Journal of Statistical Planning and Inference 219 (2022) 85–97

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

# Optimal designs for mean–covariance models with missing observations

Si-Yu Yi [a,b], Yong-Dao Zhou [a,*], Wei Zheng [c]

[a] School of Statistics and Data Science LPMC & KLMDASR, Nankai University, Tianjin 300071, China
[b] College of Mathematics, Sichuan University, Chengdu 610064, China
[c] Haslam College of Business, University of Tennessee, Knoxville, TN 37996, United States

ARTICLE INFO

*Article history:*
Received 28 May 2020
Received in revised form 24 November 2021
Accepted 5 December 2021
Available online 14 December 2021

*Keywords:*
D-optimality criterion
Fisher information matrix
Robustness
Sampling windows
Symmetrization

ABSTRACT

This paper focuses on the optimal design of time point allocation in repeated measurements experiments with applications to longitudinal studies. Most design literature mainly focus on the estimation of the mean responses of each subject whereas we try to find the design that aids the estimation of both the mean and the correlation structures of longitudinal observations. Our design criterion also takes into account the missing data issue, which is very common in practice. Instead of the local optimal design approach, which relies on a good guess of the unknown parameter, we adopt the Bayesian optimal design approach to protect for the parameter uncertainty. To allow for operational errors such as time delays, we discuss the sampling windows which allow flexibility in timing the data collection. In other words, our design is robust against the missingness, parameter uncertainty, and operational errors. Simulation studies and a real data analysis are carried out to demonstrate the proposed criterion as well as the resulting designs.



## 1. Introduction

Longitudinal data arise frequently in the biomedical, epidemiological, social, and economical fields. A salient feature of longitudinal studies is that subjects are measured repeatedly over time. Thus, the responses between subjects may be independent but the repeated measurements within subjects are very likely to be correlated. Ignoring such correlation could result in invalid statistical inferences. Based on the modified Cholesky decomposition, Pourahmadi (1999) and Pourahmadi (2000) reparameterized the marginal covariance matrix by a mean–covariance model where the mean and the covariance matrix were modeled jointly and estimated by the maximum likelihood estimation (MLE) method. Afterwards, much literature work has devoted to extending the approach. Ye and Pan (2006) estimated the parameters in mean–covariance models within the framework of generalized estimating equations. Leng et al. (2010) proposed a data-driven approach based on semiparametric regression models for the mean and the covariance simultaneously, motivated by the modified Cholesky decomposition. Xu et al. (2019) developed a maximum $L_q$-likelihood estimation for the mean–covariance model, which could yield robust and consistent estimators of the mean regression coefficients.

In practical longitudinal studies, the measurements of interest can be missing due to subjects' non-response, dropout, or other reasons. In fact, it is rare to have complete data. There is a rich statistical literature on the analysis of missing data (Rubin, 1976; Little and Rubin, 1994). Two types of missing patterns are generally considered. One is called

* Corresponding author.
*E-mail address:* ydzhou@nankai.edu.cn (Y.-D. Zhou).

https://doi.org/10.1016/j.jspi.2021.12.004
0378-3758/

'nonmonotone missing', where a subject may miss particular visits during the course of study and return at later scheduled visits. The other is 'monotone missing', where a subject may leave the study at some point and never return. For both missing patterns, the missing mechanisms can be classified into three categories. If the missingness is independent of both observed and unobserved data, it is missing completely at random (MCAR). Given the observed data, if the missingness is independent of the unobserved data, it is missing at random (MAR). If the missing probability depends on the unobserved data, it is not missing at random (NMAR). Many studies also have been done to handle incomplete longitudinal data based on the mean–covariance models. Pan and MacKenzie (2006) presented new computational algorithms which can handle unbalanced longitudinal data with missingness, thereby extending existing methods. Huang et al. (2012) embedded the covariance matrix of the observed data in a larger covariance matrix and employed the EM algorithm for both monotone and nonmonotone missingness. Garcia et al. (2012) adopted data-based and graphical methods to handle missing data.

From the viewpoint of experimental designs, it is well known that a carefully designed experiment can substantially improve statistical inferences. Many criteria have been proposed in optimal experimental designs. For example, D-optimality criterion maximizes the determinant of the information matrix of the estimates for the parameters and thus minimized the volume of the confidence region of the parameters of interest. However, in previous studies of longitudinal data, most focused on the estimation of the mean or the fixed effects and few on the covariance matrix, e.g. Ouwens et al. (2002), Tekle et al. (2008), Zhou et al. (2021). Instead, we hope to focus on the estimation of mean and covariance concurrently. Moreover, we also try to address multiple issues simultaneously. (a) The designs which satisfy the optimality criterion without considering the missing observations may be not optimal when there exists missingness indeed. Hence the optimality criteria need to be modified to accommodate for the missingness. Herzberg and Andrews (1976) and Andrews and Herzberg (1979) added the random variables to characterize the potential missingness in the information matrix and maximized its expected determinant. Imhof et al. (2004) introduced a known probability function to depict the probability of valid observation at a trial and put it into the original D-optimality criterion. Alrweili et al. (2019) considered the minimax loss response surface designs which is robust to one missing design point. (b) For a nonlinear model including the one that we will use, the information matrix and hence the optimal design depends on the true value of the parameter. However, the latter is unknown during the planning stage of experiment. One mainstream approach is to derive the optimal design for a particular guessed value of the parameter, and hence the local optimal design. Here, we would adopt the Bayesian optimal design, which is to find a design that is a good compromise over a distribution of the parameter as a prior information. (c) In longitudinal studies, accurate timing for taking measurements is difficult or even impossible. For example, the precisely prearranged times may be not likely to be adhered to in clinical experiments where patients have to attend a clinic for treatment. To allow an experimenter some flexibility in timing the data collection and assure a required design efficiency for parameter estimation, a sampling window approach has been used previously in pharmacokinetic studies and other clinical trials. Graham and Aarons (2006) proposed an approach to pharmacokinetic study design which determined the optimal time windows around the D-optimal pharmacokinetic sampling times. Bogacka et al. (2008) calculated the sampling windows based on the equivalence theorem for D-optimality which makes the widths of the windows related to the parameter sensitivities.

This paper is organized as follows. Section 2 gives a brief introduction of the mean–covariance models and the technique to characterize missingness. Section 3 proposes the optimality criterion and shows the superiority about the symmetrized design. Section 4 discusses the sampling windows and gives the algorithm to search it. Section 5 considers different missing probabilities and shows some simulation studies to compare different kinds of designs. The optimal window widths for certain target efficiencies are also obtained. Section 6 applies the proposed criterion to a real dataset. Some conclusions and discussions are given in Section 7. The proofs of the propositions and theorems are all given in Appendix A.

## 2. Preliminaries based on mean–covariance models

Like Ouwens et al. (2002), we consider designs where all subjects are measured at the same time points, say $(t_1, \ldots, t_n)$. Under a criterion, the optimal time allocation is denoted by $d_{ro} = (t_1^*, t_2^*, \ldots, t_n^*)$. That means the numbers of the repeated measurements for all the subjects are the same, say $n$. If there is no missing data, the complete data is balanced. The covariance matrices of the responses are assumed to be the same for all the subjects. For coherence, we introduce the mean–covariance models for the balanced longitudinal data, as discussed in Pourahmadi (2000). Let $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{in})'$ be an $n \times 1$ stacked vector of $n$ responses made typically at times $t_1, t_2, \ldots, t_n$ for the $i$th subject, $i = 1, \ldots, m$. It is assumed that $\boldsymbol{y}_i \sim N_n(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$ where $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{in})'$ and $\boldsymbol{\Sigma} = (\sigma_{ij})$ are the $n \times 1$ mean vector and the $n \times n$ covariance matrix of $\boldsymbol{y}_i$ respectively. Without loss of generality, the matrix $\boldsymbol{\Sigma}$ is assumed to be positive definite.

Pourahmadi (1999) showed that the modified Cholesky decomposition of $\boldsymbol{\Sigma}^{-1}$ offers a simple unconstrained and statistically meaningful reparameterization of the covariance matrix. In fact, there exist a unique lower triangular matrix $\boldsymbol{T}$ with 1's as diagonal entries and a unique diagonal matrix $\boldsymbol{D}$ with positive diagonals such that $\boldsymbol{\Sigma}^{-1} = \boldsymbol{T}'\boldsymbol{D}^{-1}\boldsymbol{T}$. This decomposition has a simple statistical interpretation. The below-diagonal entries of $\boldsymbol{T}$ are the negatives of the autoregressive coefficients $\phi_{jg}$ in $\hat{y}_{ij} = \mu_{ij} + \sum_{g=1}^{j-1} \phi_{jg}(y_{ig} - \mu_{ig})$, the linear least-squares predictor of $y_{ij}$ based on its predecessors $y_{i(j-1)}, \ldots, y_{i1}$. In other words, $\boldsymbol{\phi}_j = (\phi_{j1}, \ldots, \phi_{j(j-1)})'$ minimizes $E(\epsilon_{ij} - \sum_{g=1}^{j-1} c_g \epsilon_{ig})^2$ with respect to the $c_g$'s where $\epsilon_{ik} = y_{ik} - \mu_{ik}, k = 1, \ldots, j$. The diagonal entries of $\boldsymbol{D}$ are the prediction error (innovation) variances

$\sigma_j^2 = \mathrm{var}(y_{ij} - \hat{y}_{ij})$, $i = 1, \ldots, m, j = 1, \ldots, n$. Since $\phi_{jg}$ and $\log \sigma_j^2$ are unconstrained, they may be further modeled by covariates. For $1 \le i \le m$, $1 \le j \le n$ and $1 \le g \le j-1$, Pourahmadi (2000) considered the mean–covariance models

$$\mu_{ij} = \boldsymbol{x}_{ij}'\boldsymbol{\beta}, \qquad \log \sigma_j^2 = \boldsymbol{z}_j'\boldsymbol{\lambda}, \qquad \phi_{jg} = \boldsymbol{z}_{jg}'\boldsymbol{\gamma}, \tag{1}$$

where $\boldsymbol{x}_{ij}$ is a $r \times 1$ vector of covariates which may contain baseline covariates, polynomials of time and their interactions as well, $\boldsymbol{z}_j$ and $\boldsymbol{z}_{jg}$ are $d \times 1$ and $q \times 1$ vectors of polynomials of time, $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$ are called mean parameter, variance parameter and correlation parameter, respectively. For example, when $\boldsymbol{x}_{ij}$ contains covariates *sex*, *age* and polynomials of time, they may take the forms as follows,

$$\begin{aligned}
\boldsymbol{x}_{ij} &= (1, sex_i, age_i, t_j, \ldots, t_j^{r-3})', \\
\boldsymbol{z}_j &= (1, t_j, \ldots, t_j^{d-1})', \\
\boldsymbol{z}_{jg} &= (1, (t_j - t_g), \ldots, (t_j - t_g)^{q-1})'.
\end{aligned}$$

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\lambda}', \boldsymbol{\gamma}')' = (\boldsymbol{\beta}', \boldsymbol{\alpha}')'$. The Fisher information matrix can be obtained as (Pourahmadi, 2000; Ye and Pan, 2006)

$$I_{\boldsymbol{\theta}}(n, t_1, \ldots, t_n, m, \boldsymbol{\alpha}) = \begin{pmatrix} \sum_{i=1}^m \boldsymbol{X}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_i & & \\ & \frac{m}{2}\boldsymbol{Z}'\boldsymbol{Z} & \\ & & m\boldsymbol{W} \end{pmatrix},$$

where $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{in})'$, $\boldsymbol{Z} = (\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_n)'$ and $\boldsymbol{W} = \sum_{j=1}^n \sigma_j^{-2}(\sum_{k=1}^{j-1}\sum_{l=1}^{j-1}\sigma_{kl}\boldsymbol{z}_{jk}\boldsymbol{z}_{jl}')$. The Fisher information matrix depends on $\boldsymbol{\alpha}$ through $\boldsymbol{\Sigma}$ implicitly. Moreover, under some regularity conditions, Pourahmadi (2000) showed that $I_{\boldsymbol{\theta}}^{-1}(n, t_1, \ldots, t_n, m, \boldsymbol{\alpha}_0)$ is the asymptotic covariance matrix of the MLE $\hat{\boldsymbol{\theta}}$ where $\boldsymbol{\alpha}_0$ is the true value of $\boldsymbol{\alpha}$. The block-diagonal form of $I_{\boldsymbol{\theta}}(n, t_1, \ldots, t_n, m, \boldsymbol{\alpha})$ implies that $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\lambda}}$ and $\hat{\boldsymbol{\gamma}}$ are asymptotically independent.

Due to the complexity of the form of $I_{\boldsymbol{\theta}}(n, t_1, \ldots, t_n, m, \boldsymbol{\alpha})$, it is not easy to derive the optimal design based on the general equivalence theorem, since the direction derivatives of the criteria based on $I_{\boldsymbol{\theta}}(n, t_1, \ldots, t_n, m, \boldsymbol{\alpha})$, such as A-optimality criterion and D-optimality criterion, are difficult to derive. In the following, we focus on exact designs. Let the design space be $\Delta$ and $\tau = (t_1, \ldots, t_n) \in \Delta$ with $t_i \in [-T, T]$, $i = 1, \ldots, n$. Our goal is to arrange a design schedule, such that the estimations of the parameters in models (1) are as accurate as possible in the presence of missingness. Denote the probability of missing the observation at time point $t$ for subject $i$ by $p_i^t$ and let $\boldsymbol{p} = \{p_i^t, i = 1, \ldots, m, t \in [-T, T]\}$. As discussed in Section 1, for any $i$ and $t$, if $p_i^t$ is independent of the time points and other data, no matter whether observed, the missing mechanism is MCAR. If $p_i^t$ only relies on the time points or other observed data, it is MAR. If $p_i^t$ depends on the unobserved data, it is NMAR. Moreover, we assume that the missingness of each subject is independent of other subjects.

Inspired by Herzberg and Andrews (1976), for a design $\tau \in \Delta$, we add the 0-1 random variables $h_{ij}$ (*resp.* $h_{ijk}$) to the Fisher information matrix $I_{\boldsymbol{\theta}}(n, \tau, m, \boldsymbol{\alpha})$ to characterize the situation where the observation at $t_j$ (*resp.* $t_j$ or $t_k$) may be missing, $j, k = 1, \ldots, n$ and $j \ne k$, i.e.,

$$h_{ij} = \begin{cases} 0 & \text{with probability } p_i^{t_j}, \\ 1 & \text{with probability } 1 - p_i^{t_j}, \end{cases}$$

$$h_{ijk} = \begin{cases} 0 & \text{with probability } p_i^{t_j} + p_i^{t_k} - p_i^{t_j}p_i^{t_k}, \\ 1 & \text{with probability } (1 - p_i^{t_j})(1 - p_i^{t_k}). \end{cases}$$

Let $\boldsymbol{H}_i = \mathrm{diag}(h_{i1}, h_{i2}, \ldots, h_{in})$, $1 \le i \le m$. We consider a variant of $I_{\boldsymbol{\theta}}(n, \tau, m, \boldsymbol{\alpha})$,

$$I_{\boldsymbol{\theta}}^H(n, \tau, m, \boldsymbol{\alpha}) = \begin{pmatrix} \sum_{i=1}^m \boldsymbol{X}_i'\boldsymbol{H}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{H}_i\boldsymbol{X}_i & & \\ & \frac{1}{2}\sum_{i=1}^m \boldsymbol{Z}'\boldsymbol{H}_i\boldsymbol{Z} & \\ & & \sum_{i=1}^m \boldsymbol{W}_i^H \end{pmatrix},$$

where $\boldsymbol{W}_i^H = \sum_{j=1}^n \sigma_j^{-2}(\sum_{k=1}^{j-1}\sum_{l=1}^{j-1}\sigma_{kl}h_{ijk}h_{ijl}\boldsymbol{z}_{jk}\boldsymbol{z}_{jl}')$. Thus $I_{\boldsymbol{\theta}}^H(n, \tau, m, \boldsymbol{\alpha})$ is a random matrix with its expectation as

$$E(I_{\boldsymbol{\theta}}^H(n, \tau, m, \boldsymbol{\alpha})) = \begin{pmatrix} \sum_{i=1}^m E(\boldsymbol{X}_i'\boldsymbol{H}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{H}_i\boldsymbol{X}_i) & & \\ & \frac{1}{2}\sum_{i=1}^m E(\boldsymbol{Z}'\boldsymbol{H}_i\boldsymbol{Z}) & \\ & & \sum_{i=1}^m E(\boldsymbol{W}_i^H) \end{pmatrix}. \tag{2}$$

We shall obtain the explicit expressions of $E(I_{\boldsymbol{\theta}}^H(n, \tau, m, \boldsymbol{\alpha}))$ for different missing patterns. Let $\Omega = \{1, 2, \ldots, n\}$ and $\Omega_J = \{J, J+1, \ldots, n\}$. It is clear that $\Omega_1 = \Omega$ and $\Omega_J = \emptyset$ when $J \ge n+1$. For any subset $\omega \subseteq \Omega$, $\boldsymbol{M}_{(-\omega)}$ (*resp.* $\boldsymbol{M}^{(-\omega)}$) represents the submatrix of $\boldsymbol{M}$ by removing its rows (*resp.* both rows and columns) indexed by $\omega$. For monotone missingness, once a missingness occurs, all the subsequent observations are missing. That is, the condition $h_{ij} = 0$ implies $h_{ij} = 0, j > J$ for any $J \in \Omega$. There are only $n+1$ possible values of the matrix $\boldsymbol{H}_i$ characterizing the missingness of subject $i$. Thus the missing probability $p_i^t$ discussed above is a conditional probability,

$$\begin{cases} p_i^{t_1} = p(h_{i1} = 0), \\ p_i^{t_J} = p(h_{iJ} = 0 \mid h_{i1} = \cdots = h_{i(J-1)} = 1), \quad J \in \Omega_2, \end{cases}$$

and the joint probabilities of $(h_{i1},\ldots,h_{in})$ are

$$\begin{cases} U_i^m(J-1) \triangleq p(h_{i1}=1,\ldots,h_{i(J-1)}=1,h_{iJ}=0,\ldots,h_{in}=0)=p_i^{t_J}\prod_{v=1}^{J-1}(1-p_i^{t_v}), & J\in\Omega,\\ U_i^m(n) \triangleq p(h_{i1}=\cdots=h_{in}=1)=\prod_{v=1}^{n}(1-p_i^{t_v}). \end{cases}$$

By convention, we have $\prod_{v=1}^{0}p_v=1$ and $\sum_{v=1}^{0}p_v=0$, where $p_v$ can be any real number or matrix. For monotone missingness, the explicit expressions of the diagonal blocks of $E(I_{\boldsymbol{\theta}}^H(n,\tau,m,\boldsymbol{\alpha}))$ in (2) are given as follows,

$$\begin{aligned} E(\boldsymbol{X}_i'\boldsymbol{H}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{H}_i\boldsymbol{X}_i)&=\sum_{J=1}^{n+1}U_i^m(J-1)(\boldsymbol{X}_{i(-\Omega_J)})'\left(\boldsymbol{\Sigma}^{(-\Omega_J)}\right)^{-1}(\boldsymbol{X}_{i(-\Omega_J)}),\\ E(\boldsymbol{Z}'\boldsymbol{H}_i\boldsymbol{Z})&=\sum_{J=1}^{n+1}U_i^m(J-1)(\boldsymbol{Z}_{(-\Omega_J)})'(\boldsymbol{Z}_{(-\Omega_J)}),\\ E(\boldsymbol{W}_i^H)&=\sum_{J=1}^{n+1}U_i^m(J-1)\sum_{j=1}^{J-1}\sigma_j^{-2}\sum_{k=1}^{j-1}\sum_{l=1}^{j-1}\sigma_{kl}\boldsymbol{z}_{jk}\boldsymbol{z}_{jl}'. \end{aligned}$$

For nonmonotone missingness, we assume that the missingness of the observation for any time point does not affect others. Thus, there are $2^n$ possible values of the matrix $\boldsymbol{H}_i$ and $p_i^{t_j}=p(h_{ij}=0), j=1,\ldots,n$. The joint probabilities of $(h_{i1},\ldots,h_{in})$ are

$$p(h_{i1}=i_1,\ldots,h_{in}=i_n)=\prod_{j=1}^{n}p(h_{ij}=i_j),\ \ i_j\in\{0,1\}.$$

For nonmonotone missingness, the corresponding explicit expressions of the diagonal blocks of $E(I_{\boldsymbol{\theta}}^H(n,\tau,m,\boldsymbol{\alpha}))$ in (2) are

$$\begin{aligned} E(\boldsymbol{X}_i'\boldsymbol{H}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{H}_i\boldsymbol{X}_i)&=\prod_{j=1}^{n}(1-p_i^{t_j})\boldsymbol{X}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_i+\sum_{j_1=1}^{n}p_i^{t_{j_1}}\prod_{j\in\Omega\setminus\{j_1\}}(1-p_i^{t_j})(\boldsymbol{X}_{i(-j_1)})'\left(\boldsymbol{\Sigma}^{(-j_1)}\right)^{-1}(\boldsymbol{X}_{i(-j_1)})+\\ &\quad\sum_{j_1<j_2}^{n}p_i^{t_{j_1}}p_i^{t_{j_2}}\prod_{j\in\Omega\setminus\{j_1,j_2\}}(1-p_i^{t_j})(\boldsymbol{X}_{i(-\{j_1,j_2\})})'\left(\boldsymbol{\Sigma}^{(-\{j_1,j_2\})}\right)^{-1}(\boldsymbol{X}_{i(-\{j_1,j_2\})})+\cdots,\\ E(\boldsymbol{Z}'\boldsymbol{H}_i\boldsymbol{Z})&=\prod_{j=1}^{n}(1-p_i^{t_j})\boldsymbol{Z}'\boldsymbol{Z}+\sum_{j_1=1}^{n}p_i^{t_{j_1}}\prod_{j\in\Omega\setminus\{j_1\}}(1-p_i^{t_j})(\boldsymbol{Z}_{(-j_1)})'(\boldsymbol{Z}_{(-j_1)})+\\ &\quad\sum_{j_1<j_2}^{n}p_i^{t_{j_1}}p_i^{t_{j_2}}\prod_{j\in\Omega\setminus\{j_1,j_2\}}(1-p_i^{t_j})(\boldsymbol{Z}_{(-\{j_1,j_2\})})'(\boldsymbol{Z}_{(-\{j_1,j_2\})})+\cdots\\ &=\sum_{j=1}^{n}(1-p_i^{t_j})\boldsymbol{z}_j\boldsymbol{z}_j',\\ E(\boldsymbol{W}_i^H)&=\prod_{j=1}^{n}(1-p_i^{t_j})\left(\sum_{w=1}^{n}\sigma_w^{-2}\sum_{k=1}^{w-1}\sum_{l=1}^{w-1}\sigma_{kl}\boldsymbol{z}_{wk}\boldsymbol{z}_{wl}'\right)+\\ &\quad\sum_{j_1=1}^{n}p_i^{t_{j_1}}\prod_{j\in\Omega\setminus\{j_1\}}(1-p_i^{t_j})\left(\sum_{w\neq j_1}^{n}\sigma_w^{-2}\sum_{k\neq j_1}^{w-1}\sum_{l\neq j_1}^{w-1}\sigma_{kl}\boldsymbol{z}_{wk}\boldsymbol{z}_{wl}'\right)+\\ &\quad\sum_{j_1<j_2}^{n}p_i^{t_{j_1}}p_i^{t_{j_2}}\prod_{j\in\Omega\setminus\{j_1,j_2\}}(1-p_i^{t_j})\left(\sum_{w\neq j_1,j_2}^{n}\sigma_w^{-2}\sum_{k\neq j_1,j_2}^{w-1}\sum_{l\neq j_1,j_2}^{w-1}\sigma_{kl}\boldsymbol{z}_{wk}\boldsymbol{z}_{wl}'\right)+\cdots\\ &=\sum_{j=1}^{n}\sigma_j^{-2}\left(\sum_{k=l}^{j-1}(1-p_i^{t_j})(1-p_i^{t_k})\sigma_{kl}\boldsymbol{z}_{jk}\boldsymbol{z}_{jl}'+\sum_{k\neq l}^{j-1}(1-p_i^{t_j})(1-p_i^{t_k})(1-p_i^{t_l})\sigma_{kl}\boldsymbol{z}_{jk}\boldsymbol{z}_{jl}'\right). \end{aligned}$$

Hence for both monotone and nonmonotone missingness, the expectation $E(I_{\boldsymbol{\theta}}^H(n,\tau,m,\boldsymbol{\alpha}))$ incorporates all possible cases of missingness. We shall propose the optimality criteria based on it in the following section.

## 3. Optimality criteria for parameter estimation

Based on the explicit expression of $E(I_{\boldsymbol{\theta}}^H(n,\tau,m,\boldsymbol{\alpha}))$, we can regard its determinant, $|E(I_{\boldsymbol{\theta}}^H(n,\tau,m,\boldsymbol{\alpha}))|$, as a variant of the regular D-optimality criterion which is robust to missing observations. Note that $|E(I_{\boldsymbol{\theta}}^H(n,\tau,m,\boldsymbol{\alpha}))|$ still depends on $\boldsymbol{\alpha}=(\boldsymbol{\lambda}',\boldsymbol{\gamma}')'$ through the unknown marginal covariance matrix $\boldsymbol{\Sigma}$. The prior information about $\boldsymbol{\alpha}$, is needed. Instead of the local optimal design approach, which heavily depends on a good guess of the parameter to be estimated, we adopt the Bayesian optimal design approach to protect for the parameter misspecification or uncertainty. Let $\Phi$ be the prior parameter space of $\boldsymbol{\alpha}$. We propose the optimality criterion as

$$R_0(n,\tau,m,\boldsymbol{p})=\frac{1}{|\Phi|}\sum_{\boldsymbol{\alpha}\in\Phi}|E(I_{\boldsymbol{\theta}}^H(n,\tau,m,\boldsymbol{\alpha}))|^{\frac{1}{r+d+q}}, \tag{3}$$

where $|\Phi|$ is the cardinality of $\Phi$ and we take the $(r+d+q)$th root of $|E(I_{\boldsymbol{\theta}}^H(n,\tau,m,\boldsymbol{\alpha}))|$ to eliminate the influence of the order of the Fisher information matrix. Moreover, we consider the feasible region as $D_n=\{(t_1,t_2,\ldots,t_n):-T<t_1<t_2<\cdots<t_n\le T,\ \min_{i=1:(n-1)}\{t_{i+1}-t_i\}\ge\ell\}$. The minimal interval of the $n$ variables is required to be not less than $\ell$ for practical application. A design $d_{ro}=(t_1^*,t_2^*,\ldots,t_n^*)$ is called an optimal design if it maximizes the $R_0$-criterion in (3) among $D_n$. Hence this optimality criterion not only ensures the accuracy of parameter estimations from the perspective of asymptotic variance, but also is robust against missing observations and parameter uncertainty.

Based on the proposed $R_0$-criterion in (3), the optimal designs are derived numerically. Assume that the contained baseline covariates are not changed with the considered time points. That is, for each subject and any $\tau_1,\tau_2\in\Delta$, the values of these baseline covariates under $\tau_1$ equal those under $\tau_2$. To reduce the amount of computational work, the following proposition is used which relates $|E(I_{\boldsymbol{\theta}}^H(n,\tau,m,\boldsymbol{\alpha}))|$ under the design $\tau$ to that under its symmetric design $\tilde{\tau}$, a design obtained from $\tau$ by multiplying each time point by $-1$ in $\Delta$.

**Proposition 1.** *Given a mean–covariance model, let* $\boldsymbol{\alpha}_1=(\boldsymbol{\lambda}_1',\boldsymbol{\gamma}_1')'$ *and* $\boldsymbol{\alpha}_2=(\boldsymbol{\lambda}_2',\boldsymbol{\gamma}_2')'$ *satisfy* $\boldsymbol{\alpha}_2=\boldsymbol{A}\boldsymbol{\alpha}_1$, *where* $\boldsymbol{A}=\begin{pmatrix}\boldsymbol{A}_1 & \\ & \boldsymbol{A}_2\end{pmatrix}$, $\boldsymbol{A}_1$ *and* $\boldsymbol{A}_2$ *are* $d\times d$ *and* $q\times q$ *diagonal matrices, respectively, with odd diagonal entries as 1 and even diagonal entries as -1. If* $p_i^t$ *is origin-symmetric on* $[-T,T]$ *for any subject i,* $i=1,\ldots,m$, *we have*

$$|E(I_{\boldsymbol{\theta}}^H(n,\tau,m,\boldsymbol{\alpha}_1))|=|E(I_{\boldsymbol{\theta}}^H(n,\tilde{\tau},m,\boldsymbol{\alpha}_2))|. \tag{4}$$

Based on Proposition 1, it $\tau^*$ maximizes $|E(I_{\boldsymbol{\theta}}^H(n,\tau,m,\boldsymbol{\alpha}_1))|$, then $\tilde{\tau}^*$ maximizes $|E(I_{\boldsymbol{\theta}}^H(n,\tau,m,\boldsymbol{\alpha}_2))|$ with $\boldsymbol{\alpha}_2=\boldsymbol{A}\boldsymbol{\alpha}_1$. If $\tau$ is symmetric, i.e. $\tilde{\tau}=\tau$, then $|E(I_{\boldsymbol{\theta}}^H(n,\tau,m,\boldsymbol{\alpha}_1))|$ is equal to $|E(I_{\boldsymbol{\theta}}^H(n,\tau,m,\boldsymbol{\alpha}_2))|$. Thereafter, we consider $\Phi$ such that for any $\boldsymbol{\alpha}\in\Phi$, we have $\boldsymbol{A}\boldsymbol{\alpha}\in\Phi$. The following theorem states that for any design we can find a better design according to symmetrization under $R_0$-criterion.

**Theorem 1.** *Assume that* $p_i^t$ *is origin-symmetric on* $[-T,T]$ *for any subject i,* $i=1,\ldots,m$. *For any design* $\tau\in\Delta$, *we have*

$$R_0(n,\xi,m,\boldsymbol{p})\ge R_0(n,\tau,m,\boldsymbol{p}),$$

*where* $\xi=\{(\tau,\frac{1}{2}),(\tilde{\tau},\frac{1}{2})\}$ *and* $\tilde{\tau}$ *is the symmetric design of* $\tau$ *in* $\Delta$.

Theorem 1 is in the spirit of Ouwens et al. (2002) who however discussed the maximin designs only for the mean estimation of the polynomial random slope model. We consider the Bayesian designs for both the mean and the covariance matrix estimations in the presence of missingness. According to Theorem 1, based on a selected optimal design $\tau^*$, we can obtain a symmetrized optimal design where half of the subjects are measured according to $\tau^*$ and half of the subjects are measured according to $\tilde{\tau}^*$. Then this derived design is better than $\tau^*$ under $R_0$-criterion. Further, if $\tau$ itself is a symmetric design, i.e. $\tilde{\tau}=\tau$, we have $\xi=\tau$. It provides a possible search direction to symmetric designs, which may lead to good designs.

Define

$$R_1(n,\tau,m,\boldsymbol{p})=\frac{1}{|\Phi|}\sum_{\boldsymbol{\alpha}\in\Phi}E(|I_{\boldsymbol{\theta}}^H(n,\tau,m,\boldsymbol{\alpha})|)^{\frac{1}{r+d+q}}, \tag{5}$$

where the order of taking expectation and taking determinant is exchanged compared with $R_0$-criterion. The form of taking expectation for the determinant is used in the literature, e.g. Herzberg and Andrews (1976) and Zheng (2013). Both $R_0$-criterion and $R_1$-criterion are feasible and the proposed $R_0$-criterion may be computed more easily with less determinants. For $i=0,1$, let $\tau_i^*$ be an optimal design under $R_i$-criterion. Define $\mathrm{eff}_i(\tau)=R_i(n,\tau,m,\boldsymbol{p})/R_i(n,\tau_i^*,m,\boldsymbol{p})$, $i=0,1$, as the efficiency of $\tau$ under $R_i$-criterion and the gap function $g(\tau)=R_1(n,\tau,m,\boldsymbol{p})/R_0(n,\tau,m,\boldsymbol{p})$. Actually, according to Lemma 2 in Zheng (2013), we can obtain that $R_1$-efficiency $\mathrm{eff}_1(\tau)$ could be bounded by $\mathrm{eff}_0(\tau)g(\tau)$.

**Proposition 2.** *For any design* $\tau$, *we have* $R_1(n,\tau,m,\boldsymbol{p})\le R_0(n,\tau,m,\boldsymbol{p})$. *Further, we have* $\mathrm{eff}_1(\tau)\ge\mathrm{eff}_0(\tau)g(\tau)$. *In particular, for any* $R_0$*-optimal design* $\tau$, *we have* $\mathrm{eff}_1(\tau)\ge g(\tau)$.

The proof of Proposition 2 can be referred to Zheng (2013) and we omit it here. By Proposition 2, we have $g(\tau) \leq 1$ and $g(\tau_0^*) \leq \text{eff}_1(\tau_0^*)$. It means that if we could select a $R_0$-optimal design, then the value of the gap function $g$ evaluated at this design serves as a lower bound of its $R_1$-efficiency. When the missing probability is not large, the value of the gap function $g$ would be close to unity. Thus the $R_0$-optimal design could be highly efficient under $R_1$-criterion.

For $R_0$-criterion, we define the relative efficiency ($RE$) to compare the performance of the designs. Actually, the values of $n$ and $m$ are not replaceable, e.g. $n_1 = 2n_2$ cannot be offset by $m_1 = m_2/2$, because of the within-subject correlation. Doubling up the number of time points for each subject does not double up the information, while doubling up the number of subjects indeed doubles up the amount of the information. Thus for any two designs $\eta_1 = (n, \tau_1, m_1)$ and $\eta_2 = (n, \tau_2, m_2)$ with the same number of repeated measurements, the $RE$ are defined as

$$RE(\eta_1, \eta_2) = \frac{m_2}{m_1} \cdot \frac{R_0(\eta_1, \boldsymbol{p})}{R_0(\eta_2, \boldsymbol{p})},$$

where the term $m_2/m_1$ is used to eliminate the impact of the number of subjects.

The derivatives of $R_0$-criterion are not easy to calculate. As a result, many optimization algorithms which need to use derivatives are not applicable. Alternatively, we use the MSNTO algorithm proposed in Yi et al. (2021), which is a sequential algorithm for optimization and does not need the derivatives. It combines the SNTO algorithm (Fang and Wang, 1994) with a transformation, denoted by $\boldsymbol{h}$-transformation, to convert the points on the regular hypercube $[0, 1]^n$ to the irregular region $D_n$. Thus MSNTO can be used to find the global maximum point of a continuous function over the closed, bounded and irregular domain $D_n$. MSNTO scatters points on the domain using uniform design (Fang et al., 2018) coupled with $\boldsymbol{h}$-transformation. According to the $R_0$-criterion value calculated on the uniform design points, we select the current maximum point $\boldsymbol{t}$. Let $\boldsymbol{o} = (o_1, \ldots, o_n)$ be the point on $[0, 1]^n$ with $\boldsymbol{t} = \boldsymbol{h}(\boldsymbol{o})$. The domain is contracted to $[\boldsymbol{a}^{(w+1)}, \boldsymbol{b}^{(w+1)}]$ by $a_i^{(w+1)} = \max(o_i - \gamma_0 c_i^{(w)}, 0)$, $b_i^{(w+1)} = \min(o_i + \gamma_0 c_i^{(w)}, 1)$, $i = 1, \ldots, n$, where $\boldsymbol{a}^{(w+1)} = (a_1^{(w+1)}, \ldots, a_n^{(w+1)})$, $\boldsymbol{b}^{(w+1)} = (b_1^{(w+1)}, \ldots, b_n^{(w+1)})$, $c_i^{(w)} = (b_i^{(w)} - a_i^{(w)})/2$ and $\gamma_0$ is a predefined contraction ratio. Then we obtain the subdomain of $D_n$ which corresponds to $\boldsymbol{h}([\boldsymbol{a}^{(w+1)}, \boldsymbol{b}^{(w+1)}])$. The same uniform design is used on the subdomain and it is further contracted by the same way. The process is repeated until the volume of the domain is small enough. Moreover, if the functions in $\boldsymbol{p}$ are all even functions, i.e. origin-symmetric, we can restrict the design space to the symmetric designs. It may bring designs with good performance and can reduce the computation complexity.

## 4. $R_0$-Efficient sampling windows

Section 3 discussed the search of the design $d_{ro} = (t_1^*, t_2^*, \ldots, t_n^*)$ which achieved the precision for parameter estimation and the robustness to missing observations and parameter misspecification. However, in clinical experiments, if the patients are in an uncontrolled environment, accurate timing for taking samples is problematic. Even though there is no missingness, there may be some reasons for delays, caused by the patients or the clinic personnel. Then an optimal design in Section 3 can be much less informative for the purpose when it was set up. For this purpose, we consider the approach of sampling windows to allow the flexibility in the data collection and guarantee a certain efficiency for the estimates of the parameters.

Since we do not know the specific impact of each optimal time point on the estimates of the parameters in advance, we treat every point equally and set the width of each window the same. We define the sampling windows as

$$\Im_\kappa = [t_1^*, t_1^* + \kappa] \times [t_2^*, t_2^* + \kappa] \cdots \times [t_n^*, t_n^* + \kappa],$$

where any randomly sampled time point on $[t_i^*, t_i^* + \kappa]$ is uniformly distributed for $1 \leq i \leq n$. For any design $\tau = (t_1, \ldots, t_n) \in \Im_\kappa$, we can calculate the $R_0$-efficiency $\text{eff}_0(\tau)$. Our goal is to select the optimal window width $\kappa^*$, such that the minimal $R_0$-efficiency of all the designs on $\Im_{\kappa^*}$ is equal to a given $R_0$-efficiency level $\text{Eff}_0$, where an allowed loss of efficiency is $1 - \text{Eff}_0$. That is, for all the subjects, the $\Im_{\kappa^*}$ allows the missingness and the measurements not to be taken at the precisely original time points, which still guarantees a certain estimation effects of parameters. It can be regarded as a robust treatment to both missingness and time disturbance. Then $\Im_{\kappa^*}$ is called $R_0$-efficient sampling windows.

Similar to the idea of the MSNTO, we can also apply the uniform designs to search the global minimal $R_0$-efficiency on $\Im_\kappa$. It is also worth noting that we should add constraints to avoid the sampling windows overlapping, i.e. $t_i^* + \kappa > t_{i+1}^*$, since if a measurement is taken at the time point $t_i$ from the sampling window $[t_i^*, t_i^* + \kappa]$ and $t_i > t_{i+1}^*$, it is impossible to take the next measurement at any time point in the interval $[t_{i+1}^*, t_i]$ from the window $[t_{i+1}^*, t_{i+1}^* + \kappa]$. Fortunately, under suitable settings of the parameters $T$, $\ell$ and $\text{Eff}_0$, it can be avoided. Here we give the general steps of the algorithm to search the optimal window width $\kappa^*$ as follows:

Step 1 Using MSNTO to search the optimal design $d_{ro}$ on $D_n$ under $R_0$-criterion and construct the sampling windows $\Im_\kappa$;
Step 2 Set the threshold $\delta$, the iteration step size $\zeta$ and the target efficiency $\text{Eff}_0$. Choose an initial value $\kappa = \kappa_0 (< \ell)$;
Step 3 Calculate the minimal $R_0$-efficiency $\text{eff}_{0\min}$ by constructing a uniform design on $\Im_\kappa$;

Step 4 If $|\text{eff}_{0\min} - \text{Eff}_0| < \delta$, stop, else,

(a) if $\text{eff}_{0\min} - \text{Eff}_0 > \delta$, increase $\kappa$ ($\kappa = \kappa + \zeta$) satisfying $\kappa < \ell$ and go to Step 3,
(b) if $\text{eff}_{0\min} - \text{Eff}_0 < (-\delta)$, decrease $\kappa$ ($\kappa = \kappa - \zeta$) and go to Step 3;

In Step 2, one can randomly choose the initial value $\kappa_0$ from the interval $(0, \ell)$ first and run the algorithm. If the number of iterations is considered to be large, we can adjust $\kappa_0$ according to the change trend of $\kappa$ and rerun the algorithm. Hence, this algorithm can be accelerated in the right direction through a rough exploration of $\kappa_0$. The choice of the target efficiency $\text{Eff}_0$ decides for the width of the windows: larger $\text{Eff}_0$ will lead to narrower windows and smaller $\text{Eff}_0$ will lead to wider windows.

## 5. Simulation studies

We use polynomials of time to model the mean, the innovation variances and the autoregressive coefficients, i.e. the covariates in the mean–covariance models take the forms

$$\boldsymbol{x}_{ij} = (1, t_j, \ldots, t_j^{r-1})', \quad \boldsymbol{z}_j = (1, t_j, \ldots, t_j^{d-1})', \quad \boldsymbol{z}_{jg} = (1, (t_j - t_g), \ldots, (t_j - t_g)^{q-1})'. \tag{6}$$

Certainly, $\boldsymbol{x}_{ij}$ can also take other form, which is decided by practical analysis. To illustrate, we consider the case of nonmonotone missingness. We fix the number of subjects $m = 100$ and the number of the repeated measurements $n = 8$. We also set the maximal time point $T = 15$, the number of the parameters in models (1)$r = d = q = 3$ and the minimal time interval $\ell = 2$. We assume the prior ranges of both $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$ are $[-1, 1] \times [-0.1, 0.1] \times [-0.01, 0.01]$. Any randomly sampled point in this cuboid is uniformly distributed. In each simulation, we assume that each subject shares the same missing probability $p_i^t = p^t$, $i = 1, \ldots, m$ and consider different kinds of it. Then we compare the performance of the optimal design under $R_0$-criterion ($d_{ro}$), the optimal design without considering missingness ($d_o$) and the equidistant design ($d_{equi}$). When $p^t$ is origin-symmetric on $[-T, T]$, we also add the comparison with the symmetric optimal design ($d_{rosym}$) and the symmetrized optimal design ($d_{rosym2}$). The former is obtained based on the symmetric design space and the latter is derived from the optimal design by symmetrization as in Theorem 1.

We first consider two different origin-symmetric forms of $p^t$. For each case, we discuss different values of the parameter in it,

- Case 1 :

$$p^t = a, t \in [-T, T],$$

(i) $a = 0.2$, (ii) $a = 0.4$, (iii) $a = 0.6$, (iv) $a = 0.8$;

- Case 2 :

$$p^t = b\left|\cos\left(\frac{3\pi}{2T}t\right)\right|, t \in [-T, T],$$

(i) $b = 0.2$, (ii) $b = 0.4$, (iii) $b = 0.6$, (iv) $b = 0.8$.

The results are shown in Fig. 1. The four subplots in the first row depict the results of Case 1 and the other four subplots are related to Case 2. The plots of $p^t$'s for the two cases are shown in Fig. 1(a)–(b). The optimal design $d_{ro}^1$ (*resp.* $d_{ro}^2$) and symmetric optimal design $d_{rosym}^1$ (*resp.* $d_{rosym}^2$) are drawn in Fig. 1(c) (*resp.* Fig. 1(d)). The time points at both ends are nearly the same and the time intervals are exactly $\ell$. For each $a$ of Case 1, $d_{rosym}^1$ is almost the same due to the invariance of the missing probability. For Case 2, compared with $d_o$ which does not consider the missingness, $d_{ro}^2$ and $d_{rosym}^2$ both avoid the point with high missing probability and choose the point with acceptable missing probability. For Case 1, the logarithm of the $R_0$-criterion values ($\log R_0$-values) are 11.9661, 11.4982, 10.8610, 9.7373 for $d_{ro}^1$ and 11.9739, 11.5103, 10.8204, 9.6876 for $d_{rosym}^1$ in turn. For Case 2, the $\log R_0$-values are 12.1411, 12.0663, 11.9188, 11.7905 for $d_{ro}^2$ and 12.1450, 12.0796, 11.9953, 11.8868 for $d_{rosym}^2$ in turn. For both cases, it occurs that the $\log R_0$-value of the symmetric optimal design is larger than that of the optimal design. The reason may be that with the same sized uniform design used in MSNTO, the precision of the searching is higher when the searching space is restricted to the symmetric designs. Moreover, we calculate the *RE*'s of $d_{ro}^1$ and $d_{equi}$, $d_{ro}^1$ and $d_o$ for Case 1 and *RE*'s of $d_{ro}^2$ and $d_{equi}$, $d_{ro}^2$ and $d_o$ for Case 2 in Fig. 1(e)–(f). Both of them show that the optimal design is much better than the optimal design without considering missingness and the equidistant design. And the optimal design without considering missingness is still better than the equidistant design in the presence of missingness. Also, we compare the performance of $d_{ro}^1$, $d_{rosym}^1$, $d_{rosym2}^1$ and $d_{ro}^2$, $d_{rosym}^2$, $d_{rosym2}^2$ for the two cases through *RE* in Fig. 1(g)–(h) respectively. $RE(d_{rosym}^1, d_{ro}^1)$ and $RE(d_{rosym}^2, d_{ro}^2)$ are around 1 which implies that the performance of the two kinds of designs is similar. Moreover, $RE(d_{rosym2}^1, d_{ro}^1)$, $RE(d_{rosym2}^1, d_{rosym}^1)$, $RE(d_{rosym2}^2, d_{ro}^2)$ and $RE(d_{rosym2}^2, d_{rosym}^2)$ are always close to or larger than 2. It illustrates that the trick of symmetrization is feasible and effective.

Next, we consider the cases in which $p^t$ increases with the time $t$ strictly. Similarly, we also discuss different settings of the parameters in them, i.e.

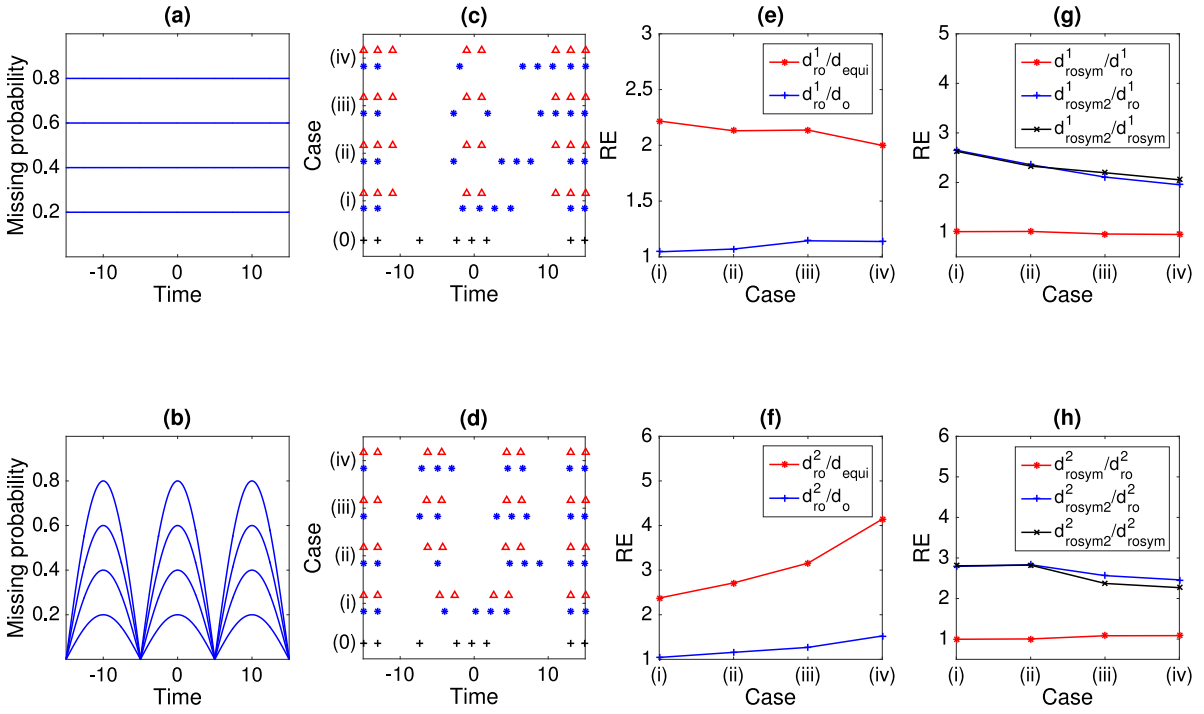


**Fig. 1.** (a)–(b): the $p^t$'s in Case 1 and Case 2; (c)–(d): the optimal designs $d^1_{ro}$, $d^2_{ro}$ with symbol '$*$' and the symmetric optimal designs $d^1_{rosym}$, $d^2_{rosym}$ with symbol '$\triangle$'. Case (0) corresponds to the design $d_o$; (e)–(f): the *RE*'s of $d^1_{ro}$ and $d_{equi}$, $d^1_{ro}$ and $d_o$, $d^2_{ro}$ and $d_{equi}$, $d^2_{ro}$ and $d_o$; (g)–(h): the *RE*'s of any two designs in $d^1_{ro}$, $d^1_{rosym}$, $d^1_{rosym2}$ and $d^2_{ro}$, $d^2_{rosym}$, $d^2_{rosym2}$.

**Table 1**
The gap function $g$ for the selected optimal designs.

| Case | $d^1_{ro}$ | $d^1_{rosym}$ | $d^2_{ro}$ | $d^2_{rosym}$ | $d^3_{ro}$ | $d^4_{ro}$ |
|---|---|---|---|---|---|---|
| (i) | 0.9770 | 0.9786 | 0.9900 | 0.9946 | 0.9215 | 0.9225 |
| (ii) | 0.9272 | 0.9210 | 0.9736 | 0.9640 | 0.8941 | 0.8614 |
| (iii) | 0.8460 | 0.8454 | 0.9493 | 0.9365 | 0.8413 | 0.7969 |
| (iv) | 0.5835 | 0.5868 | 0.9069 | 0.9036 | 0.7949 | 0.7223 |

- Case 3 :

$$p^t = \frac{f-e}{900}t^2 + \frac{f-e}{30}t + \frac{3e+f}{4}, t \in [-T, T],$$

(i) $e = 0.2, f = 0.5$, (ii) $e = 0.2, f = 0.6$, (iii) $e = 0.2, f = 0.7$, (iv) $e = 0.2, f = 0.8$;

- Case 4 :

$$p^t = -\frac{f-e}{900}t^2 + \frac{f-e}{30}t + \frac{e+3f}{4}, t \in [-T, T],$$

(i) $e = 0.2, f = 0.5$, (ii) $e = 0.2, f = 0.6$, (iii) $e = 0.2, f = 0.7$, (iv) $e = 0.2, f = 0.8$;

where $e$ and $f$ are the missing probabilities at the time points $-T$ and $T$, i.e., $p^{-T} = e$ and $p^T = f$.

The results are shown in Fig. 2. Fig. 2(a)–(b) show the plots of $p^t$'s for the two cases. Fig. 2(c)–(d) draw the optimal design $d^3_{ro}$ for Case 3 and $d^4_{ro}$ for Case 4. Comparing with $d_o$, it seems that $d^3_{ro}$ and $d^4_{ro}$ prefer the time points at both ends. However, for both cases, it can be seen that the time points move to the left with the increase of $f$ since they all tend to choose the time point with smaller missing probability and thus guarantee the robustness to missingness. The trend is more obvious for Case 4 since the change is faster when the missing probability is small. The $\log R_0$-values respectively are 11.6229, 11.4827, 11.3156, 11.0844 for Case 3 and 11.4786, 11.2499, 10.9785, 10.6260 for Case 4 in turn. Similarly, the value becomes smaller with the increase of the missing probability for both cases. Finally, the *RE*'s of $d^3_{ro}$ and $d_{equi}$, $d^3_{ro}$ and $d_o$ for Case 3 and *RE*'s of $d^4_{ro}$ and $d_{equi}$, $d^4_{ro}$ and $d_o$ for Case 4 are shown in Fig. 2(e)–(f), they are always larger than 1. It implies that the performance of the optimal designs under $R_0$-criterion is always better than the other two designs.

Further, based on the discussion in Proposition 2, we also calculate the values of the gap function $g(\tau) = R_1(n, \tau, m, \boldsymbol{p})/R_0(n, \tau, m, \boldsymbol{p})$ for the optimal designs in Cases 1–4 where $R_1$-criterion is defined in (5). The results are shown in Table 1.

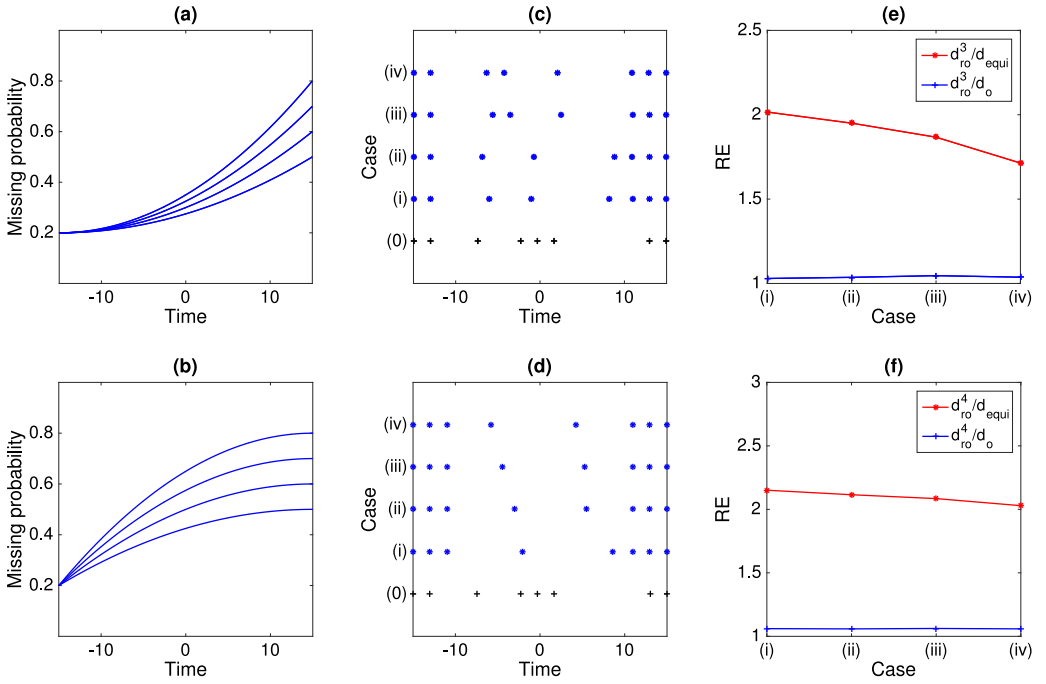


**Fig. 2.** (a)–(b): the $p^t$s in Case 3 and Case 4; (c)–(d): the optimal designs $d_{ro}^3$ and $d_{ro}^4$. Case (0) corresponds to the design $d_o$; (e)–(f): the *RE*s of $d_{ro}^3$ and $d_{equi}$, $d_{ro}^3$ and $d_o$, $d_{ro}^4$ and $d_{equi}$, $d_{ro}^4$ and $d_o$.

**Table 2**
The biases and standard errors of the estimators of parameters based on $d_{ro}^2$, $d_o$, $d_\beta$ and $d_{equi}$ with missing probability as Case 2(iv).

| Design | det_OC | se_bta | se_lamb | se_gam | bia_bta | bia_lamb | bia_gam |
|---|---|---|---|---|---|---|---|
| $d_{ro}^2$ | 2.2482e−38 | 2.0308e−05 | 8.4509e−07 | 3.0674e−06 | 0.0016 | 0.9843 | 0.4915 |
| $d_o$ | 8.5898e−35 | 1.6041e−04 | 1.9769e−06 | 1.0986e−05 | 0.0386 | 2.5183 | 0.4914 |
| $d_\beta$ | 4.7125e−34 | 1.6412e−04 | 1.6480e−06 | 4.0670e−05 | 0.0137 | 2.1905 | 1.0957 |
| $d_{equi}$ | 4.2736e−33 | 9.3022e−05 | 2.7225e−06 | 5.3664e−05 | 0.0078 | 1.9890 | 1.2115 |

For each case, when the missing probability decreases, the gap function increases and thus the optimal design is more effective under $R_1$-criterion. Moreover, we reduce the value of the number of repeated measurement $n$ and select the optimal designs under $R_0$-criterion. We calculate the gap function again based on the optimal designs with reduced $n$. The values of the gap function become larger, e.g. when $n = 5$, most of them are larger than 90%. Hence, when the missing probability is not large or the number of the repeated measurement is small, the optimal design under $R_0$-criterion can be an alternative with good performance under $R_1$-criterion.

In order to investigate the properties of different designs for the parameter estimations of the mean–covariance models, we also conduct some simulation studies. To illustrate, we compare the optimal design $d_{ro}^2$ with missing probability as Case 2(iv) with $d_o$, $d_{equi}$ and $d_\beta$. Here $d_\beta$ is obtained by the previous D-optimality criterion which only considers the mean and does not take the missingness into account. Assume that the true parameters $\boldsymbol{\beta}_t = (50, 0.3, 0.1)'$, $\boldsymbol{\lambda}_t = (0.8268, -0.0675, 0.0006)'$ and $\boldsymbol{\gamma}_t = (0.8896, 0.0066, -0.0054)'$, where $\boldsymbol{\lambda}_t$ and $\boldsymbol{\gamma}_t$ are chosen from the prior samples. We carry out 100 repetitions to generate the observations based on each design. With these generated unbalanced longitudinal data $\boldsymbol{y}_i$, $i = 1, \ldots, m$, we use the R package *jmcm* to estimate the parameters in the mean–covariance models with MLE method and show the results in Table 2, where (i) det_OC denotes for the determinant of the observed covariance matrix of the estimators (the generalized variance); (ii) se_bta, se_lamb and se_gam denote for the products of the elements of the standard errors for the estimators $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\lambda}}$ and $\hat{\boldsymbol{\gamma}}$ respectively; (iii) bia_bta, bia_lamb and bia_gam denote for the summations of the elements of the differences between the true values and the averages for the estimators $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\lambda}}$ and $\hat{\boldsymbol{\gamma}}$ respectively. As shown in Table 2, the optimality of the generalized variance of the estimators based on $d_{ro}^2$ is guaranteed by $R_0$-criterion. All the estimation results based on $d_{ro}^2$ are also smaller than those based on other designs. $d_o$ and $d_\beta$ are still better than $d_{equi}$ in terms of the results about the variation, but in the presence of missingness, the optimal design without considering missingness is not optimal any more. Also, $d_\beta$ cannot derive good estimation even for $\boldsymbol{\beta}$ any more. Thus our proposed $R_0$-criterion is more valid and effective than previous criteria.

Finally, based on the obtained $d_{ro}^4$ with missing probability as Case 4(i)-(iv), we search the optimal window width $\kappa^*$ discussed in Section 4. Set the threshold $\delta = 10^{-3}$ and iteration step size $\zeta = 10^{-2}$. We construct the uniform designs on

**Table 3**
The optimal window width $\kappa^*$ for $d_{ro}^4$ with missing probability as Case 4.

| Case | $\text{Eff}_0$ | | | |
|---|---|---|---|---|
| | 0.95 | 0.90 | 0.85 | 0.80 |
| (i) | 0.42 | 0.90 | 1.41 | 2.00 |
| (ii) | 0.38 | 0.79 | 1.22 | 1.70 |
| (iii) | 0.34 | 0.70 | 1.08 | 1.50 |
| (iv) | 0.30 | 0.61 | 0.95 | 1.31 |

$\Im_\kappa$ by good lattice point method with a power generator (Korobov, 1959) under mixture discrepancy (Zhou et al., 2013), a measure of uniformity of a design. In Table 3, we show the $\kappa^*$ for different values of $\text{Eff}_0$, i.e. $\text{Eff}_0 = 0.95, 0.90, 0.85, 0.80$. In general, whatever the missing probability is, the $\kappa^*$ increases with the decrease of $\text{Eff}_0$. Moreover, given $\text{Eff}_0$, the $\kappa^*$ tends to be smaller with the increase of the missing probability, which implies that larger missing probability leads to faster decline in efficiency. The $\kappa^*$ for a specific $\text{Eff}_0$ can also serve as a guidance for the choice of the initial value $\kappa_0$ with smaller or larger $\text{Eff}_0$ to save the search time.

## 6. Real data illustration

In this section, we apply the proposed $R_0$-criterion to the fruit fly mortality (FFM) data (Zimmerman and Núñez Antón, 2010) for a follow-up study. The FFM data consist of age-specific mortality measurements from 112 cohorts (subjects) of Drosophila melanogaster (a common fruit fly). Mortality measurements were obtained by replicating 56 recombinant inbred lines to get 500 to 1000 fruit flies for each cohort. Everyday, dead flies were counted and removed for each cohort, and these counts were pooled into 11 five-day intervals. The raw mortality rate was recorded as $-\log(N(t+1)/N(t))$, where $N(t)$ is the number of flies in the cohort living at the beginning of time interval $t (t = 1, 2, \ldots, 11)$. To make the responses more normally distributed, these final raw mortality rates were log-transformed. For unknown reasons, approximately 22% of the data were missing nonmonotonically. Specifically, the missing rates (the number of missing cohorts/the number of the total cohorts) at the recorded time points, respectively, were 0.2411, 0.2054, 0.1607, 0.1250, 0.0804, 0.0625, 0.1071, 0.2411, 0.3482, 0.4018, 0.5000. So the missing rate is quite different at different time points. Larger missing rate leads to less accuracy in the subsequent modeling and statistical inference. Thus, if a follow-up study is needed and this missing case is considered to be crucial, the original equidistant time points may be not a desirable design. Instead, we can adopt the robust optimal design under the proposed $R_0$-criterion.

Following Garcia et al. (2012) and Huang et al. (2012), the mean–covariance models are appropriate for analyzing the FFM data and the third-degree polynomials in time are suitable for modeling $\mu_{ij}$, $\log(\sigma_j^2)$ and $\phi_{jg}$, where the forms $\boldsymbol{x}_{ij}$, $\boldsymbol{z}_j$ and $\boldsymbol{z}_{jg}$ are specified by (6) with $r = d = q = 4$. Here we re-analyze the data using the mean–covariance models with the original times scaled to $-5, -4, \ldots, 4, 5$. The MLEs are $\hat{\boldsymbol{\theta}}_0 = (\hat{\boldsymbol{\beta}}_0', \hat{\boldsymbol{\lambda}}_0', \hat{\boldsymbol{\gamma}}_0')'$, where $\hat{\boldsymbol{\beta}}_0 = (-1.6700, 0.6313, -0.0152, -0.0122)'$, $\hat{\boldsymbol{\lambda}}_0 = (-0.3286, -0.3135, -2.3112\text{e}{-4}, 8.6461\text{e}{-3})'$ and $\hat{\boldsymbol{\gamma}}_0 = (1.1321, -0.7265, 0.1401, -8.2571\text{e}{-3})'$. Based on the MLEs, we construct the prior space $\Phi$ with $\boldsymbol{\lambda} \in [-0.4, 0.4] \times [-0.4, 0.4] \times \{-2.3112\text{e}{-4}\} \times \{8.6461\text{e}{-3}\}$, $\boldsymbol{\gamma} \in [-1.3, 1.3] \times [-1, 1] \times [-0.2, 0.2] \times \{-8.2571\text{e}{-3}\}$, and any randomly sampled prior is independently and uniformly distributed. For the robust optimal design selection, we approximate the missing probabilities by fitting the missing rates with a quadratic polynomial, i.e., $p^t = 0.0114t^2 + 0.0264t + 0.1113$. Let $n = 11$, $m = 100$, $\ell = 0.3$ and $T = 5$. Similar to Section 5, we also compare the performance of the optimal design under $R_0$-criterion ($d_{ro}^{\text{FFM}}$), the optimal design without considering missingness ($d_o^{\text{FFM}}$), the optimal design which only considers the mean and does not take the missingness into account ($d_\beta^{\text{FFM}}$) and the equidistant design ($d_{equi}^{\text{FFM}}$), by $RE$ and the parameter estimation for demonstrating the validity of the $R_0$-criterion. Those optimal designs are all found by the MSNTO algorithm.

In Fig. 3, we show the approximate missing probability and the diagrams for $d_{ro}^{\text{FFM}}$, $d_o^{\text{FFM}}$, $d_\beta^{\text{FFM}}$ and $d_{equi}^{\text{FFM}}$. It can be seen that $d_{ro}^{\text{FFM}}$ yields the strongest sensitivity to avoid the locations with high missing probability. Similar to Figs. 1 and 2, those selected designs are inclined to contain the endpoints, even though the missing probabilities are relatively large at those points. The $RE$'s between $d_{ro}^{\text{FFM}}$ and other three designs are $RE(d_{ro}^{\text{FFM}}, d_o^{\text{FFM}}) = 1.1382$, $RE(d_{ro}^{\text{FFM}}, d_\beta^{\text{FFM}}) = 1.1989$ and $RE(d_{ro}^{\text{FFM}}, d_{equi}^{\text{FFM}}) = 2.9912$, respectively. It shows that the optimal design under the $R_0$-criterion $d_{ro}^{\text{FFM}}$ has the best performance and the equidistant design $d_{equi}^{\text{FFM}}$ shows the worst performance. Let the threshold $\delta = 0.0015$ and the iteration step size $\zeta = 0.005$. To deal with the operational errors, by the algorithm in Section 4, we also obtain the optimal windows $\kappa^* = 0.05, 0.085, 0.12, 0.165$ for $\text{Eff}_0 = 0.97, 0.95, 0.93, 0.91$, respectively. Assume that the true parameters $\boldsymbol{\theta}_t = (\boldsymbol{\beta}_t', \boldsymbol{\lambda}_t', \boldsymbol{\gamma}_t')' = (\hat{\boldsymbol{\beta}}_0', \hat{\boldsymbol{\lambda}}_0', \hat{\boldsymbol{\gamma}}_0')'$. The true mean and true covariance matrix are obtained by $\boldsymbol{\mu}_t = \boldsymbol{X}\boldsymbol{\beta}_t$ and $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_t(\boldsymbol{\theta}_t) = \boldsymbol{T}_t^{-1}\boldsymbol{D}_t(\boldsymbol{T}_t')^{-1}$. Once the model structure and design are given, the true mean and covariance matrix are also determined. Then, we carry out 100 repetitions to generate the observations based on each design, where the missing probability is embedded in the data generation. With the unbalanced data, we also use the R package *jmcm* to estimate the parameters in the mean–covariance models and the results are shown in Table 4. In terms of the generalized variance of the estimators, $d_{ro}^{\text{FFM}}$ gives the best estimations for these parameters. Except for the estimation of $\boldsymbol{\beta}$, all the other results based on $d_{ro}^{\text{FFM}}$ are better than those based on other designs. Hence, the data illustration also shows that in the presence of missingness, the robust optimal design based on the $R_0$-criterion outperforms those designs which do not take the missingness into account.

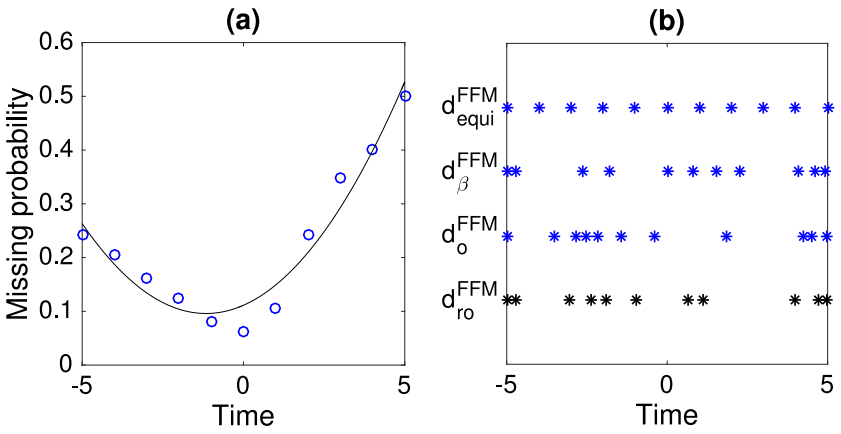


**Fig. 3.** (a): the curve depicts the missing probability and 'o' denotes for the missing rate; (b): the diagrams for $d^{FFM}_{ro}$, $d^{FFM}_{o}$, $d^{FFM}_{\beta}$ and $d^{FFM}_{equi}$.

**Table 4**
The biases and standard errors of the estimators of parameters based on $d^{FFM}_{ro}$, $d^{FFM}_{o}$, $d^{FFM}_{\beta}$ and $d^{FFM}_{equi}$. The evaluation indexes are the same as those in Table 2.

| Design | det_OC | se_bta | se_lamb | se_gam | bia_bta | bia_lamb | bia_gam |
|---|---|---|---|---|---|---|---|
| $d^{FFM}_{ro}$ | 3.5475e−54 | 6.5910e−08 | 4.7577e−08 | 1.2015e−08 | 0.0349 | 0.1197 | 0.0661 |
| $d^{FFM}_{o}$ | 2.9408e−53 | 4.3243e−08 | 2.3587e−07 | 1.6311e−08 | 0.0144 | 0.6509 | 0.2165 |
| $d^{FFM}_{\beta}$ | 3.8769e−52 | 4.7594e−08 | 6.6021e−08 | 4.9915e−08 | 0.0245 | 0.1503 | 0.0846 |
| $d^{FFM}_{equi}$ | 3.1317e−51 | 8.5523e−08 | 7.9240e−08 | 1.2273e−07 | 0.0305 | 0.2748 | 0.4789 |

## 7. Conclusion and discussion

In this paper, we propose a novel criterion to select the optimal designs for the mean–covariance models which is robust against the missingness, parameter uncertainty and operational errors. We take the possible cases of missingness into account so that even though not every measurement can be obtained, the mean and the covariance matrix can still be estimated as accurately as possible. We prove that the trick of symmetrization can bring better performance under the proposed criterion if the missing probability is origin-symmetric. We also discuss the sampling windows to guarantee a certain efficiency in the presence of operational errors. We conduct some simulation studies and a real data analysis based on the FFM data which consider different missing probabilities. The time points always avoid the locations with high missing probability and select those with acceptable one instead. In terms of the relative efficiency, the optimal designs under the proposed criterion perform better than the optimal designs without considering missingness and the equidistant designs. When the missing probability is origin-symmetric, the symmetrized optimal designs perform the best and the symmetric optimal designs are similar to the general optimal designs. That is, symmetric designs can not only reduce the complexity of the algorithm, but also bring good performance. Moreover, we also compare the estimation effects of different designs. It is shown that the results based on the optimal design are more accurate in terms of the biases and the standard errors of the estimators. In the presence of missingness, our proposed criterion is more effective compared with the previous criteria which just consider the mean and (/or) do not take the missingness into account.

Additionally, in the literature, there are several more valid and more efficient modeling methods to handle the missing data, such as multiple imputation and other likelihood-based methods. But the corresponding asymptotic covariance matrix of the estimates of the parameters would be more complicated and it may be more difficult to deal with. For example, for some likelihood-based methods, the EM algorithm may be used to estimate the parameters, in which the asymptotic covariance matrix of the estimators of the parameters, i.e. the inverse of the observed information matrix, is more complex and may rely on the values of the estimators. It is beyond the scope of the current paper but worthy for further investigations.

## Acknowledgments

Dr. Zhou's work was supported by the National Natural Science Foundation of China (11871288) and the Natural Science Foundation of Tianjin, China (19JCZDJC31100). Dr. Zheng's work was supported by the National Science Foundation, United States (DMS-1830864).

## Appendix A

**Proof of Proposition 1.** Without loss of generality, we consider that for all subjects, the model matrices for the mean vector are all the same only containing polynomials of time. Consider two designs $\tau$ and $\tilde{\tau}$. $\Sigma$ $(= \boldsymbol{T}^{-1}\boldsymbol{D}(\boldsymbol{T}')^{-1})$, $\boldsymbol{H}_i$, $\boldsymbol{W}_i^H$, $\boldsymbol{X}$, $\boldsymbol{Z}$, $\boldsymbol{z}_{jg}$ correspond to $\tau$ and $\tilde{\Sigma}$ $(= \tilde{\boldsymbol{T}}^{-1}\tilde{\boldsymbol{D}}(\tilde{\boldsymbol{T}}')^{-1})$, $\tilde{\boldsymbol{H}}_i$, $\tilde{\boldsymbol{W}}_i^H$, $\tilde{\boldsymbol{X}}$, $\tilde{\boldsymbol{Z}}$, $\tilde{\boldsymbol{z}}_{jg}$ correspond to $\tilde{\tau}$. The entries of $\boldsymbol{D}$ and $\tilde{\boldsymbol{D}}$ are denoted by $\sigma_j^2$ and $\tilde{\sigma}_j^2$ respectively, which satisfy $\log\sigma_j^2 = \boldsymbol{z}_j'\boldsymbol{\lambda}_1$ and $\log\tilde{\sigma}_j^2 = \tilde{\boldsymbol{z}}_j'\boldsymbol{\lambda}_2$. The entries of $\boldsymbol{T}$ and $\tilde{\boldsymbol{T}}$ are denoted by $\phi_{jg}$ and $\tilde{\phi}_{jg}$ respectively, which satisfy $\phi_{jg} = \boldsymbol{z}_{jg}'\boldsymbol{\gamma}_1$ and $\tilde{\phi}_{jg} = \tilde{\boldsymbol{z}}_{jg}'\boldsymbol{\gamma}_2$. If the designs are mutually symmetric, i.e. $\tilde{\tau} = -\tau$, we have $\tilde{\boldsymbol{z}}_j = \boldsymbol{A}_1\boldsymbol{z}_j$ and $\tilde{\boldsymbol{z}}_{jg} = \boldsymbol{A}_2\boldsymbol{z}_{jg}$. Then under the condition $\boldsymbol{\lambda}_2 = \boldsymbol{A}_1\boldsymbol{\lambda}_1$ and $\boldsymbol{\gamma}_2 = \boldsymbol{A}_2\boldsymbol{\gamma}_1$, we have $\log\tilde{\sigma}_j^2 = \log\sigma_j^2$, $\tilde{\phi}_{jg} = \phi_{jg}$ and thus $\tilde{\Sigma} = \Sigma$. Let $\boldsymbol{A}_3$ be a $r \times r$ diagonal matrix with odd diagonal entries being 1 and even being -1. We also have $\tilde{\Sigma}^{(-\omega)} = \Sigma^{(-\omega)}$, $\tilde{\boldsymbol{X}}_{(-\omega)} = \boldsymbol{X}_{(-\omega)}\boldsymbol{A}_3$, $\tilde{\boldsymbol{Z}}_{(-\omega)} = \boldsymbol{Z}_{(-\omega)}\boldsymbol{A}_1$ for any $\omega \subseteq \Omega$. Further, if $p_i^t$ is origin-symmetric for any subject $i$, $i = 1, \ldots, m$, then for both monotone and nonmonotone missingness,

$$\sum_{i=1}^{m} E(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{H}}_i\tilde{\Sigma}^{-1}\tilde{\boldsymbol{H}}_i\tilde{\boldsymbol{X}}_i) = \boldsymbol{A}_3\sum_{i=1}^{m} E(\boldsymbol{X}_i'\boldsymbol{H}_i\Sigma^{-1}\boldsymbol{H}_i\boldsymbol{X}_i)\boldsymbol{A}_3,$$

$$\sum_{i=1}^{m} E(\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{H}}_i^2\tilde{\boldsymbol{Z}}) = \boldsymbol{A}_1\sum_{i=1}^{m} E(\boldsymbol{Z}'\boldsymbol{H}_i^2\boldsymbol{Z})\boldsymbol{A}_1, \quad \sum_{i=1}^{m} E(\tilde{\boldsymbol{W}}_i^H) = \boldsymbol{A}_2\sum_{i=1}^{m} E(\boldsymbol{W}_i^H)\boldsymbol{A}_2.$$

It follows immediately that Eq. (4) holds. If the model matrices for the mean vector also contain other baseline covariates and the interaction terms with time. According to the assumption, for each subject, the values of these baseline covariates under $\tau$ equal those under $\tilde{\tau}$. Then the corresponding $\boldsymbol{A}_3$ is still a diagonal matrix with entries as 1 or -1. It is easy to show that Eq. (4) also holds with similar arguments and we omit it here.

**Proof of Theorem 1.** Let $\tau \in \Delta$. It follows from Proposition 1 that $|E(I_{\boldsymbol{\theta}}^H(n, \tau, m, \boldsymbol{\alpha}_1))|$ is equal to $|E(I_{\boldsymbol{\theta}}^H(n, \tilde{\tau}, m, \boldsymbol{\alpha}_2))|$. Thus, $\sum_{\boldsymbol{\alpha}\in\Phi} |E(I_{\boldsymbol{\theta}}^H(n, \tau, m, \boldsymbol{\alpha}))|^{\frac{1}{r+d+q}} = \sum_{\boldsymbol{\alpha}\in\Phi} |E(I_{\boldsymbol{\theta}}^H(n, \tilde{\tau}, m, \boldsymbol{\alpha}))|^{\frac{1}{r+d+q}}$. Consider the design $\xi = \{(\tau, \frac{1}{2}), (\tilde{\tau}, \frac{1}{2})\}$ for which half of the subjects are measured according to $\tau$ and half of the subjects are measured according to $\tilde{\tau}$. It is obvious that $\xi$ is symmetric. Then due to the fact that

$$|\boldsymbol{G} + \boldsymbol{K}|^{\frac{1}{s}} \geq |\boldsymbol{G}|^{\frac{1}{s}} + |\boldsymbol{K}|^{\frac{1}{s}}$$

where $\boldsymbol{G}$ and $\boldsymbol{K}$ are positive definite matrices of order $s$, we have

$$\begin{aligned}
\sum_{\boldsymbol{\alpha}\in\Phi} |E(I_{\boldsymbol{\theta}}^H(n, \xi, m, \boldsymbol{\alpha}))|^{\frac{1}{r+d+q}} &= \sum_{\boldsymbol{\alpha}\in\Phi} \left|\frac{1}{2}E(I_{\boldsymbol{\theta}}^H(n, \tau, m, \boldsymbol{\alpha})) + \frac{1}{2}E(I_{\boldsymbol{\theta}}^H(n, \tilde{\tau}, m, \boldsymbol{\alpha}))\right|^{\frac{1}{r+d+q}} \\
&\geq \frac{1}{2}\sum_{\boldsymbol{\alpha}\in\Phi} |E(I_{\boldsymbol{\theta}}^H(n, \tau, m, \boldsymbol{\alpha}))|^{\frac{1}{r+d+q}} + \frac{1}{2}\sum_{\boldsymbol{\alpha}\in\Phi} |E(I_{\boldsymbol{\theta}}^H(n, \tilde{\tau}, m, \boldsymbol{\alpha}))|^{\frac{1}{r+d+q}} \\
&= \sum_{\boldsymbol{\alpha}\in\Phi} |E(I_{\theta}^H(n, \tau, m, \boldsymbol{\alpha}))|^{\frac{1}{r+d+q}}.
\end{aligned}$$

## Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jspi.2021.12.004.

## References


Alrweili, H., Georgiou, S., Stylianou, S., 2019. Robustness of response surface designs to missing data. Qual. Reliab. Eng. Int. 35, 1288–1296.
Andrews, D.F., Herzberg, A.M., 1979. The robustness and optimality of response surface designs. J. Statist. Plann. Inference 3, 249–257.
Bogacka, B., Johnson, P., Jones, B., Volkov, O., 2008. D-efficient window experimental designs. J. Statist. Plann. Inference 138, 160–168.
Fang, K.T., Liu, M.Q., Qin, H., Zhou, Y.D., 2018. Theory and Application of Uniform Experimental Designs. Springer, Singapore.
Fang, K.T., Wang, Y., 1994. Number-Theoretic Methods in Statistics. Chapman and Hall, London.
Garcia, T.P., Kohli, P., M., Pourahmadi., 2012. Regressograms and mean-covariance models for incomplete longitudinal data.. Amer. Statist. 66, 85–91.
Graham, G., Aarons, L., 2006. Optimum blood sampling time windows for parameter estimation in population pharmacokinetic experiments. Stat. Med. 25, 4004–4019.
Herzberg, A.M., Andrews, D.F., 1976. Some considerations in the optimal design of experiments in non-optimal situations. J. R. Stat. Soc. Ser. B Stat. Methodol. 38, 284–289.
Huang, J.H., Chen, M., Maadooliat, M., Pourahmadi, M., 2012. A cautionary note on generalized linear models for covariance of unbalanced longitudinal data. J. Statist. Plann. Inference 142, 743–751.
Imhof, L.A., Song, D., Wong, W.K., 2004. Optimal design of experiments with anticipated pattern of missing observations. J. Theoret. Biol. 228, 251–260.
Korobov, N.M., 1959. The approximate computation of multiple integrals. Dokl. Akad. Nauk SSSR 124, 1207–1210.
Leng, C.L., Zhang, W.P., Pan, J.X., 2010. Semiparametric mean-covariance regression analysis for longitudinal data. J. Amer. Statist. Assoc. 105, 181–193.
Little, R.J.A., Rubin, D.B., 1994. Statistical Analysis with Missing Data, Second ed. Wiley, New York.

Ouwens, M.J.N.M., Tan, F.E.S., Berger, M.P.F., 2002. Maximin D-optimal designs for longitudinal mixed effects models. Biometrics 58, 735–741.
Pan, J.X., MacKenzie, G., 2006. Regression models for covariance structures in longitudinal studies. Stat. Model. 6, 43–57.
Pourahmadi, M., 1999. Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. Biometrika 86, 677–690.
Pourahmadi, M., 2000. Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. Biometrika 87, 425–435.
Rubin, D.B., 1976. Inference and missing data. Biometrika 63, 581–592.
Tekle, F.B., Tan, F.E.S., Berger, M.P.F., 2008. D-optimal cohort designs for linear mixed-effects models. Stat. Med. 27, 2586–2600.
Xu, L., Xiang, S.J., Yao, W.X., 2019. Robust maximum $L_q$-likelihood estimation of joint mean-covariance models for longitudinal data. J. Multivariate Anal. 171, 397–411.
Ye, H.J., Pan, J.X., 2006. Modelling of covariance structures in generalised estimating equations for longitudinal data. Biometrika 93, 927–941.
Yi, S.Y., Zhou, Y.D., Pan, J.X., 2021. D-optimal designs of mean-covariance models for longitudinal data. Biom. J. 63, 1072–1085.
Zheng, W., 2013. Universally optimal crossover designs under subject dropout. Ann. Statist. 1, 63–90.
Zhou, Y.D., Fang, K.T., Ning, J.H., 2013. Mixture discrepancy for quasi-random point sets. J. Complexity 29, 283–301.
Zhou, X.D., Wang, Y.J., Yue, R.X., 2021. Optimal designs for discrete-time survival models with random effects. Lifetime Data Anal. 27, 300–332.
Zimmerman, D.L., Núñez Antón, V., 2010. Antedependence Models for Longitudinal Data. Chapman and Hall, London.