Verbeke (2005) to include autocorrelations of the errors in the standard formulation of GLMM. The asymptotic variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ for the logistic mixed effects models (5.1) with autocorrelation, based on the extension of the GEE approach is approximated by:

$$\text{var}(\hat{\boldsymbol{\beta}}) \approx \left(\sum_{i=1}^N \frac{\partial \mathbf{P}_i'}{\partial \boldsymbol{\beta}} \mathbf{u}_i^{-1} \frac{\partial \mathbf{P}_i}{\partial \boldsymbol{\beta}'} \right)^{-1}, \quad (5.7)$$

where $\hat{\boldsymbol{\beta}}$ is the estimator of $\boldsymbol{\beta}$ for model (5.1), $\mathbf{P}_i = (p(y_{i1} | \mathbf{b}_i), \dots, p(y_{iq} | \mathbf{b}_i))'$, and the working variance-covariance matrix \mathbf{u}_i of the responses is given by:

$$\mathbf{u}_i \approx \mathbf{w}_i^{1/2} \mathbf{R}(\rho) \mathbf{w}_i^{1/2} + \mathbf{w}_i \mathbf{Z} \mathbf{D} \mathbf{Z}' \mathbf{w}_i . \quad (5.8)$$

When there are no residual correlations in $\mathbf{R}(\rho)$, a conditional independence model or purely random effects model results and equation (5.8) reduces to

$$\mathbf{u}_i \approx \mathbf{w}_i + \mathbf{w}_i \mathbf{Z} \mathbf{D} \mathbf{Z}' \mathbf{w}_i . \quad (5.9)$$

where w_i is the diagonal matrix of the conditional variances of untransformed responses given the random effects $b_i = 0$, which is given by:

$$w_i = \text{diag}(w_{i1}^{b_i=0}, \dots, w_{iq}^{b_i=0}), \quad (5.10)$$

where $w_{ij}^{b_i=0} = \text{var}(y_{ij} | b_i = 0)$, for $i = 1, \dots, N$, $j = 1, \dots, q$ (Molenberghs and Verbeke, 2005 page 443).

In the present chapter, a first order autoregressive (AR1) correlation structure is considered for the residual correlation matrix $R(\rho)$ i.e., the (j, h) element of the residual correlation matrix $R(\rho)$ is equal to $\rho^{|t_j - t_h|}$, where $j, h = 1, \dots, q$, and so ρ is the autocorrelation between two residuals for which $|t_j - t_h| = 1$. This autocorrelation structure implies that repeated measurements which are closer in time are more highly correlated and that the correlation decreases as the distance between the time points increases.

5.4 Optimal design and relative efficiency

5.4.1 Bayesian D-optimality criterion

To introduce the notation for the optimality criterion, suppose that the study to be designed will have q time points at which measurements are taken for all N subjects. The design space \mathcal{E} then contains all designs with $q \times 1$ time vectors, i.e.,

$$\mathcal{E} = \left\{ (t_1, t_2, \dots, t_q)' : t_j \in [a, b] \right\}, \quad (5.11)$$

where t_1, t_2, \dots, t_q are the q time points, and $q \geq p$, to make these fixed effects identifiable with p being the number of fixed parameters of the model. The time interval $[a, b]$ is assumed to be fixed by substantive constraints within the field of application, for instance the total follow-up time in a cohort study of smoking or unemployment. Note that $p = 3$ for our model (5.2). A design ξ_q is an element of the design space \mathcal{E} if it has q time points within the time interval $[a, b]$. The goal of optimal design is to maximize parameter information by minimizing variability of the parameter estimators. Optimal designs are usually selected by minimizing a real-valued function of the variance-covariance matrix, which is known as optimality criterion (see, e.g., Silvey 1980). There are numerous optimality criteria proposed in the literature, such as the D-, A- or G-optimality criteria. In this chapter, we will focus on the most well known and widely used optimality criterion, i.e., the D-optimality criterion. This optimality criterion has some nice properties: 1. It minimizes the volume of the asymptotic confidence ellipsoid of the parameters. 2. It does not depend on the coding used for the endpoints of the chosen time interval $[a, b]$. This means that if the coding for the time interval is transformed linearly, a D-optimal design for the new time interval is obtained by applying the same linear transformation to the D-optimal design for the old interval (See

Ouwens et al. 2006). The D-optimal design ξ_q^* is the design among all possible designs ξ_q with q time points for which the determinant of the variance-covariance matrix is minimized or equivalently, for which the determinant of the inverse of variance-covariance matrix is maximized (Atkinson et al. 2007 page 151). The approximate variance-covariance matrix depends on the unknown parameter vector β . The Bayesian approach resolves this dependency by taking the expectation of a function of the approximate variance-covariance matrix over a prior distribution for parameter vector β . Thus, the Bayesian D-optimality criterion (by analogy with the non-Bayesian optimal design) is defined as follows:

$$\phi_D(\xi|\pi) = E_{\beta} \log \left| \text{var}(\hat{\beta}) \right|^{-1} = \int \log \left| \text{var}(\hat{\beta}) \right|^{-1} \pi(\beta) d(\beta), \quad (5.12)$$

where $\pi(\beta)$ is the prior distribution for β and $\text{var}(\hat{\beta})$ is the variance-covariance matrix of $\hat{\beta}$. The design which maximizes (5.12) is the Bayesian D-optimal design.

5.4.2 Cost function and relative efficiency

The precision of estimating the fixed effect parameters increases by taking more measurements and sampling more subjects. However, the addition of subjects and of measurements per subject will increase the costs of the study and these are usually limited by budget constraints. Therefore, it is reasonable to take into account the costs of a longitudinal study when designs are compared with each other. There are two main components of these costs. These are the costs for the recruitment of subjects and the costs of the measurements once a subject has been recruited. Let the cost of recruiting a subject be C_1 and the cost of one measurement per subject be C_2 . Then the total costs of a longitudinal study with q time points and N subjects, excluding overhead cost, is given by the linear cost function:

$$C = C_1 N + C_2 N q = N C_2 (k + q), \quad (5.13)$$

where $k = C_1/C_2$ is the ratio of the cost of adding a new subject to the cost of an additional measurement per subject. To study the effect of this cost ratio on the Bayesian optimal design, we will vary the cost ratio from $k = 0$ to $k = 5$.

To compare different designs, we will use their relative efficiencies while fixing the total costs C . This means that the designs can differ in terms of the number of subjects N and the number and timing of the measurements q . First we compute the Bayesian D-optimal designs using N fixed and then we correct for costs and different q and N as follows. Let $\phi_D(\xi_q^*|\pi)$ denote the value of design criterion (5.12) for the optimal design ξ_q^* with q time points. Then the relative efficiency (RE) of an arbitrary design ξ_s with s time points relative to the optimal design ξ_q^* is defined as:

$$\text{RE}(\xi_s; \xi_q^* | \pi) = \frac{k+q}{k+s} \left[\exp \left\{ \frac{\phi_D(\xi_s | \pi) - \phi_D(\xi_q^* | \pi)}{p} \right\} \right], \quad (5.14)$$

where $(q, s) \geq p$ and π is the prior distribution for the fixed effects. If the value of this relative efficiency is close to unity, then the design ξ_s is about equally efficient as the optimal design ξ_q^* for a given prior π . The inverse of this relative efficiency is the number of times that a design ξ_s must be replicated to have the same efficiency as the optimal design ξ_q^* . Note that this relative efficiency (5.14) is an extension of the relative efficiency given by Chaloner and Larntz (1989). When the ratio $\frac{k+q}{k+s}$ is one, it reduces to their relative efficiency. This ratio follows from (5.13) and takes into account that the design with the smaller number of time points can have more subjects for the same budget.

5.5 Numerical approach

5.5.1 Method of Optimization

The Bayesian D-optimal designs for the logistic mixed effect model (5.2) were found numerically by maximization of the criterion value (5.12) among all candidate designs for a given prior distribution of the parameters. Details of this will be given in the next subsections.

5.5.1.1 Sampling parameter values from priors to compute the criterion

To construct Bayesian designs for continuous prior distributions, all candidate designs must be evaluated in terms of their criterion values as defined by (5.12). However, evaluation of the integration over the prior distribution is very complicated and can not easily be done analytically. A numerical approximation of the integral is necessary. Numerical approximations can be done by sampling parameter values from the prior distribution and then replacing the integral in (5.12) with a summation over the sample (Atkinson et al. 2007; Chaloner and Verdinelli 1995). Estimating (5.12) using the traditional sampling (pseudo Monte Carlo) method requires very large samples from the prior to minimize sampling error. Thus, this approach is costly in terms of computing time. In this chapter, we will use an Adaptive Rejection Metropolis Sampling (ARMS) algorithm (Gilks and Wild 1992; Gilks et al. 1995), which is a more efficient sampling algorithm that requires a smaller sample to obtain a good approximation of the design criterion (5.12).

5.5.1.2 Optimization algorithm for finding an optimal design

To find candidate designs and in particular the optimal design, the program uses the FMINCON function of MATLAB version 7.10.0499 (R2010a). This function performs constrained non-linear optimization and requires an initial design ξ_0 . Without loss of

generality, we coded the time interval as $[-1, 1]$, and we used equally spaced time points as initial designs. The following global search algorithm was used to find the Bayesian D-optimal designs for a given prior distribution of the parameters:

1. Take samples from the prior distribution of the parameters using ARMS.
2. Compute the Bayesian D-optimal allocation of q time points, using $q = p$ equidistant time points as initial design, where p is the number of fixed parameters of the model.
3. Increase the number of time points q by one and perform step 2 again to find the Bayesian optimal design (allocation) for the new value of q . Repeat step 2 and 3 until the maximum number of time points q (user specified) is reached.
4. Afterwards, select the optimal number of time points q for the Bayesian D-optimal design by computing the relative efficiencies of the Bayesian optimal designs with different numbers of time points against each other for a given subject-to-measurement cost ratios and then select the design with the highest relative efficiency. Do this for each cost ratio considered to obtain one Bayesian optimal design per cost ratio for a chosen prior distribution.

5.5.2 Priors and covariance structures

We considered the logistic mixed model of equation (5.2) with a quadratic ($p = 3$) time effect, and a number of different priors of the fixed parameters and different covariance structures \mathbf{D} of the random effects. We assumed an independent trivariate normal prior for the fixed effects β_0 , β_1 and β_2 , with two different prior mean values, i.e., $\mu = [0, 2]$, and two prior variances, i.e., $\sigma^2 = [1, 5]$ for each of the fixed effects (β_0 , β_1 and β_2). All 8 combinations of the prior mean values were used to generate 8 different prior distributions per σ^2 . Thus in total there were 16 normal prior distributions. We used $N = 100$ subjects to estimate the determinant in (5.12) as part of the search for Bayesian D-optimal designs, because Abebe et al. (2014b) found that the estimated determinant of the variance-covariance matrix of parameter estimates becomes stable when N is at least 100.

The following three random effects covariance structures \mathbf{D} were considered:

- $\mathbf{D}_1 = 0.5$, i.e., a random intercept model with variance of the random intercept 0.5.
- $\mathbf{D}_2 = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$, i.e., a random intercept and random linear slope model with variances $d_{11} = 0.5$ for the intercept and $d_{22} = 0.5$ for the slope, and no slope-intercept covariance.
- $\mathbf{D}_3 = \begin{bmatrix} 0.5 & 0.25 \\ 0.25 & 0.5 \end{bmatrix}$, i.e., the same as \mathbf{D}_2 except that there is now a slope-intercept covariance of 0.25, giving a correlation of 0.5.

Because in practice the number of repeated measurements in a longitudinal study is usually less than seven, we restricted the number of time points q to a maximum of seven measurements. This also kept the computation time manageable. Since the optimal combination of N and q depends on the relative costs of subjects and measurements, the

following cost ratios $k = 0, 0.5, 1, 2$ and 5 were considered. This means that, e.g. when $k = 5$, the costs of recruiting a new subject is five times as expensive as the cost of one measurement per subject. To deal with the uncertainty of the size of the autocorrelation parameter ρ , we considered values of ρ in the range $[0, 0.9]$. We performed all computations in MATLAB R2010a (32bit) using a computer with a 2.67 GHz Intel processor and 6 GB RAM. The total computing time to generate all the Bayesian optimal time points $3 \leq q \leq 7$ based on PQL1, is about 22 hours per covariance structure for a prior with $\sigma^2 = 1$, and about 52 hours for a prior with $\sigma^2 = 5$. Whereas, based on extended GEE method, the total computing time is about 15 minutes for a prior with $\sigma^2 = 1$, and about 1 hour for a prior with $\sigma^2 = 5$. This is due to the fact that extended GEE does not incorporate the random effects \mathbf{b}_i , i.e., the diagonal matrix of the conditional variances of the responses \mathbf{w}_i in equation (5.10) are evaluated under $\mathbf{b}_i = 0$ and thus the variance-covariance matrix \mathbf{u}_i of responses equation (5.8) becomes the same for all subjects N .

5.6 Results

5.6.1 Bayesian optimal allocations of time points

Firstly, Bayesian D-optimal designs were computed for each of the different number of time points $3 \leq q \leq 7$, for each of the 16 different prior distributions, for each of the three covariance structures, and for different values of the autocorrelation, i.e., $\rho \in [0, 0.9]$, using the approximated variance-covariance matrix based on both PQL1 and extended GEE methods. Thereafter, Bayesian D-optimal designs for these two methods were compared per q , per prior and per covariance structure \mathbf{D} , for different autocorrelations ρ . As there was no substantive difference between different values of autocorrelation ρ , with respect to the comparison PQL1 and extended GEE, we will only present results for $\rho = 0.1$ here, postponing the systematic comparison between different values of ρ to section 5.6.3. Moreover, since the plots of Bayesian optimal designs for models with the different covariance structures \mathbf{D} of the random parameters have a very similar pattern, we will only present Bayesian D-optimal designs for the covariance structure \mathbf{D}_2 .

Figure 5.1 and Figure 5.2 show the Bayesian D-optimal allocations of designs with $q = 3, 4, \dots, 7$ time points for each of the 8 different prior distributions and with $\rho = 0.1$, respectively, for the priors with $\sigma^2 = 1$ and $\sigma^2 = 5$. In Figure 5.1 and Figure 5.2, the prior of β_0 varies across the columns and the priors of β_1 and β_2 vary across the rows. Within each plot, the vertical axis shows the number of time points q , and the horizontal axis shows the allocations of the Bayesian optimal design points for a given q , for each of the two methods with separate lines. It can be seen from these plots that PQL1 and extended GEE give very similar Bayesian optimal allocations of time points for a given q and prior. Moreover, there

are other several interesting features in this plot. Firstly, the Bayesian D-optimal allocation of time points becomes approximately equally spaced, when the number of time points q increases. As a result the Bayesian D-optimal designs become similar under different priors as q increases (see Figure 5.1 and Figure 5.2). For the priors with variance $\sigma^2 = 1$ (Figure 5.1), the Bayesian D-optimal designs vary more between the priors for $q = 3, 4$ and 5 , than for $q = 6, 7$, whereas, for the priors with variance $\sigma^2 = 5$ (Figure 5.2), the designs vary between the priors more for $q = 3$, than for $q = 4, 5, 6$ and 7 . This hold for other ρ values in a similar way. In other words the prior with large variance is more robust against a change of prior means because, due to the larger uncertainty, the choice of the prior mean values becomes less important. Secondly, the Bayesian D-optimal allocations of time points are always symmetric around $t = 0$ when $\mu_{\beta_1} = 0$, and also symmetric when both μ_{β_0} and μ_{β_2} are zero. Atkinson and Haines (1996) and Abebe et al. (2014a) have found a similar effect for the fixed logistic model. This follows from the fact that the information, which is a function of $p(y_{ij} = 1 | \mathbf{b}_i)(1 - p(y_{ij} = 1 | \mathbf{b}_i))$ is the same at time point t as at time point $-t$, if either $\mu_{\beta_1} = 0$ or both μ_{β_0} and μ_{β_2} are zero. Thirdly, the Bayesian optimal allocation of design points includes the two end points of the time interval, that is, $t = -1$ and $t = +1$, for every number of measurements q , except for $q = 3, 4$ and 5 for some priors with $\sigma^2 = 1$ (Figure 5.1), and except for $q = 3$ for some priors with $\sigma^2 = 5$ (Figure 5.2), where one or both of the extreme design points moves away from the end points of the time interval when $\mu_{\beta_2} = 2$. In short the optimal designs are not always at the extreme values of the time interval. The non-linear relation between the response probability and the predictor may cause this move. Atkinson Haines (1996), Chipman and Welch (1996) and Abebe et al. (2014a) have found a similar phenomenon in their comparison of D-optimal designs for logistic fixed effects. The optimal design points ‘move in’ from the edge of the design region as the values of the parameters increase. Recently, Abebe et al. (2014b) and Tekle et al. (2008) found a similar result for logistic mixed effects models.

From Figure 5.1 and Figure 5.2, it seems that the Bayesian D-optimal allocations of time points are very similar for both methods, PQL1 and extended GEE. To check this we computed the relative efficiency of the Bayesian D-optimal design based on PQL1 compared to a Bayesian optimal design based on extended GEE and vice versa, for a given number of time points q , prior, covariance structure \mathbf{D} and autocorrelation ρ (these results are not included here). These relative efficiencies were all very high, i.e., above 0.98 for all q and all priors, \mathbf{D} and ρ . This indicates that Bayesian D-optimal designs based on the PQL1 and those based on the extended GEE do not differ noticeably either in terms of design points or in terms of efficiency.

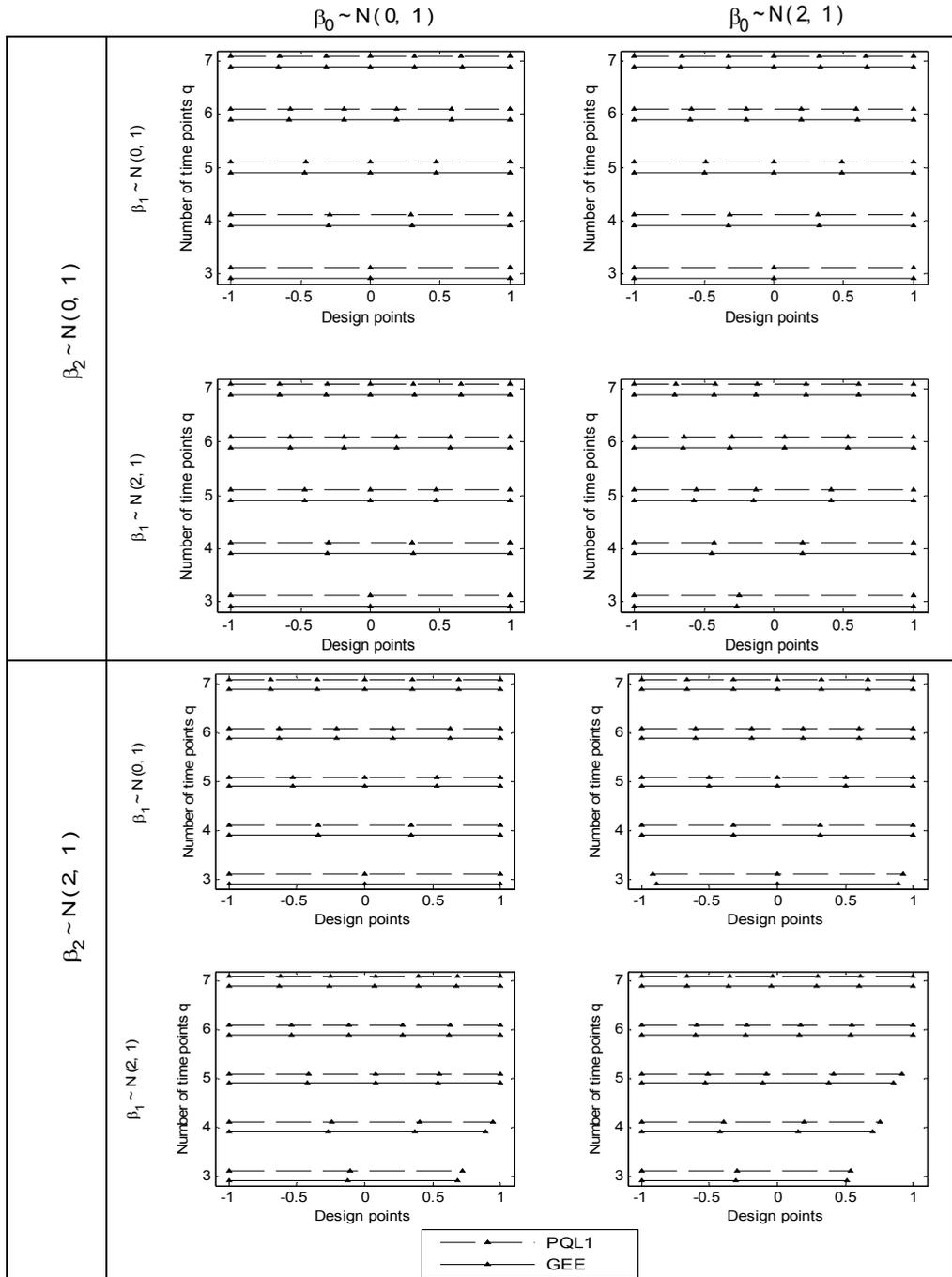


Figure 5.1 Bayesian D-optimal designs using PQL1 and extended GEE approach for the logistic mixed model with quadratic time effects with covariance structure D_2 and $\rho = 0.1$ for the priors with $\sigma^2 = 1$.

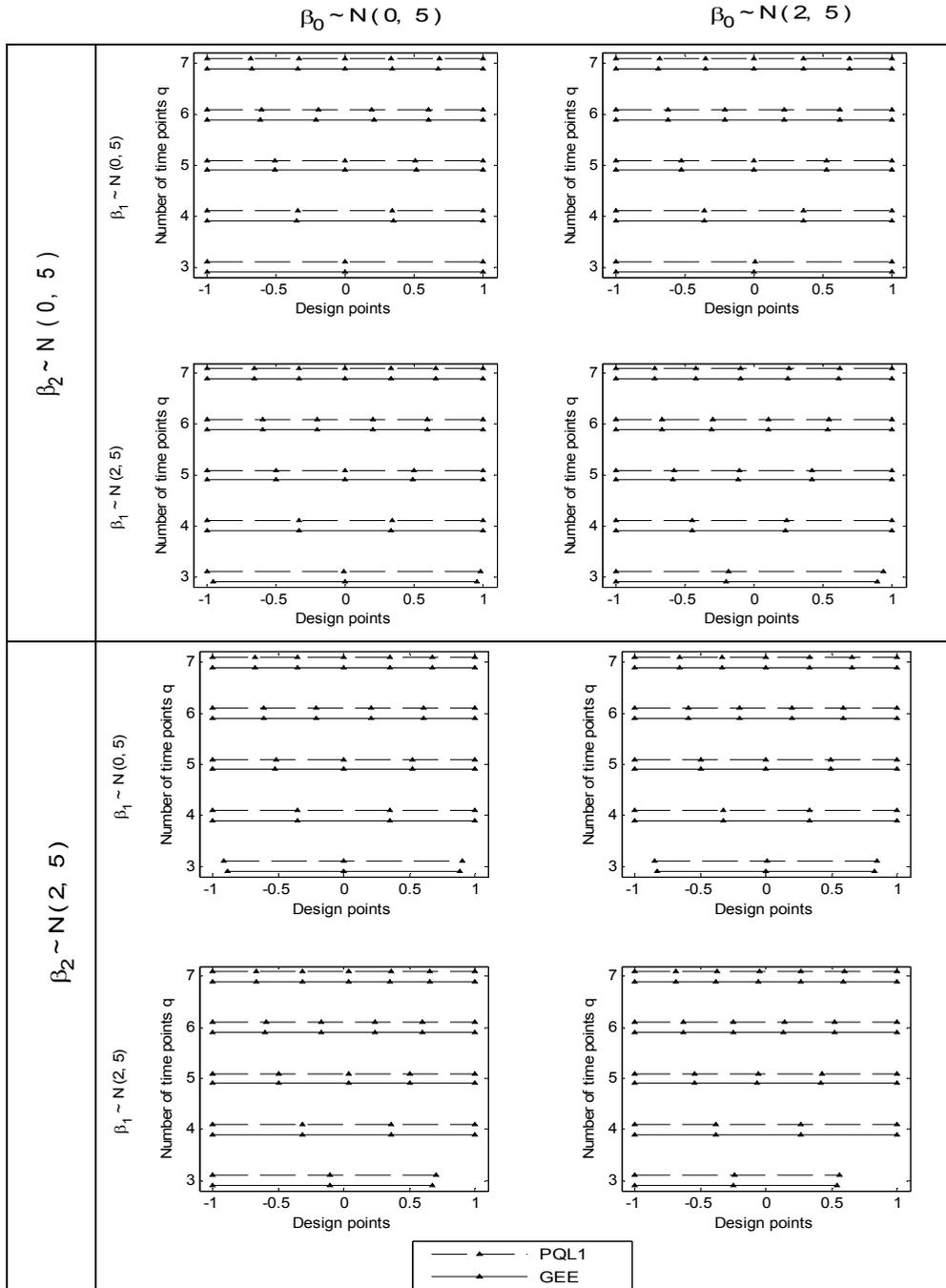


Figure 5.2 Bayesian D-optimal designs using PQL1 and extended GEE approach for the logistic mixed model with quadratic time effects with covariance structure D_2 and $\rho = 0.1$ for the priors with $\sigma^2 = 5$.

5.6.2 Selection of the optimal number of time points for a given cost ratio k

The results in the previous section show that PQL1 and extended GEE give almost the same Bayesian D-optimal allocation of design (time) points, given the same number of time points q for both methods. As will be demonstrated, both methods also give the same Bayesian D-optimal number of time points for a given subject-to-measurement cost ratio k , prior, covariance structure \mathbf{D} and autocorrelation ρ . For each method, we computed the relative efficiencies of designs with $3 \leq q \leq 7$ time points compared to a design with $q = 7$ time points for a given prior, covariance structure \mathbf{D} , cost ratio k and autocorrelation ρ . Then the q with the highest relative efficiency (RE) was selected, or simply $q = 7$ if all RE were < 1 (to save space the results are not included here). It turns out that the optimal number of time points was the same for both methods, for a given cost ratio k , prior distribution, covariance structure \mathbf{D} and autocorrelation ρ , except in a small percentage of the cases where PQL1 gave either one time point more or one time point less than the extended GEE and this always occurred for cost ratios of 1 and higher. However, for these cases the maximum losses in efficiencies of PQL1 relative to extended GEE and vice versa were less than 1% (detailed results are not included here to save space).

In conclusion, PQL1 and extended GEE give very similar allocations of Bayesian D-optimal design points, given the number of time points, and PQL1 and extended GEE also give (almost) the same optimal number of time points for a given cost ratio k . Therefore, in the remainder of this chapter, we will only present results based on the extended GEE approach which takes much less computing time.

5.6.3 Effect of autocorrelation on Bayesian D-optimal designs

To save space we will only present the effect of autocorrelation on the Bayesian D-optimal allocations of time points for some choices of q . Moreover, since the pattern of the plots was similar for all priors, we will only present results for the extreme prior means, i.e., for the prior means $\mu = [0, 0, 0]$ and $\mu = [2, 2, 2]$. Figure 5.3a and Figure 5.3b show the Bayesian D-optimal allocations of time points for different values of the autocorrelation, respectively, for the prior mean $\mu = [0, 0, 0]$ and $\mu = [2, 2, 2]$ with $\sigma^2 = 1$. In the plots of Figure 5.3a and 5.3b, the number of time points q varies across the columns and the covariance structure \mathbf{D} varies across the rows. Each plot shows the optimal design points (x-axis) as a function of the autocorrelation ρ (y-axis). Note that in each plot the optimal allocations of design points for $\rho = 0$ is given by diamond marker. It can be seen from these plots that for $\rho > 0$, the Bayesian D-optimal designs are fairly independent of ρ under all q , except for $q = 3$, where the optimal design points shift a bit for larger values of ρ (in Figure 5.3b). For $q = 5$ (in Figure 5.3b), the last design point moves away a bit from the end point of the time interval for small value of ρ . Moreover, when $\rho = 0$, the number of coinciding time points increases with q , and more

for the prior mean $\mu = [0, 0, 0]$ than for the prior mean $\mu = [2, 2, 2]$, i.e., the number of distinct time points is three for the prior mean $\mu = [0, 0, 0]$ and four for the prior mean $\mu = [2, 2, 2]$.

Figure 5.4a and Figure 5.4b show the Bayesian D-optimal allocations of time points for different values of the autocorrelation, respectively, for the prior mean $\mu = [0, 0, 0]$ and $\mu = [2, 2, 2]$ with prior variance $\sigma^2 = 5$. From the Figures, it can be seen that for $\rho > 0$, the Bayesian D-optimal designs are fairly independent of ρ under all q , except for $q = 3$, where the optimal design points shift a bit for large value of ρ . For $\rho = 0$ and $q = 7$, the number of distinct time points at which measurements are taken is five for the prior mean $\mu = [0, 0, 0]$, and six for the prior mean $\mu = [2, 2, 2]$.

Comparing Figures 5.3a-5.3b with Figures 5.4a-5.4b, the Bayesian D-optimal designs vary more between the priors with $\sigma^2 = 1$ (Figures 5.3a-5.3b) than for the priors with $\sigma^2 = 5$ (Figures 5.4a-5.4b). Moreover, when $\rho = 0$, the number of distinct time points is larger for the priors with $\sigma^2 = 5$ (Figures 5.4a-5.4b) than for the priors with $\sigma^2 = 1$ (Figures 5.3a-5.3b). Furthermore, some calculations for $q \geq 4$ confirm that the Bayesian optimal design is close to equidistance, and suggest that the relative efficiencies of equidistance versus the Bayesian optimal designs are very high (i.e., $RE > 0.96$) for all values of $\rho \geq 0.001$, whereas, for $q = 3$ the Bayesian optimal design differs more from equidistance as the prior mean increases, and so the relative efficiency of equidistance compared to the Bayesian optimal design becomes very small, i.e., $RE > 0.59$.

In general, the Bayesian D-optimal designs are fairly independent of ρ , and are approximately equally spaced, provided that the autocorrelation ρ is not too small, $\rho \geq 0.001$. The autocorrelation pushes all design points away from each other toward equidistance for all $q \geq 4$. Moreover, the allocations of time points are essentially the same for all three covariance structures \mathbf{D} of the random parameters. Furthermore, in the case of a zero autocorrelation, the Bayesian D-optimal design is affected by the choice of a prior mean.

From Figures 5.3a-5.3b and Figures 5.4a-5.4b, it can be seen that the Bayesian D-optimal allocations of time points are hardly affected by the height of the autocorrelation for $\rho > 0$, for a given number of time points q . To check this we first studied the effect of very small autocorrelation values ($0.001 \leq \rho \leq 0.1$) on the Bayesian D-optimal designs for each number of time points $q = 3, 4, \dots, 7$, by computing the relative efficiencies of the Bayesian D-optimal designs for $0.001 < \rho \leq 0.1$ compared to a Bayesian optimal design for $\rho = 0.001$.

Secondly, we computed the relative efficiencies of the Bayesian D-optimal designs for $0.1 < \rho \leq 0.9$ compared to a Bayesian optimal design for $\rho = 0.1$ for each number of time points $q = 3, 4, \dots, 7$. Table 5.1 shows the range of those relative efficiencies, respectively, in the first and second row for the prior means $\mu = [0, 0, 0]$ and $\mu = [2, 2, 2]$ with $\sigma^2 = 1$ and $\sigma^2 = 5$, for the covariance structure \mathbf{D}_2 . It can be seen that the relative efficiencies are all very high, above 0.99 for all q when $\rho \in [0.001, 0.1]$ and above 0.95 for all q when $\rho \in [0.1, 0.9]$. Table 5.1 also shows that the relative efficiency tends to increase with the number of time points q .

This is due to the fact that the Bayesian D-optimal allocation of time points becomes approximately equally spaced when the number of time points q increases, irrespective the autocorrelation as long as $\rho \geq 0.001$. To summarize, the Bayesian D-optimal allocations of design (time) points are hardly affected by the value of the autocorrelations at least within the range $\rho \in [0.001, 0.9]$.

Thirdly, we computed the relative efficiencies of the Bayesian D-optimal designs for $0 < \rho \leq 0.9$ compared to a Bayesian optimal design for $\rho = 0$ for each time points $q = 3, 4, \dots, 7$. Table 5.2 shows the range of those relative efficiencies for the prior means $\mu = [0, 0, 0]$ and $\mu = [2, 2, 2]$ with $\sigma^2 = 1$ and $\sigma^2 = 5$, for the covariance structure \mathbf{D}_2 . As Table 5.2 shows the relative efficiencies are larger for the priors with variance $\sigma^2 = 5$, than for the priors with $\sigma^2 = 1$ for all q , except for $q = 3$. The maximum loss in efficiency is about 17% for the priors with $\sigma^2 = 1$, whereas, for the priors with $\sigma^2 = 5$, it is about 7%. In other words the prior with large variance is more robust against misspecification of the autocorrelation. Table 5.2 also shows that the relative efficiency decreases as the number of time points q increases. The smallest relative efficiency is obtained by $q = 7$ in all priors. This is because, when $\rho = 0$, the number of coinciding time points increases as q increases. For $q \leq 5$, if we erroneously assume an autocorrelation $\rho > 0$ instead of $\rho = 0$, the loss in efficiency is at most 10%, which can be compensated by sampling about 11% more subjects. For larger number of time points ($q > 5$), if we assume $\rho > 0$ instead of $\rho = 0$, the loss in efficiency is at most 17%, which can be compensated by sampling about 21% more subjects.

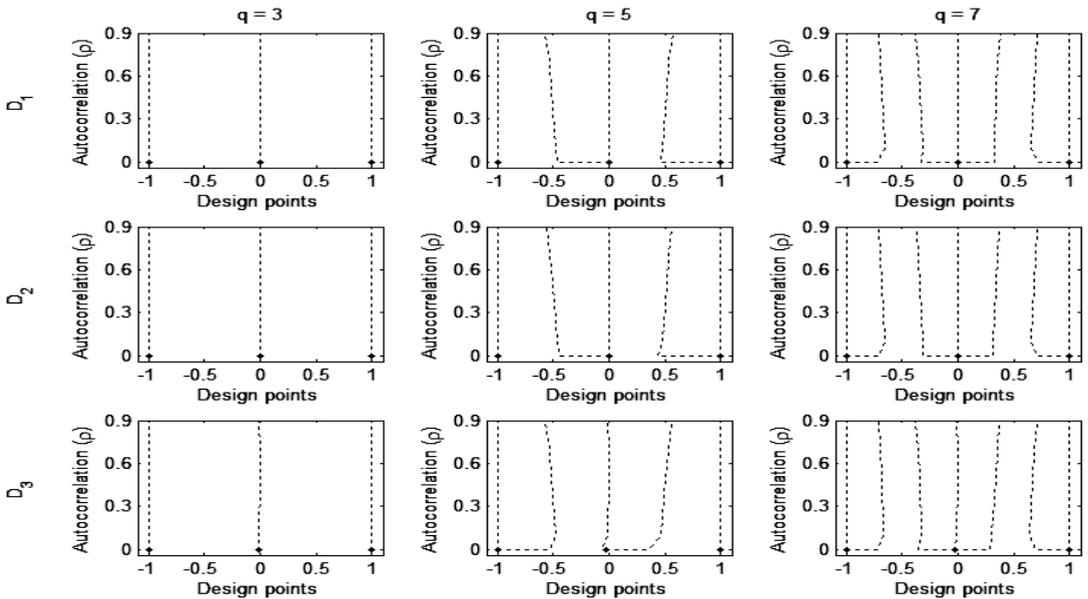


Figure 5.3a Bayesian D-optimal allocations of time points for the logistic mixed effect model with quadratic time effects for different values of autocorrelation for the prior mean $[0, 0, 0]$ with $\sigma^2 = 1$. ♦ Optimal allocations of design points for $\rho = 0$.

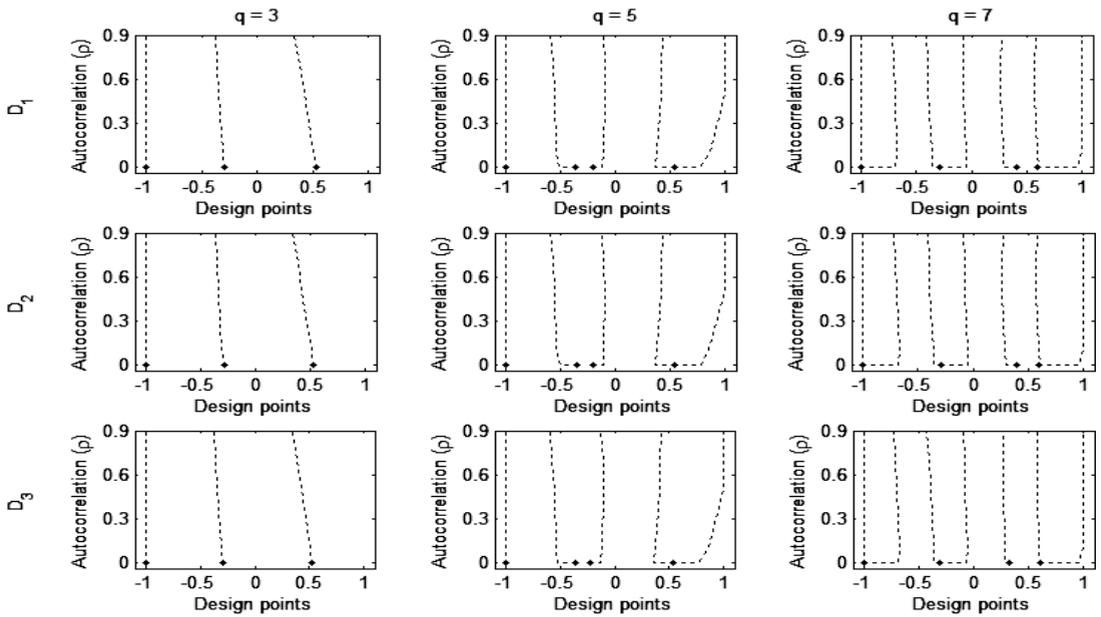


Figure 5.3b Bayesian D-optimal allocations of time points for the logistic mixed effect model with quadratic time effects for different values of autocorrelation for the prior mean $[2, 2, 2]$ with $\sigma^2 = 1$.

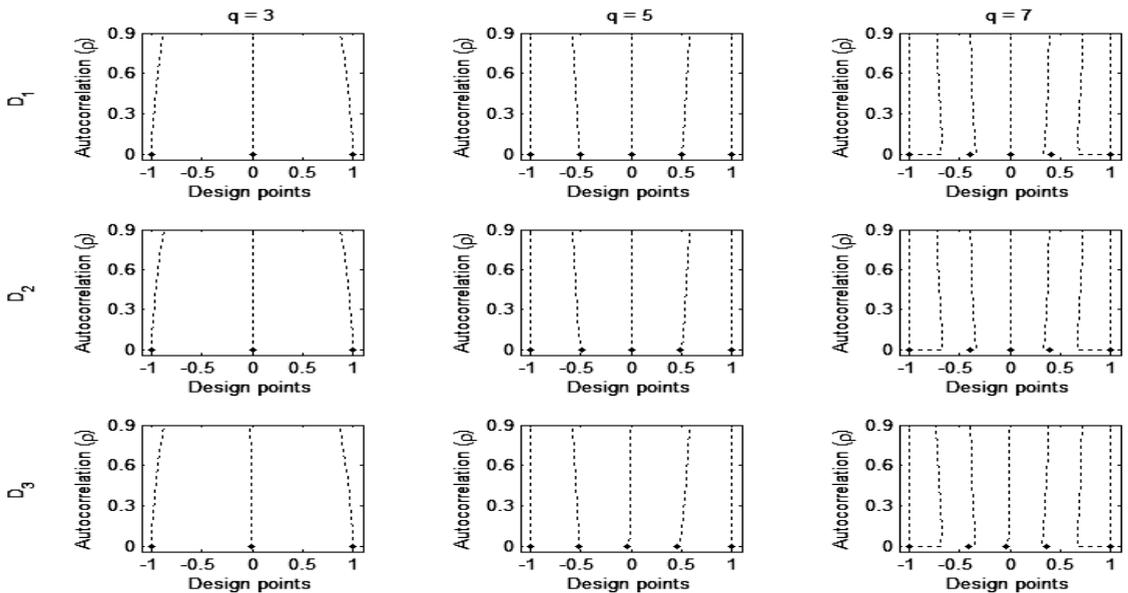


Figure 5.4a Bayesian D-optimal allocations of time points for the logistic mixed effect model with quadratic time effects for different values of autocorrelation for the prior mean $[0, 0, 0]$ with $\sigma^2 = 5$.

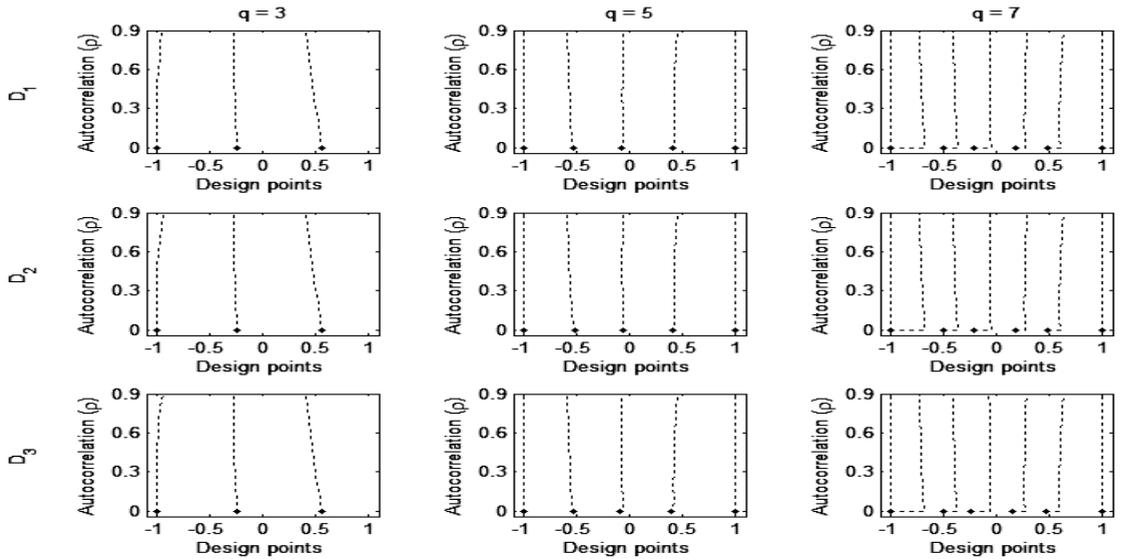


Figure 5.4b Bayesian D-optimal allocations of time points for the logistic mixed effect model with quadratic time effects for different values of autocorrelation for the prior mean $[2, 2, 2]$ with $\sigma^2 = 5$.

Table 5.1 First row: range of relative efficiencies of the Bayesian D-optimal designs for $0.001 < \rho \leq 0.1$ compared to a Bayesian optimal design for $\rho = 0.001$ for D_2 .

Second row: range of relative efficiencies of the Bayesian D-optimal designs for $0.1 < \rho \leq 0.9$ compared to a Bayesian optimal design for $\rho = 0.1$ for D_2 .

		Range of relative efficiencies			
		$\sigma^2 = 1$		$\sigma^2 = 5$	
Number of time points	Values of ρ	$\beta \sim N(0, 0, 0)$	$\beta \sim N(2, 2, 2)$	$\beta \sim N(0, 0, 0)$	$\beta \sim N(2, 2, 2)$
		$q = 3$	$0.001^* < \rho \leq 0.1$	(0.9999, 1.0000)	(0.9992, 1.0000)
	$0.1^{**} < \rho \leq 0.9$	(0.9999, 1.0000)	(0.9653, 0.9993)	(0.9540, 1.0000)	(0.9561, 0.9995)
$q = 4$	$0.001 < \rho \leq 0.1$	(0.9946, 0.9986)	(0.9978, 0.9994)	(0.9990, 0.9999)	(0.9987, 0.9995)
	$0.1 < \rho \leq 0.9$	(0.9908, 0.9998)	(0.9772, 0.9997)	(0.9929, 0.9998)	(0.9986, 0.9999)
$q = 5$	$0.001 < \rho \leq 0.1$	(0.9992, 0.9997)	(0.9949, 0.9984)	(0.9995, 1.0000)	(0.9995, 0.9999)
	$0.1 < \rho \leq 0.9$	(0.9942, 0.9999)	(0.9936, 0.9999)	(0.9939, 0.9998)	(0.9964, 0.9999)
$q = 6$	$0.001 < \rho \leq 0.1$	(0.9987, 0.9994)	(0.9944, 0.9981)	(0.9998, 1.0000)	(0.9998, 1.0000)
	$0.1 < \rho \leq 0.9$	(0.9975, 0.9999)	(0.9982, 0.9999)	(0.9964, 0.9999)	(0.9975, 0.9999)
$q = 7$	$0.001 < \rho \leq 0.1$	(0.9971, 0.9990)	(0.9988, 0.9995)	(0.9997, 0.9999)	(0.9998, 1.0000)
	$0.1 < \rho \leq 0.9$	(0.9983, 0.9999)	(0.9990, 0.9999)	(0.9975, 0.9999)	(0.9982, 1.0000)

* Used as reference value when $\rho \leq 0.1$.

** Used as reference value when $0.1 < \rho \leq 0.9$.

Table 5.2 Range of relative efficiencies of the Bayesian D-optimal designs for $\rho > 0$ relative to a Bayesian optimal design for $\rho = 0$ for the extreme prior means and covariance structure \mathbf{D}_2 .

Number of time points	Values of ρ	Range of relative efficiencies			
		$\sigma^2 = 1$		$\sigma^2 = 5$	
		$\beta \sim N(0, 0, 0)$	$\beta \sim N(2, 2, 2)$	$\beta \sim N(0, 0, 0)$	$\beta \sim N(2, 2, 2)$
$q = 3$	$0^* < \rho \leq 0.1$	(0.9999, 1.0000)	(0.9991, 1.0000)	(0.9999, 1.0000)	(0.9995, 1.0000)
	$0.1 < \rho \leq 0.9$	(0.9999, 1.0000)	(0.9526, 0.9967)	(0.9472, 1.0000)	(0.9432, 0.9978)
$q = 4$	$0 < \rho \leq 0.1$	(0.9787, 0.9892)	(0.9947, 0.9988)	(0.9987, 1.0000)	(0.9976, 0.9999)
	$0.1 < \rho \leq 0.9$	(0.9555, 0.9761)	(0.9462, 0.9944)	(0.9863, 0.9975)	(0.9954, 0.9976)
$q = 5$	$0 < \rho \leq 0.1$	(0.9175, 0.9201)	(0.9651, 0.9785)	(0.9994, 0.9999)	(0.9990, 0.9997)
	$0.1 < \rho \leq 0.9$	(0.9089, 0.9167)	(0.9364, 0.9648)	(0.9926, 0.9988)	(0.9929, 0.9985)
$q = 6$	$0 < \rho \leq 0.1$	(0.8437, 0.8481)	(0.9630, 0.9796)	(0.9769, 0.9772)	(0.9986, 0.9988)
	$0.1 < \rho \leq 0.9$	(0.8437, 0.8457)	(0.9598, 0.9623)	(0.9760, 0.9769)	(0.9944, 0.9986)
$q = 7$	$0 < \rho \leq 0.1$	(0.8276, 0.8402)	(0.8969, 0.9020)	(0.9377, 0.9477)	(0.9698, 0.9704)
	$0.1 < \rho \leq 0.9$	(0.8274, 0.8323)	(0.8967, 0.8977)	(0.9340, 0.9388)	(0.9680, 0.9704)

* Reference value $\rho=0$.

5.7 Summary and Conclusion

The main purpose of this chapter was to investigate the robustness of Bayesian D-optimal designs for the logistic mixed effects model with quadratic time effect against incorrect specification of the autocorrelation. Due to the random effects in the logistic mixed effects models, the log-likelihood cannot be written down in closed form. To overcome this problem, we used two popular approximation methods which are implemented in commercially available software packages: The first order penalized quasi likelihood (PQL1) and an extended version of generalized estimating equations (GEE). We first computed numerically Bayesian D-optimal designs for $q = 3, 4, \dots, 7$ time points, for different values of the autocorrelation, i.e., $\rho \in [0, 0.9]$, and three different covariance structures under various different prior distributions, by optimizing a scalar function of the approximate variance-covariance matrix based on PQL1 and on extended GEE. Then, we compared PQL1 with extended GEE in terms of the resulting Bayesian D-optimal designs for a given number of time points q , prior and covariance structure \mathbf{D} for different autocorrelations ρ . We also compared the two methods in terms of the optimal number of time points for a given subject-to-measurement cost ratio. Finally, the effect of different values of the autocorrelation ρ on

the Bayesian D-optimal designs and the robustness of the latter against incorrect specification of the autocorrelation were studied.

With respect to the comparison of the two methods, the results show that Bayesian D-optimal designs based on PQL1 and those based on extended GEE do not differ substantially in terms of allocations of design (time) points or in terms of efficiency, and that PQL1 and extended GEE also give (almost) the same optimal number of time points for a given subject-to-measurement cost ratio. Further findings are that the Bayesian D-optimal designs are hardly affected by the choice of a covariance structure for the random effects, and also hardly affected by the choice of prior means, provided that the autocorrelation is not too close to zero ($\rho \geq 0.001$). This is different from the case of a zero autocorrelation, where the Bayesian optimal design is affected by the choice of a prior (see Figure 5.3 and 5.4).

The Bayesian D-optimal allocations of a given number time points q is hardly affected by the height of the autocorrelation at least within the range $[0.001, 0.9]$ and the maximum loss in efficiency was found to be below 5%, when the value of the autocorrelation is incorrectly specified (see Table 5.1). When we use an autocorrelation $\rho > 0$ when the true $\rho = 0$, the loss in efficiency is at most 10%, for a number of time points $q \leq 5$, whereas for $q > 5$, the maximum loss in efficiency is about 17%. These losses can be compensated by sampling about 11% and 21% more subjects, respectively. Moreover, when the number of time points increases and the autocorrelation is at least ≥ 0.001 , the Bayesian optimal allocations of time points become approximately equally spaced, i.e., the autocorrelation pushes all the design points away from each other, toward equidistance. Furthermore, some calculations for $q \geq 4$ confirm that the Bayesian D-optimal designs are close to equidistance, and suggest that the relative efficiencies of equidistance versus the Bayesian optimal designs are very high (i.e., $RE > 0.96$) for all values of $\rho \geq 0.001$. In contrast, for $q = 3$ the Bayesian optimal design differs more from equidistance as the value of the autocorrelation parameter increases, and as a result the relative efficiency of equidistance compared to the Bayesian optimal design becomes very small, i.e., $RE > 0.59$.

In conclusion, the Bayesian D-optimal allocations of time points are robust against incorrect specification of the autocorrelation, and are approximately equally spaced, provided that the autocorrelation ρ is non-zero ($\rho \geq 0.001$). Moreover, the Bayesian D-optimal designs are hardly affected either by the choice of a covariance structure, or by the choice of a prior distribution, provided that the autocorrelation is ≥ 0.001 . Furthermore, PQL1 and extended GEE give essentially the same Bayesian D-optimal allocation of time points for a given number of time points q , and also the same optimal number of time points for a given subject-to-measurement cost ratio. As a final conclusion, for the purpose of optimal designs, if there is limited computational time, our recommendation is that researchers should use extended GEE instead of PQL1, as extended GEE is preferable in terms of computational time.

CHAPTER 6

A computer program for Bayesian D-optimal binary repeated measurements designs

Submitted for publication.

Abstract

In the planning stage of longitudinal studies, the choice of the number and timing of repeated measurements is a crucial issue because the precision of the model parameter estimates of interest, such as the rate of change, depends on these choices. Optimal designs for longitudinal studies with binary responses have been shown useful to improve the precision of the model parameter estimates. However, due to absence of a computer program, planners of longitudinal studies of binary responses in psychology, health sciences and medicine have not yet benefitted from optimal designs. Usually the number and the allocation of the time points are determined by non-statistical criteria, and the time points are frequently chosen to be equally spaced. To help researchers to identify the optimal number and allocations of time points for a given subject-to-measurement cost ratio, this chapter presents a new interactive computer program for Bayesian optimal binary repeated measurements designs, which also computes the loss in efficiency of equidistant time points compared to the optimal allocation. Moreover, it helps to assess the effect of autocorrelation on optimal allocations of design points. The computer program is described and illustrated using a practical example.

Keywords: Bayesian D-optimal designs; logistic mixed effects models; subject-to-measurement cost ratio; relative efficiency; number of time points, autocorrelation.

6.1 Introduction

Designs of longitudinal studies for dichotomous data have been used in different disciplines of science to study the change of a particular outcome variable over time. In a clinical trial, for example, patients are studied over a period of time and classified in terms of presence or absence of a clinical symptom. The generalized linear mixed model (GLMM) is the most frequently used model for the analysis of longitudinal dichotomous data. Optimal designs for longitudinal studies have been shown useful to improve the precision of the model parameter estimates of interest, such as the rate of change. For the generalized linear mixed model optimal designs were studied extensively in literature by Han and Chaloner (2004); Ouwens et al. (2006); Tekle et al. (2008); Niaparast (2009); Sinha and Xu (2011); Niaparast and Schwabe (2013) and Abebe et al. (2014b), among others.

Optimal designs depend on the unknown parameter values in the case of nonlinear models. A common approach is to design an experiment for a best guess of the parameter values, which leads to locally optimal designs (see, e.g., Chernoff, 1953). However, such designs may not be efficient for other parameter values. To overcome this problem, various methods have been proposed in the literature (see, e.g., Atkinson et al., 2007). The Bayesian approach is one way that has been shown to be useful when the uncertainty of the parameter values of the statistical model is considered (Chaloner and Larntz, 1989; Atkinson et al., 2007; Abebe et al., 2014a; among others). The Bayesian design literature is vastly restricted to binary response models. However, no user friendly software has been developed so far for longitudinal studies with binary responses.

Due to absence of a computer program, planners of longitudinal studies in psychology, health sciences and medicine, face the problem of choosing the best number and timing of the repeated measurements. Usually the number and the allocation of the time points at which the measurements are taken are determined by non-statistical criteria, and the time points are frequently chosen to be equally spaced. However, these designs may be statistically inefficient when compared with their optimal counterparts.

In this chapter, we present a new interactive computer program that computes Bayesian optimal repeated measurements designs for mixed effects logistic models with polynomial time effects. To the authors' knowledge, this is the first program that computes optimal designs for such studies under cost constraints. The computer program that we will describe helps users to identify the optimal number and optimal allocations of time points for a given subject-to-measurement cost ratio. Moreover, it computes the loss in efficiencies of equidistance compared to the optimal allocation. Finally, it produces a plot of optimal allocations of time points under different values of autocorrelation.

This chapter is organized as follows. In section 6.2, we briefly describe the logistic mixed-effects models with polynomial time effects and the variance-covariance matrix of parameter estimators. The optimality criterion and the relative efficiency as a measure for the comparison of designs are given in section 6.3. In section 6.4, the graphical user interface will

be introduced. Thereafter, a longitudinal study will be used to illustrate the application of the program in section 6.5. Finally, conclusions and recommendations are provided in section 6.6.

6.2 Model specification and estimation methods

6.2.1 The logistic mixed effects model

Let the $q \times 1$ vector $\mathbf{y}_i = (y_{i1}, \dots, y_{iq})'$ be binary responses y_{ij} of subject i at q time points, $i = 1, 2, \dots, N$ and $j = 1, \dots, q$. It is assumed that all subjects have measurements at the same time points, and that, conditional on the subject-specific random effect vector \mathbf{b}_i , the binary responses y_{ij} of \mathbf{y}_i are assumed to be Bernoulli distributed with probability of success $p(y_{ij} = 1 | \mathbf{b}_i)$. These probabilities are related to the fixed and random effects via the logit link function. The corresponding logistic mixed effects model is given by:

$$\text{logit}(p(y_{ij} = 1 | \mathbf{b}_i)) = \log\left(\frac{p(y_{ij} = 1 | \mathbf{b}_i)}{1 - p(y_{ij} = 1 | \mathbf{b}_i)}\right) = \mathbf{x}_j' \boldsymbol{\beta} + \mathbf{z}_j' \mathbf{b}_i, \quad (6.1)$$

where the $p \times 1$ vector \mathbf{x}_j is the design vector of the explanatory variables at the j th measurement for subject i , $\boldsymbol{\beta}$ is the corresponding $p \times 1$ vector of fixed polynomial time effects, and \mathbf{z}_j is the $r \times 1$ design vector for the random effects that is usually a subset of vector \mathbf{x}_j . The vector \mathbf{b}_i is the corresponding $r \times 1$ vector of random effects, which is assumed to have a multivariate normal distribution with mean zero and covariance matrix \mathbf{D} .

6.2.2 Estimation methods

Due to the random effects in model (6.1), the log-likelihood cannot be written down in closed form. Hence, either numerical methods or approximations to the log-likelihood must be used. Numerical methods require large computational resources and more importantly they require full knowledge of the data (Moerbeek et al. 2003; Han and Chaloner 2004), making them computationally inconvenient for optimal design procedures. To overcome this problem, approximation methods are employed. Optimal designs obtained by approximation methods were discussed by Moerbeek et al. (2001), Moerbeek and Maas (2005), Ouwens et al. (2006), Woods and Van de Ven (2006), Tekle et al. (2008), Niaparast (2009) and Niaparast and Schwabe (2013). There is a large statistical literature on various approximation methods, but here, for the purpose of obtaining optimal designs, we will focus on the two most frequently used ones, which are implemented in commercially available software packages: Penalized quasi-likelihood (PQL) (Breslow and Clayton 1993; Jang and Lim 2009) and an extended version of generalized estimating equations (GEE) (Molenberghs and Verbeke 2005). Details of these methods will be given in the next sub-sections.

6.2.2.1 First order penalized quasi-likelihood (PQL1)

We use an approximate information matrix which is the inverse of the asymptotic variance-covariance matrix of the parameters based on first order penalized quasi-likelihood (PQL1) estimation. See Breslow and Clayton (1993) and Jang and Lim (2009) for details. PQL1 performs well in terms of point estimation since it produces the smallest mean squared error (MSE) and the bias of the estimators decreases as the sample size increases (Breslow and Clayton 1993; Moerbeek et al. 2003; Jang and Lim 2009). The variance-covariance matrix of the estimator $\hat{\beta}$ for the logistic mixed effects models (6.1) is approximated in PQL1 by:

$$\text{var}(\hat{\beta}) \approx (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}, \quad (6.2)$$

where \mathbf{X} is the $Nq \times p$ design matrix formed by stacking $\{\mathbf{x}_j'\}$ for N subjects and q time points, and the $Nq \times Nq$ block-diagonal matrix \mathbf{V} has N blocks of $q \times q$ variance-covariance matrices given by:

$$\mathbf{v}_i \approx \mathbf{w}_i^{-1/2} \mathbf{R} \mathbf{w}_i^{-1/2} + \mathbf{Z} \mathbf{D} \mathbf{Z}', \quad (6.3)$$

where \mathbf{R} is the $q \times q$ correlation matrix that captures autocorrelation of the responses over time, \mathbf{Z} is the $q \times r$ design matrix with rows \mathbf{z}_j' , $j = 1, \dots, q$, the $r \times r$ matrix \mathbf{D} is the variance-covariance matrix of the random effects, and \mathbf{w}_i^{-1} is the diagonal matrix of the conditional variances of the transformed responses given the random effects \mathbf{b}_i , which is equal to the inverse of the diagonal matrix of the conditional variances of the untransformed responses given the random effects \mathbf{b}_i (see for detail Moerbeek et al. 2001; Molenberghs and Verbeke 2005 page 270). The diagonal matrix \mathbf{w}_i is given by:

$$\mathbf{w}_i = \text{diag}(w_{i1}^{b_i}, \dots, w_{iq}^{b_i}), \quad (6.4)$$

where $w_{ij}^{b_i} = p(y_{ij} = 1 | \mathbf{b}_i) (1 - p(y_{ij} = 1 | \mathbf{b}_i))$, for $i = 1, \dots, N$ and $j = 1, \dots, q$. Since the random effects are unknown in the design stage, we will generate \mathbf{b}_i from a multivariate normal distribution with mean zero and variance-covariance \mathbf{D} .

6.2.2.2 Extension of generalized estimating equations (GEE)

Another approximation method is the extended GEE. The variance-covariance matrix of the estimator $\hat{\beta}$ for model (6.1), based on the extension of the GEE approach (Zeger et al., 1988; Molenberghs and Verbeke, 2005), is approximated by:

$$\text{var}(\hat{\beta}) \approx \left(\sum_{i=1}^N \frac{\partial \mathbf{P}_i'}{\partial \beta} \mathbf{u}_i^{-1} \frac{\partial \mathbf{P}_i}{\partial \beta'} \right)^{-1}, \quad (6.5)$$

where $\mathbf{P}_i = (p(y_{i1}|\mathbf{b}_i), \dots, p(y_{iq}|\mathbf{b}_i))'$, $\frac{\partial \mathbf{P}_i'}{\partial \boldsymbol{\beta}} = \mathbf{X}_i' \mathbf{W}_i$, the $p \times q$ matrix \mathbf{X}_i with rows $\{\mathbf{x}_j'\}$ for $j = 1, \dots, q$ and the working variance-covariance matrix \mathbf{u}_i of the responses is given by:

$$\mathbf{u}_i \approx \mathbf{w}_i^{1/2} \mathbf{R} \mathbf{w}_i^{1/2} + \mathbf{w}_i \mathbf{Z} \mathbf{D} \mathbf{Z}' \mathbf{w}_i', \quad (6.6)$$

where \mathbf{w}_i is the diagonal matrix of the conditional variances of responses given the random effects $\mathbf{b}_i = 0$ and is given by:

$$\mathbf{w}_i = \text{diag}(w_{i1}^{b_i=0}, \dots, w_{iq}^{b_i=0}), \quad (6.7)$$

where $w_{ij}^{b_i=0} = p(y_{ij} = 1 | \mathbf{b}_i = 0) (1 - p(y_{ij} = 1 | \mathbf{b}_i = 0))$, for $i = 1, \dots, N$ and $j = 1, \dots, q$ (Molenberghs and Verbeke, 2005 page 443).

In the present chapter, a first order auto regressive (AR1) correlation structure is considered for the correlation matrix \mathbf{R} , i.e., the (j, l) element of the correlation matrix \mathbf{R} is equal to $\rho^{|t_j - t_l|}$, where $l = 1, \dots, q$, and so ρ is the autocorrelation coefficient between two responses at a time distance of one, that is, $\rho = \text{Corr}(y_{ij}, y_{il})$ for which $|t_j - t_l| = 1$. This autocorrelation structure implies that repeated measurements closer in time are more highly correlated and that the correlation decreases as the distance between the time points increases.

6.3 Optimal design and relative efficiency

6.3.1 Bayesian D-optimality criterion

To introduce the notation for the optimality criterion, suppose that the study to be designed will have q time points at which measurements are taken for all N subjects. The design space \mathcal{E} then contains all designs with $q \times 1$ time vectors, i.e.,

$$\mathcal{E} = \left\{ (t_1, t_2, \dots, t_q)' : t_j \in [a, b] \right\}, \quad (6.8)$$

where t_1, t_2, \dots, t_q are the q time points, and $q \geq p$, to make these fixed effects identifiable with p being the number of fixed parameters of the model. The time interval $[a, b]$ is assumed to be fixed by substantive constraints within the field of application, for instance the total follow-up time in a cohort study of smoking or unemployment. A design ξ_q is an element of the design space \mathcal{E} if it has q time points within the time interval $[a, b]$. The goal of optimal design is to maximize parameter information by minimizing variability of the parameter estimators. Optimal designs are usually selected by minimizing a real-valued function of the variance-covariance matrix, which is known as optimality criterion (see, e.g., Silvey 1980). There are numerous optimality criteria proposed in the literature, such as the D-, A- or G-

optimality criteria. In this chapter, we will focus on the best-known and most popular optimality criterion, i.e., the D-optimality criterion. This optimality criterion has some nice properties: 1. It minimizes the volume of the asymptotic confidence ellipsoid of the parameters. 2. It does not depend on the coding used for the endpoints of the chosen time interval $[a, b]$. This means that if the coding for the time interval is transformed linearly, a D-optimal design for the new time interval is obtained by applying the same linear transformation to the D-optimal design for the old interval (See Ouwens et al. 2006). The D-optimal design ξ_q^* is the design among all possible designs ξ_q with q time points for which the determinant of the variance-covariance matrix is minimized or equivalently, for which the determinant of the information matrix is maximized (Atkinson et al. 2007 page 151).

The approximate variance-covariance matrix depends on the unknown parameter vector β . The Bayesian approach resolves this dependency by taking the expectation of a function of the approximate variance-covariance matrix over a prior distribution for parameter vector β . Thus, the Bayesian D-optimality criterion (by analogy with the non-Bayesian optimal design) is defined as follows:

$$\phi_D(\xi|\pi) = E_{\beta} \log \left| \text{var}(\hat{\beta}) \right|^{-1} = \int_{\beta} \log \left| \text{var}(\hat{\beta}) \right|^{-1} \pi(\beta) d(\beta), \quad (6.9)$$

where $\pi(\beta)$ is the prior distribution for β and $\text{var}(\hat{\beta})$ is the variance-covariance matrix of $\hat{\beta}$. The design which maximizes (6.9) is the Bayesian D-optimal design.

6.3.2 Cost function and relative efficiency

The precision of estimating the fixed effects parameters increases by taking more measurements and sampling more subjects. However, the addition of subjects and of measurements per subject will increase the costs of the study and these are usually limited by budget constraints. Therefore, it is reasonable to take into account the costs of a longitudinal study when designs are compared with each other. There are two main components of these costs. These are the costs for the recruitment of subjects and the costs of the measurements once a subject has been recruited. Let the cost of recruiting a subject be C_1 and the cost of one measurement per subject be C_2 . Then the total cost of a longitudinal study with q time points and N subjects, excluding overhead cost, is given by the linear cost function:

$$C = C_1 N + C_2 N q = N C_2 (k + q), \quad (6.10)$$

where $k = C_1/C_2$ is the ratio of the cost of adding a new subject to the cost of an additional measurement per subject.

To compare different designs, we will use their relative efficiencies while fixing the total costs C . This means that the designs can differ in terms of the number of subjects N and the number and timing of the measurements q . First, we compute the Bayesian D-optimal designs

using N fixed and then we correct for costs and different q and N as follows. Let $\phi_D(\xi_q^*|\pi)$ denote the value of design criterion (6.9) for the optimal design ξ_q^* with q time points. Then the relative efficiency (RE) of an arbitrary design ξ_s with s time points relative to the optimal design ξ_q^* is defined as:

$$\text{RE}(\xi_s; \xi_q^*|\pi) = \frac{k+q}{k+s} \left[\exp \left\{ \frac{\phi_D(\xi_s|\pi) - \phi_D(\xi_q^*|\pi)}{p} \right\} \right], \quad (6.11)$$

where $(q,s) \geq p$ and π is the prior distribution for the fixed effects. If the value of this relative efficiency is close to unity, then the design ξ_s is about equally efficient as the optimal design ξ_q^* for a given prior π . The inverse of this relative efficiency is the number of times that a design ξ_s must be replicated to have the same efficiency as the optimal design ξ_q^* . Note that this relative efficiency (6.11) is an extension of the relative efficiency given by Chaloner and Larntz (1989). When the ratio $\frac{k+q}{k+s}$ is one, it reduces to their relative efficiency. This ratio follows from (6.10) and takes into account that the design with the smaller number of time points can have more subjects for the same budget.

6.3.3 Method of Optimization

The Bayesian D-optimal designs for the logistic mixed effect model are found by our computer program numerically by maximization of the criterion value (6.9) among all candidate designs for a given prior distribution of the parameters. Details of this will be given in the next sub-sections.

6.3.3.1 Sampling parameter values from priors to compute the criterion

To construct Bayesian designs for continuous prior distributions, all candidate designs must be evaluated in terms of their criterion values as defined by (6.9). However, evaluation of the integration over the prior distribution is very complicated and cannot easily be done analytically. A numerical approximation of the integral is necessary. Numerical approximations can be done by sampling parameter values from the prior distribution and then replacing the integral in (6.9) with a summation over the sample (Atkinson et al. 2007; Chaloner and Verdinelli 1995). Estimating (6.9) using the traditional sampling (pseudo Monte Carlo) method requires very large samples from the prior to reduce the sample-to-sample variability to the point where different samples do not lead to different design choices. Thus, this approach is costly in terms of computing time. In our computer program, we will use an Adaptive Rejection Metropolis Sampling (ARMS) algorithm (Gilks and Wild 1992; Gilks et al. 1995), which is a more efficient sampling algorithm that requires a smaller sample to obtain a good approximation of the design criterion (6.9).

6.3.3.2 Optimization algorithm for finding an optimal design

To find candidate designs and in particular the optimal design, the program uses the FMINCON function of MATLAB version 7.10.0499 (R2010a). This function performs constrained non-linear optimization and requires an initial design ξ_0 . Without loss of generality, we coded the time interval as $[-1, 1]$, and we used equally spaced time points as initial designs. The following global search algorithm was used to find the Bayesian D-optimal designs for a given prior distribution of the parameters:

1. Take samples from the prior distribution of the parameters using ARMS.
2. Compute the Bayesian D-optimal allocation of q time points, using $q = p$ equidistant time points as initial design, where p is the number of fixed parameters of the model.
3. Increase the number of time points q by one and perform step 2 again to find the Bayesian optimal design (allocation) for the new value of q . Repeat step 2 and 3 until the maximum number of time point q (user specified) is reached.
4. Thereafter, select the optimal number of time points q for the Bayesian D-optimal design by computing the relative efficiencies of designs with different number of time points against each other for a given subject-to-measurement cost ratio. Do this for each cost ratio considered to obtain one optimal design per cost ratio for a chosen prior distribution.

6.4 BODMixed_Logistic program

In this section, we will introduce the graphical user interface of the computer program that computes optimal designs for longitudinal studies with a binary response. Bayesian Optimal Design for Mixed effects Logistic models with polynomial time effect (BODMixed_Logistic) is an interactive computer program that helps researchers to plan their longitudinal studies in an optimal way. The program is designed to enable interactive input specifications. The program runs in a MATLAB environment.

6.4.1 Description and use of the program

To start the program type BODMixed_Logistic in the command window of MATLAB and press the \downarrow Enter key. Then, the menu shown in Figure 6.1 will appear. Note that the current directory on the MATLAB desktop must contain the BODMixed_Logistic program. The user should give the necessary input parameters. The program has five panels. Details of these panels are given below:

The first panel allows the user:

- To *choose model type*: The user can choose the degree of the polynomial of the model, i.e., a linear, quadratic or cubic model describing the trend over time. This choice can be based on knowledge from previous studies, a pilot study or expert opinion. The default polynomial of the mixed logistic model is a linear time effect but other time effects can be selected.

• To enter values of the *variance–covariance parameters* (\mathbf{D}): By clicking on this button, the user will find a sub-menu to enter the input values for \mathbf{D} . The fixed and random effects can be chosen up to cubic. For example, if a linear growth model is chosen, the user will find the sub-menu, which is shown in Figure 6.2, for the input of the values of the variance components in \mathbf{D} matrix. A fixed effects logistic model is obtained by setting all values in \mathbf{D} to zero.

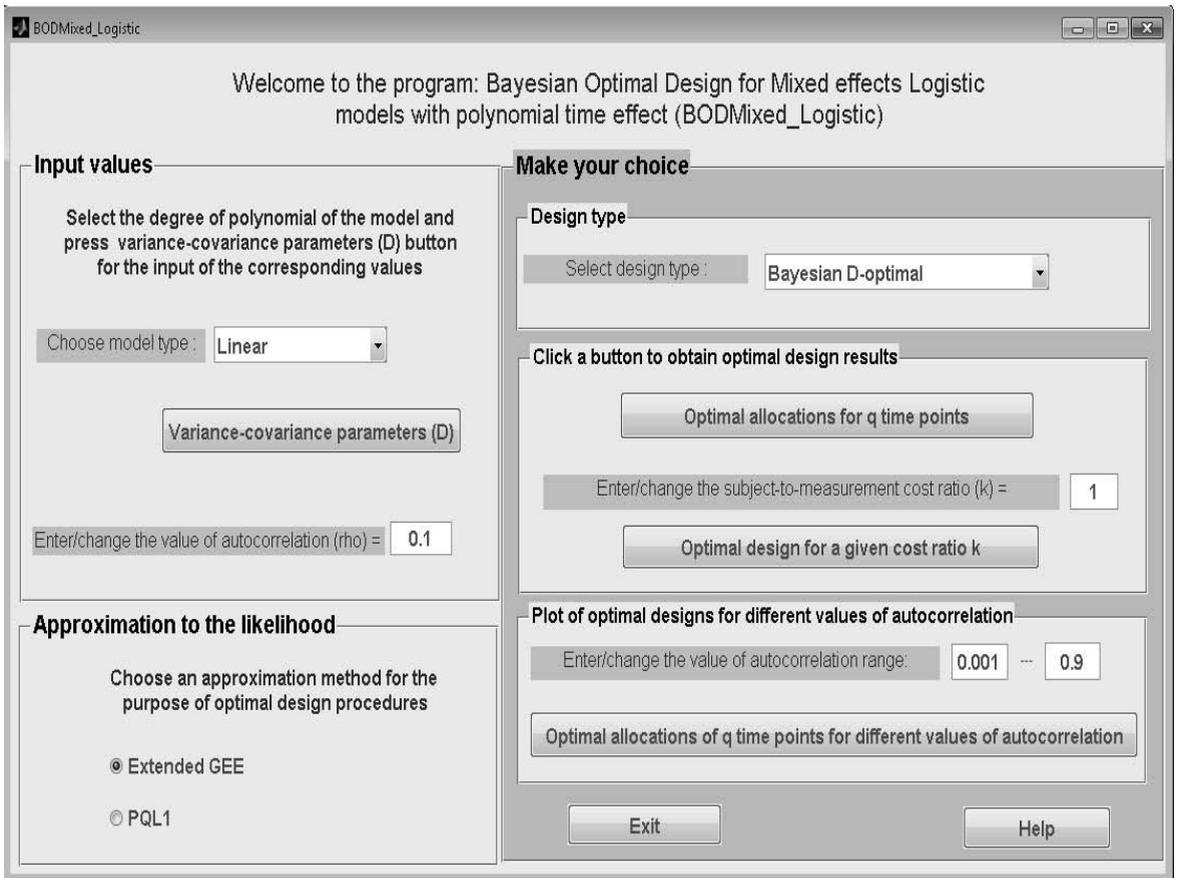


Figure 6.1 Layout of the main menu with the default input values for BODMixed_Logistic program.

• To *enter/change the value of autocorrelation* (ρ): This is the size of the autocorrelation coefficient that the user expects between two repeated measurements at a time distance of one, i.e., $\rho = \text{Corr}(y_{ij}, y_{il})$ for which $|t_j - t_l| = 1$, keeping in mind that the total follow-up time is scaled to the interval $[-1, +1]$ so that a time distance of 1 corresponds to half the follow-up time.

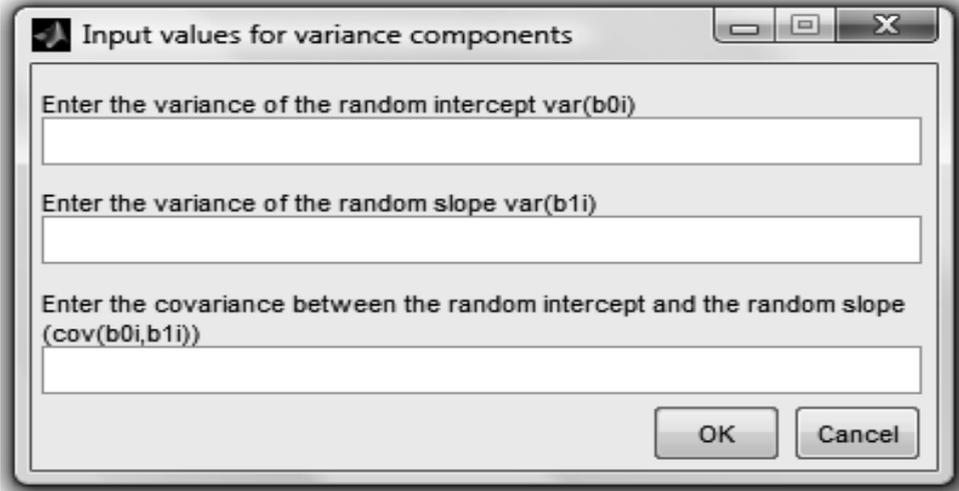


Figure 6.2 A sub-menu of BODMixed_Logistic for input values for variance components in the D matrix for the mixed logistic model with linear time effects.

The second panel in Figure 6.1 allows the user:

- To *choose an approximation to the likelihood method*: The user can choose an approximation method for the computation of optimal designs, i.e., either Extended GEE or PQL1. The default method is the Extended GEE, which is computationally less intensive than PQL1 and usually gives almost the same optimal design as PQL1.

The third panel in Figure 6.1 allows the user:

- To *select the design type*: Either Bayesian D-optimal, or locally D-optimal, or Maximin D-optimal design. When the user selects a design type, a sub-menu to fill in the input values for the relevant parameters will appear. For example, if a linear time effect model is chosen in the first panel and Bayesian D-optimal design is selected in this panel, then the following sub-menu (Figure 6.3) will appear to fill in the input values of priors:

After the user has selected a design type in panel three, the user has to choose between panel four and panel five, and in panel four the user has to choose between either computing the optimal allocation for a given number of time points q , or finding the optimal design for a given subject-to-measurement cost ratio k .

Figure 6.3 The sub-menu of BODMixed_Logistic for input values for the priors for parameters of the logistic model with linear time effect in the case of Bayesian design.

The fourth panel in Figure 6.1 allows the user:

- To compute the *optimal allocations for q time points*: The user will find a dialog box to fill in a specific number of time points q (see, Figure 6.4(a)). Then, the optimal allocations of time points within the time interval $[-1, 1]$ will be found for the specified number of time points q , and the relative efficiency of equidistance compared to the optimal allocation will also be computed.
- To *enter/change the subject-to-measurement cost ratio (k)*: This is the ratio of the cost of adding a new subject to the cost of an additional measurement per subject. This ratio is assumed to be greater than or equal to zero. The user has to decide on the value of the subject-to-measurement cost ratio k .
- To determine the *optimal design for a given cost ratio k* : This helps researchers to identify the optimal number of repeated measurements q , as well as the optimal allocation of those q time points, for a given subject-to-measurement cost ratio k . To find the optimal number and allocations of time points for a given cost ratio k , the user must specify the maximum allowable number of time points, see Figure (4b) for the sub-menu shown by clicking on ‘*Optimal design for a given cost ratio k* ’. Note that the minimum number of time points is two for a linear, three for quadratic and four for cubic polynomial time effect model. These minima have been implemented in the program already.

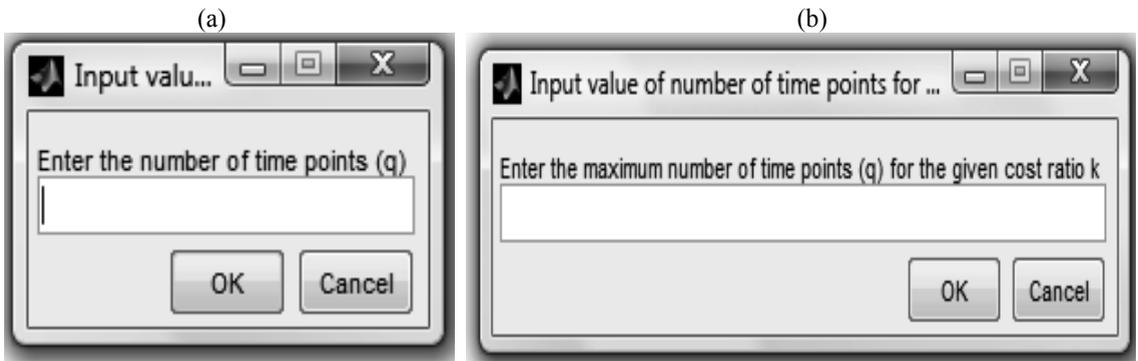


Figure 6.4 A sub-menu of BODMixed_Logistic for input values of number of time points (q).

The fifth panel in Figure 6.1 allows the user:

- To *enter/change the value of autocorrelation range*: Since in general, a user will not know the size of the autocorrelation parameter in the design stage, the user can enter a lower and upper bound for the autocorrelation parameter.
- To plot the *optimal allocations of q time points for different values of autocorrelation*: This is used to assess the effect of different values of the autocorrelation on optimal designs. The optimal allocations within the time interval $[-1, 1]$ for a given number of time points q are computed and plotted against the autocorrelation within the autocorrelation range chosen by the user. By clicking on this button, the user will see the sub-menu of Figure 6.4(a). Any value with $q \geq p$ can be filled in, where p is the number of fixed parameters of the model ($p = 2, 3$ or 4 for the linear, quadratic or cubic model, respectively).

Note that computer time increases exponentially as the number of time points increases. The total computing time to generate the Bayesian optimal designs for all q from p to 7 , per prior with $\sigma^2 = 5$, per covariance structure \mathbf{D} and per autocorrelation, is about 1 hour and 30 minutes irrespective the order of the polynomial (linear, quadratic, cubic), if based on extended GEE. However, for Bayesian optimal design based on PQL1, the total computing time is substantially longer and may take days. This is due to the fact that extended GEE does not incorporate the random effects \mathbf{b}_i , i.e. the diagonal matrix of the conditional variances of the responses \mathbf{w}_i in equation (6.7) is evaluated under $\mathbf{b}_i = 0$ and thus the variance-covariance matrix \mathbf{u}_i of responses equation (6.6) becomes the same for all subjects N in the extended GEE method.

Once the BODMixed_Logistic computer program has started, a user can change input values or obtain results by pressing the corresponding buttons on the main menu (Figure 6.1) as many times as he/she wishes. A ‘Help’ button is also available for guidance. The ‘Exit’ button in the main menu stops the program.

6.5 Example: The respiratory infection study

The Indonesian Children's Health study was a prospective longitudinal study of preschool children aged up to 6 years, conducted in 6 rural villages of the Purwakarta district, West Java, Indonesia between 1977 and 1978 (Sommer et al., 1984; Diggle et al., 1994). More than 3000 preschool children were re-examined medically for up to six consecutive quarters to assess whether they suffered from respiratory or diarrhoeal infection and xerophthalmia, an ocular manifestation of vitamin A deficiency. Diggle et al. (1994) focused on the question of whether vitamin A deficient children are at increased risk of respiratory infection, and found a curvilinear pattern and fitted a random intercept logistic model with a quadratic time effect for the longitudinal data on respiratory infection of 275 Indonesian children who were examined quarterly up to six visits. We adopt this model and some of the results from the analysis of the Indonesian children by Diggle et al. (1994) to illustrate the practical application of our BODMixed_Logistic program in guiding researchers for a similar future study.

After starting the BODMixed_Logistic program (see Section 6.4.1), the user will find the main menu of the BODMixed_Logistic program as shown in Figure 6.1. In the case of the respiratory infection study, the choices might be made as follows:

- *Choose model type*: A quadratic growth function, as this is more flexible than linear and can handle monotonic trends as well as u-shaped trends, see also the results in Diggle et al. (1994).
- *Variance-covariance parameters (\mathbf{D})*: Values reported by Diggle et al. (1994) for the 275 Indonesian children can be used as input values of the variance-covariance parameters. This is $\text{var}(b_0) = 0.5184$ for the variance of the random intercept, and 0 for all other parameters.
- *Value of autocorrelation (ρ)*: For this illustration, we will assume the default value of 0.1 given in Figure 6.1 for the value of autocorrelation.
- *An approximation method*: For this illustration, we will use the default method, i.e., Extended GEE, which is computationally less intensive than PQL1 and usually gives almost the same optimal design as PQL1 (Abebe et al., 2014c).

After choosing the model type, and approximation method, and providing the input values for the variance-covariance parameters (\mathbf{D}) and for the autocorrelation on the left side of the main menu in Figure 6.1, the user has to make choices on the right side of the main menu to obtain results that will help in the design of the study. For the present example we choose:

- *Select design type*: Bayesian D-optimal, with as input priors for the fixed effects parameters an independent normal with prior means $\mu = [1, 2, 3]$ and a prior variance $\sigma^2 = 5$ for each fixed effect.
- *Optimal allocations for q time points*: We use $q = 6$ time points, as the design in the present example had 6 (equidistant) repeated measurements. The user gets the output in the command window of MATLAB. For the present example, this output is given in Figure 6.5, showing the Bayesian optimal allocation of the 6 time points and the relative efficiency of equidistance compared to the Bayesian optimal allocation.

```

Command Window
Bayesian D-optimal design for the given time points (q):
Bayesian_D_optimal =

    -1.0000    -0.6096    -0.2073     0.1883     0.5477     1.0000

Equidistance designs with time points (q):
Equidistance_designs =

    -1.0000    -0.6000    -0.2000     0.2000     0.6000     1.0000

Relative efficiency of equidistance compared to the Bayesian D-optimal design with q time points:
RE_equidistance_compared_to_the_Bayesian_design =

    0.9979

fx >>

```

Figure 6.5 The Bayesian optimal design for time points $q = 6$ for the random intercept logistic model with quadratic time effects using extended GEE.

- *Subject-to-measurement cost ratio (k):* For this illustration, we will assume the default cost ratio given in Figure 6.1, $k = 1$, i.e., we assume equal costs for recruiting a subject and for measurement per a subject. However, the user can change this value if the cost for a subject and the cost for measurement per subject are known at the time of the study.
- *Optimal design for a given cost ratio k :* To find the optimal number of repeated measurements and allocations of time points for a given subject-to-measurement cost ratio (k), the user has to click the button ‘Optimal design for a given cost ratio k ’ and the user must specify the maximum number of time points. Here we specify the number of time points q to a maximum of seven measures, which covers the number of time points in most longitudinal studies. The results are given in Figure 6.6, showing the Bayesian optimal designs for each of the number of time points $q = 3, 4, \dots, 7$, and the relative efficiency of each Bayesian optimal designs compared to the Bayesian optimal design with $q = 7$ time points, for the chosen cost ratio, here $k=1$. The output also indicates which number of repeated measures q is optimal for the given cost ratio $k=1$, that is, which q gives the highest relative efficiency. In this example, $q = 4$ is optimal, giving a relative efficiency of 1.2282 compared to $q = 7$. Moreover, the output gives the relative efficiency of equidistance compared to the optimal design, which in this example is 0.9779. Finally, it gives a plot of relative efficiencies of optimal designs with different numbers of time points compared to the optimal design with the maximum number of time points under different cost ratios k (see, Figure 6.7).

```

Command Window
Bayesian D-optimal allocations of time points for each of the different number of time points (q) :
ans =
    -0.9763    -0.2048     0.5117

ans =
    -1.0000    -0.4073     0.1975     0.7597

ans =
    -1.0000    -0.5200    -0.0163     0.4490     1.0000

ans =
    -1.0000    -0.6096    -0.2073     0.1883     0.5477     1.0000

ans =
    -1.0000    -0.6735    -0.3455    -0.0097     0.3058     0.6111     1.0000

REs of optimal designs compared to the optimal design with maximum time points (q):
REs =
    1.1022     1.2282     1.1924     1.0985     1.0000

The selected optimal Bayesian design for the given cost ratio (k):
Optimal_design_for_a_given_cost_ratio =
    -1.0000    -0.4073     0.1975     0.7597

Equidistance designs with time points (q):
Equidistance_designs =
    -1.0000    -0.3333     0.3333     1.0000

Relative efficiency of equidistance compared to the optimal Bayesian design:
RE_equidistance_compared_to_the_Bayesian_design =
    0.9779

fx >>

```

Figure 6.6 Bayesian optimal allocations of time points for cost ratio $k = 1$ with a maximum number of time points $q = 7$ for the random intercept logistic model with quadratic time effects.

In our example we assumed a single value 0.1 for the autocorrelation. However, the user can also use the `BODMixed_Logistic` program to assess the effect of autocorrelation on optimal allocations of design points for a given q . This additional result is obtained through the panel ‘*Plot of optimal designs for different values of autocorrelation*’ in the main menu in Figure 6.1, where the user can specify both a range for the autocorrelation and the number of time points.

- *Value of autocorrelation range*: The user can enter the lower and upper bounds of the autocorrelation parameter. For this illustration we will use the default values in Figure 6.1.
- *Optimal allocations of q time points for different values of autocorrelation*: By clicking on this button, the user gets the sub-menu of Figure 6.4(a), to choose the number of time points (q). We use $q = 6$ time points.

The results of these choices are displayed in Figure 6.8, showing the Bayesian D-optimal allocation of $q = 6$ time points (horizontal axis) against the autocorrelation (vertical axis) for the random intercept logistic model with quadratic time effects. From the Figure it is observed that the Bayesian D-optimal allocation for $q = 6$ is fairly independent of the size of the autocorrelation at least within the range from 0.001 to 0.9.

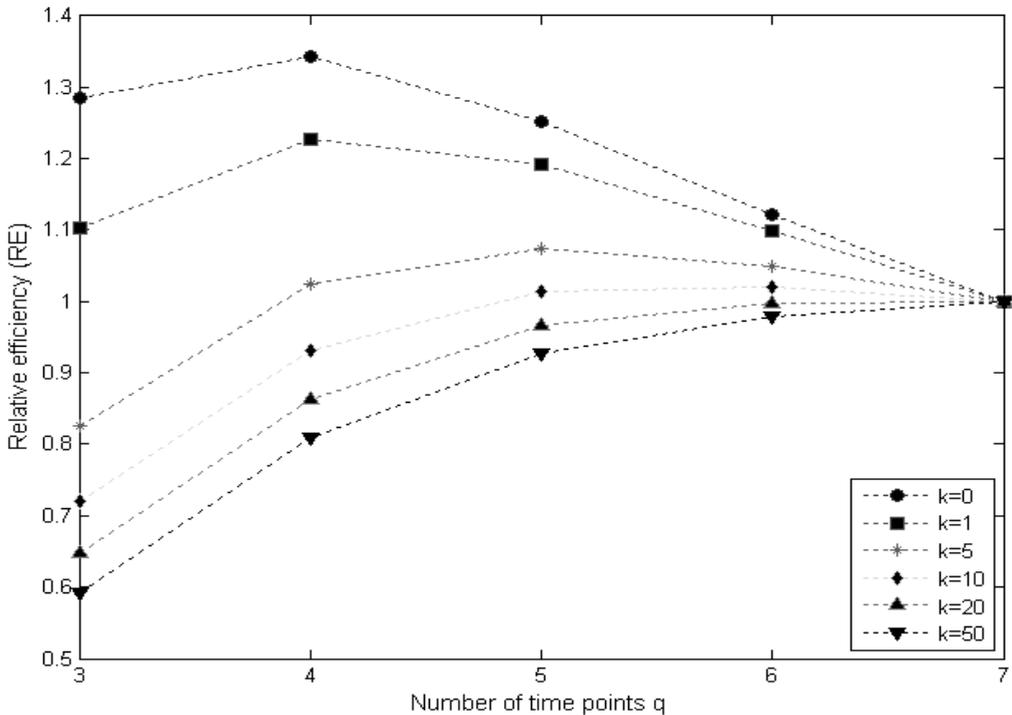


Figure 6.7 Relative efficiency of Bayesian optimal designs with 3 up to q time points compared to the Bayesian optimal design with $q=7$ time points for different cost ratios k for the random intercept logistic model with quadratic time effects.

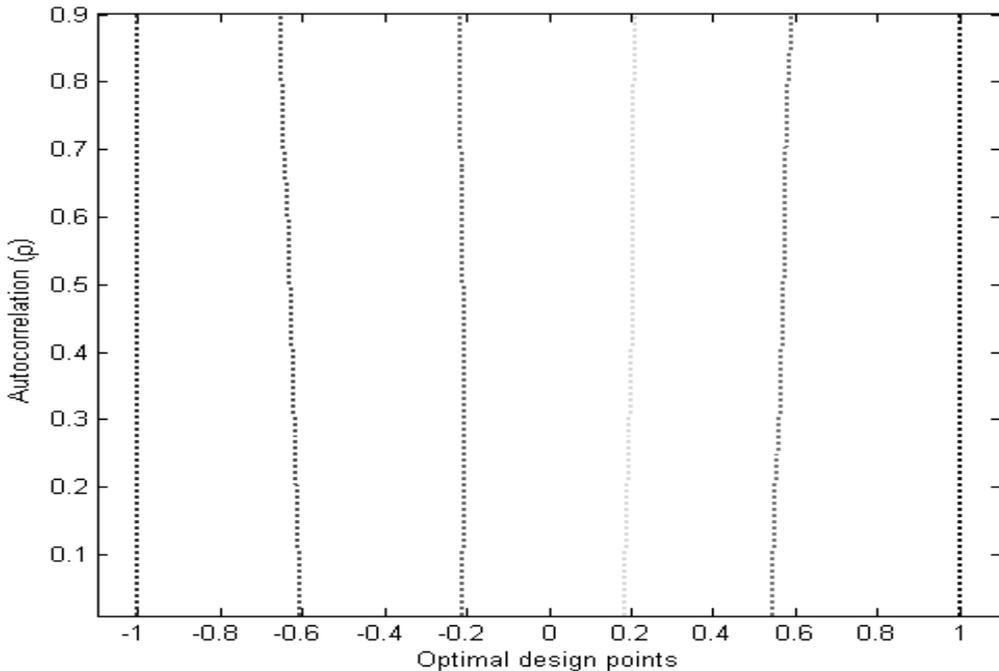


Figure 6.8 Bayesian D-optimal allocations of designs for time points $q = 6$ for a random intercept logistic model with quadratic time effects.

6.6 Conclusions

Optimal designs for longitudinal studies have been shown useful to improve the precision of the model parameter estimates of interest, such as the rate of change. Due to absence of a computer program for optimal design of longitudinal studies with a binary response, planners of such studies in psychology, health sciences and medicine have not yet benefitted from optimal design theory. In this chapter, we presented a user friendly computer program that computes Bayesian optimal designs for mixed effects logistic models with polynomial time effects. This computer program helps researchers to identify the optimal number and allocations of time points of measurements for a given subject-to-measurement cost ratio, and computes the loss in efficiency of equidistance compared to the optimal allocation. Moreover, it helps to assess the effect of autocorrelation on optimal allocations of design points.

The current version of the computer MATLAB program can be obtained from the authors upon request. The current version of the program considers designs based on the D-optimality criterion and assumes that all subjects are available over the total study period and that there is no dropout. Future work may therefore aim at the inclusion of other optimality criteria, and at taking into account anticipated dropout in the design stage.

CHAPTER 7

Summary and recommendations

7.1 Summary

In this thesis, Bayesian optimal designs for cross-sectional studies and longitudinal studies are presented, with the focus on time-structured repeated measures of a binary outcome under costs constraints. Below, the objectives of this thesis and a summary of each chapter are provided. Thereafter, overall conclusions and some suggestions for future research are made.

The aims of this thesis are:

- To provide guidance on the choice of the prior distribution for Bayesian designs and to develop a procedure for obtaining a Bayesian optimal design which is robust against a change of prior distribution.
- To identify an optimal design in terms of the number and allocation of time points within a study period for longitudinal studies with binary response, with constraints of the study cost.
- To investigate how sensitive Bayesian D-optimal designs are against incorrect specification of the autocorrelation and the covariance structure of the random effect parameters.
- To develop a computer program algorithm that produces the optimal number and allocations of time points for a given subject-to-measurement cost ratio.

Chapter 1 is introductory and has three purposes: (1) to explain the need of optimal designs for longitudinal and cross-sectional studies, and the scope of the research; (2) to introduce the logistic mixed models; the Bayesian optimality criterion and suitable cost function for longitudinal studies with corresponding relative efficiency; and (3) to introduce the organization of this thesis.

Chapter 2 concerns how researchers could choose a prior distribution which leads to a Bayesian D-optimal design that remains highly efficient under other prior distributions. The Bayesian optimal designs are computed numerically for various hyperparameter values of a normal and a uniform prior for the fixed logistic regression model with two parameters. Moreover, the loss of efficiency of equidistant equally weighted design points against the Bayesian D-optimal designs is investigated. The main conclusions are: (a) a prior with large variance leads to a design which remains highly efficient under other priors; (b) uniform and normal priors lead to equally efficient designs; and (c) designs with four or five equidistant equally weighted design points are highly efficient relative to the Bayesian D-optimal designs.

Chapter 3 presents the Bayesian optimal design for binary longitudinal responses analyzed with mixed logistic regression and describing a linear time effect in the absence of autocorrelation. A scalar function of the approximated information matrix based on first order penalized quasi likelihood (PQL1) is optimized to find the optimal number and allocations of time points, for different priors, cost constraints and covariance structures of the random effects. Maximin Bayesian D-optimal designs are proposed to overcome the problem of dependence of Bayesian designs on the choice of prior distribution. The results show that the optimal number of time points depends on the subject-to-measurement cost ratio, and

maximin Bayesian D-optimal designs are highly efficient and robust against a change of priors. Locally D-optimal designs are also investigated and maximin locally D-optimal designs are found to have much lower minimum relative efficiency and a smaller number of time points than maximin Bayesian D-optimal designs. Comparing the efficiencies of designs with equidistant time points with the Bayesian D-optimal designs, it was found that three or four equidistant time points are advisable for small cost ratios, and five or six equidistant time points for large cost ratios.

Chapter 4 concerns the Bayesian designs for a mixed logistic model with linear and quadratic changes over time. Bayesian D-optimal allocations of time points are computed for different priors, subject-to-measurement cost ratios, covariance structures and values of the autocorrelation. The results show that: (1) the optimal number of time points increases with the subject-to-measurement cost ratio, and neither the optimal number nor the optimal allocation of time points appear to depend strongly on the prior, the covariance structure, or on the size of the autocorrelation; (2) for subject-to-measurement cost ratios up to five, four equidistant time points, and for larger cost ratios, five or six equidistant time points are highly efficient; (3) our results are compared with the actual design of a respiratory infection study in Indonesia and it is shown that selection of a Bayesian optimal design will increase efficiency, especially for small cost ratios.

Chapter 5 deals with the robustness of Bayesian D-optimal designs for the logistic mixed effects model for longitudinal data with a linear or quadratic time effect against incorrect specification of the autocorrelation. To find the Bayesian D-optimal allocations of time points for different values of the autocorrelation, under different priors for the fixed effects and different covariance structures of the random effects, a scalar function of the approximate variance-covariance matrix of the fixed effects is optimized. Two approximations are compared; one based on a first order penalized quasi likelihood (PQL1) and one based on an extended version of the generalized estimating equations (GEE). The results show that Bayesian D-optimal allocations of time points are robust against misspecification of the autocorrelation, and PQL1 and extended GEE give essentially the same Bayesian D-optimal allocation of time points for a given subject-to-measurement cost ratio. Moreover, Bayesian optimal designs are hardly affected either by the choice of a covariance structure or by the choice of a prior distribution.

Chapter 6 introduces a software tool for researchers. `BODMixed_Logistic` is an interactive user-friendly computer program to obtain the optimal number and allocations of time points for a given subject-to-measurement cost ratio. It also computes the loss in efficiency of equidistant time points compared to the optimal allocation, and it helps to assess the effect of autocorrelation on optimal allocations of design points. This computer program is written in Matlab and is described and illustrated using a practical example.

7.2 Overall conclusions and recommendations

In this research we first investigated how researchers could choose a prior distribution that leads to a Bayesian optimal design that remains robust under other prior distributions. We considered various hyperparameter values of a normal and a uniform prior for the fixed logistic model and computed numerically Bayesian optimal designs. Secondly, we studied the problem of choosing the optimal number and allocation of time points for logistic mixed effects models for longitudinal studies with constraints on the study costs. Our findings suggest that the choice between normal and uniform prior has little consequences for Bayesian D-optimal designs. However, Bayesian optimal designs based on a prior distribution with relatively small variances may be very inefficient due to misspecification of the prior means. So if a researcher has little information about these mean values, he/she should choose a Bayesian design based on a prior distribution with relatively large variance. In contrast, if reasonable information is available about the mean values of the prior distribution, designs based on a prior distribution with relatively large variance may lead to a loss of efficiency. Therefore the prior distribution needs to be very carefully chosen before the design is implemented. Furthermore, Maximin Bayesian optimal designs have a high minimum relative efficiency and improve the robustness against a change of prior means.

The Bayesian optimal *number* of time points depends on the subject-to-measurement cost ratio, and four or five time points are highly efficient for cost ratios less than or equal to ten, whereas, for cost ratios greater than ten, six or seven time points are most efficient. The Bayesian optimal *allocation* of time points is hardly affected by the size of the autocorrelation at least within the range [0.001, 0.9], and is approximately equally spaced. The Bayesian optimal designs are also hardly affected by the choice of a covariance structure for the random effects if there is autocorrelation. In contrast, if there is no autocorrelation, the Bayesian optimal design is affected by the choice of a covariance structure and is no longer approximately equally spaced. Moreover, neither the optimal number nor the optimal allocation of time points appears to depend strongly on the prior means, at least if the prior variance is large. Finally, PQL1 and extended GEE give essentially the same Bayesian D-optimal allocation of time points for a given number of time points q , and also the same optimal number of time points for a given subject-to-measurement cost ratio.

7.3 Ideas for future research

In this thesis, the most commonly used D-optimality criterion, which is based on minimization of the volume of the confidence ellipsoid for all fixed effects parameters β , is considered. An investigation of other optimality criteria for longitudinal studies might be a topic of interest for future work. Further, it is assumed in this thesis that all subjects are available over the total study period and that there is no dropout. However, in longitudinal studies missing data due to dropout are common and it is difficult at the design stage to predict the pattern and rate of the dropout. The impact of anticipated dropout on power as

well as ways to allow dropouts in linear mixed models at the design stage were studied by Galbraith and Marschner (2002) and Verbeke and Lesaffre (1999). For continuous outcomes, optimal designs that do take dropout into account have been studied by Azurduy et al. (2008) and Imhof et al. (2004). They showed that under certain conditions, D-optimal designs are robust against dropout and the loss in efficiency is moderate. Further research is needed to see whether their results will still hold in the Bayesian context and for binary outcomes. Other important issues for future work are Bayesian design for models using splines and for models comparing two or more groups over time. Since the variance of the difference between two independent samples equals the sum of their variances, we expect the present Bayesian optimal design results for polynomial change in a single group to apply to the comparison between groups with respect to such change as well. But again, more theoretical work is needed to check this. A further extension is the optimal allocation of subjects to the groups in such group comparisons, among others the issue of balanced versus unbalanced allocation.

CHAPTER 8

Samenvatting en aanbevelingen (Dutch Summary)

8.1 Samenvatting

In dit proefschrift worden ‘Bayesian optimal designs’ voor cross-sectional studies en longitudinale studies gepresenteerd. Dit met de nadruk op tijd-gestructureerde herhaalde metingen van een binair resultaat met kostenbeperkingen.

Hieronder worden de doelstellingen van dit proefschrift en een samenvatting van elk hoofdstuk verstrekt. Daarna worden algemene conclusies en enkele suggesties voor toekomstig onderzoek gedaan.

De doelstellingen van dit proefschrift zijn:

- Een leidraad geven voor de keuze van de prior verdeling voor ‘Bayesian designs’, en om een procedure te ontwikkelen voor het verkrijgen van een Bayesiaanse optimale proefopzet (design), die robuust is tegen een verandering van de prior verdeling.
- Om een optimale proefopzet in termen van het aantal en de verdeling van de tijdstippen binnen een studieperiode voor longitudinale studies met binaire respons en met kostenbeperkingen te identificeren.
- Om te onderzoeken hoe gevoelig D-optimale Bayesiaanse proefopzetten zijn tegen onjuiste specificatie van de autocorrelatie en de structuur van de covariantie van de random-effect parameters.
- Om een computer programma algoritme te ontwikkelen, dat het optimale aantalen toewijzingen van tijdstippen voor een bepaald ‘subject-to-measurement’ cost ratio’ produceert.

Hoofdstuk 1 is inleidend en heeft drie doeleinden: (1) om de noodzaak van optimale proefopzetten voor longitudinale en cross-sectional studies en de reikwijdte van het onderzoek uit te leggen; (2) om de logistische mixed modellen, de Bayesiaanse optimaliteitscriterium en de geschikte kostenfunctie voor longitudinale studies met hun bijbehorende relatieve efficiëntie in te voeren en (3) om de organisatie van dit proefschrift te introduceren.

Hoofdstuk 2 betreft hoe onderzoekers kunnen kiezen tussen een prior verdeling, die leidt tot een Bayesiaanse D-optimale proefopzet en die zeer efficiënt blijft onder andere prior verdelingen. De Bayesiaanse optimale proefopzetten zijn numeriek berekend voor verschillende hyperparameter-waarden van een normale en een uniforme prior voor het ‘fixed’ logistische regressiemodel met twee parameters. Bovendien wordt het verlies van de efficiëntie van equidistante, gelijk gewogen, designpunten tegen de Bayesiaanse D-optimale proefopzetten onderzocht. De belangrijkste conclusies zijn: (a) een prior met grote afwijking leidt tot een model dat zeer efficiënt blijft onder andere priors (b) uniforme en normale priors leiden tot gelijke efficiënte proefopzetten; en (c) proefopzetten met vier of vijf equidistante, gelijk gewogen, designpunten zijn zeer efficiënt ten opzichte van de Bayesiaanse D-optimale proefopzetten.

Hoofdstuk 3 presenteert de Bayesiaanse optimale proefopzet voor longitudinale binaire respons, geanalyseerd met mixed logistische regressie en een beschrijving van de lineaire tijd effect in de afwezigheid van autocorrelatie. Een scalaire functie van de benaderde informatiematrix, gebaseerd op eerste order penalized quasi likelihood (PQL1), is

geoptimaliseerd om te zoeken naar het optimale aantal en allocatie van tijdspunten) voor de verschillende priors, kosten beperkingen en covariantie structuren van de 'random effects'. Maximin Bayesiaanse D-optimale proefopzetten worden voorgesteld voor het probleem van de afhankelijkheid van Bayesiaanse proefopzetten op de keuze van prior verdeling. De resultaten tonen aan dat het optimale aantal tijdspunten afhangt van het 'subject-to-measurement cost ratio', en dat maximin D-optimale Bayesiaanse proefopzetten zeer efficiënt en robuust zijn tegen een wijziging van priors. Lokaal D-optimale proefopzetten worden ook onderzocht, en maximin lokaal D-optimale proefopzetten bleken een veel lagere minimale relatieve efficiëntie en een kleiner aantal tijdstippen te hebben dan maximin D-optimale Bayesiaanse proefopzetten. In de vergelijking van de efficiëntie tussen proefopzetten met equidistante tijdspunten en de Bayesiaanse D-optimale proefopzetten, bleek dat drie of vier equidistante tijdstippen aan te raden zijn voor kleine 'cost ratios' en vijf of zes equidistante tijdstippen voor grote 'cost ratios'.

Hoofdstuk 4 betreft de Bayesiaanse proefopzetten voor een mixed logistisch model met lineaire en kwadratische veranderingen over tijd. Bayesian D-optimale toewijzingen van tijdspunten worden berekend voor de verschillende priors, subject-to-measurement cost ratios, covariantie structuren en waarden van de autocorrelatie. De resultaten tonen aan dat: (1) het optimale aantal tijdspunten toeneemt met het 'subject-to-measurement cost ratio', en noch het optimale aantal, noch de optimale toewijzing van tijdspunten, lijken sterk af te hangen van de prior, de covariantie structuur, en de grootte van de autocorrelatie; (2) voor 'subject-to-measurement cost ratios', waarden tot vijf, zijn vier equidistante tijdstippen, en voor grotere 'cost ratio' waarden, vijf of zes equidistante tijdstippen zeer efficiënt; (3) onze resultaten worden vergeleken met een actuele proefopzet van een infectie van de luchtwegen onderzoek in Indonesië. Hierbij is aangetoond dat de keus voor een Bayesiaanse optimale proefopzet de efficiëntie zal vergroten, met name voor kleine 'cost ratio' waarden.

Hoofdstuk 5 behandelt de robuustheid van Bayesiaanse D-optimale proefopzetten voor het logistische mixed-effects model voor longitudinale gegevens, met een lineaire en kwadratische time effect tegen onjuiste specificatie van de autocorrelatie. Om de Bayesian D-optimale toewijzing van tijdspunten voor verschillende waarden van de autocorrelatie te kunnen beschouwen onder verschillende priors voor de fixed effecten en verschillende covariantie structuren van de willekeurige effecten, is een scalaire functie van de geschatte variantie-covariantie matrix van de 'fixed-effects' geoptimaliseerd. Twee benaderingen worden vergeleken; één op basis van een eerste orde 'penalized quasi likelihood' (PQL1) en één op basis van een uitgebreide versie van de 'generalized estimating equations' (GEE). De resultaten laten zien dat de Bayesian D-optimale toewijzing van tijdspunten robuust zijn tegen misspecificatie van de autocorrelatie, en dat PQL1 en GEE in essentie dezelfde Bayesian D-optimale toewijzing van tijdspunten voor een 'subject-to-measurement cost ratio' geven. Bovendien worden Bayesiaanse optimale proefopzetten nauwelijks beïnvloed door de keuze van een covariantie structuur of door de keuze van een prior verdeling.

Hoofdstuk 6 introduceert een 'software tool' voor onderzoekers. BODMixed_Logistic is een interactieve gebruikersvriendelijk computerprogramma om het optimale aantal, en de

toewijzingen van tijdpunten voor een bepaald ‘subject-to-measurement cost ratio’ te bepalen. Het berekent ook het verlies van de efficiëntie van equidistante tijdpunten ten opzichte van de optimale toewijzing en het helpt om het effect van autocorrelatie op optimale toewijzing van designpunten te beoordelen. Dit computerprogramma is geschreven in MATLAB en wordt beschreven en geïllustreerd met behulp van een praktisch voorbeeld.

8.2 Globale conclusies en aanbevelingen

In dit onderzoek hebben we ten eerste onderzocht hoe onderzoekers kunnen kiezen tussen een prior verdeling die tot een Bayesiaanse optimale proefopzet leidt, welke robuust blijft onder andere prior verdelingen. We overwogen verschillende hyperparameter waarden van een normale en een uniforme prior voor het vaste logistieke model, en berekende numerieke Bayesiaanse optimale proefopzetten. Ten tweede hebben wij het probleem bestudeerd van de keuze van het optimale aantal en de toewijzing van tijdpunten voor logistieke fixed effecten modellen voor longitudinale studies met kosten beperkingen. Onze bevindingen suggereren dat de keuze tussen de normale en de uniforme prior weinig gevolgen heeft voor Bayesiaanse D-optimale proefopzetten. Echter, Bayesiaanse optimale proefopzetten, op basis van een prior verdeling met relatief kleine afwijkingen, kunnen zeer inefficiënt zijn, dit te wijten aan misspecificatie van de prior means. Dus als een onderzoeker weinig informatie over deze gemiddelde waarden heeft, zou hij / zij een Bayesiaanse proefopzet moeten kiezen op basis van een prior verdeling met relatief grote afwijking. Indien echter redelijke informatie beschikbaar is over de gemiddelde waarden van de prior verdeling, kunnen proefopzetten op basis van een prior verdeling met relatief grote afwijking leiden tot een verlies aan efficiëntie. Daarom moet de prior verdeling zeer zorgvuldig worden gekozen voordat de proefopzet wordt geïmplementeerd. Bovendien, Maximin Bayesiaanse optimale proefopzetten hebben een hoge minimale relatieve efficiëntie en verbeteren de robuustheid tegen een verandering van prior means.

Het Bayesian optimale aantal tijdpunten hangt af van het ‘subject-to-measurement cost ratio’ en vier of vijf tijdpunten zijn zeer efficiënt voor ‘cost ratio’ waarden minder dan of gelijk aan tien, terwijl voor ‘cost ratio’ waarden groter dan tienzes of zeven tijdpunten het meest efficiënt zijn. De Bayesiaanse optimale toewijzing van tijdpunten is nauwelijks beïnvloed door de omvang van de autocorrelatie, ten minste binnen het bereik $[0.001, 0.9]$ en is ongeveer gelijk verdeeld. De Bayesiaanse optimale proefopzetten worden ook nauwelijks beïnvloed door de keuze van een covariantie structuur voor de willekeurige effecten indien er autocorrelatie is. Als er echter geen autocorrelatie is, wordt de optimale proefopzet van Bayesiaanse beïnvloed door de keuze van een covariantie structuur en is niet meer ongeveer gelijk verdeeld. Bovendien, het optimale aantal noch de optimale toewijzing van tijdpunten lijkt sterk afhankelijk te zijn van de prior means, ten minste als de prior variantie groot is. Tot slot, PQL1 en ‘extended’ GEE geven in wezen de dezelfde Bayesian D-optimale toewijzing van tijdpunten voor een gegeven aantal tijd punten q , en ook dezelfde optimale tijdpunten voor een bepaald subject-to-measurement kosten ratio.

References

- Abebe, H.T., Tan F.E.S., van Breukelen G.J.P., Serroyen J., and Berger M.P.F. (2014a). On the choice of a prior for Bayesian D-optimal designs for the logistic regression model with a single predictor. *Communications in Statistics-Simulation and Computation (in press)*. DOI:10.1080/03610918.2012.745556.
- Abebe, H.T., Tan, F.E.S., van Breukelen, G.J.P., and Berger, M.P.F. (2014b). Bayesian D-optimal designs for the two parameter logistic mixed effects model. *Computational Statistics and Data Analysis*, **71**, 1066-1076.
- Abebe, H.T., Tan, F.E.S., van Breukelen, G.J.P., and Berger, M.P.F. (2014c). Robustness of Bayesian D-optimal design for the logistic mixed model against misspecification of autocorrelation. Under review.
- Abdelbasit, K.M., and Plackett, R.L. (1983). Experimental design for binary data. *Journal of the American Statistical Association*, **78**, 90-98.
- Atkinson, A.C., and Haines, L.M. (1996). Designs for nonlinear and generalized linear models. *Handbook of Statistics*, **13**, 437-475.
- Atkins, J.E., and Cheng, C.S. (1999). Optimal regression designs in the presence of random block effects. *Journal of Statistical Planning and Inference*, **77**, 321-335.
- Atkinson, A.C., Donev, A.N., and Tobias, R.D. (2007). *Optimum Experimental Designs, with SAS*. Clarendon, Oxford.
- Azurduy, S.A.O., Tan, F.E.S., and Berger M.P. (2008). The effect of dropout on the efficiency of D-optimal designs of linear mixed models. *Statistical in Medicine*, **27**, 2601-2617.
- Berger, J.O. (1985). *Statistical decision theory and Bayesian analysis*. Springer, New York.

- Berger, M.P.F., and Tan, F.E.S. (2004). Robust designs for linear mixed effects models. *Applied Statistics*, **53**, 569-581.
- Berger, M.P.F., and Wong, W.K. (2009). *An introduction to Optimal designs for social and Biomedical Research*. John Wiley and Sons Ltd.
- Biedermann, S., Dette, H., and Woods, D.C. (2011). Optimal design for additive partially nonlinear models. *Biometrika*, **98**, 449-458.
- 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.
- Bloch, D.A. (1986). Sample size requirements and the cost of a randomized clinical trial with repeated measurements. *Statistical in Medicine*, **47**, 386-399.
- Braess, D., and Dette, H. (2007). On the number of support points of maximin and Bayesian D-optimal designs. *Institute of Mathematical Statistics*, **35**, 772-792.
- Breslow, N.E., and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9-25.
- Carrière, I., and Bouyer, J. (2002). Choosing marginal or random-effects models for longitudinal binary studies: application to self-reported disability among older persons. *BMC Medical research Methodology*, **2**:15.
- Chaloner, K. (1989). Bayesian designs for estimating the turning point of a quadratic regression. *Communications in Statistics, Theory and Methods*, **18**, 1385-1400.
- Chaloner, K., and Larntz, K. (1989). Optimal Bayesian designs applied to logistic regression experiments. *Journal of Statistical Planning and Inference*, **18**, 191-208.
- Chaloner, K., and Verdinelli, I. (1995). Bayesian experimental design. A review. *Statistical Science*, **10**, 273-304.
- Chernoff, H. (1953). Locally optimal designs for estimating parameters. *The Annals of Mathematical Statistics*, **24**, 586-602.
- Chipman, H., and Welch, W. (1996). D-optimal design for Generalized Linear Models, unpublished manuscript. <http://math.acadiau.ca/chipmanh/publications.html>.
- Dargent-Molina, P., Favier, F., Grandjean, H., Baudoin, C., Schott, A.M., Hausherr, E., Meunier, P.J., and Bréart, G. (1996). Fall-related factors and risk of hip fracture: the EPIDOS prospective study. *The Lancet*, **348**, 145-149.
- Dette, H. (1996). A note on Bayesian c- and D-optimal designs in nonlinear regression models. *Annals of Statistics*, **24**, 1225-1234.

- Dette, H. (1997). Designing of experiments with respect to standardized optimality criteria. *Journal of the Royal Statistical Society, Series B*, **59**, 97-110.
- Dette, H., and Saha, M. (1997). Standardized optimal designs for binary response experiments. *South African Statistical Journal*, **31**, 271-298.
- Diggle, P.J., Liang, K.-Y., and Zeger, S. (1994). *Analysis of longitudinal data*. Clarendon Press Oxford.
- Dragalin, V., Fedorov, V., and Wu, Y. (2008). Adaptive designs for selecting drug combinations based on efficacy-toxicity response. *Journal of Statistical Planning and Inference*, **138**, 352-373.
- Dror, H., and Steinberg, D. (2006). Robust experimental design for multivariate generalized linear models. *Technometrics*, **48**, 520-529.
- Drovandi, C.C., McGree, J.M., and Pettitt, A. N. (2013). Sequential Monte Carlo for Bayesian sequentially designed experiments for discrete data. *Computational Statistics and Data Analysis*, **57**, 320-335.
- Firth, D., and Hinde, J.P. (1997). On Bayesian D-optimum Design criteria and the Equivalence Theorem in Non-linear models. *Journal of the Royal Statistical Society, Series B* **59**, 793-797.
- Ford, I., Torsney, B., and Wu, C.-F.J. (1992). The use of a canonical form in the construction of locally optimal designs for nonlinear problems. *Journal of the Royal Statistical Society, Series B* **54**, 569-583.
- Galbraith, S., and Marschner, I.C. (2002). Guidelines for the design of clinical trials with longitudinal outcomes. *Controlled Clinical Trials*; **23**, 257-273.
- Gilks, W.R., and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, **41**, 337-348.
- Gilks, W.R., Best, N.G., and Tan, K.K.C. (1995). Adaptive rejection metropolis sampling within Gibbs sampling. *Applied Statistics*, **44**, 455-472.
- Goos, P. (2002). *The Optimal Design of Blocked and Split-Plot Experiments*. Springer.
- Gotwalt, C., Jones, B., and Steinberg, D. (2009). Fast Computation of Designs Robust to Parameter Uncertainty for Nonlinear Settings. *Technometrics*, **51**, 88-95.
- Han, C., and Chaloner, K. (2004). Bayesian experimental designs for nonlinear mixed models with application to HIV dynamics. *Biometrics*, **60**, 25-33.
- Imhof, L.A. (2001). Maximin designs for exponential growth models and heteroscedastic polynomial models. *Annals of Statistics*, **29**, 561-576.

- Imhof, L.A., Song, D., and Wong, W.K. (2004). Optimal designs of experiment with anticipated pattern of missing observations. *Journal of Theoretical Biology*, **228**, 251-260.
- Jang, W., and Lim, J. (2009). A numerical study of PQL estimation biases in generalized linear mixed models under heterogeneity of random effects. *Communication in Statistics-Simulation and Computation*, **38**, 692-702.
- Khuri, A.I., Mukherjee, B., Sinha, B.K., and Ghosh, M. (2006). Design issues for generalized linear models: A Review. *Statistical Science*, **21**, 376-399.
- King, J., and Wong, W.K. (2000). Minimax D-optimal designs for the logistic model. *Biometrics*, **56**, 1263-1267.
- Lindley, D.V. (1956). On a measurement of the information provided by an experiment. *Annals of Mathematical Statistics*, **27**, 986-1005.
- Matlab (2010). *Handbook Matlab, version 7.10.0.499(R2010a)*. MathWorks Inc, MA, USA.
- Mentře F., Mallet, A., and Baccar, D. (1997). Optimal design in random effect regression models. *Biometrika*, **84**, 429-442.
- Moerbeek, M., Breukelen, G.J.P., and Berger, M.P.F. (2001). Optimal experimental designs for multilevel logistic models. *The Statistician*, **50**(1), 17-30.
- Moerbeek, M., Breukelen, G.J.P., and Berger, M.P.F. (2003). A comparison of estimation methods for multilevel logistic models. *Computational Statistics*, **18**(1), 19-37.
- Moerbeek, M., and Maas, C.J.M. (2005). Optimal experimental designs for multilevel logistic models with two binary predictors. *Communications in Statistics-Theory and Methods*, **34**, 1151-1167.
- Molenberghs, G., and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer.
- Müller, C. (1995). Maximin efficient designs for estimating nonlinear aspects in linear models. *Journal of Statistical Planning and Inference*, **44**, 117-132.
- Müller, C., and Pazman, A. (1998). Application of necessary and sufficient conditions for maximin efficient designs. *Metrika*, **48**, 1-19.
- Niaparast, M. (2009). On optimal design for a Poisson regression model with random intercept. *Statistics and Probability Letters*, **79**, 741-747.
- Niaparast, M., and Schwabe, R. (2013). Optimal design for quasi-likelihood estimation in Poisson regression with random coefficients. *Journal of Statistical Planning and Inference*, **143**, 296-306.

- 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.
- Ouwens, M.J.N.M., Tan, F.E.S., and Berger, M.P.F. (2006). A maximin criterion for the logistic random intercept model with covariates. *Journal of Statistical Planning and Inference*, **136**, 962-981.
- Retout, S., and Mentré, F. (2003). Further Developments of the Fisher information matrix in nonlinear mixed effects models with evaluation in population pharmacokinetics. *Journal of Biopharmaceutical Statistics*, **13**(2), 209-227.
- Sebastiani, P., and Settini, R. (1998). First-order Optimal Design for Nonlinear Problems. *Journal of Statistical Planning and Inference*, **74**, 177-192.
- Sinha, S.k., and Xu, X. (2011). Sequential D-optimal designs for generalized linear mixed models. *Journal of Statistical Planning and Inference*, **141**, 1394-1402.
- Silvey, S.D. (1980). *Optimal design: An introduction to the theory for parameter estimation*. Chapman and Hall, New York.
- Sommer, A., Katz, J., and Tarwotjo, I. (1984). Increased risk of respiratory infection and diarrhea in children with pre-existing mild vitamin A deficiency. *American Journal of clinical Nutrition*, **40**, 1090-1095.
- Song, D., and Wong, W.K. (1998). Optimal two-point designs for the Michaelis-Menten model with heteroscedastic errors. *Communications in Statistics, Theory and Methods*, **27**(6), 1503-1516.
- Tan, F.E.S., and Berger, M.P.F. (1999). Optimal allocation of time points for random effects models. *Communication in Statistics-Simulation and computation*, **28**, 517-540.
- Tan, F.E.S. (2011). Conditions for D_A -Maximin Marginal Designs for Generalized Linear Mixed Models to be Uniform. *Communications in Statistics-Theory and Methods*, **40**, 255-266.
- 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**, 5253-5262.
- Tekle, F.B. (2008). *D-optimal designs for prospective cohort studies*. Ph.D. thesis, Department of Methodology and Statistics, University of Maastricht, The Netherlands.
- 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*, **64**(12), 1383-1390.

-
- Tommasi, C., and López-Fidalgo, J. (2010). Bayesian optimum designs for discriminating between models with any distribution. *Computational Statistics and Data Analysis*, **54**, 143-150.
- Verbeke, G., and Lesaffre, E. (1999). The effect of drop-out on the efficiency of longitudinal experiments. *Applied Statistics*; **48**(3), 363-375.
- 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.
- Woods, D.C., and van de Ven, P. (2006). Blocked designs for experiments with correlated non-normal response. *Technometrics*, **53**, 173–182.
- Woods, D.C., and van de Ven, P. (2011). Blocked designs for experiments with non-normal response. *Technometrics*, **53**, 173-182.
- Wu, C–F.J. (1985). Efficient sequential designs with binary data. *Journal of the American Statistical Association*, **80**, 974-984.
- Yu, J., Goos, P., and Vandebroek, M. (2008). A comparison of different Bayesian design criteria to compute efficient conjoint choice experiments. Available at: https://lirias.kuleuven.be/bitstream/123456789/196127/1/KBI_0817.pdf.
- Zeger, S.L., Liang, K-Y., and Albert, P.S. (1988). Model for Longitudinal Data: A generalized estimating equation approach. *Biometrics*, **44**, 1049-1060.

Acknowledgements

During the journey I started four years ago, it was clear that, although the destination would guarantee satisfaction, the journey would never be a one-man show. This journey would have been impossible without the support and help of many persons.

First of all, I would like to express my gratitude to my supervisors Prof. dr. Martijn Berger, Dr. Frans Tan and Dr. Gerard van Breukelen for their outstanding supervision throughout these four years. It has been a delightful privilege and honour to have learnt from, and been mentored by, such great minds. I am forever grateful. Martijn, thank you for your wonderful ideas, advice, support, kindness and words of encouragement. I have been truly fortunate to work with you, your guidance and valuable comments made my work easier. Frans, thank you for your assistance, discussions, mentoring, and patience with my questions. You always encouraged me to work hard and made time for me to iron out outstanding issues and kept me on track day by day. You are a good friend and fun to work with. Gerard, thank you for your critical, thoughtful and detailed corrections on my research, I really appreciate the effort you put in to improving all my articles, although sometimes I needed time to catch your brilliant ideas. Your support, suggestions and contributions are beyond par.

I would also like to thank the members of the assessment committee for their helpful feedback and discussions on the manuscript of this thesis. I am grateful you made time to read this thesis and also agreed to take part in the public defense.

I owe a great depth of gratitude to all the colleagues at the Department of Methodology and Statistics, Maastricht University, for their distinguishable support over these years in various forms. In particular, I want to thank Marga Doyle for all the help, administrative assistance throughout the project and for translating my summary, Mickey Chenault for willingness to correct the English on my article, Jimmie Leppink for all help and discussions we had, Hans Hilberink for the technical assistance and Jan Serroyen for the technical assistance with the simulation computer and for the discussions we had on optimal designs.

To my former and current B2.068 office mates Baerbel, Stephanie, Francesca and Abu, it has been a pleasure to share the office with you. Many thanks for all your encouragement and interesting wide ranging discussions.

I am grateful for the encouragement and support I have received over the past four years during my stay in The Netherlands. My special thanks go to: Dr. Mulu Gebreyesus, Letish

Gebremedhin, Etetu Ayele, Math Seegers, Dr. Fetene Bekele, Mihiret Anteneh, Mehreteab Aregay (Dn.), Aklilu Habteab, bro. Alfred Fest, Helen Aklilu, Bezawit Tefera, John van, Abiot Tessema (the whole family), Merdia Issa, Hayelom kebede, Araya Medhanyie and Berhe Weldearegwai. I also owe my special thanks to my close friends Sisay Gebremedihin, Niguse Girmay, Assefa G/tsadik and Gidena Teklu for their continuous moral support and for looking after my parents while I was away.

I am deeply grateful to my immediate family, relatives and my in-laws for their unconditional love and support. My parents: Temesgen Abebe and Beriha Yigzaw, my beloved brothers (Alem and Daniel) and my lovely sisters (Alganesh, Zufan and Rezenesh) who have always encouraged me to pursue my studies. I also wish to acknowledge my extended family, Gultmesh, Degef, Alemtsehay, Assefa (the whole family), Abraha, Ataklit (the whole family), Eyasu, Mulu, Selam, Kiros, Tirhas, Hiluf, Yikum, Yonas, Nunish, Aradesh, Ftisum and Teklay for their love, support, encouragement and advice. I am also thankful to my in-laws, especially Abay Mesfin, Tsige Alemayo (the whole family), Eden Girmay, Zerya Gebremedhin and kasech Hailu.

Finally, my greatest realization throughout this long journey is the understanding of the importance of family and the value of time. I thank God every day for blessing me with such a wonderful wife. Words fail me to express my appreciation of my adored wife Desta Abay who accompanied me and took care of our beloved daughter Rina and son Simon. This thesis could not have been accomplished without her constant support and encouragement. Without her I would be a very different person today, and it would certainly have been much harder to finish a PhD. She is the greatest story to ever enter my life. I love you Desta. Thanks for the gracious understanding of my daughter Rina, “Daddy no work stay home,” is finally a reality.

Last in this list but first in my heart, I owe it all to the Almighty God for being my guide and for granting me the wisdom, health and strength to undertake this research task and enabling me to its completion.

About the author

Haftom Temesgen Abebe was born in Wukro, Ethiopia, on September 13, 1982. He attended secondary school at Wukro from 1996 to 2001. Subsequently, he started studying statistics with computer science as minor subject at the Addis Ababa University, Ethiopia. He obtained his Bachelor of Science Degree in Statistics from the same university in 2005 and started working as assistance lecturer at Mekelle University (MU), Ethiopia in the same year. He won the Flemish Inter-University Council (VLIR-UOS) scholarship and joined Hasselt University, Belgium in September 2007. After he specialized and obtained his master degree in Biostatistics from the Hasselt University in 2009, he started his PhD study at the department of Methodology and Statistics of the Maastricht University, The Netherlands. His research focused on Bayesian optimal designs for longitudinal and cross-sectional studies and led to this dissertation.