9-17-2004

Optimal Sampling Times in Bioequivalence Studies Using a Simulated Annealing Algorithm

Leena Choi
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics, lchoi@jhsph.edu

Brian Caffo
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics, bcaffo@jhsph.edu

Charles Rohde
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics, crohde@jhsph.edu

Optimal Sampling Times in Bioequivalence Studies Using a Simulated Annealing Algorithm

Leena Choi, Brian Caffo, and Charles Rohde*

September 17, 2004

Abstract

In pharmacokinetic (PK) studies, blood samples are taken over time on subjects after the administration of a drug to measure the time-course of the plasma drug concentrations. In bioequivalence studies, the trapezoidal rule on the sampled time points is often used to estimate the area under the plasma concentration-time curve, a quantity of principle interest. This manuscript investigates the choice of sampling time points to estimate the area under the curve. In particular, we explore the relative merits of several objective functions, those functions which are minimized with respect to the sampling times to obtain an optimal study design. We propose an objective function which overcomes some of the deficits of existing choices. We also present a simulated annealing algorithm to perform the minimization. The main benefits of the simulated annealing algorithm are the ease in which it can handle constraints on the sampling schedules and its ability to accommodate a variety of models and objective functions. The manuscript presents optimal sampling times for some key examples of true underlying models.

1 Introduction

In pharmacokinetic (PK) studies, blood samples are taken over time on subjects after the administration of a drug to measure the time-course of the plasma drug concentrations. A variety of PK parameters are then estimated using the observed drug concentration profile for each subject. The area under the (plasma/serum/blood) concentration time-curve (called AUC) is an important parameter, since it is known to be a measure of the amount of drug absorbed, a quantity of special interest in bioavailability (BA) and bioequivalence (BE) studies. The bioavailability of a drug is quantified using AUC, the maximum drug concentration (called C_{max}) and the associated time

*Department of Biostatistics,Bloomberg School of Public Health, Johns Hopkins University, lchoi@jhsph.edu

where maximum concentration is reached (called t_{max}). Bioequivalence studies are performed as part of the process of getting approval for a new drug formulation or a generic drug. The goal of BE studies is to ascertain whether or not the bioavailability of the new formulation and an existing approved formulation are equivalent. Here, AUC is used as the primary metric with C_{max} and t_{max} as secondary metrics. Therefore, accurately estimating AUC is an important aspect of the drug development and evaluation process. In this manuscript, we explore the choice of the placement of sampling times to better estimate AUC in BE studies.

Westlake (1979) summarized the selction of sampling times in BE studies with: "no universal rule is apparent and a pragmatic approach is usually taken". After a single dose of administration, a rule of thumb is that blood samples are drawn at several times during the absorption phase of the drug, then several times near the peak and at relatively fewer times in the elimination phase. Usually, 10-15 total sampling times are employed. For example, for a drug with a half-life of 4-5 hours, a typical sampling schedule might be: 0, .5, 1, 1.5, 2, 3, 4, 6, 8, 10, 12, 15 and 24 hours following administration (Westlake, 1979). A diagramatic example of the estimated AUC using 11 sampling times for a PK model with two sets of PK parameter is shown in Figure 1.

Before continuing to the specific problem of sampling time selection in BE studies, we emphasize the distinction between our focus and related areas of choosing pharmacokinetic sampling time points. For example, consider the distinction between choosing sampling times for population and individual PK studies on human subjects. Specifically, population PK studies focus on investigating the population characteristics of PK parameters using a modelling approach. In these studies, it is common to use sparse sampling times within each subject. Thus, it is critical to reduce the bias and variance in estimating the PK model parameters. In contrast, as the name would suggest, individual PK studies focus on individual PK characteristics. Some examples of individual PK studies include preclinical PK studies, BA and BE studies. Unlike population PK studies, individual PK studies employ many sampling times for moderate or small number of subjects. In preclinical PK studies, pharmacologists are often interested in estimating individual PK model parameters using the PK modelling approach. In BA and BE studies, however, nonparametric estimation of AUC is the principal goal. It follows then, that methodologies for choosing sampling time points are likely not the same in these areas. Furthermore, it should be noted that the design of toxicokinetics/pharmacokinetics studies on animal subjects would also be different still from any of the designs of PK studies on human subjects if "destructive sampling" is used. That is, replicates of animals may be assigned at each time point, but sampled only once often after sacrificing them. A review methodology for study design in toxicokinetics/pharmacokinetics can be found in Beatty and Piegorsch (1997).

Here we review some of the related research on sampling time selection in PK studies. For the design of population PK studies, the performance of several design strategies was evaluated using simulation (Al-Banna et al., 1990). Wang and Endrenyi (1992) proposed a simulation approach

which they claim is more efficient than standard simulation approaches. Recently, several authors have explored analytic solutions (Mentré et al., 1997; Tod et al., 1998; Retout et al., 2001). These solutions are usually based on some variant of D-optimality (introduced in Box and Lucas, 1959) where a design is said to be D-optimal if the resulting estimate minimizes the determinant of the inverse of the Fisher information matrix.

With regard to individual PK studies, several approaches have been proposed to find optimal sampling time points. Atkinson et al. (1993) proposed an optimization method using a D-optimality criteria for BA parameters such as AUC, C_{max} and t_{max}. Katz and D'Argenio (1983) proposed an algorithm which used the mean squared error (MSE) of the AUC estimate as an objective function for finding optimal sampling times. Here, an objective function is a summary function of the underlying true concentration time-curve and sampling times which is minimized with respect to the sampling times to select an optimal placement of time points. Katz and D'Argenio (1983) used quasi-Newton methods implemented in FORTRAN to minimize their objective function. Wang (2001) used MSE for estimating the concentration time-curve using linear interpolation instead of MSE of the AUC estimate. Kong and Gonin (2000) also based their objective function on the error in linear interpolation, however using the squared bias as a summary rather than MSE. They also accounted for between-subject (population level) heterogeneity by numerically integrating their objective function over an assumed random effect distribution. Their solution was implemented using a sequential quadratic programming algorithm.

Simulated annealing (SA) has been used sparingly in this area with some notable exceptions. For example, Jones and Wang (1999) applied a simulated annealing algorithm to find starting values, which are then used within a more locally efficient algorithm under a D-optimality criterion. Duffull et al. (2002) used a simulated annealing algorithm for finding sampling time points in population PK setting, also using a D-optimality criterion. Furthermore, they critically compared simulated annealing with several other optimization algorithms.

In this manuscript we suggest a new objective function as well as a simulated annealing algorithm to perform the minimization for choosing sampling time points to estimate AUC in BE studies. The proposed objective function is conceptually based on using the squared bias in estimating AUC. We note that a correction is necessary since the use of the direct estimate of bias in optimization algorithms often results in convergence to sampling time points with little face validity. To avoid this defect, we propose a simple correction that divides up the time space. As a result of this correction, our objective function finds optimal sampling time points taking into account not only the error in AUC estimate, but also the error in linear interpolation. As in Kong and Gonin (2000) we employ both single subject and population averaged versions of this objective function. We use Monte Carlo integration within our optimization algorithm to simplify calculations. Finally, we propose a simulated annealing algorithm to perform the optimization. Our choice of using simulated annealing was largely dominated by SA's ability to

handle constraints on the sampling schedules and its ability to accommodate a variety of models and objective functions.

The remainder of the paper is outlined as follows. In Section 2 we set up the problem and notation while in Section 3 we review objective functions and propose a refinement of using the squared bias of AUC estimate. Section 4 describes the proposed simulated annealing algorithm. In Section 5, we compare the estimated sampling times using our proposed method with those from other methods and also show potential problems depending on the models. A summary and discussion follows in Section 6.

2 Estimating AUC

To better conceptualize estimating the AUC, we define a general PK model. Let $f(t;\boldsymbol{\beta})$ denote the underlying blood concentration for a subject at time t after the administration of a drug for parameters $\boldsymbol{\beta}$. Suppose that a given subject is sampled at time points t_j for $j = 0, \ldots, m+1$. We assume that the responses, y_j, for this subject satisfy

$$y_j = f(t_j;\boldsymbol{\beta}) + e_j, \qquad \boldsymbol{e} \mid \boldsymbol{\beta} \sim \{\mathbf{0}, \boldsymbol{R}(\boldsymbol{\beta};\boldsymbol{\lambda},\sigma)\} \qquad \boldsymbol{\beta} \sim \{\boldsymbol{\mu}, \boldsymbol{D}(\boldsymbol{\xi})\},$$

where the $\boldsymbol{e}$ is a vector comprised of the e_j and the notation "$\sim \{a, b\}$" is short-hand for "distributed with mean a and covariance b". Here $\boldsymbol{\beta}$ is a vector of subject-specific PK parameters, $\boldsymbol{R}(\boldsymbol{\beta};\boldsymbol{\lambda},\sigma)$ represents the within-subject covariance matrix for $\boldsymbol{e}$ and $\boldsymbol{D}(\boldsymbol{\xi})$ is the between-subject covariance matrix for $\boldsymbol{\beta}$. We assume that $\boldsymbol{R}(\boldsymbol{\beta};\boldsymbol{\lambda},\sigma)$ is a diagonal matrix with entries $\sigma_j^2 \equiv \lambda_1 + \sigma^2 f(t_j;\boldsymbol{\beta})^{\lambda_2}$ where $\lambda_i > 0$.

The primary parameter of interest is the area under the concentration time-curve for a particular subject,

$$A_{t_0}^{\infty} = \int_0^{\infty} f(t;\boldsymbol{\beta})dt.$$

Note that we are ommiting the dependence of $A_{t_0}^{\infty}$ on $\boldsymbol{\beta}$ and recall that this model refers to a specific subject. Of course, in practice data is collected on several subjects where the above model holds assuming conditional independence given the seperate $\boldsymbol{\beta}$ for each individual while the parameters $\boldsymbol{\lambda}$, σ, $\boldsymbol{\mu}$ and $\boldsymbol{\xi}$ would be held fixed across subjects.

2.1 Approximation of AUC

There are two major approaches for estimating the AUC in PK studies: model based and non-parametric. Model based estimation involves specifying a PK model for the observed data. A common choice is to assume a compartmental model for $f(t,\boldsymbol{\beta})$, Gaussian errors, a diagonal matrix for $\boldsymbol{R}(\boldsymbol{\beta};\boldsymbol{\lambda},\sigma)$, and a Gaussian random effect distribution. In model based estimation, one

first estimates subject-specific PK parameters (using empirical Bayes methods, for example), then integrates a plug-in estimate of the concentration time-curve. We point the reader to Davidian and Giltinan (1995) for more information on PK model based approaches.

To avoid the problems of model misspecification, nonparametric approaches are more commonly used to estimate AUC in individual PK studies than parametric PK modelling approaches, and nonparametric approaches are also the recommended method by the FDA. In what follows, we investigate the optimal selection of sampling time points using the "trapezoidal rule" to estimate AUC.

The trapezoidal rule is a numerical quadrature rule that is the most common nonparametric estimate of the area under the drug concentration-time curve. This choice is well founded because the trapezoidal rule has been shown to yield a good approximation to AUC when $f(t;\boldsymbol{\beta})$ is decreasing exponentially (Piegorsch and Bailer, 1989; Schempp, 1981), which is the case when the true concentration time-curve follows one of the basic compartmental PK models. Bailer and Piegorsch (1990) compared several quadrature methods for approximating AUC using MSE as an objective function. They found that the trapezoidal rule had a relatively smaller variance. It is worth noting that Bailer and Piegorsch (1990) showed the variance dominates in MSE criteria unless the measurement error is very small.

Given the total number of sampling time points, the accuracy of the trapezoidal approximation of the AUC depends on the placement of the time points. We denote the AUC between time t_0 to t_{m+1} for a subject as

$$A_{t_0}^{t_{m+1}} \equiv \int_{t_0}^{t_{m+1}} f(t;\boldsymbol{\beta})dt.$$

The trapezoidal approximation to $A_{t_0}^{t_{m+1}}$ is

$$\hat{A}_{t_0}^{t_{m+1}} = \sum_{j=1}^{m+1} \hat{A}_{t_{j-1}}^{t_j} = \frac{1}{2}\sum_{j=1}^{m+1} (y_j + y_{j-1})\,(t_j - t_{j-1})\,.$$

The trapezoidal approximation defines the integral only between the first and last sampling time points, t_0 and t_{m+1}. In most PK studies, these times are fixed prior to the study with t_0 being at the before administration of the drug and t_{m+1} being late enough to ensure the accuracy of the approximation; usually, t_{m+1} is taken to be greater than 3 terminal half-lifes. The AUC after the last sampling time point is typically approximated by $\hat{A}_{t_{m+1}}^{\infty} = y_{m+1}/k$ where k is the subject specific elimination rate constant. The elimination rate constant can be either estimated using a PK modelling approach or estimated using the last three sampling points of the current data. Then we define

$$\hat{A}_{t_0}^{\infty} = \hat{A}_{t_0}^{t_{m+1}} + \hat{A}_{t_{m+1}}^{\infty}.$$

Conventional wisdom suggests that to otain an accurate estimates of AUC, $\hat{A}_{t_{m+1}}^{\infty}/\hat{A}_{t_0}^{\infty}$ should be less than 3% for all subjects. However, it should be noted that the last sampling time cannot

always be chosen to satisfy this criteria as it is often fixed to 24 hours in BE studies, since it is difficult to keep subjects beyond that time (Kong and Gonin, 2000). In our current study, we focus only on accurately approximating the AUC up to the last sampling point under the constraint that t_0 and t_{m+1} are fixed.

2.2 Approximation of the underlying function

Another way to visualize the trapezoidal rule approximation of AUC comes from approximating the underlying function using linear interpolation (connecting adjacent points). That is, the trapezoidal rule is the integral of the estimate of the concentration time-curve obtained by interpolation. This interpretation is useful, as it suggests that objective functions for assessing the error in AUC can be based on the error of the linear interpolation approximation.

The linear interpolated function $\hat{f}_j(t)$ from t_{j-1} to t_j can be expressed as:

$$\hat{f}_j(t) = a_j + b_j t, \quad \text{for} \quad t_{j-1} \leq t \leq t_j, \tag{1}$$

where

$$a_j = -\frac{y_j t_{j-1} - y_{j-1} t_j}{t_j - t_{j-1}}, \tag{2}$$

$$b_j = \frac{y_j - y_{j-1}}{t_i - t_{j-1}}, \tag{3}$$

and the integration of $\hat{f}_j(t)$ over time interval results in:

$$\hat{A}_{t_0}^{t_{m+1}} = \sum_{j=1}^{m+1} \hat{A}_{t_{j-1}}^{t_j} = \sum_{j=1}^{m+1} \int_{t_{j-1}}^{t_j} \hat{f}_j(t) dt.$$

3 The Objective Functions

Assuming that the true concentration-time curve $f(t; \boldsymbol{\beta})$ is known and t_0 and t_{m+1} are fixed, in this section we explore objective functions which will be minimized with respect to the remaining m time points, $T = (t_1, \ldots, t_m)$. We distinguish between selection of optimal sampling time points for a single subject, with a fixed value of $\boldsymbol{\beta}$, versus choosing sampling time points for a population of $\boldsymbol{\beta}$ values. In the latter case, we average the single subject objective functions over an assumed random effect distribution.

Two major techniques for defining the objective functions have been discussed in the literature. The first uses MSE for estimating AUC while the second uses MSE for estimating the concentration-time curve using linear interpolation. We summarize three realizations of these techniques:

1. Katz and D'Argenio (1983) proposed the following MSE based objective function for estimating AUC:

$$\begin{aligned} Obj_1(t;\boldsymbol{\beta},\boldsymbol{\lambda},\sigma) &= \mathsf{E}\left\{\sum_{j=1}^{m+1}\int_{t_{j-1}}^{t_j}\left[\hat{f}_j(t)-f(t;\boldsymbol{\beta})\right]dt\right\}^2 \\ &= [\mathsf{Bias}]_1^2+[\mathsf{Var}]_1 \\ &= \left\{\mathsf{E}\left(\hat{A}_{t_0}^{t_{m+1}}\right)-A_{t_0}^{t_{m+1}}\right\}^2 \\ &+\frac{1}{4}\left\{\sigma_0^2(t_1-t_0)^2+\sum_{j=1}^{m}\sigma_j^2(t_{j+1}-t_{j-1})^2+\sigma_{m+1}^2(t_{m+1}-t_m)^2\right\}, \quad (4) \end{aligned}$$

where $\mathsf{E}\left(\hat{A}_{t_0}^{t_{m+1}}\right)=\frac{1}{2}\sum_{j=1}^{m+1}\left(f(t_j;\boldsymbol{\beta})+f(t_{j-1};\boldsymbol{\beta})\right)(t_j-t_{j-1})$.

2. Wang (2001) proposed the following MSE based objective function for estimating the concentration-time curve using linear interpolation:

$$\begin{aligned} Obj_2(t;\boldsymbol{\beta},\boldsymbol{\lambda},\sigma) &= \sum_{j=1}^{m+1}\int_{t_{j-1}}^{t_j}\mathsf{E}\left\{\hat{f}_j(t)-f(t;\boldsymbol{\beta})\right\}^2 dt \\ &= [\mathsf{Bias}]_2^2+[\mathsf{Var}]_2 \\ &= \sum_{j=1}^{m+1}\int_{t_{j-1}}^{t_j}\left\{\mathsf{E}[\hat{f}_j(t)]-f(t;\boldsymbol{\beta})\right\}^2 dt+\frac{1}{3}\sum_{j=1}^{m+1}(t_j-t_{j-1})(\sigma_{j-1}^2+\sigma_j^2), \quad (5) \end{aligned}$$

where $\mathsf{E}[\hat{f}_j(t)]=\mathsf{E}[a_j]+\mathsf{E}[b_j]t$ so that $\mathsf{E}[a_j]=-\frac{f(t_j;\boldsymbol{\beta})t_{j-1}-f(t_{j-1};\boldsymbol{\beta})t_j}{t_j-t_{j-1}}$, $\mathsf{E}[b_j]=\frac{f(t_j;\boldsymbol{\beta})-f(t_{j-1};\boldsymbol{\beta})}{t_j-t_{j-1}}$. This objective function is the objective function for the fixed effects model in Kong and Gonin (2000), assuming no random error.

3. Another way of defining an objective function is to approximate AUC piece-wise and to base the objective function on the sum of the piece-wise MSEs:

$$\begin{aligned} Obj_3(t;\boldsymbol{\beta},\boldsymbol{\lambda},\sigma) &= \sum_{j=1}^{m+1}\mathsf{E}\left\{\int_{t_{j-1}}^{t_j}\left[\hat{f}_j(t)-f(t;\boldsymbol{\beta})\right]dt\right\}^2 \\ &= [\mathsf{Bias}]_3^2+[\mathsf{Var}]_3 \\ &= \sum_{j=1}^{m+1}\left\{\mathsf{E}\left(\hat{A}_{t_{j-1}}^{t_j}\right)-A_{t_{j-1}}^{t_j}\right\}^2+\frac{1}{4}\sum_{j=1}^{m+1}(t_j-t_{j-1})^2(\sigma_{j-1}^2+\sigma_j^2), \quad (6) \end{aligned}$$

where $\mathsf{E}\left(\hat{A}_{t_{j-1}}^{t_j}\right)=\frac{1}{2}\left(f(t_j;\boldsymbol{\beta})+f(t_{j-1};\boldsymbol{\beta})\right)(t_j-t_{j-1})$.

Several authors (Katz and D'Argenio, 1983; Katz, 1984; Bailer and Piegorsch, 1990; Wang, 2001) have noticed that the variance term tends to dominate the MSE. As a result, these objective

functions often do not give intuitively reasonable solutions, even when the measurement error is relatively small. Figures 2 and 3 demonstrate this by plotting the optimal sampling time points associated with the various objective functions assuming true parameter values that will be discussed in Section 5. Notice that the optimal sampling time points tend to be far more spread out than conventional wisdom would suggest for graphs E and F while the points appear almost randomly in graphs A-D. We use these results to motivate the practice of focussing on bias rather than mean squared error to define the objective function.

A second argument to support this conclusion lies in the fact that uncertainty due to the between-subject variability of PK parameters is usually much greater than the measurement error (the within-subject variability), even when the suggested PK model reasonably well describes the concentration profile. Thus, it seems to be more appropriate to use only the bias term as an objective function, while still accounting for the between-subjects variability. Therefore, we reexpress the previous objective functions using the bias terms only. We denote the new objective functions by $Obj_{B1}(t;\boldsymbol{\beta}) = [\text{Bias}]_1^2$, $Obj_{B2}(t;\boldsymbol{\beta}) = [\text{Bias}]_2^2$ and $Obj_{B3}(t;\boldsymbol{\beta}) = [\text{Bias}]_3^2$.

3.1 The grid-wise evaluated objective function

Unfortunately, using the bias only can produce problems where the optimal design requires placing time points far apart where over and underestimates of the area under the concentration-time curve cancel out. Figure 4 illustrates an instance of this by plotting the optimal sampling time points under objective function $Obj_{B1}(t;\boldsymbol{\beta})$ (again the true parameter values will be discussed in Section 5). In practice we have found the same flaw occurs with the objective function $Obj_{B3}(t;\boldsymbol{\beta})$. Among three objective functions, $Obj_{B2}(t;\boldsymbol{\beta})$ appears to be resistant to this problem. However, $Obj_{B2}(t;\boldsymbol{\beta})$ requires an explicit integration to construct the function, which is both difficult and error-prone if the assumed underlying model is complicated.

We propose an objective function which does not require integration yet does not suffer from the cancelling out problem described above. The basic idea is that each time interval between t_{j-1} and t_j is divided equally with the number of divisions determined by the length of the interval. The error between the true AUC and the approximated AUC is then evaluated within each division and the results are summed up. An example of creating the grid and evaluating the error in a time interval is shown in Figure 4. More formally, we define the objective function as:

$$Obj_4(t;\boldsymbol{\beta}) = \sum_{j=1}^{m+1}\Bigg\{\Big[\mathsf{E}\big(\hat{A}_{t_{j-1}}^{g_{(j-1)(1)}}\big) - A_{t_{j-1}}^{g_{(j-1)(1)}}\Big]^2 + \sum_{k=2}^{n_j}\Big[\mathsf{E}\big(\hat{A}_{g_{(j-1)(k-1)}}^{g_{(j-1)(k)}}\big) - A_{g_{(j-1)(k-1)}}^{g_{(j-1)(k)}}\Big]^2 + \Big[\mathsf{E}\big(\hat{A}_{g_{(j-1)(n_j)}}^{t_j}\big) - A_{g_{(j-1)(n_j)}}^{t_j}\Big]^2\Bigg\}, \tag{7}$$

where n_j is the number of grids within the interval (t_{j-1}, t_j) and $g_{(j-1)(l)} = g_{(j-1)(l)}(t_{j-1},\ t_j)$

which is a function of t_{j-1} and t_j, and

$$\begin{aligned}
\mathsf{E}\big(\hat{A}^{g_{(j-1)(1)}}_{t_{j-1}}\big) &= \frac{1}{2}\Big(f(t_{j-1};\boldsymbol{\beta}) + \mathsf{E}[\hat{f}_j(g_{(j-1)(1)})]\Big)\Big(g_{(j-1)(1)} - t_{j-1}\Big) \\
\mathsf{E}\big(\hat{A}^{g_{(j-1)(k)}}_{g_{(j-1)(k-1)}}\big) &= \frac{1}{2}\Big(\mathsf{E}[\hat{f}_j(g_{(j-1)(k-1)})] + \mathsf{E}[\hat{f}_j(g_{(j-1)(k)})]\Big)\Big(g_{(j-1)(k)} - g_{(j-1)(k-1)}\Big) \\
\mathsf{E}\big(\hat{A}^{t_j}_{g_{(j-1)(n_j)}}\big) &= \frac{1}{2}\Big(\mathsf{E}[\hat{f}_j(g_{(j-1)(n_j)})] + f(t_j;\boldsymbol{\beta})\Big)\Big(t_j - g_{(j-1)(n_j)}\Big).
\end{aligned}$$

To illustrate $Obj_4(t;\boldsymbol{\beta})$, suppose that the underestimated area and the overestimated area in Figure 4 are same. Then the evaluated overall error within the interval t_{j-1} and t_j would be 0 if the objective function $Obj_{B1}(t;\boldsymbol{\beta})$ or $Obj_{B3}(t;\boldsymbol{\beta})$ is used whereas the evaluated error would be the underestimated area plus the overestimated area (whole shaded area) if $Obj_4(t;\boldsymbol{\beta})$ is used. While conceivably, the same cancelling out error can occur within a grid of $Obj_4(t;\boldsymbol{\beta})$, the grid length can always be decreased until stable estimates are obtained.

3.2 A population-averaged objective function

We define $Obj_r^{PA}(t;\boldsymbol{\mu},\boldsymbol{\xi})$ as the population-averaged objective function corresponding to the individual objective function $Obj_{Br}(t;\boldsymbol{\beta})$ where $r = 1,\ldots,4$. We define $Obj_r^{PA}(t;\boldsymbol{\mu},\boldsymbol{\xi})$ as:

$$Obj_r^{PA}(t;\boldsymbol{\mu},\boldsymbol{\xi}) = \int_{\boldsymbol{\beta}} Obj_{Br}(t;\boldsymbol{\beta})\, p_{\boldsymbol{\beta}}(\boldsymbol{\beta})\, d\boldsymbol{\beta}, \tag{8}$$

where $p_{\boldsymbol{\beta}}(\boldsymbol{\beta})$ is the joint density function of random effects.

Most of PK models and corresponding objective functions are nonlinear function of parameters; hence there is no analytical form of $Obj_r^{PA}(t;\boldsymbol{\mu},\boldsymbol{\xi})$ and numerical integration is employed. Usual numerical integration methods are often technically difficult to be applied in this context. We therefore use a Monte Carlo integration scheme within our simulated annealing algorithm when using population averaged objective functions.

4 Optimization Using a Simulated Annealing Algorithm

The simulated annealing algorithm (SA) is a stochastic global optimization method. Simulated annealing gets its name as it is a conceptual analogue of the cooling (annealing) of high energy state molten metals in thermodynamics. The annealed metal reaches the global minimum energy state if the annealing process is very slow, since random fluctuation of the energy of the metal allows it to escape from local minima. If the metal is cooled quickly, it might not escape from a local minimum of energy. Like the physical annealing process, the simulated annealing algorithm allows the current state to move uphill out of a local minimum with some acceptance probability. This idea was first described in a seminal paper by Metropolis et al. (1953). The acceptance

probability in SA is controlled by a parameter called the "temperature" which gets its name from the cooling temperature schedule in the physical annealing process. The readers can find excellent reviews of SA in Bohachevsky et al. (1986) and Goffe et al. (1994).

Unlike most deterministic algorithms, since SA moves both downhill and uphill along the surface of the objective function, it theoretically converges to the global minimum for complicated functions for which derivatives do not exist, have many local minima or high dimensional discrete or continuous functions. In many cases, if run in a sensible manner, the solution from SA is insensitive to starting values, which is a major problem for many optimization algorithms. The algorithm can easily employ a variety of constraints and the implementation is relatively simple since SA does not require derivatives of the objective function. However, SA is not without its disadvantages which can include long run times and quite a bit tinkering with the cooling schedule to obtain a reasonable algorithm.

We adopted SA for the following reasons. First, our objective function is complicated, non-linear, multimodal and involves many parameters (sampling times). Secondly, SA can easily incorporate constraints such as discretizing the sampling times or modifying the grid-lengths in the objective function $Obj_4^{PA}(t; \boldsymbol{\mu}, \boldsymbol{\xi})$. Finally, the use of Monte Carlo integration to estimate the population-averaged optimal sampling time points using objective function, $Obj_4^{PA}(t; \boldsymbol{\mu}, \boldsymbol{\xi})$, fits well within the SA framework.

The description of our SA algorithm is as follows while a schematic of the algorithm is shown in Figure 5. We assume the "true" PK parameters are fixed and known as well as the first and last sampling times (to be used in the objective function). We initialize the temperature, $T^{(1)}$, and starting sampling time points $t^{(1)} = (t_0^{(1)}, \ldots, t_{m+1}^{(1)})$. Then each iteration $l = 2, \ldots, L$ is as follows:

1. set $T^{(l)} = T^{(1)}/\log(l)$;
2. randomly choose k uniformly from $1, \ldots, m$;
3. simulate t_{new} uniformly in the interval $(t_{k-1}^{(l-1)}, t_{k+1}^{(l-1)})$;
4. set $t^{(l)} = (t_0^{(l-1)}, \ldots, t_{k-1}^{(l-1)}, t_{new}, t_{k+1}^{(l-1)}, \ldots, t_{m+1}^{(l-1)})$;
5. simulate I random effects vectors, $\{\boldsymbol{\beta}_i\}_{i=1}^I$, each from $\{\boldsymbol{\mu}, \boldsymbol{D}(\boldsymbol{\xi})\}$ for $i = 1, \ldots, I$;
6. calculate $\phi_l = \sum_i Obj_4(t^{(l)}; \boldsymbol{\beta}_i)/I$;
7. set $p = \min\{\exp\big(-\frac{\Delta\phi}{T^{(l)}}\big), 1\}$ where $\Delta\phi = \phi_l - \phi_{l-1}$;
8. generate $u \sim U[0, 1]$;
9. accept the current sampling times $t^{(l)}$ if $u \leq p$; otherwise $t^{(l)} = t^{(l-1)}$.

Model	ka	k	V	D	μ_k	ξ_k
Fixed effects model 1 (FM1)	0.2	0.1	40	400		
Fixed effects model 2 (FM2)	0.8	0.2	40	400		
Random effects model 1 (RM1)	0.2		40	400	0.1	0.05

Table 1: The parameters for the examples, where μ_k and ξ_k are mean and standard deviation of the random effect k. These parameters were adopted from Kong and Gonin (2000) for comparison.

For a fixed effects model, we omit Step 5 and replace Step 6 with $\phi_l = Obj_4(t^{(l)}; \boldsymbol{\beta})$. A benefit of the algorithm is that timing constraints, such as no two sampling points being too close, can be easily incorporated into Step 3.

5 Examples

For illustration, consider a pharmacokinetic system of the one-compartment open model with first-order absorption and first-order elimination kinetics (Gibaldi and Perrier, 1982). That is, the concentration time curve is parametrized as:

$$f(t; \boldsymbol{\beta}) = \frac{k_a F D}{V(k_a - k)}(e^{-kt} - e^{-k_a t}),$$

where

$\boldsymbol{\beta}$: $\boldsymbol{\beta} = [\beta_1, \beta_2, \beta_3] = [V, ka, k]$
F : the fraction of the administered dose D that is absorbed (assume $F = 1$ in the examples)
V : the volume of distribution (a proportionality constant relating the amount of drug to the concentration of drug in the blood)
k_a : the apparent first-order absorption rate constant
k : the apparent first-order elimination rate constant.

We focus our discussion on fixed effects models for two sets of PK parameters and a random effects model where k is random; Table 1 gives the specific parameters values used. The two sets of fixed effect parameters represent two extremes of the concentration-time curve. In FM1, the curve is smoothly rising and decaying while in FM2 the curve changes very rapidly.

The optimal sampling times for fixed effects models are shown in Figures 6 and 7. The left panel shows the (estimated) optimal sampling times obtained using $Obj_4(t; \boldsymbol{\beta})$ as the objective function while the right panel shows the optimal sampling times obtained using $Obj_{B2}(t; \boldsymbol{\beta})$ both estimated by SA. The resulting sampling schedules for the two objective functions are very similar regardless of m. Furthermore, the optimal sampling times obtained using $Obj_{B2}(t; \boldsymbol{\beta})$

under SA are similar to those obtained by Kong and Gonin (2000) who used the same objective function (using a different algorithm). Of course when the concentration-time curve is smooth, such as under model FM1, the number of sampling time points can be reduced (see Figure 6). However, we note that care must be taken because estimation of the subject specific k is not represented in the objective function. With $m = 8$ an insufficient number of points (only two) are represented in the tail of the concentration-time curve when the last 3 time points need to be used to estimate the subject specific k.

For FM2, the optimal sampling time points obtained using $Obj_4(t;\boldsymbol{\beta})$ are very different from those obtained using $Obj_{B2}(t;\boldsymbol{\beta})$ unless m is large. Since the rate of absorption is very large compared to the rate of elimination and the rate of elimination is also large for the FM2, the profile is rising very quickly, has a sharp peak, then decreases very rapidly. The optimization procedure in this case is almost entirely determined by the objective function, which is unfortunate because the two functions are conceptually similar. Furthermore, stable convergence was difficult to obtain because of the many local minima of similar magnitude to the global minimum. Thus, it is probably more feasible to transfer the problem to a scale that is less dependent on the objective function.

We focus our attention on estimating the bulk of the area under the concentration-time curve, say for a time interval $[0, t_{max} + w]$ for some w. Then, we suggest that users employ the logarithmic trapezoidal method (Yeh and Kwan, 1978; Chiou, 1978) for estimating AUC. Technically, this focus is justified because when using this method, only the (fixed) final time point after $t_{max} + w$ is required if the PK model is correct and the observations are measured without error. Of course, several points should be used in practice because the PK model is never exactly correct and measurement error is always present.

Focussing on this interval, we find that the results are more robust to specification of the objective function and converged more quickly. Since t_{max} varies across subjects, it is necessary to take w to be large enough in order for the time interval $[0, t_{max} + w]$ to be able to cover the peak. Figure 8 shows the sampling time points using this approach for $w = 3.7$. The left panel shows the log of the concentration time profile while the right shows the curve on the raw scale.

To obtain population-averaged optimal sampling time points, we need to assume a distribution for the PK parameters. For model RM1, we assume that k follows a lognormal distribution with mean $\mu_k = 0.1$ and standard deviation $\xi_k = 0.05$ (similar to Kong and Gonin, 2000). Further, we assume that $k_a > k$ (also as in Kong and Gonin, 2000), which is reasonable for most oral conventional/immediate release drug products. We emphasize that our objective function, $Obj_4^{PA}(t;\boldsymbol{\mu},\boldsymbol{\xi})$, does not require this assumption.

The obtained optimal sampling time points are shown in Figure 9. It is interesting to note that the optimal sampling time points for the random effects model are quite different from those for the fixed effects model. It is also interesting to note that the sampling time points for the

random effects model have been shifted, resulting in one less point in the absorption phase and one more point in the elimination phase of the concentration-time curve. The additional point in the elimination phase seems to be needed to account for variability due to the random elimination rate constant, k. The optimal sampling time points for the random effects model of Kong and Gonin (2000) are slightly different from our result, which might be due to the different method of numerical integration.

Before concluding this section we note that a strength of the SA approach is that a model with multidimensional random effects could be implemented using our approach with little additional effort.

6 Summary and Discussion

In this manuscript, we emphasized that the within-subject variance term of objective functions based on MSE over-dominated the selection of sampling time points. In addition, since the between-subject variability in PK parameters is usually much greater than the within-subject variability, it is more important to take into account the between-subject variability. Thus, we recommend to use an objective function based on the squared bias as an exploratory procedure to get a general idea about the shape of profile. Then we suggest using a population-averaged objective function to design the sampling schedule.

We proposed a grid-wise evaluated objective function which can overcome some of the disadvantages of other objective functions in the literature. The proposed objective function is simple to implement and produced reasonable sampling time points consistent with those produced by the objective function of Kong and Gonin (2000). Because the newly developed objective function is not amenable to most non-linear optimization methods, we implemented a simulated annealing algorithm. Furthermore, for obtaining population-averaged results, we embedded Monte Carlo numerical integration within the algorithm. An appealing aspect of SA is the ease of handling the Monte Carlo integration and stochastic optimization within a unified framework. Also, the simulated annealing algorithm easily handles constraints on the sampling times.

As shown in the examples, when the profile is smooth, both the grid-wise evaluated objective function and $Obj_{B2}(t;\boldsymbol{\beta})$ gave similar results. However, when the profile is sharp, special attention is needed and one should not use any optimization method blindly. We suggested a strategy to select time points to estimate the bulk of the area under the concentration-time curve. However, it would seem that prior knowledge regarding subjects is important to select time points in these scenarios.

It should be noted in BE studies, C_{max} and t_{max} are also important in addition to AUC. A concern when estimating these parameters is that sampling times may straddle the functional maximum. We note that the error in the AUC approximation is directly related to the curvature

of the concentration-time curve. (This is seen most clearly with Wang (2001)'s approximated objective function which formally depends on the second derivative.) As a result, in many case both the grid-wise objective function, $Obj_{B2}(t;\boldsymbol{\beta})$ and their population-averaged counterparts used for estimating AUC concentrated points around t_{max} (see Figure 6-9). Encouragingly, these objective functions produced reasonable sampling time points for estimating C_{max} and t_{max}.

Broadly speaking, the optimal sampling time points change drastically depending on the underlying model. Thus, if the number of sampling times is limited for either clinical or economic reasons, it is important to understand the optimal placement of the sampling times given various underlying models. The objective function and stochastic optimization method presented in this manuscript would be useful tools to aid in the discussion of PK study design.

References

Al-Banna, M. K., Kelman, A. W., and Whiting, B. "Experimental design and efficient parameter estimation in population pharmacokinetics." *Journal of Pharmacokinetics and Biopharmaceutics*, 18:347–360 (1990).

Atkinson, A. C., Chaloner, K., Herzberg, A. M., and Juritz, J. "Optimum experimental designs for properties of a compartmental model." *Biometrics*, 49:325–337 (1993).

Bailer, A. J. and Piegorsch, W. W. "Estimating integrals using quadrature methods with an application in pharmacokinetics." *Biometrics*, 46:1201–1211 (1990).

Beatty, D. and Piegorsch, W. W. "Optimal statistical design for toxicokinetic studies." *Statistical Methods in Medical Research*, 6:359–376 (1997).

Bohachevsky, I. O., Johnson, M. E., and Stein, M. L. "Generalized simulated annealing for function optimization." *Technometrics*, 28:209–217 (1986).

Box, G. and Lucas, H. "Design of experiments in non-linear situations." *Biometrika*, 49:77–90 (1959).

Chiou, W. L. "Critical evaluation of the potential error in pharmacokinetic studies of using the linear trapezoidal rule method for the calculation of the area under the plasma level-time curve." *Journal of Pharmacokinetics and Biopharmaceutics*, 6:539–546 (1978).

Davidian, M. and Giltinan, D. M. *Nonlinear models for repeated measurement data*. Chapman & Hall (1995).

Duffull, S. B., Retout, S., and Mentré, F. "The use of simulated annealing for finding optimal population designs." *Computer Methods and Programs in Biomedicine*, 69:25–35 (2002).

Gibaldi, M. and Perrier, D. *Pharmacokinetics*. New York: Marcel Dekker, 2nd edition (1982).

Goffe, W. L., Ferrier, G. D., and Rogers, J. "Global optimization of statistical functions with simulated annealing." *Journal of Econometrics*, 60:65–99 (1994).

Jones, B. and Wang, J. "Constructing optimal designs for fitting pharmacokinetic models." *Statistics and Computing*, 9:209–218 (1999).

Katz, D. "Optimal quadrature points for approximating integrals when function values are observed with error." *Mathematics Magazine*, 59:284–290 (1984).

Katz, D. and D'Argenio, D. Z. "Experimental design for estimating integrals by numerical quadrature, with applications to pharmacokinetic studies." *Biometrics*, 39:621–628 (1983).

Kong, F. H. and Gonin, R. "Optimal sampling times in bioequivalence tests." *Journal of Biopharmaceutical Statistics*, 10:31–44 (2000).

Mentré, F., Mallet, A., and Baccar, D. "Optimal design in random-effects regression models." *Biometrika*, 84:429–442 (1997).

Metropolis, N., Rosenbluth, A., Rosenbluth, M., and Teller, A. "Equation of state calculation by fast computing machines." *Journal of Chemical Physics*, 21:1087–1092 (1953).

Piegorsch, W. W. and Bailer, A. J. "Optimal design allocations for estimating area under curves for studies employing destructive sampling." *Journal of Pharmacokinetics and Biopharmaceutics*, 17:493–507 (1989).

Retout, S., Duffull, S., and Mentré, F. "Development and implementation of the population Fisher information matrix for the evaluation of population pharmacokinetic designs." *Computer Methods and Programs in Biomedicine*, 65:141–151 (2001).

Schempp, W. "Cubature, quadrature, and group actions." *Mathematical Reports of the Academy of Science*, 3:337–342 (1981).

Tod, M., Mentré, F., and Mallet, A. "Robust optimal design for the estimation of hyperparameters in population pharmacokinetics." *Journal of Pharmacokinetics and Biopharmaceutics*, 26:689–716 (1998).

Wang, J. "Optimal design for linear interpolation of curves." *Statistics in Medicine*, 20:2467–2477 (2001).

Wang, J. and Endrenyi, L. "A computationally efficient approach for the design of population pharmacokinetic studies." *Journal of Pharmacokinetics and Biopharmaceutics*, 20:279–294 (1992).

Westlake, W. J. "Statistical aspects of comparative bioavailability trials." *Biometrics*, 35:273–280 (1979).

Yeh, K. and Kwan, K. "A comparison of numerical integrating algorithms by trapezoidal, Lagrange, and spline approximation." *Journal of Pharmacokinetics and Biopharmaceutics*, 6:79–98 (1978).

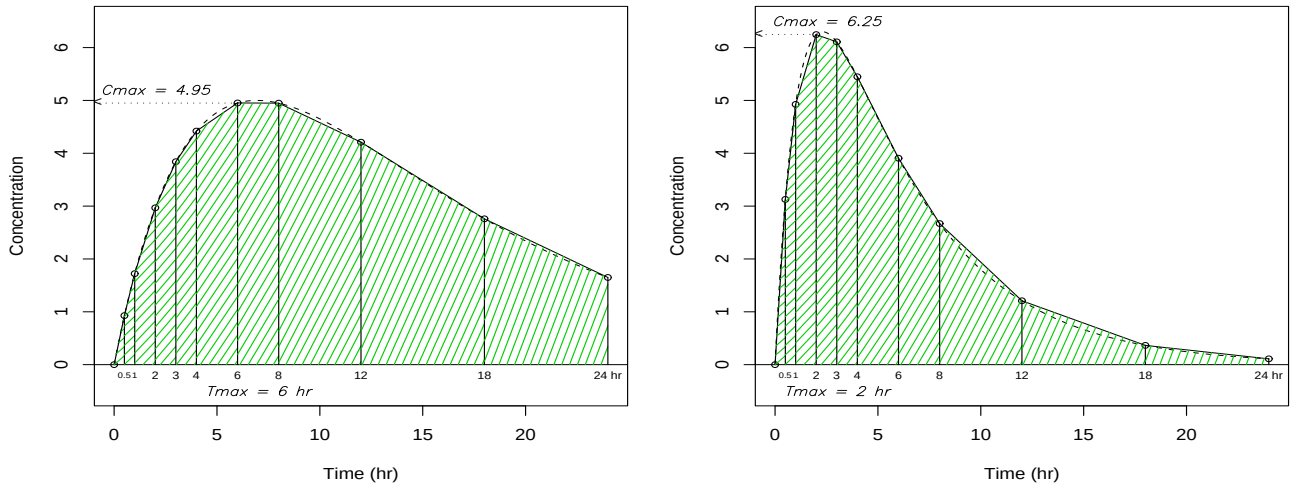
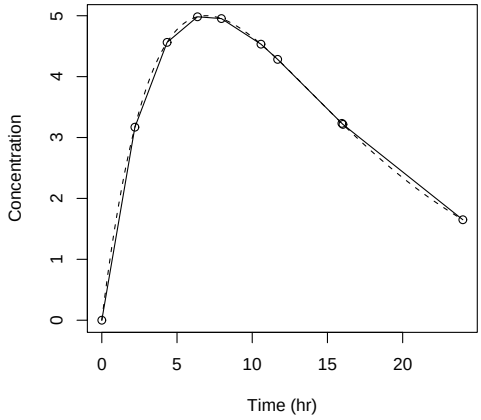
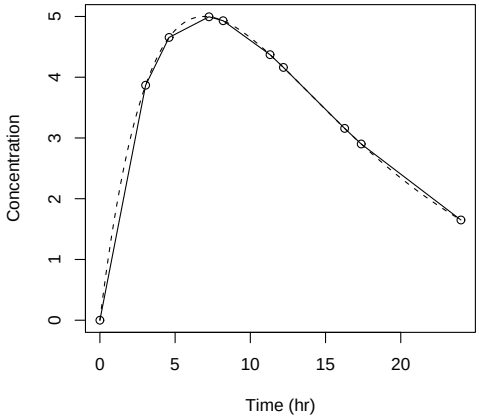


Figure 1: The estimated AUC, $Cmax$ and $tmax$ using conventional sampling time points. The dotted line and solid line present the true underlying concentration-time curve and the linear interpolation of sampling time points assuming no random errors, respectively. The estimated AUC via the trapezoidal approximation is the shaded area under the solid line.

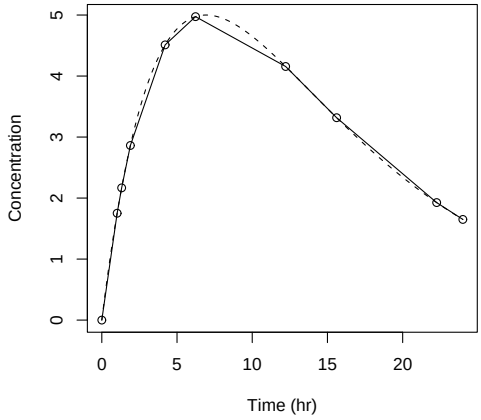
$(A)\ Obj_1(t;\boldsymbol{\beta},\boldsymbol{\lambda},\sigma),\ \sigma = 0.1$



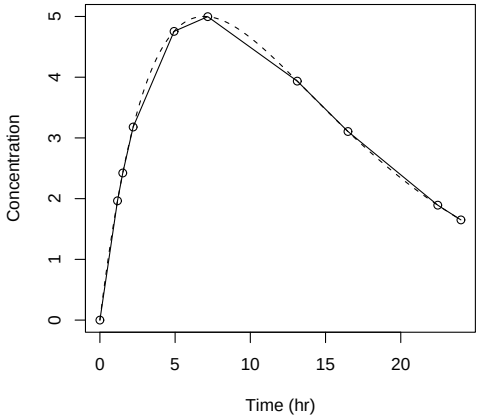
$(B)\ Obj_1(t;\boldsymbol{\beta},\boldsymbol{\lambda},\sigma),\ \sigma = 0.3$



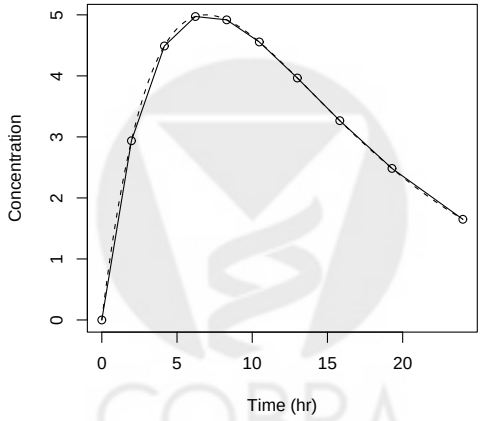
$(C)\ Obj_2(t;\boldsymbol{\beta},\boldsymbol{\lambda},\sigma),\ \sigma = 0.1$



$(D)\ Obj_2(t;\boldsymbol{\beta},\boldsymbol{\lambda},\sigma),\ \sigma = 0.3$



$(E)\ Obj_3(t;\boldsymbol{\beta},\boldsymbol{\lambda},\sigma),\ \sigma = 0.1$



$(F)\ Obj_3(t;\boldsymbol{\beta},\boldsymbol{\lambda},\sigma),\ \sigma = 0.3$

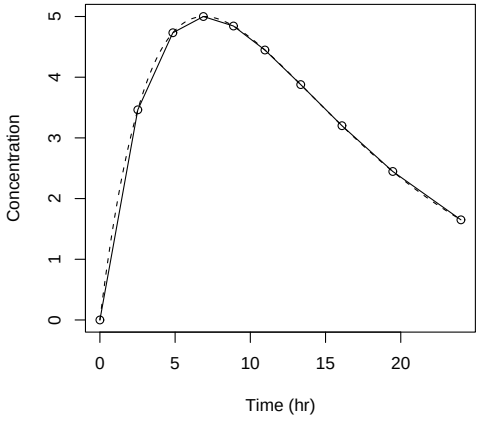


Figure 2: The estimated sampling time points using the mean squared error objective functions for fixed effects model with parameters $ka = 0.2, k = 0.1, V = 40, Dose = 400, \lambda_1 = 0, \lambda_2 = 2$ (the model is discussed in Section 5). The dotted line and solid line present the true underlying concentration-time curve and the linear interpolation of optimal sampling time points, respectively.

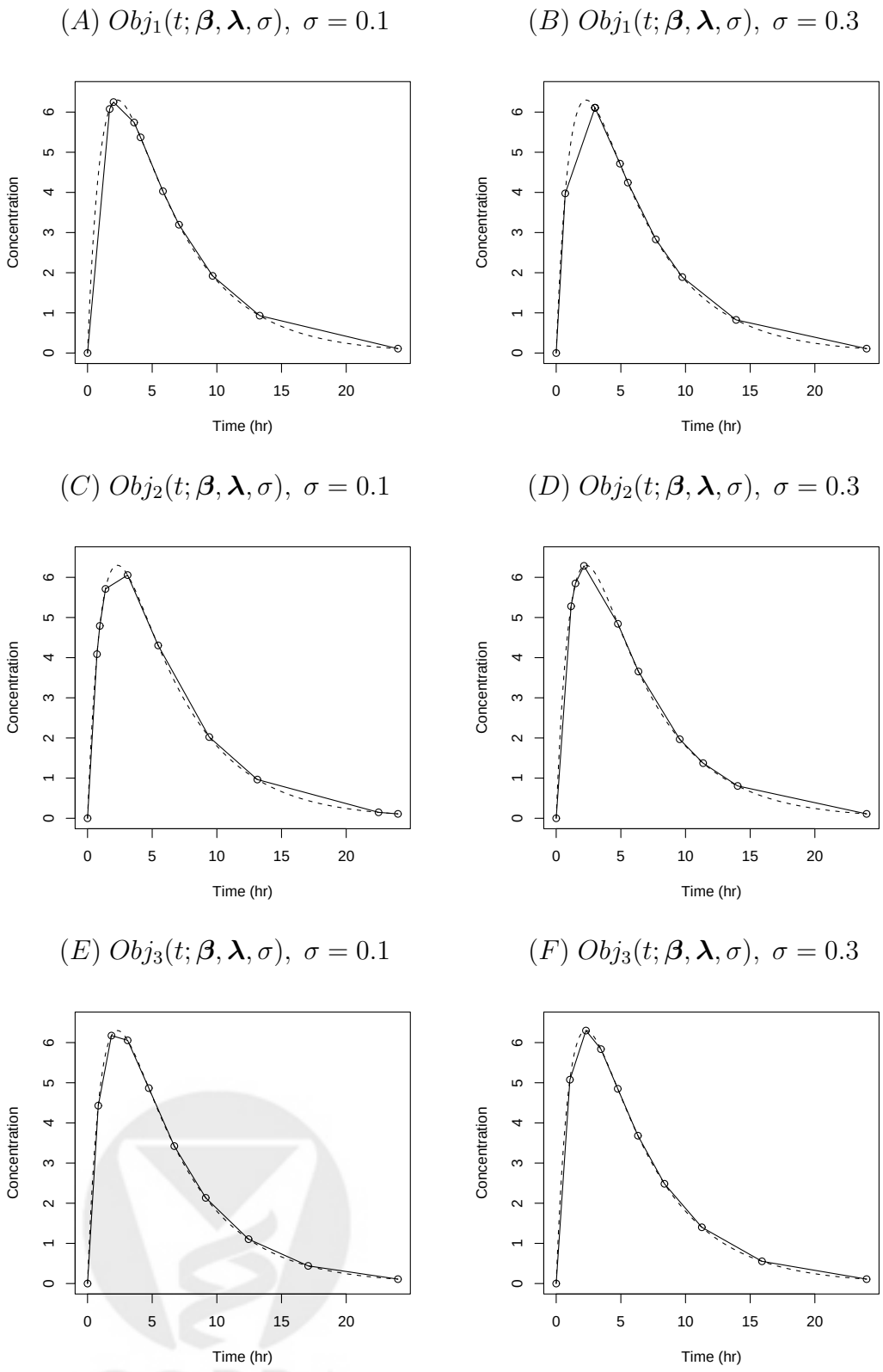


Figure 3: The estimated sampling time points using the mean squared error based objective functions for fixed effects model with parameters $ka = 0.8, k = 0.2, V = 40, Dose = 400, \lambda_1 = 0, \lambda_2 = 2$ (the model is discussed in Section 5). The dotted line and solid line present the true underlying concentration-time curve and the linear interpolation of optimal sampling time points, respectively.

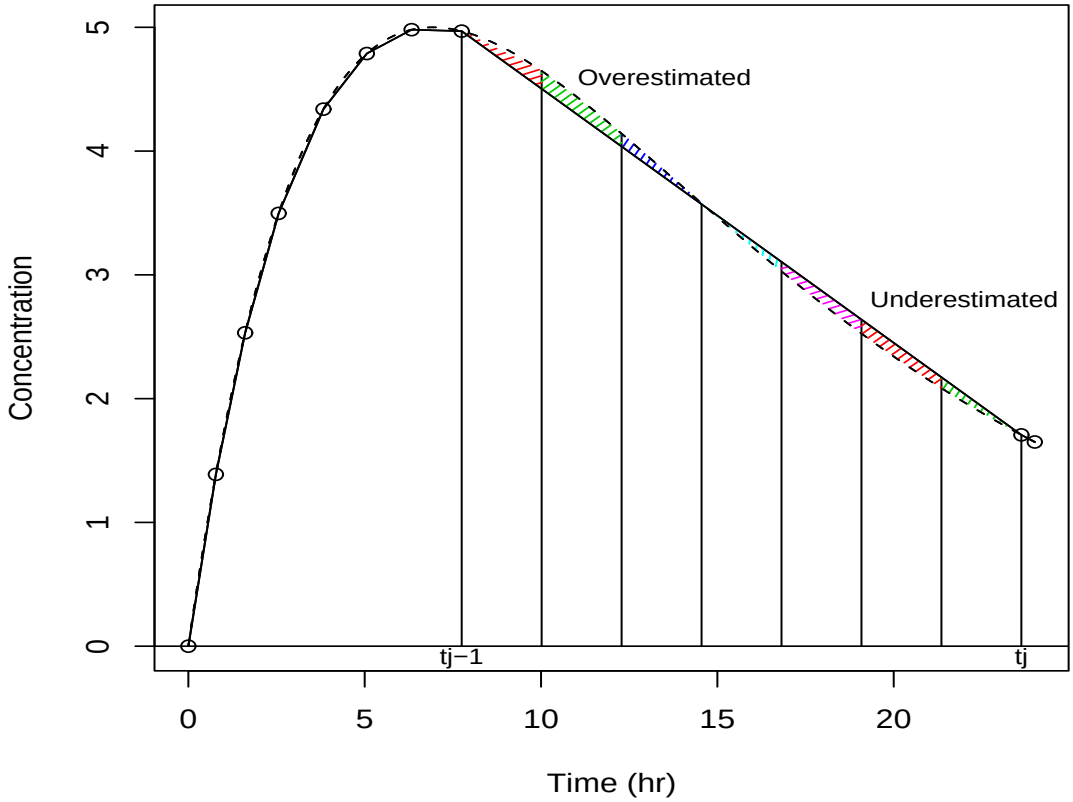


Figure 4: The demonstration of the grid-wise evaluated objective function to prevent overestimated and underestimated errors from being cancelled out. The dotted line and solid line present the true underlying concentration-time curve with parameters $ka = 0.2, k = 0.1, V = 40, Dose = 400, \lambda_1 = 0, \lambda_2 = 2$ (the model is discussed in Section 5) and the linear interpolation of optimal sampling time points under objective function $Obj_{B1}(t; \boldsymbol{\beta})$, respectively.

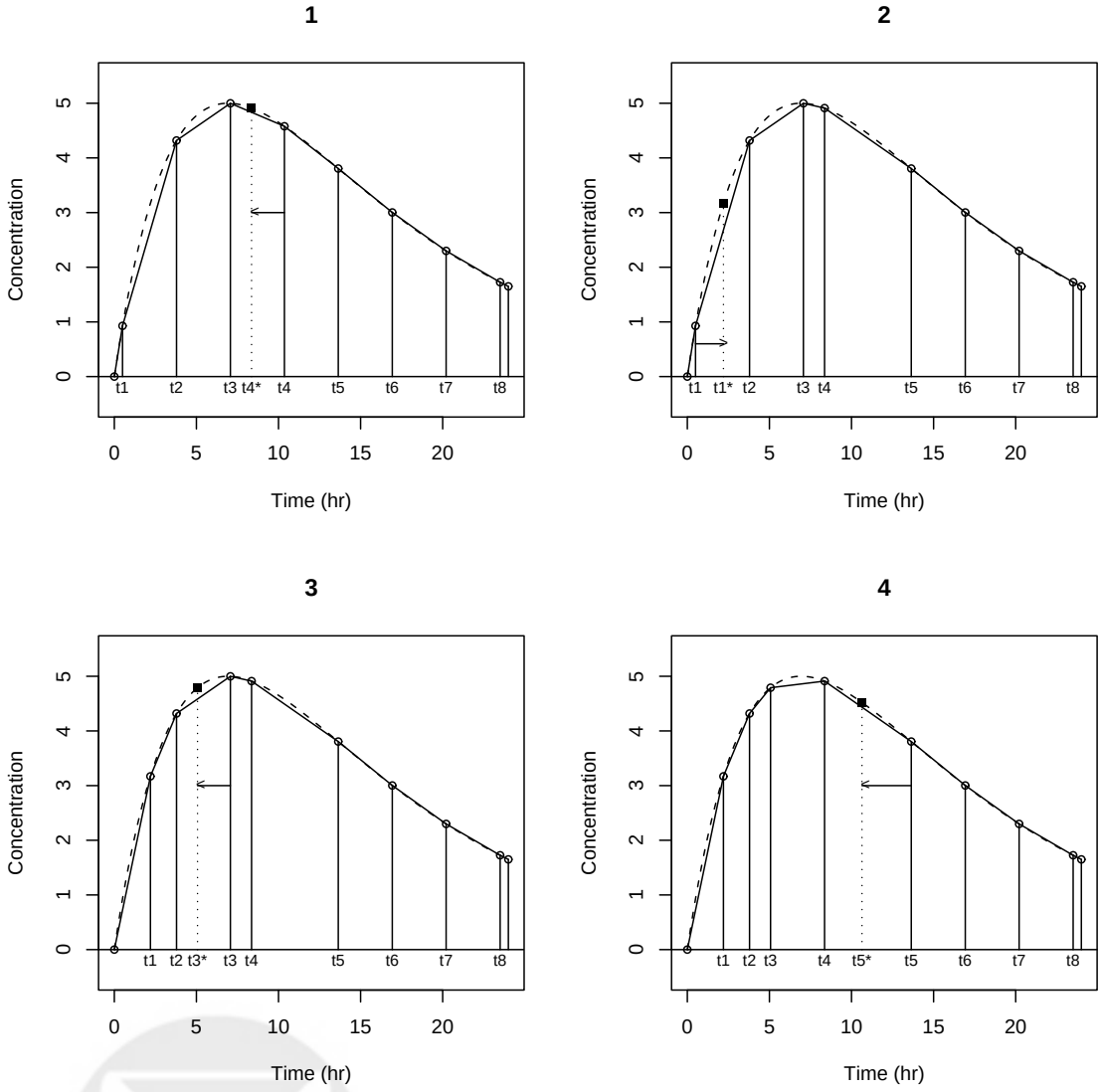


Figure 5: A demonstration of the simulated annealing algorithm from Section 4.

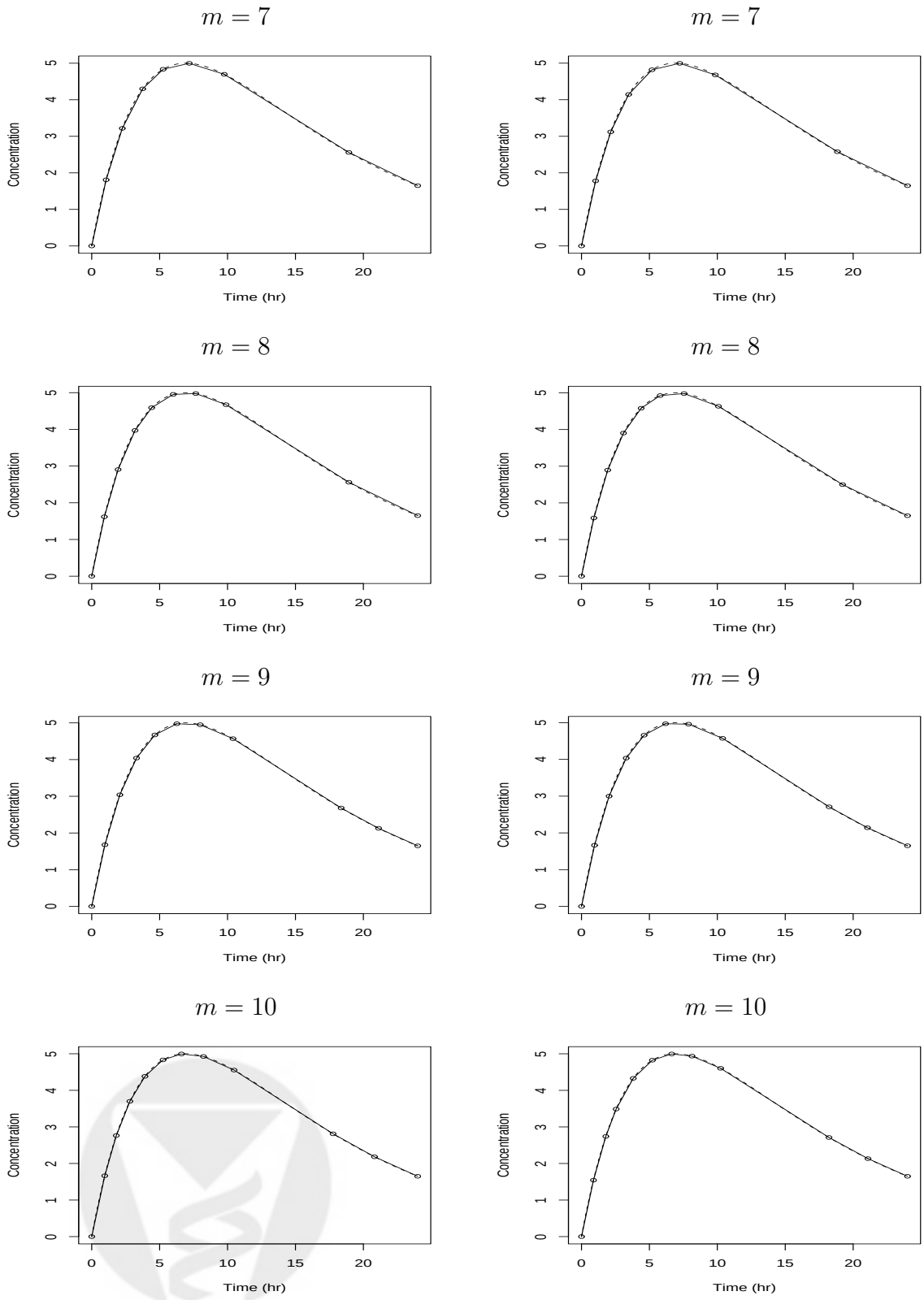


Figure 6: The optimal sampling time points for fixed effects model (FM1) with parameters $ka = 0.2, k = 0.1, V = 40, Dose = 400$. The plots on the left are obtained using the grid-wise evaluated objective function $Obj_4(t; \boldsymbol{\beta})$ while the plots on the right are obtained using $Obj_{B2}(t; \boldsymbol{\beta})$ implemented by SA. The dotted line and solid line present the true underlying concentration-time curve and the linear interpolation of optimal sampling time points, respectively.

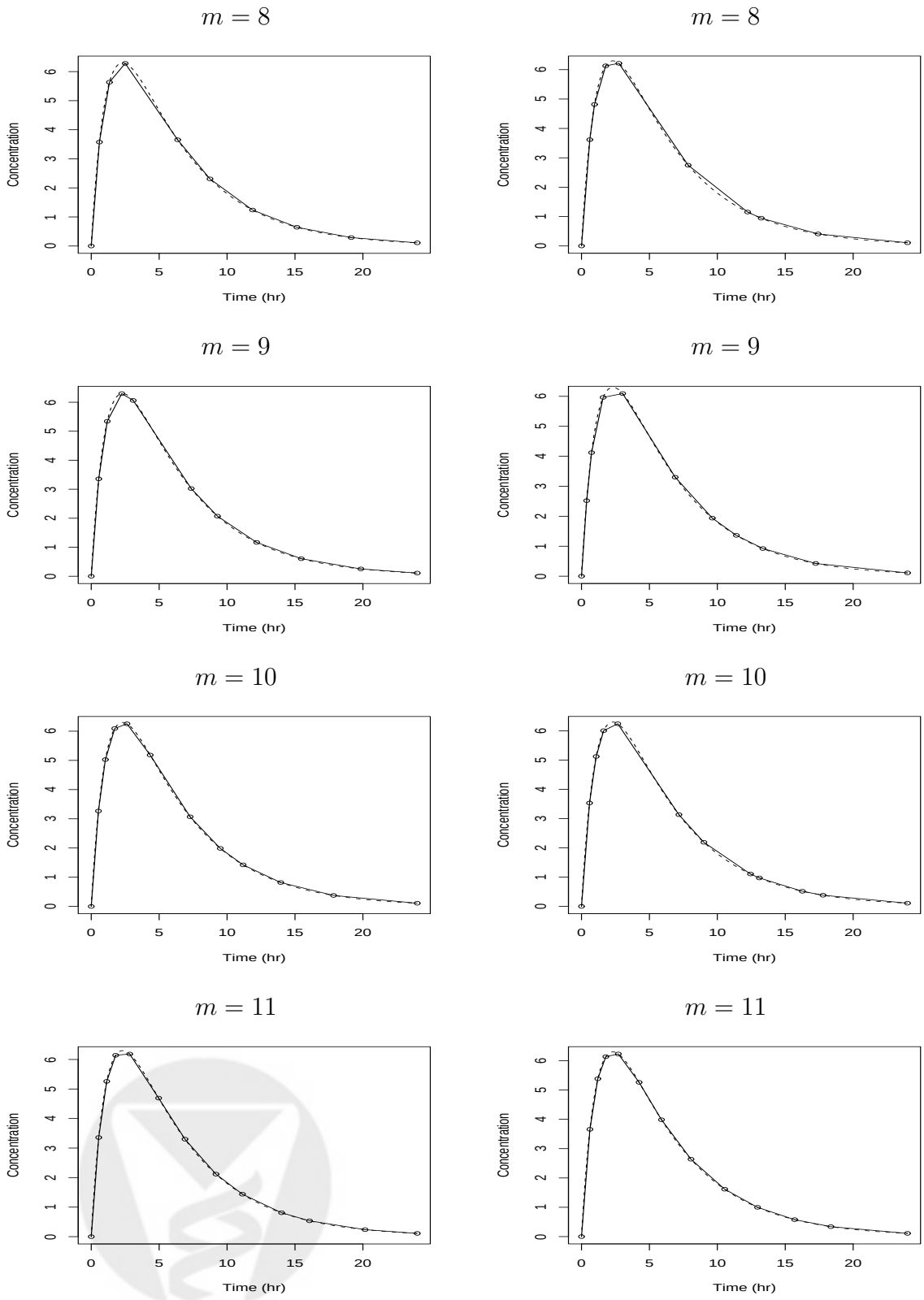


Figure 7: The optimal sampling time points for fixed effects model (FM2) with parameters $ka = 0.8, k = 0.2, V = 40, Dose = 400$. The plots on the left are the optimal sampling time points using the grid-wise evaluated objective function $Obj_4(t; \boldsymbol{\beta})$ while the plots on the right are the optimal sampling time points using $Obj_{B2}(t; \boldsymbol{\beta})$ implemented by SA. The dotted line and solid line present the true underlying concentration-time curve and the linear interpolation of optimal sampling times, respectively.

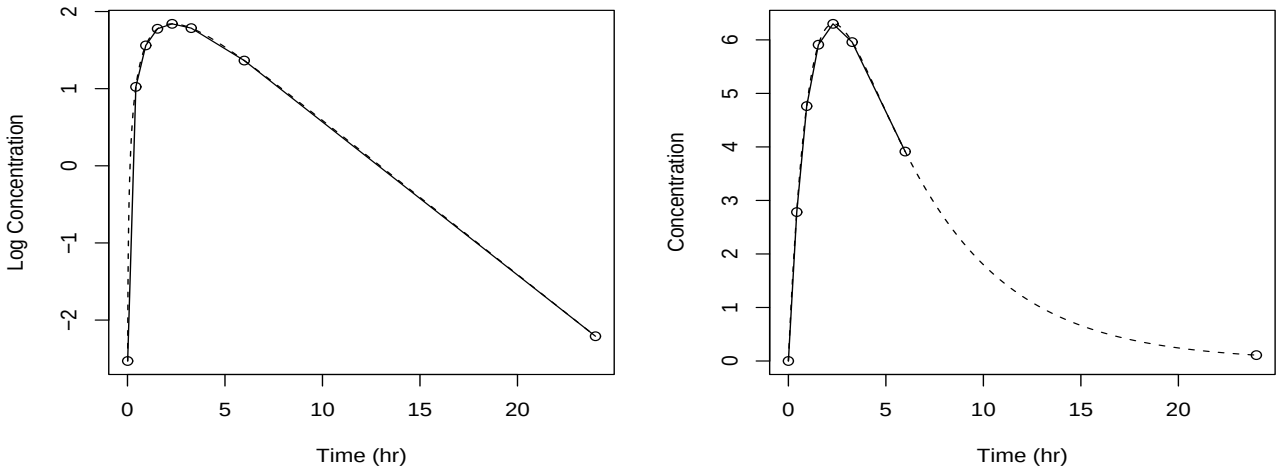
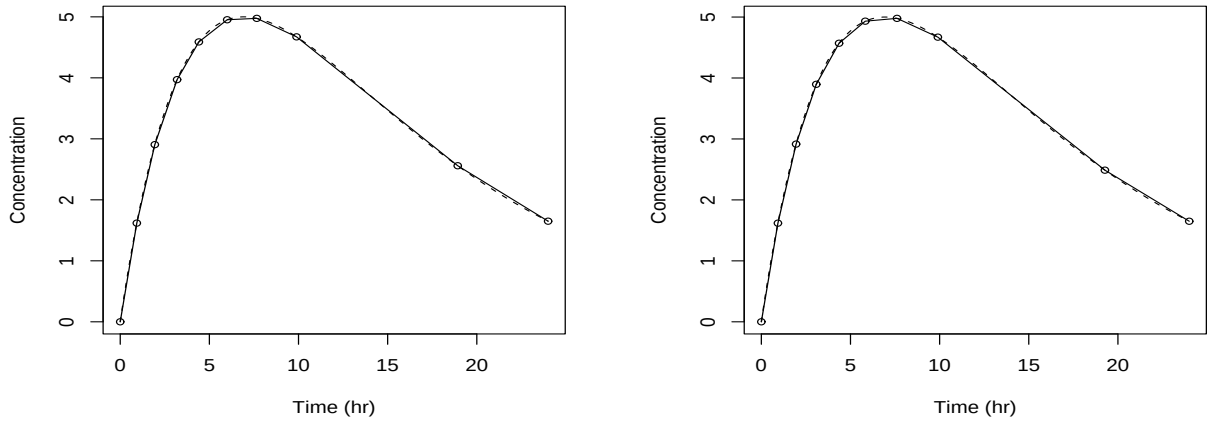


Figure 8: The optimal sampling time points for fixed effects model (FM2) with parameters $ka = 0.8, k = 0.2, V = 40, Dose = 400$ using our proposed strategy. The left panel shows the log of the concentration vs. time and the right panel shows the concentration vs. time.

$(A)\ FM1,\ Obj_4(t;\boldsymbol{\beta})$ $(B)\ FM1,\ Obj_{B2}(t;\boldsymbol{\beta})$



$(C)\ RM1,\ Obj_4^{PA}(t;\boldsymbol{\mu},\boldsymbol{\xi})$ $(D)\ RM1,\ Obj_{B2}^{PA}(t;\boldsymbol{\mu},\boldsymbol{\xi})$

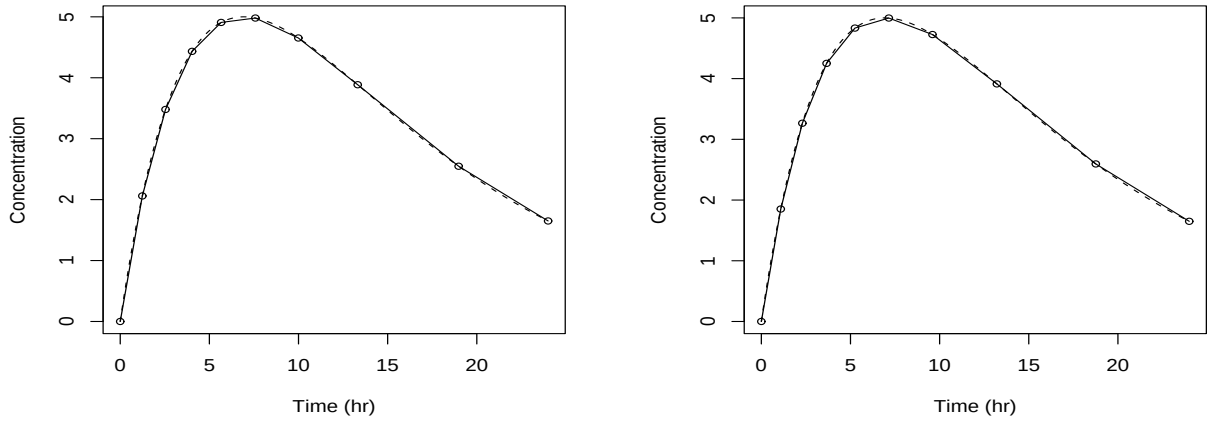


Figure 9: The comparison of the optimal sampling time points for fixed effects model (FM1) (top panel) and for random effects model (RM1) (bottom panel) where FM1 has parameters $ka = 0.2, k = 0.1, V = 40, Dose = 400$ and RM1 has fixed parameters $ka = 0.2, V = 40, Dose = 400$ and random effect k with the mean $\mu_k = 0.1$ and standard deviation $\xi_k = 0.05$ where $k_a > k$. The plots on the left are the optimal sampling time points using the grid-wise evaluated function implemented by SA while the plots on the right are the optimal sampling time points from Kong and Gonin (2000).