**Optimal cohort designs for longitudinal trials with dropout - with application to an Alzheimer's trial**

Kim May Lee (Kim.Lee@mrc-bsu.cam.ac.uk, MRC Biostatistics Unit, Cambridge), Stefanie Biedermann (S.Biedermann@soton.ac.uk), and Robin Mitra (R.Mitra@soton.ac.uk) (both University of Southampton)

**Abstract**

Dropout is a common issue in longitudinal trials, where experimental units are measured repeatedly over time. However, its impact is often ignored at the design stage of the experiment which may lead to less accurate statistical conclusions. We develop a framework for optimal cohort designs that takes potential dropout into account, and provide designs for linear mixed models where the presence of dropout follows a missing at random mechanism. Our framework is illustrated through redesigning a clinical trial on Alzheimer's disease, where we show substantial benefits. The benefits of our designs compared with standard designs are demonstrated through simulations.

**Keywords:** Available case analysis; Design of experiments; Dropouts; Linear mixed models; Missing at random.

# 1 Introduction

Alzheimer's is a degenerative disease that typically results in a gradual deterioration of a patient's condition over time. When evaluating the effectiveness of a treatment, in addition to assessing the immediate effect, it is often of interest to assess whether the rate of decline in a patient's condition has been reduced. Thus, designs of clinical trials to treat Alzheimer's disease are often longitudinal in nature with patients followed up at regular intervals over a fixed period of time. The design problem can then involve two separate decisions, the level of dose to prescribe, as well as the optimal times for follow up.

An example of such a clinical trial can be found in Howard et al. (2012), which studies the benefits of administering the treatments donepezil, memantine, and the combination of the two, to those patients who suffer from moderate to severe Alzheimer's disease. In this study, 291 patients were being followed up for the duration of 52 weeks, whereby measurements on each patient were taken at week 6, 18, 30 and 52 respectively. This type of data is

often called longitudinal data or repeated measurements in the literature. In practice, the presence of missing values is often unavoidable in longitudinal data especially when the duration of the study is long. Moreover, since the clinical studies involve human beings who are relatively old and unhealthy, it is particularly difficult to ensure that patients attend follow-up sessions and some attrition would normally be expected. This feature was present in the Howard et al. (2012) clinical trial.

In the literature on design of experiments, some authors such as Hedayat and John (1974), Ghosh (1979), Ahmad and Gilmour (2010), Ortega-Azurduy, Tan and Berger (2008) investigate the robustness of designs to missing values; Herzberg and Andrews (1976) and Hackl (1995) account for the presence of missing responses in the respective design criteria; Imhof, Song and Wong (2002) and Lee, Biedermann and Mitra (2017) propose optimal design frameworks for regression models in the presence of responses missing at random. The attention on the optimal design framework for longitudinal studies is rather limited, with Ortega-Azurduy, Tan and Berger (2008) the only study taking dropout into account. Assuming completely observed data, Ouwens, Tan and Berger (2002) and Schmelter (2007) focus on optimal design methodology for linear mixed models with a fixed number of time points.

A key feature of an optimal experimental design is its cost efficiency. For a fixed trial budget, an optimal design will provide the largest possible amount of information from the data. Ortega-Azurduy, Tan and Berger (2008) conclude that if a $D$-optimal design assuming all responses will be observed is used "...researchers can easily compensate for such a small efficiency loss by increasing the sample size by at most 15%". However, in many situations a 15% increase in sample size can already incur considerable extra costs, and there are further ways to make a clinical trial design even more efficient. In our investigation, we redesign the trial by Howard et al. (2012) assuming a fixed budget and realistic relative costs of recruiting a new patient versus measuring an existing patient at a further time point, thus finding the cost-optimal number of time points for the trial, in addition to the optimal locations of these time points.

We propose an optimal cohort design framework for longitudinal studies that consider linear mixed models and some pragmatic design constraints assuming that the dropouts in the studies are missing at random (Rubin 1976). By definition, a dropout refers to an experimental unit whose information is not being observed further once the outcome variable on the subject is not being measured at a time point. An observation is missing at random when the probability that a response being missing is dependent

on the observed information only. In our investigation, we assume at the design stage of the experiment that a linear mixed model will be fitted to the incomplete longitudinal data using the available cases. This missing data analysis approach is appealing for its simplicity of application and as it makes inferences based on all the observed data. Moreover, likelihood based inference yields valid conclusions in this framework when missing responses are MAR; see, e.g., Rubin (1976), Little (1995), Little and Rubin (2002) and Molenberghs et al. (2004).

Our framework presents a more general approach to finding optimal cohorts designs over previously proposed approaches, that allows differential missingness between different cohorts of patients (different treatment groups), and more complex design problems where a design consists not only of the optimal time points to measure patients but also of extra variables such as the number of time points (if using a fixed cost constraint) or the doses of a drug. We have thus unified the approaches by Ortega-Azurduy, Tan and Berger (2008), Ouwens, Tan and Berger (2002) and Schmelter (2007), which can be viewed as special cases of our framework, and have extended them to include further more sophisticated and pragmatic models. We further investigate designs where different cohorts of patients can have different time points versus designs where everyone is measured at the same time points. Our framework incorporates practical scenarios, and can thus find designs that may be useful to practitioners, which fills an important gap.

The structure of this paper is as follows. Section 2 introduces the concept of linear mixed models and the notion of missing data mechanisms, as well as depicting the optimal design framework for longitudinal studies that have more than one group of experimental units. Section 3 then derives the optimal design framework for longitudinal studies in the presence of drop out. Allowing for the possibility that different cohorts have different dropout probability functions, we consider two scenarios when constructing designs: (1) different cohorts are allowed to have different sets of time points for measuring the outcome variable, and (2) all experimental units are restricted to have the same set of time points for measuring the outcome variable. Section 4 revisits the Alzheimer's disease clinical trial (Howard et al., 2012) to construct optimal designs through the derived framework, using the real experimental data to elicit the dropout probability functions and the model parameters. Here the main focus is on finding the design that provides the most information on a fixed budget. Section 5 concludes this paper with some discussion and research directions for future work.

# 2 Model

## 2.1 General linear mixed models

We now present the general formulation of a linear mixed model. Let $\mathbf{y}_i^T = (y_{i1}, y_{i2}, ..., y_{iq})$ be the $q$ repeated measurements of subject $i$, $i = 1, .., N$. The responses of subject $i$ can be represented by the linear mixed model,

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

where $\boldsymbol{\beta}$ is a vector of unknown fixed parameters, $\mathbf{X}_i$ and $\mathbf{Z}_i$ are design matrices, $\mathbf{b}_i$ is a vector of unknown individual effects (also called random coefficients) that is normally distributed with mean zero and covariance matrix $\mathbf{D}$, i.e. $\mathbf{b}_i \overset{iid}{\sim} N(\mathbf{0}, \mathbf{D})$, and $\epsilon_i = (\epsilon_{i1}, \epsilon_{i2}, ..., \epsilon_{iq})^T$ is a vector of observational errors that is normally distributed with mean zero and covariance matrix $\sigma^2\boldsymbol{\Psi}$, i.e. $\epsilon_i \overset{iid}{\sim} N(\mathbf{0}, \sigma^2\boldsymbol{\Psi})$. Moreover, $\mathbf{b}_i$ and $\epsilon_i$ are assumed to be independent, $i = 1, \ldots, N$.

The maximum likelihood method provides an unbiased estimator for the fixed effect parameters $\boldsymbol{\beta}$ of the linear mixed model, i.e. $\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^N \mathbf{X}_i^T\mathbf{V}_i^{-1}\mathbf{X}_i\right)^{-1} \sum_{i=1}^N \mathbf{X}_i^T\mathbf{V}_i^{-1}\mathbf{y}_i$, with covariance matrix $\boldsymbol{cov}(\hat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^N \mathbf{X}_i^T\mathbf{V}_i^{-1}\mathbf{X}_i\right)^{-1}$, where $\mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \sigma^2\boldsymbol{\Psi}$ is the covariance matrix of $\mathbf{y}_i$ and $\sum_{i=1}^N \mathbf{X}_i^T\mathbf{V}_i^{-1}\mathbf{X}_i$ is called the Fisher information matrix. Note that this information matrix is summing the individual information that is being contributed by each experimental subject, and can be re-expressed as

$$\sum_{i=1}^{N} \mathbf{X}_i^T\mathbf{V}_i^{-1}\mathbf{X}_i = \sum_{k=1}^{c} n_k\mathbf{X}_k^T\mathbf{V}_k^{-1}\mathbf{X}_k = N\sum_{k=1}^{c} w_k\mathbf{X}_k^T\mathbf{V}_k^{-1}\mathbf{X}_k, \qquad (1)$$

where we assume units are allocated into $c$ groups or cohorts with the same values of the design variables in each cohort. We can then denote $X_k$ to be the unique design matrix of the $k$th cohort, where $k = 1, \ldots, c$, with $n_k$ ($w_k = n_k/N$) reflecting the number (proportion) of experimental units in cohort $k$, respectively.

## 2.2 Dropout mechanisms

We now present the classification of missing data mechanisms as defined by Rubin (1976). Let $y_{ij}$ denote the jth measurement taken for subject i, where $i = 1, \ldots, n$ and $j = 1, \ldots, q$. Similarly denote a binary missing data indicator $l_{ij}$ where $l_{ij} = 1$ denotes $y_{ij}$ is missing and $l_{ij} = 0$ denotes $y_{ij}$ is

observed. The missing data are said to be missing at random if $P(l_{ij} = 1)$ only depends on observed information; it is not missing at random if this probability is dependent on unobserved factors or the missing data itself. A special case of MAR is called missing completely at random, which assumes missing values arise as a random sample from the data. In practice, it is not possible to test whether a missing at random assumption is plausible, but sensitivity analysis can be implemented for making inferences. Missingness is typically assumed to be MAR as inference under NMAR is challenging and not well developed.

At the design stage of an experiment, we consider the presence of MAR that forms a monotone missingness (dropout) pattern, i.e. once an outcome $y_{ij}$ for subject $i$ at time point $j$ is not observed all outcomes at later time points will not be observed for this subject either. We further assume that the MAR mechanism depends on the cohort the subject is in, e.g. which treatment the subject receives, and that the probability that $y_{ij}$ is missing is monotonic increasing with time. Let $n_{k,j}$ be the number of subjects in cohort $k$ who remain in the experiment at time point $j$. Then the dropout process causes $n_{k,1} \geqslant n_{k,2} \geqslant ... \geqslant n_{k,q}$. For $j < q$, we have $g_{k,j} = n_{k,j} - n_{k,j+1}$ units who have exactly $j$ measurements observed and $q - j$ measurements missing, and $g_{k,q} = n_{k,q}$ are the number of subjects with no missing data. At the design stage, the exact rate of dropout will not be known and so $n_{k,j}$ and thus $g_{k,j}$ are treated as random variables. In what follows, we denote the probability of having a response observed on a subject in cohort $k$ at time point $j$ by $p_{k,obs}(t_{kj})$, which is the complement of the missing response probability. Denote $E[g_{k,j}]$ by $m_{k,j}$, then

$$E[g_{k,j}] = m_{k,j} = \begin{cases} Nw_k \; p_{k,obs}(t_{kj}) & \text{if } j = q, \\ Nw_k \; p_{k,obs}(t_{kj}) - Nw_k \; p_{k,obs}(t_{kj+1}) & \text{if } j < q, \end{cases} \tag{2}$$

where $Nw_k$ is the number of subjects originally allocated to cohort $k$.

### 2.3 Cohort design framework

In the absence of missing responses, a design framework for the linear mixed model constructs an optimal design by finding the setting of design matrix $\boldsymbol{X}_k$ such that a function of the Fisher information defined in (1) is optimised over a design region. Considering the information used in available case analysis and the impact of dropouts, the Fisher Information can be written as

$$\sum_{k=1}^{c} \sum_{j=1}^{q} g_{k,j} \boldsymbol{X}_{k[j]}^T \boldsymbol{V}_{k[j]}^{-1} \boldsymbol{X}_{k[j]}$$

where $\boldsymbol{X}_{k[j]}$ denotes the sub-design matrix of a subject in cohort $k$ measured up to the jth time point. However, at the planning stage of the experiment, the observed values of $g_{k,j}$, $k = 1, \ldots, c,$, $j = 1, \ldots q$, are not available, and we aim to maximise a function of the expected information matrix instead, i.e.

$$E\Big[\sum_{k=1}^{c}\sum_{j=1}^{q} g_{k,j}\boldsymbol{X}_{k[j]}^{T}\boldsymbol{V}_{k[j]}^{-1}\boldsymbol{X}_{k[j]}\Big] = \sum_{k=1}^{c}\sum_{j=1}^{q} m_{k,j}\boldsymbol{X}_{k[j]}^{T}\boldsymbol{V}_{k[j]}^{-1}\boldsymbol{X}_{k[j]}.$$

Ortega-Azurduy, Tan and Berger (2008) investigated the loss in efficiency of $D$-optimal designs that were found assuming the complete data set would be observed, when in fact dropouts occured. These authors, however, only considered the situation of one cohort, i.e. where $c = 1$. Here we propose a more comprehensive design framework for longitudinal studies where more than one group of experimental units are considered in the experiment. This is a common occurrence in most clinical studies where there will be two or more groups followed up, e.g. placebo and treatment. Sections 2.3.1 and 2.3.2 derive the relevant framework.

### 2.3.1 Longitudinal cohort study with comparable baseline measurements

In this section we assume that the experimental units are recruited from a homogeneous population where different groups of units have comparable baseline measurements at the onset of the study. An example of this type of study is to investigate the efficacy of different treatments over time on subjects who have the same health status. We consider a special case of the linear mixed model where a group indicator matrix, $\boldsymbol{K}_i$, is incorporated in the model formulation, giving

$$\mathbf{y}_i = \boldsymbol{X}_i\boldsymbol{K}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i,$$

as the repeated measurements of subject $i$. In what follows, we refer to this model as $\mathsf{M}_g$. Schmelter (2007) uses similar models to find optimal designs for linear mixed models, but does not take dropout into account.

To fix ideas, if there are two groups in an experiment, corresponding to e.g. placebo and an active treatment, and the regression is function is linear in time, $\mathsf{M}_g$ has $\boldsymbol{K}_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ if experimental unit $i$ is in group 1; otherwise $\boldsymbol{K}_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ , giving for the observation at the $j$th time

point

$$(\boldsymbol{X}_i\boldsymbol{K}_i\boldsymbol{\beta})_j = \begin{cases} \beta_0 + t_{1j}\beta_1 & \text{if experimental unit } i \text{ is in group 1,} \\ \beta_0 + t_{2j}\beta_2 & \text{if experimental unit } i \text{ is in group 2,} \end{cases}$$

with random effects $(\boldsymbol{Z}_i\boldsymbol{b}_i)_j = b_{0i} + t_{kj}b_{1i}$, $k = 1, 2$. The slope parameters $\beta_1$ and $\beta_2$ reflect the effect on responses of group 1 and group 2 respectively due to a unit change in time.

To construct an optimal cohort design for this special case of the linear mixed model, we seek an optimal design,

$$\xi^* = \begin{Bmatrix} \boldsymbol{t_1'} & \boldsymbol{t_2'} & \cdots & \boldsymbol{t_c'} \\ w_1 & w_2 & \cdots & w_c \end{Bmatrix}, \tag{3}$$

that optimises a function of

$$\sum_{k=1}^{c}\sum_{j=1}^{q} m_{k,j}\boldsymbol{K}_k^T\boldsymbol{X}_{k[j]}^T\boldsymbol{V}_{k[j]}^{-1}\boldsymbol{X}_{k[j]}\boldsymbol{K}_k \tag{4}$$

over the design region of possible time points, $\mathfrak{X}$, and the weights $w_k$, where $\boldsymbol{t_k'} = \{t_{k1}, t_{k2}, ..., t_{kq}\}$ is the optimal allocation of unique time points for measuring an outcome variable on cohort $k$, $k = 1, ..., c$, and $m_{k,j}$, a function of $w_k$, is obtained from (2) . The optimisation is subject to the constraints

$$t_{k1} < t_{k2} < ... < t_{kq} \tag{5}$$

and

$$0 \leq w_k \leq 1,\ k = 1, \ldots, c, \quad \sum_{k=1}^{c} w_k = 1. \tag{6}$$

Note that in order to implement such a design, the weights $w_k$, $k = 1, \ldots, c$, may need to be rounded such that $Nw_k$ is an integer for all $k = 1, \ldots, c$. The more relaxed condition (6) is commonly used in the optimal design literature to facilitate numerical design search. A design found under (6) is referred to as an approximate design, whereas a design where all $Nw_k$'s are integers is called an exact design.

### 2.3.2 Longitudinal cohort study with further design variable

In some scenarios, the experimental conditions of the different cohorts are reflected by a continuous explanatory variable $\delta$ whose levels can be set by

the experimenter. Consider, for example, a treatment that is a dose of a new drug. The experimenter may be interested in investigating $c$ different doses of the drug including placebo, where the doses can be selected from within some dose range, e.g. between placebo and maximum tolerated dose. In this situation, the variable 'dose' becomes part of the design, and it may be possible to increase the amount of information to be gleaned from the data by an efficient selection of the doses as well as the time points. In addition, the probability to observe a response at a given time point may also depend on the dose the patient received. In this section, we extend our design framework to incorporate this scenario.

In the simplest example, the $j^{th}$ repeated measurement of subject $i$ who is in dose group $k$, where $k = 1, \ldots, c$, is

$$y_{ij} = \beta_0 + t_{kj}\beta_1 + \delta_k\beta_2 + b_{0i} + t_{kj}b_{1i} + \epsilon_{ij},$$

where $\beta_0$ is the intercept, $\beta_1$ is the effect on responses of subjects due to a unit change in time, $\beta_2$ is the effect on responses due to a unit change in dose, and $\{b_{0i}, b_{1i}\}$ are random effects. We refer to this model as $\mathsf{M}_d$. This model can easily be extended to include, for example, a dose-time interaction effect to incorporate the possibility that different doses may affect the responses differently over time, but we will consider $\mathsf{M}_d$ for illustration purposes in what follows.

Denote the design matrix of this linear mixed model by $\mathbf{X}_k(t, \delta)$. A design problem is then to find

$$\xi^* = \begin{Bmatrix} (\boldsymbol{t_1'}, \delta_1) & (\boldsymbol{t_2'}, \delta_2) & \cdots & (\boldsymbol{t_c'}, \delta_c) \\ w_1 & w_2 & \cdots & w_c \end{Bmatrix} \tag{7}$$

where the elements, $\delta_k$, reflect the experimental conditions of cohort $k$, $k = 1, ..., c$, such that a function of

$$\sum_{k=1}^{c}\sum_{j=1}^{q} m_{k,j}\boldsymbol{X}_{k[j]}^T(t, \delta)\boldsymbol{V}_{k[j]}^{-1}\boldsymbol{X}_{k[j]}(t, \delta) \tag{8}$$

is optimised over the design regions, i.e. the design region of time points, $\mathfrak{X}$, and $w_k$, the weight for cohort $k$, as well as the design region of $\delta$. Similar to the model in 2.3.1, $m_{k,j}$ is obtained from (2), and this design problem is subject to constraints (5) and (6).

## 2.4 Locally optimal designs

Having chosen the number of cohorts, $c$, the number of repeated measurements, $q$, and the structure of $\boldsymbol{V}_i$, the covariance matrix of $\boldsymbol{y}_i$, for a chosen

formulation of the linear mixed model, a design problem is to find the time points of measuring an outcome variable on the cohorts and the proportion of units to allocate to each cohort, such that a function of the corresponding information matrix is optimised over the design region $\mathfrak{X}$ (and the design region of $\delta$ if $\delta$ is to be optimised in the design problem of the model introduced in 2.3.2). In practice, the structure of $\boldsymbol{V}_i$ is not known at the design stage of an experiment. To construct an optimal design for a future experiment, we employ the notion of locally optimal designs. The structure of $\boldsymbol{V}_i$ can be estimated using some historical data or the information that is obtained from some pilot studies. Moreover, the experimenters need to specify some MAR mechanisms for the different cohorts prior to finding an optimal design for the longitudinal study.

In addition to the assumption that experimental units are identical and independently distributed, we assume that $\epsilon_i$ and $\mathbf{b}_i$ are independent, and that a first order autoregressive process, AR(1), is chosen for $\epsilon_i$ to capture the serial correlation. The AR(1) process has parameter $0 \leq \rho < 1$ and covariance structure with elements $\boldsymbol{\psi}(t_j, t_{j'}) = \rho^{|t_j - t_{j'}|}$. This process is often used to model time series for experiments where observations measured closer together in time are more correlated than those measured further apart. On the other hand, the random effects $\mathbf{b}_i$ reflect the between person variation, i.e. how individuals behave distinctly in the population. By fixing the covariance matrix of $\mathbf{b}_i$, we can find locally optimal cohort designs for different classes of the above described linear mixed models.

For example, consider a linear regression model with a random intercept and slope denoted by $\boldsymbol{b}_i = (b_{0i}, b_{1i})^T$ with covariance matrix

$$\boldsymbol{D} = \begin{pmatrix} var(b_{0i}) & cov(b_{0i}, b_{1i}) \\ cov(b_{0i}, b_{1i}) & var(b_{1i}) \end{pmatrix} = \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix},$$

a fixed effects model has a zero matrix for $\boldsymbol{D}$; a random intercept model has $d_{11} > 0$, $d_{22} = d_{12} = 0$; a random intercept and slope model has $d_{11} > 0$, $d_{22} > 0$, $d_{12} = 0$; and a correlated random intercept and slope model has $d_{11}, d_{22} > 0, d_{12} \neq 0$.

Together with the fixed effects, $\boldsymbol{X}_i \boldsymbol{K}_i \boldsymbol{\beta}$ for model $\mathsf{M}_g$ that has the same baseline measurements for different cohorts, or $\boldsymbol{X}_i(t, \delta)\boldsymbol{\beta}$ for model $\mathsf{M}_d$ where cohorts additionally depend on a continuous design variable, we can consider the profile of locally optimal time points to measure units' responses for a range of different $\rho \in [0, 1)$ by optimising a function of the corresponding information matrix. We note that often a numerical optimisation procedure is required to find a solution to the design problem. For the latter model, we assume that the levels of $\delta$ have been pre-selected for illustration purposes.

We also consider two scenarios when finding an optimal design: (i) different cohorts are allowed to have different sets of time points of measuring the outcome variable, and (ii) all experimental units are restricted to have the same set of time points of measuring the outcome variable. These can be done by setting all sets of $\boldsymbol{t}_{\boldsymbol{k}}^{'}$ as free variables to be searched in the optimisation problem for scenario (i), and restricting all $\boldsymbol{t}_{\boldsymbol{k}}^{'}$ to be one set of time points in the design problem for scenario (ii). In other words, the flexible design condition allows the time points in $\boldsymbol{X}_i\boldsymbol{K}_i\boldsymbol{\beta}$ or $\boldsymbol{X}_i(t,\delta)\boldsymbol{\beta}$ to be different for different cohorts whereas the restricted design condition forces the time points in these design matrices to be the same for all experimental units. In the next section, we study the profile of the optimal time points for measuring units' responses in each cohort under both these scenarios for different classes of models $\mathsf{M}_g$ and $\mathsf{M}_d$.

# 3 Optimal cohort designs in the presence of dropouts

We now illustrate the applications of the above described optimal cohort design framework in the presence of dropouts. Consider an example where an experiment has $c = 2$ and $q = 4$ in each of the two cohorts, and that both special cases of the linear mixed model have fixed effects $\{\beta_0, \beta_1, \beta_2\}$ and random effects $\{b_{0i}, b_{1i}\}$. Note that the interpretation of the fixed effect parameters for both types of models are different, and that the class of linear mixed model is determined by the values of the covariance matrix $\boldsymbol{D}$. For model $\mathsf{M}_d$, we set $\delta_1 = 1$ and $\delta_2 = 0$ for illustration purposes and ease of comparison with model $\mathsf{M}_g$ where the two cohorts are also identified by values zero and one in the matrix $K_i$.

Assume that the design region of time points $\mathfrak{X} = [-1, 1]$, $\sigma^2 = 1$, and that a linear response probability function,

$$p_{1,obs}(t) = 0.65 - 0.35t, \tag{9}$$

and a quadratic response probability function,

$$p_{2,obs}(t) = 0.5 - 0.35t + 0.15t^2, \tag{10}$$

are sufficient to capture the presence of missing responses in each cohort respectively. With this missing mechanism we are assuming units in cohort 1 are more likely to be observed for longer than units in cohort 2. For model $\mathsf{M}_d$, this is an example of differential missingness probabilities for different doses, i.e. there is a probability function $p_{obs}(\delta, t)$ depending on $\delta$ and on

$t$ such that $p_{k,obs}(t) = p_{obs}(\delta_k, t)$, $k = 1, 2$, for $t \in \mathfrak{X}$. Note that the design region $\mathfrak{X} = [-1, 1]$ has been chosen for better comparability of our results with Ouwens, Tan and Berger (2002) and Ortega-Azurduy, Tan and Berger (2008). Since time cannot be negative, we interpret this as an appropriate transformation, e.g. log-time, similar to the common use of log-dose rather than dose in pharmacological studies.

An example of a design problem is to find $\boldsymbol{t_1'} = (t_{11}, t_{12}, t_{13}, t_{14})$, $\boldsymbol{t_2'} = (t_{21}, t_{22}, t_{23}, t_{24})$ and $w_1$, such that the determinant of (4) for model $\mathsf{M}_g$ or (8) for model $\mathsf{M}_d$ is maximised over $\mathfrak{X}$. Note that this is equivalent to minimising the determinant of the covariance matrix for the estimator of the population parameter vector $\hat{\boldsymbol{\beta}}$ and is defined as $D$-optimality; see, e.g. Atkinson, Donev and Tobias (2007).

With the above choices for $p_{k,obs}(t_{kj})$, which lie between zero and one for design region $\mathfrak{X} = [-1, 1]$, this design problem is subject to constraints (2), (5), (6) and the flexible/restricted design condition on the time points. We note that these $p_{k,obs}(t_{kj})$ have been employed in Ortega-Azurduy, Tan and Berger (2008) for the investigation of efficiency loss of $D$-optimal designs due to the presence of dropouts in one cohort, and our design framework is compatible with a wide class of missing at random mechanisms that have monotone response/missing response probability functions.

To find the optimal time points, we choose the lower bound of $\mathfrak{X}$ as the first time point for both groups, i.e. $t_{11} = t_{21} = -1$ in this example, because it has the highest response probability rate over $\mathfrak{X}$; and choose the upper bound of $\mathfrak{X}$ as the last time point, i.e. $t_{14} = t_{24} = 1$, for practical reasons (e.g. pre-selected end of study time). The middle time points of measuring the outcome variable on the groups and the corresponding weights can be found by using optimisation algorithms. In our investigation, we use the function *fmincon* in *Matlab* to find the optimal cohort designs, with equidistant time points and $w_1 = 0.5$ as the initial values for the optimisation problem. For each class of the linear mixed model, we employed $d_{11} = 1,\ d_{22} = 3,\ d_{12} = 0.8\sqrt{3}$ respectively, unless the model required some (or all of these) to be set to 0, to study the profile of the optimal time points across a range of $\rho$ from 0 to 1 in steps sizes of 0.1.

For the experiment that has two cohorts and four repeated measurements in each of the two cohorts, Figure 1 and Figure 2 show the second and the third time points of locally $D$-optimal cohort designs for model $\mathsf{M}_d$ and $\mathsf{M}_g$ respectively, with quadratic response probability function (10) in one group, and linear response probability function (9) in the other group. In these figures, each plot corresponds to the middle time points of the optimal cohort designs for each class of the models; the pair of dotted-lines in the first

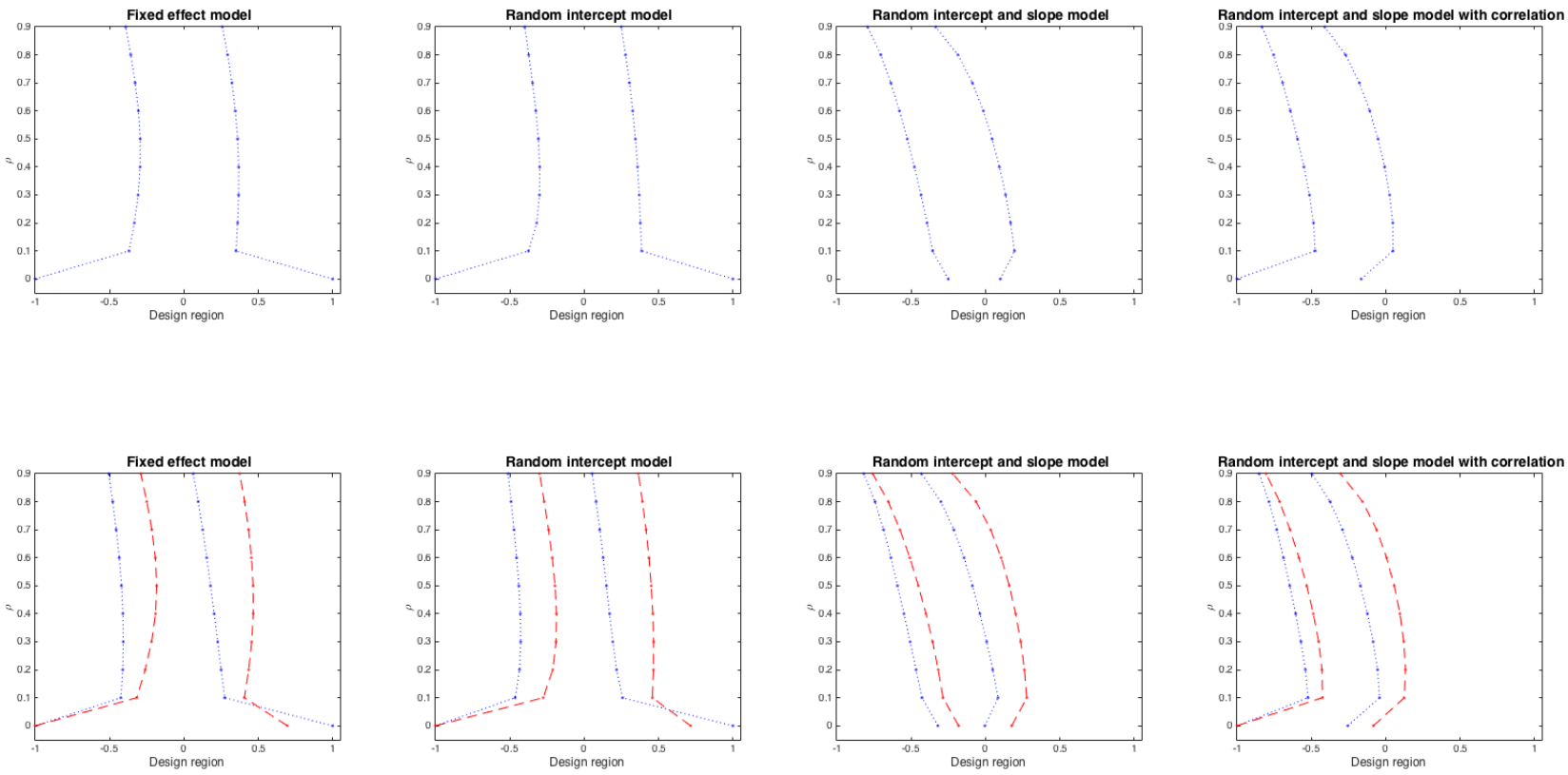


Figure 1: The middle two $D$-optimal time points for model $\mathsf{M}_d$ with $c = 2$, $q = 4$, restricted design condition (top row) and flexible design condition (bottom row) respectively. In the bottom plots, Cohort 1 (blue-dotted lines) has $\delta_1 = 1$ and quadratic response probability function (10); Cohort 2 (red-dashed lines) has $\delta_2 = 0$ and linear response probability function (9).

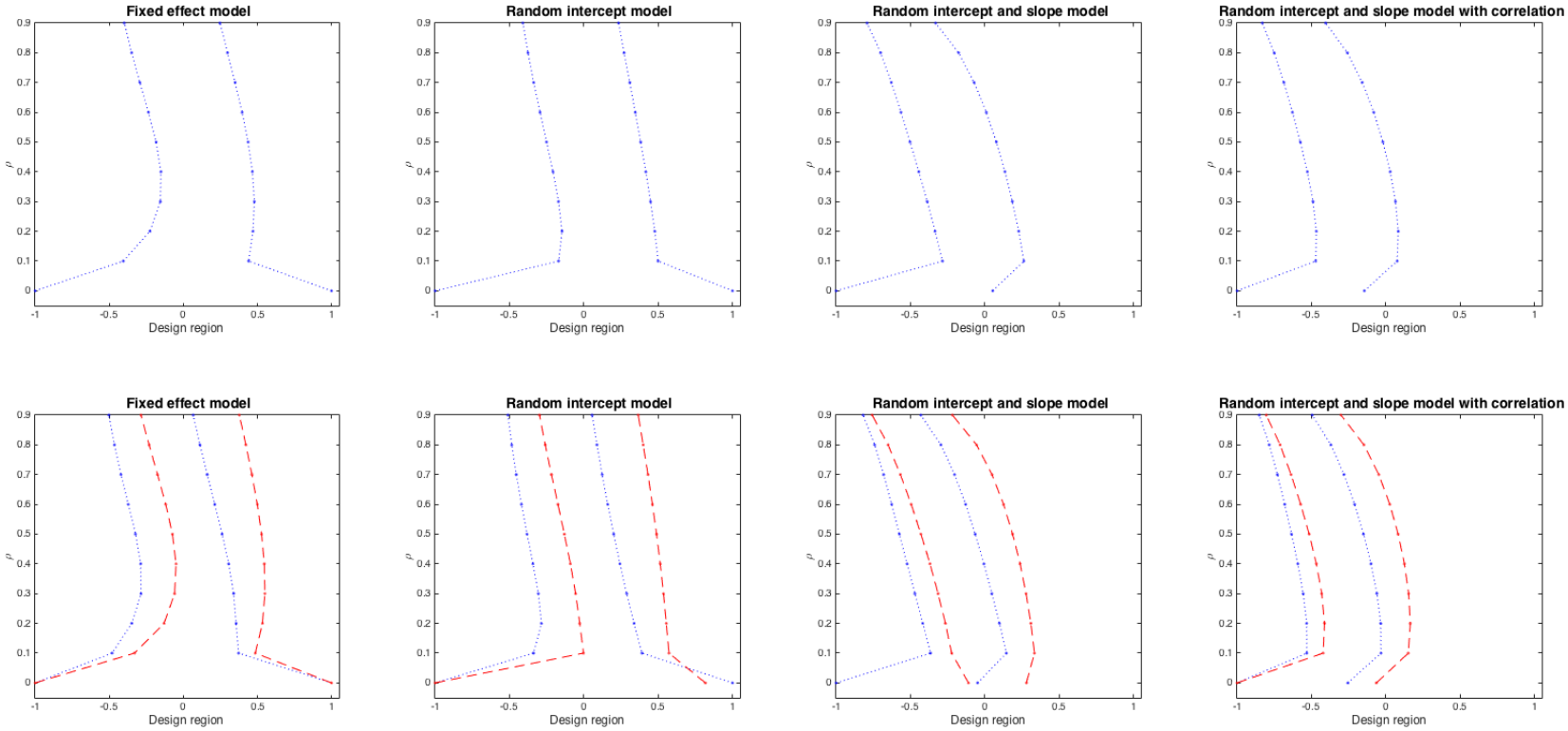


Figure 2: The middle two $D$-optimal time points for model $\mathsf{M}_g$ with with $c = 2$, $q = 4$, restricted design condition (top row) and flexible design condition (bottom row) respectively. In the bottom plots, Group 1 (blue-dotted lines) has quadratic response probability function (10); Group 2 (red-dashed lines) has linear response probability function (9).

row of plots correspond to the second and the third time points assuming both cohorts are measured at the same set of time points (scenario 1). The two pairs of lines in the second row of plots correspond to the sets of time points of measuring the outcome variable in each cohort, which are found assuming these can differ between cohorts (scenario 2). The $y$-axis in each plot shows the considered value of $\rho$ (with 0.1 between each case) in the design problems.

Looking at the trend of the time points, we find that the optimal time points for the experiments with the two response probability functions do not converge to the equidistant design as $\rho$ approaches to one, in particular for the model with independent random intercept and slope parameters, and the model with correlated random intercept and slope parameters. To be more specific, consider the second row of plots in Figure 1 and Figure 2, we find that the optimal time points of measuring the outcome variable on the experimental units who remain longer in the study, i.e. the group with the linear response probability function, are larger than those of the subjects who are expected to be dropping out earlier from the study, i.e. the group with the quadratic response probability function. Intuitively this is reasonable as we would like to observe units' follow up measurements before they drop out.

Comparing the first and the second row of plots in these figures, we learn that there may be considerable differences between the optimal time points in the different cohorts that are found under the respective design conditions. Hence when planning a longitudinal study in practice, it should be investigated if using different time points for different cohorts is feasible (e.g. would not violate double-blindness) since it would increase the amount of information that can be gathered from the data.

We now consider the weights of the locally $D$-optimal cohort designs for the above described experiments, i.e. the proportions of experimental units to allocate to each cohort. Table 1 shows the maximum and the minimum weights across the ten considered cases with $\rho$, $0 \leq \rho \leq 0.9$ (and difference of 0.1 between each case). For the experiments with $\rho > 0$, the design framework maximises the expected total information by having more experimental units in the cohort that has a higher response rate within $\mathfrak{X}$, i.e. the group that has the linear response probability function (red-dashed line) in the plots in Figure 1 and Figure 2 respectively.

On the other hand, for the experiment with $\rho = 0$ in scenario 1, the locally $D$-optimal cohort designs for the corresponding fixed effect models and the random intercept models have $w_1 = 0.5 = w_2$. This is because all the experimental units have the same response rate at the third/ fourth

Table 1: The weight under two design schemes corresponds to the second row of plots whereas the weight under one design schemes corresponds to the first row of plots in Figure 1 and 2.

| | Two design schemes | | One design schemes | |
|---|---|---|---|---|
| | Max $w_1$ | Min $w_1$ | Max $w_1$ | Min $w_1$ |
| $\mathsf{M}_d$ in Figure 1, $w_1$ for the cohort with quadratic response | | | | |
| FE | 0.5064* | 0.4808 | 0.5000* | 0.4804 |
| RI | 0.4984 | 0.4803 | 0.5000* | 0.4799 |
| RIRS | 0.4948 | 0.4885 | 0.4948 | 0.4906 |
| RIRSc | 0.4938 | 0.4869 | 0.4939 | 0.4890 |
| $\mathsf{M}_g$ in Figure 2, $w_1$ for the group with quadratic response | | | | |
| FE | 0.5000* | 0.4821 | 0.5000* | 0.4828 |
| RI | 0.4981 | 0.4901 | 0.5000* | 0.4878 |
| RIRS | 0.4921 | 0.4624 | 0.4921 | 0.4781 |
| RIRSc | 0.4907 | 0.4761 | 0.4907 | 0.4773 |

* *these maximum weights are obtained under the case with* $\rho = 0$.

optimal time points, i.e. at the end points of $\mathfrak{X}$ (see the first two plots from the left in the first row of plots in Figure 1 and in Figure 2 respectively). The same reason also applies to the locally $D$-optimal cohort design for model $\mathsf{M}_g$ with fixed effect parameters and $\rho = 0$ in scenario 2, see the first plot from the left in the second row of plots in Figure 2).

In general, the optimal design framework that accounts for the impact of dropouts overcomes the challenges of incompletely observable responses by allocating more subjects to the group that has a higher response rate within $\mathfrak{X}$. Here we have presented some examples of locally $D$-optimal cohort designs under a given model for dropout for several classes of model $\mathsf{M}_d$ and model $\mathsf{M}_g$ and different values of $\rho$, $0 \leq \rho \leq 0.9$, which are subject to the respective conditions. In this illustration, we have used the values of $\boldsymbol{D}$ given in Section 2.4. In results not presented here, we considered the same set-up of the design problem but with different sets of values for $\boldsymbol{D}$, and we found that the trends of the optimal time points across $0 \leq \rho \leq 0.9$ for each class of the models were similar to those that are illustrated here. All optimal designs were verified by using different initial points in the optimisation algorithm.

# 4 Application: Redesigning a study on Alzheimer's disease

We now illustrate an application of our design framework using the data from an Alzheimer's disease study. The study considered the effectiveness of two drugs to treat Alzheimer's disease, donepezil and memantine, as well as receiving both drugs. The study randomised patients into four groups given by the factorial design, with each group of patients followed up at four subsequent times after baseline. Full details of the study are depicted in Howard et al. (2012). Each patients had five repeated measurements in each group at week 0, 6, 18, 30, and 52 respectively. The aim of the study is to explore the changes from baseline measurement in each group over a period of 52 weeks. For illustration purposes, we only consider the experimental units in the placebo group and the donepezil-memantine group, who were included in the primary intention-to-treat sample. Here we treat the primary outcome measure, SMMSE score (higher score indicates better cognitive function), as the response variable of our model. The total sample size of the data used in this illustration is $N = 144$ (72 in each group). Since the subjects enter the study at similar stages of cognitive decline, we employ model $\mathsf{M}_g$ (which has the same intercept parameter but different slope parameters for the different groups) for this study.

Fitting the SMMSE score of the two groups to the possible classes of model $\mathsf{M}_g$ with an AR(1) process to capture the within-subject correlation in the *R* software, we find that the random intercept model with $\{\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2\} = \{9.876, -0.113 - 0.0716\}$, $\{\hat{d}_{11}, \hat{d}_{22}, \hat{d}_{12}\} = \{2.742^2, 0, 0\}$, $\hat{\sigma}^2 = 2.521^2$ and $\hat{\rho} = 0.2801$ has the smallest AIC and BIC values among the possible classes of models. To obtain realistic response probability functions for the two groups, we use the numbers of subjects who remain in the study over the period of 52 weeks (extracted from Howard et al., 2012) to fit logistic regression models for the two groups respectively, which provide a better fit than the polynomial models used in the illustration in Section 3. We obtain the following response probability

$$P_{obs}(t_{ij}) = \frac{1}{1 + exp(\alpha_0 + \alpha_1 t_{ij})}$$

with $\{\alpha_0, \alpha_1\} = \{-2.3403, 0.07418\}$ for the placebo group, and $\{\alpha_0, \alpha_1\} = \{-3.3514, 0.06427\}$ for the donepezil-memantine group.

Employing this information in the design framework, we find the locally *D*-optimal cohort design for model $\mathsf{M}_g$ with a random intercept for five time

Table 2: Middle time points of several designs for model $\mathsf{M}_g$ with random intercept, $N$=144, $t_{11} = 0 = t_{21}$ and $t_{14} = 52 = t_{24}$ and expected number of patients remaining at each time point. The penultimate column shows the determinant of the inverse of the empirical values of the Fisher information derived in (4); the last column shows the empirical values of $|\boldsymbol{cov}(\hat{\boldsymbol{\beta}})|$ that are averaged over 300 000 simulated sets.

| | $t_{12}$ | $t_{13}$ | $t_{14}$ | $t_{22}$ | $t_{23}$ | $t_{24}$ | $n_1$ | $n_2$ | $\lvert(4)^{-1}\rvert$ | $\lvert\boldsymbol{cov}(\hat{\boldsymbol{\beta}})\rvert$ |
|---|---|---|---|---|---|---|---|---|---|---|
| | placebo group | | | treatment group | | | | | (e-10) | (e-10) |
| $\xi_{ori}$ | 6 | 18 | 32 | 6 | 18 | 32 | 72 | 72 | 2.580 | 2.567 |
| | 62.6 | 52.7 | 38.1 | 68.5 | 64.8 | 58.0 | | | | |
| $\xi^*_{D,2}$ | 2.683 | 34.776 | 43.108 | 2.835 | 44.843 | 48.835 | 71 | 73 | **1.370** | **1.361** |
| | 63.5 | 31.3 | 21.1 | 70.1 | 44.9 | 40.4 | | | | |
| $\xi^*_{D,1}$ | 2.782 | 39.510 | 47.575 | 2.782 | 39.510 | 47.575 | 71 | 73 | 1.434 | 1.430 |
| | 63.5 | 25.3 | 16.6 | 70.1 | 50.6 | 41.8 | | | | |
| $\xi_{eq}$ | 13 | 26 | 39 | 13 | 26 | 39 | 72 | 72 | 2.136 | 2.123 |
| | 57.5 | 43.3 | 26.3 | 66.6 | 60.7 | 50.4 | | | | |

points over the design region $\mathfrak{X} = [0, 52]$, where we fixed week 0 (baseline measurement) and week 52 (end of study) as the first and the last time point of measuring the SMMSE scores on both groups.

Table 2 shows the original design that is used in the clinical study, $\xi_{ori}$, the $D$-optimal cohort designs $\xi^*_{D,2}$ and $\xi^*_{D,1}$ that correspond to the design that allows for having different sets of time points for different cohorts and the restricted design that has the same set of time points for both groups, and $\xi_{eq}$ that has equally spaced time points. The number under each time point reflects the expected number of subjects who remain in the study at each respective time point, which is computed using the corresponding fitted inverse logit link response probability function. If $\xi^*_{D,2}$ was employed in the Alzheimer's disease clinical trial, the experimental units in the placebo group would be measured at week 2.7, 34.8, 42.3 and 52 respectively, and the donepezil-memantine group would be measured at week 2.8, 39.5, 47.6 and 52 respectively after the baseline measurement in week 0. In this case, the responses of the placebo group would have been observed earlier (before week 52) to avoid having large proportions of dropouts in the study. For a restricted experiment that employs $\xi^*_{D,1}$, the experimenters would collect measurements on both groups at week 2.4, 44.6, 48.3 and 52 respectively after week 0.

To compare the performance of the designs, for each design in Table 2,

we simulate

$$y_{ij} = \begin{cases} \hat{\beta}_0 + t_{1j}\hat{\beta}_2 + b_{0i} + \epsilon_{ij} & \text{for subject } i \text{ in placebo group,} \\ \hat{\beta}_0 + t_{2j}\hat{\beta}_3 + b_{0i} + \epsilon_{ij} & \text{for subject } i \text{ in treatment group,} \end{cases}$$

where $t_{1j}$ and $t_{2j}$ correspond to the $j^{th}$ time point of measuring the outcome variable on placebo group that has size $n_1$ and treatment group that has size $n_2$ respectively, $\epsilon_{ij} \sim N(0, \hat{\sigma}^2\hat{\boldsymbol{\psi}})$, $\hat{\boldsymbol{\psi}}(t_j, t_{j'}) = \hat{\rho}^{|t_j - t_{j'}|}$, and $b_{0i} \sim N(0, \hat{d}_{11})$. Missing values in the vector of observed responses of subject $i$, i.e. $\boldsymbol{y}_i = \{y_{i1}, ..., y_{i5}\}$, are then introduced by specifying the corresponding response probability functions for each group of experimental units. We repeatedly simulate the incomplete data 300 000 times as described, and for each incomplete data set, we compute the sample estimates for the fixed effect coefficients, $\{\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2\}$ from the available cases. The elements of $\boldsymbol{cov}(\hat{\boldsymbol{\beta}})$ are then computed using the sample estimates that are obtained from each simulated data set, and the expected values of the information matrix for each design are obtained empirically.

In the last two columns of Table 2, we show the empirical value of the determinant of $\boldsymbol{cov}(\hat{\boldsymbol{\beta}})$ and of the inverse information matrix (4). Comparing these values, we see that $\xi_{ori}$ is not performing well under the $D$-optimality criterion even though the expected number of subjects who remain in the study at each time point is larger than those of other competing designs. Consider the relative $D$-efficiency of a design $\xi$, with respect to a design $\xi^*$, $RE_D(\xi, \xi^*) = \left(\frac{|I_{\xi^*}|}{|I_\xi|}\right)^{1/p}$, where $I_\xi$ denotes the inverse of the expected information matrix of the design $\xi$, and $p$ is the number of fixed effect coefficients in the model. We find that the $D$-efficiency of the original design, $\xi_{ori}$, relative to $\xi^*_{D,2}$ or $\xi^*_{D,1}$ is $(1.370/2.580)^{1/3} = 0.81$ and $(1.434/2.580)^{1/3} = 0.82$, respectively, implying that approximately five replicates of $\xi_{ori}$ would be as efficient as four replicates of the $D$-optimal cohort designs, $\xi^*_{D,2}$ or $\xi^*_{D,1}$.

### 4.1 Cost-saving design

From Table 2, we saw that some of the later time points of the $D$-optimal designs $\xi^*_{D,2}$ or $\xi^*_{D,1}$ were relatively close together. In this section, we investigate if measurements in such points are needed, or if we could make an Alzheimer's trial more cost efficient for the same set-up but with only four time points. Table 3 shows the second and third time points of measuring outcome variables on the different groups, and the simulation output for the design objective function. If the experiment was restricted to have four visits due to a cost constraint, our optimal designs would suggest to

Table 3: Middle time points of four-point optimal designs for model $\mathsf{M}_g$ with random intercept, $N$=144, $t_{11} = 0 = t_{21}$ and $t_{14} = 52 = t_{24}$. The last column shows the empirical values of $|\boldsymbol{cov}(\hat{\boldsymbol{\beta}})|$ that are averaged over 300 000 simulated sets.

| | $t_{12}$ | $t_{13}$ | $t_{22}$ | $t_{23}$ | $n_1$ | $n_2$ | $\lvert(4)^{-1}\rvert$ | $\lvert\boldsymbol{cov}(\hat{\boldsymbol{\beta}})\rvert$ |
|---|---|---|---|---|---|---|---|---|
| | placebo group | | treatment group | | | | (e-10) | (e-10) |
| $\xi^*_{D,2}$ | 32.0740 | 43.2139 | 2.7882 | 47.8880 | 67 | 77 | 2.19109 | 2.18248 |
| $\xi^*_{D,1}$ | 2.7398 | 43.2427 | 2.7398 | 43.2427 | 71 | 72 | 2.47401 | 2.46866 |

Table 4: Middle time points of four-point optimal designs for model $\mathsf{M}_g$ with random intercept, $N$=172, $t_{11} = 0 = t_{21}$ and $t_{14} = 52 = t_{24}$. The last column shows the empirical values of $|\boldsymbol{cov}(\hat{\boldsymbol{\beta}})|$ that are averaged over 300 000 simulated sets.

| | $t_{12}$ | $t_{13}$ | $t_{22}$ | $t_{23}$ | $n_1$ | $n_2$ | $\lvert(4)^{-1}\rvert$ | $\lvert\boldsymbol{cov}(\hat{\boldsymbol{\beta}})\rvert$ |
|---|---|---|---|---|---|---|---|---|
| | placebo group | | treatment group | | | | (e-10) | (e-10) |
| $\xi^*_{D,2}$ | 32.0740 | 43.2139 | 2.7882 | 47.8880 | 80 | 92 | 1.27932 | 1.27714 |
| $\xi^*_{D,1}$ | 2.7398 | 43.2427 | 2.7398 | 43.2427 | 85 | 87 | 1.44071 | 1.43209 |

measure the placebo group at week 32.1, 43.2, 52, and the treatment group at week 2.8, 47.9, 52 after collecting the baseline measurement in week 0; or measure all experimental units at week 2.7, 43.2 and 52 after week 0, if both groups had to be measured at the same time points. Comparing the simulation output of these two optimal designs with those of the original design, $\xi_{ori}$ (first row in Table 2), we found that both four-point $D$-optimal designs perform better than $\xi_{ori}$. Moreover, the four-point $\xi^*_{D,2}$ performs as efficiently as the five-point equal distance design.

Assuming that it costs approximately twice as much to recruit a patient to the study (and take measurements at baseline) as it does to take a follow up measurement, we see that the cost saving afforded from moving from a five time point to a four time point design (including baseline) allows the number of patients able to be recruited to the study to increase from 144 to approximately 172. Table 4 presents optimal designs similar to Table 3 but now allowing for a sample size of 172. We see that performance of the design improves as expected, and in particular $\xi^*_{D,2}$ outperforms all of the five time point optimal designs found earlier. The cost ratios using in this analysis are consistent with typical values associate with these types of studies but our analysis could be redone for a range of different cost scenarios.

# 5 Conclusion and discussion

In this work, we have developed an optimal cohort design framework in the presence of dropouts for a large class of linear mixed models, and have illustrated our methodology through assessing the optimal cohort designs for two special formulations of the linear mixed model, model $\mathsf{M}_g$ and model $\mathsf{M}_d$, where we extended the current literature by allowing differential missingness for different cohorts. This assumption is more realistic since not only time but also the treatment may have an effect on dropout. Moreover, we have studied two different experimental conditions, i.e. a restricted condition where all experimental subjects must have the same set of optimal time points of measuring the outcome variable, and a flexible condition that allows for having different sets of optimal time points of measuring the outcome variable on different cohorts. We found that, as expected, the flexible designs are more efficient, so recommend their use in practice unless this is prevented by problems with double blinding.

We have applied our methodology to a real world example by redesigning a clinical trial for Alzheimer's disease. To this end, we investigated the design problem from two different angles. First, we generated optimal designs for the exact scenario (five time points) of the trial under consideration and found that if an optimal design had been employed in the original study, almost 20% of the experimental effort (and thus experimental costs) could have been saved while obtaining the same amount of information from the resulting data.

The optimal designs for this scenario had several time points clustered relatively close together, which may be impractical (repeated visits to the clinic at short time intervals) and suggested that we may not need five time points for the trial. Hence, we investigated this problem within a cost-efficiency framework, and found that optimal designs with four time points and increased sample size such that the overall cost of the trial are kept fixed can lead to more efficient designs, and thus more information to be gleaned from the data.

We have shown that the presence of dropouts has considerable impact on the locally optimal cohort designs for longitudinal studies. Only locally optimal designs are available for the linear mixed models as the covariance structure of the repeated measurements is often unknown at the design stage of an experiment. Using data from historical/pilot studies, we can estimate the covariance structure for constructing an optimal design for a future experiment. In our investigations, we assume an AR(1) process for the observational errors of the experimental subjects to capture the within-

subject correlation. By trying different sets of values for the variance of random coefficients, we find that the structure of $\boldsymbol{D}$ rather than the values of its elements has more impact on the trends of the optimal time points across the range of realistic values for the serial correlation parameter $\rho$. Further investigation of robustness to nominal parameter values could for example be conducted through an extension of our framework to Bayesian designs.

For future research, we suggest to consider different structures for the serial correlations and the impact of an intermittent missing data pattern on the optimal cohort designs for linear mixed models. Besides that, further investigation of the effect of the extra design variable (e.g. dose) that differentiates the conditions of different cohorts in $\mathsf{M}_d$, and extensions to model $\mathsf{M}_d$ that incorporate a joint dose-time effect, could be considered. It may also be interesting to relax the usual convention (see, e.g. Ortega-Azurduy, Tan and Berger, 2008) of automatically selecting the end of trial as the latest time point for measurements. In particular in the situation where dropout is high, an optimal design may select an earlier time point to be the last, so more responses will be observed. Feasibility of this approach would, of course, depend on the research question of the trial.

We note that other missing data analysis approaches, such as multiple imputation or pattern mixture models (see, e.g., Little and Rubin, 2002), could also be applied to the longitudinal data for making inferences. Developing a design framework for these approaches would be substantially more challenging. These problems have not even been tackled yet for the simpler setting of fixed effects models. A further interesting extension of our work would be to consider generalized linear mixed models in the presence of dropouts. The key challenge to considering these suggestions is to find (a good approximation to) the expected information matrix (or to the covariance matrix) for the corresponding models, having accounted for the features of the missing data analysis approach at the design stage of an experiment. Furthermore, more sophisticated optimisation algorithms might be required to solve the potentially considerably more complex optimisation problem.

## Acknowledgements

The first author's research has been funded by the Institute for Life Sciences at the University of Southampton. We would like to acknowledge Clive Holmes; Robert Howard and Patrick Philips for supplying us with the data

from the Domino study RCTN49545035 which was funded by the MRC and Alzheimer's Society UK.

## References

[1] Ahmad, T., and Gilmour, S. G. (2010). Robustness of subset response surface designs to missing observations.*Journal of Statistical Planning and Inference* **140(1)**, 92-103.

[2] Atkinson, A.C., Donev. A.N. and Tobias, R.D. (2007), *Optimum experimental designs, with SAS*, Oxford University Press.

[3] Baek, I., Zhu, W., Wu, X., and Wong, W. K. (2006). Bayesian optimal designs for a quantal dose-response study with potentially missing observations. *Journal of Biopharmaceutical Statistics* **16(5)**, 679-693.

[4] Ghosh, S. (1979). On robustness of designs against incomplete data. Sankhyā: *The Indian Journal of Statistics, Series B*, 204-208.

[5] Hackl, P. (1995). Optimal design for experiments with potentially failing trials. In *Proc. of MODA4: Advances in Model-Oriented Data Analysis* (Edited by C. P. Kitsos and W. G. Müller), 117-124. Physica Verlag, Heidelberg.

[6] Hedayat, A. and John, P. W. M. (1974). Resistant and Susceptible BIB Designs. *Ann. Statist. 2* **1**, 148–158.

[7] Herzberg, A. M. and Andrews, D. F. (1976). Some considerations in the optimal design of experiments in non-optimal situations. *J. Roy. Statist. Soc. Ser. B* **38**, 284-289.

[8] Howard, R., McShane, R., Lindesay, J., Ritchie, C., Baldwin, A., Barber, R., ... and Phillips, P. (2012). Donepezil and memantine for moderate-to-severe Alzheimer's disease. *New England Journal of Medicine* **366(10)**, 893-903.

[9] Imhof, L. A and Song, D. and Wong, W. K. (2002). Optimal design of experiments with possibly failing trials. *Statistica Sinica* **12**, 1145-1155.

[10] Lee, K. M., Biedermann, S. and Mitra M. (2017). Optimal design for experiments with possibly incomplete observations. *Under revision.*

[11] Little, R. J. A. (1995). Modeling the Drop-Out Mechanism in Repeated-Measures Studies. *Journal of the American Statistical Association* **90(431)**, 1112-1121.

[12] Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data (Second Edition)*. Wiley-Interscience.

[13] Ortega-Azurduy, S. A., Tan, F. E. S. and Berger, M. P. F. (2008). The effect of dropout on the efficiency of $D$-optimal designs of linear mixed models. *Statist. Med.* **27**, 2601-2617.

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

[15] Molenberghs, G., Thijs, H., Jansen, I., Beunckens, C., Kenward, M. G., Mallinckrodt, C. and Carroll, R. J. (2004). Analyzing incomplete longitudinal clinical trial data. *Biostatistics*, **5(3)**, 445-464.

[16] Rubin, D.B (1976). Inference and missing data. *Biometrika* **63**, 581-592.

[17] Schmelter, T. (2007). Considerations on group-wise identical designs for linear mixed models. *Journal of Statistical Planning and Inference* **137**, 4003-4010.