

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.

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.