

Simulation-Based Methods for Optimal Sampling Times in Population PK/PD Studies

Jonathan R. Stroud †

Department of Statistics, University of Chicago, Chicago, IL 60637, U.S.A.

Peter Müller

Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27708-0251, U.S.A.

Gary L. Rosner

Biometry Division, Community & Family Medicine, Duke University Medical Center, Durham NC 27710, U.S.A., and Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27708-0251, U.S.A.

Summary. We propose a simulation-based approach to decision theoretic Bayesian optimal design. The underlying probability model is a population pharmacokinetic model which allows for correlated responses (drug concentrations) and patient-to-patient heterogeneity. We consider the problem of choosing sampling times for the anticancer agent paclitaxel, using criteria related to total area under the curve, time above a critical threshold and sampling cost.

Key Words: Clinical trial; Limited sampling strategies; Longitudinal data; Markov chain Monte Carlo; Optimal design; Population model; Random-effects regression.

1. Introduction

Most anticancer drugs in use today are cytotoxic (kill cells) and have a narrow therapeutic index, meaning there is a narrow range, at best, between an efficacious dose and a dose too toxic for clinical use. Additionally, people respond to different pharmaceutical agents differently. Among cancer patients, some experience tumor response to chemotherapy while others do not; some experience more toxicity than others. Understanding the reasons behind this heterogeneity should contribute to safer and more effective chemotherapy for each patient. Thus, it is important to understand the pharmacology of cancer chemotherapeutic agents and relationships with outcome. Achieving this goal requires clinical studies involving large numbers of cancer patients.

Population pharmacokinetic (PK) and pharmacodynamic (PD) studies seek to identify within a large sample of patients those characteristics that significantly alter a drug's disposition and a patient's response to the drug. Population PK studies often presuppose some sort of model that relates drug concentration to time and then evaluate relationships between patient characteristics and the parameters in the model. PD studies of treated patients examine relationships of clinical outcomes, such as toxicity or tumor response, to characteristics of the drug's pharmacokinetic profile

†*Address for correspondence:* Jonathan R. Stroud, Department of Statistics, University of Chicago, 5734 University Avenue, Chicago, IL 60637, U.S.A. E-mail: stroud@galton.uchicago.edu

in each patient. In this paper, we consider optimal design in the context of such studies. Specifically, we discuss the choice of optimal sampling times in a population PK study of paclitaxel conducted by the Cancer and Leukemia Group B (CALGB).

Paclitaxel (Taxol[®], Bristol-Myers Squibb) is a widely used chemotherapeutic agent. The pharmacokinetics of paclitaxel are complex and highly nonlinear (Kearns et al., 1995). The three-compartment model proposed by Kearns et al. (1995) to describe paclitaxel's pharmacokinetics includes two nonlinear pathways. The model yields no closed-form solution to the corresponding system of differential equations that describes the relationship between time and the concentration of the drug in the patient's blood (Gibaldi and Perrier, 1982).

The CALGB sought to investigate the role of paclitaxel for treating advanced breast cancer. The trial would enroll a large number of women, creating the opportunity to study paclitaxel's complex clinical pharmacology in a large group of women. Earlier studies showed an association between toxicity or tumor response and the time during which the concentration of paclitaxel in a patient's blood exceeds a threshold value (Gianni et al., 1995). The CALGB wished to confirm this observation among the hundreds of women entering the new study, as well as look for relationships between patient characteristics and the PK or PD of paclitaxel.

The study's protocol compared doses of the drug infused over three hours. Patients typically wait in the clinic for a few hours after the end of the infusion, just in case unexpected problems develop. After this waiting period, the patients go home and do not have to return to the clinic, in general, until their next course of chemotherapy four weeks later. Therefore, it would be very inconvenient for the patient if the sampling schedule would require her to wait in the clinic longer than usual or, worse still, return to the clinic the following day for, say, a blood sample drawn twenty-four hours after the start of the three-hour infusion.

Optimal choice of sampling times is important when designing large population-based studies. One usually wants to avoid excessively complex designs. For example, if many blood samples are required from each treated patient when studying a particular drug, then the study will be costly, inconvenient for the patient and trial participants, and more likely to result in some mistakes leading to a greater chance that some samples may drop, be lost, or not be stored properly prior to shipment to the central laboratory. A goal, then, in the design of such population PK/PD studies, is to estimate the model parameters as precisely as possible, while minimizing the cost inherent in sampling patients. There are few papers in the literature addressing this design problem. Huizing et al. (1995) use stepwise regression in an effort to determine optimal sampling times for cancer patients treated with paclitaxel. Similar so-called limited sampling strategies based on choosing sampling times in a stepwise fashion according to their correlation with the desired endpoint, such as area under the concentration-time curve (AUC), are discussed in Ratain and Vogelzang (1987), Ratain et al. (1988), and Egorin et al. (1988). Beatty and Piegorsch (1997) provide a general review of optimal design in toxicokinetic studies.

In this paper, we consider simulation-based methods for Bayesian optimal design in the CALGB study. The underlying probability model is a hierarchical nonlinear regression model, where the regression mean function (or theoretical concentration-time curve) is given by a compartment model. The mean function is not available in closed form, and must be obtained as the numerical solution to a system of differential equations. The design problem is to choose the optimal sampling times

$d = (d_1, \dots, d_n)$ for the next patient $i = N + 1$, given observed data from previous patients $i = 1, \dots, N$, and a specified set of covariates x_{N+1} for the new patient. The design criterion is to minimize expected posterior variance for some PK quantities of interest, such as the area under the concentration-time curve, or time above a critical threshold concentration, subject to a cost constraint. The cost will represent the inconvenience for the patient if she has to stay longer than usual or even return to the clinic after the drug infusion. Our goal, then, is to devise a limited sampling schedule that will allow us to estimate the PK quantities as precisely as possible while accounting for any potential inconvenience (cost) to the patient.

Simulation-based methods for Bayesian optimal design problems, including medical decision problems, have been considered in several recent papers. Typical examples are Carlin et al. (1998), who use forward simulation to make optimal decisions about continuation of a clinical trial with interim analyses; Wakefield (1996), who derives optimal dosage in a population PK/PD problem; Bois et al. (1999), who suggest a backward-forward approach for selecting sampling times; or Palmer and Müller (1998), who find optimal apheresis schedules. Other Bayesian approaches to optimal design in PK/PD models have been proposed by D'Argenio (1990), who uses historical data to improve designs for estimating parameters in a study of a drug's pharmacokinetics; Atkinson et al. (1993), who find optimal designs for estimating AUC, maximum concentration, and half-life of a drug having a pharmacokinetic model; Merlé et al. (1994), who consider optimal design with a discrete, nonparametric prior based on historical data; Merlé and Mentré (1995), who compare several Bayesian optimal design criteria when designing PK/PD studies. Berry (1993), Berry and Stangl (1996), and Spiegelhalter et al. (1994) discuss general issues of using Bayesian methods in clinical trials. Chaloner and Verdine (1995) provide a general review of Bayesian experimental design.

We propose two simulation-based methods for obtaining the Bayesian optimal design. We use smoothing through Monte Carlo simulations to borrow strength across similar design choices; and we use Markov chain Monte Carlo simulation from an artificially augmented auxiliary probability model to focus simulation at interesting design choices. Related optimal design strategies based on smoothing through Monte Carlo simulations are discussed and applied in Müller and Parmigiani (1995), Erkanli et al. (1998), and Sun et al. (1996), among others. Approaches based on simulation in artificially augmented probability models similar to those proposed here are considered in Bielza et al. (1999), Kuo et al. (1999), and Müller (1999).

Section 2 describes a hierarchical model for the population and individual parameters. In Section 3, we introduce the utility function and describe methods for maximizing expected utility. In Section 4, we implement our model using data from a cohort of 10 patients who received paclitaxel as part of an earlier Phase I study (Venook et al., 1998). We determine optimal sampling strategies for future patients given the data collected on these patients. Finally, conclusions are given in Section 5.

2. The Probability Model

To describe the PK behavior of paclitaxel within patients in the population, we assume a three-stage hierarchical model. At the first stage of the hierarchy, we model the concentration versus time curve for each patient as a nonlinear regression. Let y_{ij} denote the observed log-concentration for patient

$i, i = 1, \dots, N$, at time $t_{ij}, j = 1, \dots, n_i$. The model is:

$$y_{ij} = f(\theta_i, t_{ij}, x_i) + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2), \quad (1)$$

where $f(\cdot)$ is a nonlinear mean function arising from a compartment model, θ_i is a $p \times 1$ parameter vector, and x_i is the drug dosage. The observation errors ϵ_{ij} are assumed to be i.i.d. normal with variance σ^2 . Collectively, we define $y_i = \{y_{ij}\}_{j=1}^{n_i}$ and $f(\theta_i) = \{f(\theta_i, t_{ij}, x_i)\}_{j=1}^{n_i}$ as the observation and mean vector, respectively, for patient i .

At the second level of the hierarchical model, we specify a population distribution for the individual parameter vectors $\theta_i, i = 1, \dots, N$. Let μ and Σ denote the population mean and variance-covariance matrix, respectively. Conditional on (μ, Σ) , we assume:

$$\theta_i \sim N(\mu, \Sigma), \quad (2)$$

$i = 1, \dots, N$, independently. At the third level of the hierarchy, we specify prior distributions for observation variance σ^2 and the population parameters (μ, Σ) . We assume a conjugate inverse-gamma prior for σ^2 , and a conjugate normal-Wishart distribution for (μ, Σ^{-1}) :

$$\sigma^2 \sim IG(a_0, b_0), \quad \mu | \Sigma \sim N(m_0, c_0^{-1} \Sigma), \quad \Sigma^{-1} \sim W(d_0, A_0^{-1}). \quad (3)$$

Here $(a_0, b_0, m_0, c_0, d_0, A_0)$ are hyperparameters selected by the user; a_0 and b_0 are shape and scale parameters for σ^2 ; m_0 is the prior mean vector for μ ; c_0 is a scale factor; d_0 is the degrees of freedom; and $d_0 A_0^{-1}$ is the prior mean of the precision matrix Σ^{-1} . Fig. 1. gives a graphical representation of the full hierarchical model.

2.1. Posterior Simulation

To implement posterior inference, we use Markov chain Monte Carlo (MCMC) simulation to generate samples from the posterior distribution of the model parameters. The sample is obtained by iterating through a list of full conditional posterior distributions, generating a value for each parameter, and using it to replace the current value of the respective parameter. Repeating such iterative sampling from the full conditional posteriors, we eventually obtain an (approximate) Monte Carlo sample from the joint posterior. Appropriate ergodic averages are used to evaluate any desired posterior integral. See, for example, Gelfand and Smith (1990), Tierney (1994), or Gilks et al. (1995) for a review of MCMC methods.

Let $\Theta = (\theta_1, \dots, \theta_N)$ and $Y = (y_1, \dots, y_N)$. We simulate from the joint posterior distribution $p(\Theta, \sigma^2, \mu, \Sigma | Y)$, by iterating over the following list of full conditionals:

$$p(\theta_i | \sigma^2, \mu, \Sigma, Y) \propto N[y_i | f(\theta_i), \sigma^2 I] N(\theta_i | \mu, \Sigma) \quad (4)$$

$$\sigma^2 | \Theta, \mu, \Sigma, Y \sim IG(a, b) \quad (5)$$

$$\mu | \Theta, \sigma^2, \Sigma, Y \sim N(m, c^{-1} \Sigma) \quad (6)$$

$$\Sigma^{-1} | \Theta, \sigma^2, Y \sim W(d, A^{-1}) \quad (7)$$

Here, I denotes the $n_i \times n_i$ identity matrix. For the individual PK parameters in (4), direct simulation of θ_i is impossible due to the nonlinear mean function $f(\theta_i)$ in $N[y_i | f(\theta_i), \sigma^2 I]$. To sample from this distribution, we consider two Metropolis algorithms (see Tierney, 1994): a single-state scheme

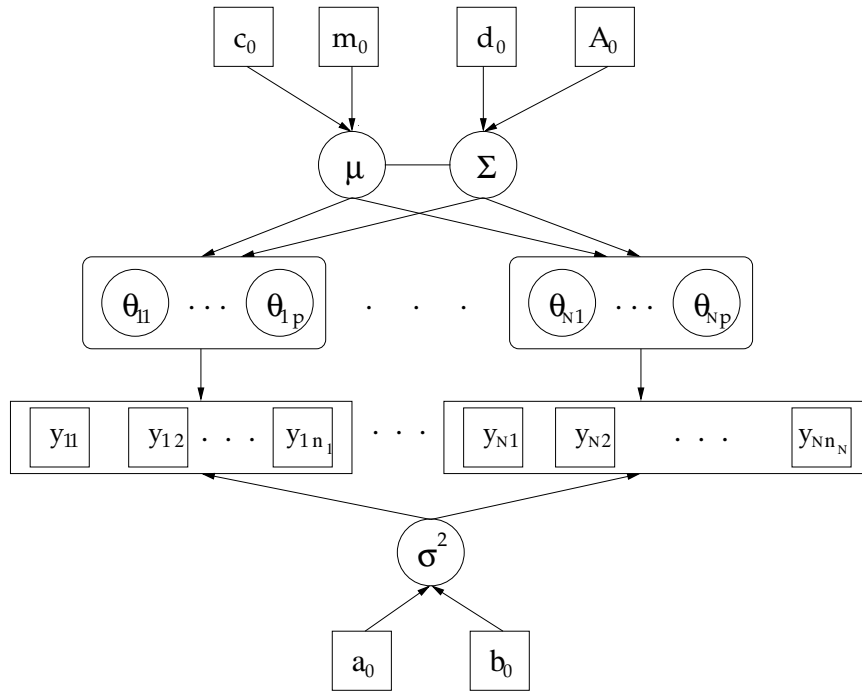


Fig. 1. Graphical representation of the hierarchical model. Circles represent unknown random quantities; squares denote fixed quantities.

where each element of θ_i is updated separately; and a block scheme where the whole vector θ_i is updated at once. We also consider a Gaussian approximation to the full conditional, obtained by combining the prior $N(\theta_i|\mu, \Sigma)$ with a normal likelihood approximation centered at the MLE, $\hat{\theta}_i$, and variance-covariance matrix given by the inverse of the observed Fisher information matrix. For the application in Section 4, we use the Gaussian approximation, since the Metropolis schemes both lead to unacceptably slow MCMC convergence.

The remaining parameters, (σ^2, μ, Σ) , can all be drawn directly from their full conditionals, which are available in closed form. The conditional posterior for σ^2 is an inverse gamma (5), with parameters $a = a_0 + \sum_{i=1}^N n_i/2$ and $b = b_0 + \sum_{i=1}^N \{y_i - f(\theta_i)\}' \{y_i - f(\theta_i)\}/2$. The population parameters (μ, Σ) can be generated jointly from their full conditional distribution by first drawing Σ from (7), and then $\mu|\Sigma$ from (6). The parameters (c, m, d, A) are: $c = c_0 + N$, $m = (c_0 m_0 + N\bar{\theta})/c$, $d = d_0 + N$, and $A = A_0 + S + c_0 N(m_0 - \bar{\theta})(m_0 - \bar{\theta})'/c$, where $\bar{\theta} = \sum_{i=1}^N \theta_i/N$ and $S = \sum_{i=1}^N (\theta_i - \bar{\theta})(\theta_i - \bar{\theta})'/N$ are the sample mean and covariance of the θ_i .

3. Design Methodology

The general decision theoretic setup for a Bayesian design problem is as follows (see DeGroot, 1970, chapter 12). Let $p(\theta|Y)$ denote the distribution of the model parameters θ at the time of the decision, given past data $Y = (y_1, \dots, y_N)$, and let $p_d(y|\theta, Y)$ denote the sampling distribution of the future observations $y = y_{N+1}$ under design d , given θ and Y . To simplify notation, we stop explicitly indicating conditioning on the historical data Y and write $p(\theta)$ and $p_d(y|\theta)$ from now on.

Let $u(d, y, \theta)$ denote the utility function. The utility function represents the reward we would obtain under action d if the values for y and θ were known. A rational decision maker chooses the design d^* which maximizes expected utility. Formally, the optimal design is $d^* = \arg \max_d U(d)$, where

$$U(d) = \iint u(d, y, \theta) p_d(y|\theta) p(\theta) dy d\theta \quad (8)$$

is the expected utility of design d , with the expectation taken with respect to the joint distribution of θ and y , conditional on Y .

3.1. The Utility Function

The utility function is often expressed as payoff minus cost. In a clinical study, the payoff might be represented by the gain in information about the target parameter resulting from the experiment, while cost typically corresponds to the expense incurred by collecting the sample. In our application, we represent payoff by the posterior precision for the next patient's target parameter, $\phi(\theta)$. The sampling cost, $C(d, \alpha, \delta)$, is given by a quadratic function which penalizes sampling times scheduled after a given time δ when the patient leaves the hospital. This formalizes our view that the incremental cost of obtaining samples jumps at time δ and increases steeply after the patient leaves the clinic, while samples taken before that point are essentially cost-free. Combining payoff and cost, the utility function is

$$\underbrace{u(d, y)}_{\text{utility}} = \underbrace{V\{\phi(\theta) | d, y\}^{-1}}_{\text{posterior precision}} - \underbrace{\alpha \sum_{i=1}^n d_i^2 \mathbf{1}\{d_i > \delta\}}_{\text{sampling cost}} \quad (9)$$

where $\theta = \theta_{N+1}$ and $y = y_{N+1}$ are the parameter and observation vector for the next patient, respectively, and $\mathbf{1}\{d_i > \delta\}$ is an indicator for $\{d_i > \delta\}$. We assume that the time δ is known, while the quadratic coefficient α is treated as a tuning parameter that converts sampling cost onto the precision scale.

Since the posterior $p(\theta|d, y)$ is unavailable in closed form, the precision in (9) must be computed numerically. We use an importance sampling approach (see Geweke, 1989) to evaluate $V\{\phi(\theta) | d, y\}$, using the next patient's prior $p(\theta)$ as the importance density. The estimated posterior variance is:

$$\hat{V}\{\phi(\theta) | d, y\} = \sum_{i=1}^{M_p} \left\{ \phi(\theta^i) - \hat{\phi} \right\}^2 w^i / \sum_{i=1}^{M_p} w^i \quad (10)$$

where $\{\theta^i\}_{i=1}^{M_p} \sim p(\theta)$ are samples from the next patient's prior, $\{w^i\}_{i=1}^{M_p} = p(\theta^i|d, y) / p(\theta^i) \propto p_d(y|\theta^i)$ are the importance weights, and $\hat{\phi} = \sum_i \phi(\theta^i) w^i / \sum_i w^i$ is an estimate for the posterior mean, $E\{\phi(\theta) | d, y\}$. The prior samples θ^i are obtained by generating $\theta^i \sim N(\mu^i, \Sigma^i)$, where (μ^i, Σ^i) are taken from the MCMC output.

Given an observation y , the utility $\hat{u}(d, y)$ can be obtained by subtracting the sampling cost from the estimated posterior precision, $\hat{u}(d, y) = 1/\hat{V}\{\phi(\theta) | d, y\} - C(d, \alpha, \delta)$. The expected utility is then obtained by averaging the utilities with respect to the marginal distribution of y :

$$\hat{U}(d) = \frac{1}{M_y} \sum_{i=1}^{M_y} \hat{u}(d, y^i) \quad (11)$$

where $\{y^i\}_{i=1}^{M_y}$ are samples from the next patient's marginal distribution $p_d(y)$.

3.2. Maximizing Expected Utility

We rely on two general techniques to obtain the optimal design. For low-dimensional design spaces, we use a grid-based scheme, where expected utility is computed on a specified grid of designs (see Müller and Parmigiani, 1995). For each design on the grid, we simulate future experiments and calculate expected utilities as in (11). The optimal design is determined by selecting the point with the largest \hat{U} , or by fitting a smooth curve through the expected utilities and graphically obtaining the maximum. Müller and Parmigiani (1995) discuss various choices for parametric and non-parametric smoothers, including nonlinear regression, kernel smoothing, and mixture of Dirichlet processes. In our example, we rely on local-regression (LOESS) smoothers implemented with the S-plus function GAM. The curve-fitting approach formalizes borrowing strength across similar design choices and allows the evaluations of \hat{U} to be based on fewer simulations.

As the dimension of the design space increases, grid-based schemes become computationally infeasible. To increase computational efficiency, we use a Metropolis-Hastings algorithm proposed by Bielza et al. (1999). The key idea is to introduce an artificial probability model, $h(d, \theta, y) \propto u(d, \theta, y) p_d(y|\theta) p(\theta)$, subject to the assumption that the utility function is non-negative and $h(d, \theta, y)$ is integrable. By construction of $h(\cdot)$, the marginal distribution $h(d) = \int h(d, \theta, y) d\theta dy$ is proportional to $U(d)$. Thus, simulation from the joint distribution $h(d, \theta, y)$ produces designs in proportion to the expected utility surface. This allows us to generate better designs more frequently. Following Müller (1999), the sampling algorithm proceeds as follows:

1. Choose an initial design d^1 .
2. Simulate $(\theta^1, y^1) \sim p_{d^1}(\theta, y) = p(\theta) p_{d^1}(y|\theta)$, and evaluate $u^1 = u(d^1, y^1)$.
3. For $i = 1, \dots, M_d$:
 - (a) Generate a candidate design \tilde{d} from a proposal distribution $g(d|\tilde{d}^i)$.
 - (b) Simulate $(\tilde{\theta}, \tilde{y}) \sim p_{\tilde{d}}(\theta, y) = p(\theta) p_{\tilde{d}}(y|\theta)$, and evaluate $\tilde{u} = u(\tilde{d}, \tilde{y})$.
 - (c) Compute:

$$\alpha = \min \left\{ 1, \frac{h(\tilde{d}, \tilde{\theta}, \tilde{y})}{h(d^i, \theta^i, y^i)} \frac{g(d^i|\tilde{d}) p_{d^i}(\theta^i, y^i)}{g(\tilde{d}|\tilde{d}^i) p_{\tilde{d}}(\tilde{\theta}, \tilde{y})} \right\} = \min \left\{ 1, \frac{\tilde{u}}{u^i} \frac{g(d^i|\tilde{d})}{g(\tilde{d}|\tilde{d}^i)} \right\}.$$

- (d) Set:

$$(d^{i+1}, u^{i+1}) = \begin{cases} (\tilde{d}, \tilde{u}), & \text{with probability } \alpha; \\ (d^i, u^i), & \text{with probability } 1 - \alpha. \end{cases}$$

By construction of the Markov chain, the asymptotic distribution is $h(d, \theta, y)$; i.e., sampling focuses on designs with high expected utility. Fitting a smooth curve to simulated pairs (d, u) , we find the optimal design. Note that *all* simulated pairs (d, u) are used for this fit, including those candidates \tilde{d} that were rejected in the Markov chain. A similar approach has been proposed by Lee (2000) in the context of model selection. MCMC simulation is used to focus sampling on models with high posterior probability, but final inference is based on appropriate averages over *all* considered

models, including rejected candidates. In Step 3(a) of the algorithm, the proposal distribution $g(\cdot|\cdot)$ is selected by the user. This function determines how the proposed designs are generated from the current design. If $g(\cdot|\cdot)$ is symmetric, i.e. $g(d^i|\tilde{d}) = g(\tilde{d}|d^i)$, then the acceptance probability in Step 3(c) reduces to $\alpha = \min\{1, \tilde{u}/u^i\}$.

In many examples, the expected utility surface can be quite flat. This weakens the advantage of the Metropolis-Hastings scheme described above for finding the optimal design. Müller (1999) proposed a simulated annealing-type modification of the above algorithm that defines $h_J(d) \propto \{U(d)\}^J$ as the marginal distribution instead of $h(d) \propto U(d)$. The power transformation leaves the solution of the decision problem unchanged, but results in a more peaked criterion surface. To sample from $h_J(d)$, we replace u in steps 1 and 2 of the algorithm by $u_J = \prod_{j=1}^J u(d, y_j)$, where $\{(\theta_j, y_j)\}_{j=1}^J \sim p_d(\theta, y)$, are experiments simulated under design d . See Müller (1999) for details.

4. The Application: CALGB Phase I Study

As part of the Phase I study, plasma samples were collected from 10 patients receiving three-hour infusions of paclitaxel. All of the patients had some degree of liver dysfunction; thus, they were given a relatively low hourly dosage of either 50, 75, or 100 mg/m². Fig. 2 shows the observed log drug concentrations in the plasma for three patients, along with fitted curves based on posterior means from the analysis described in this section. The different curvature of each of the lines highlights the nonlinear PK effects of the drug. Given the observed data, the design problem is to choose the optimal sampling times for the next patient. To solve this problem, we use the expected utility framework described in Section 3.

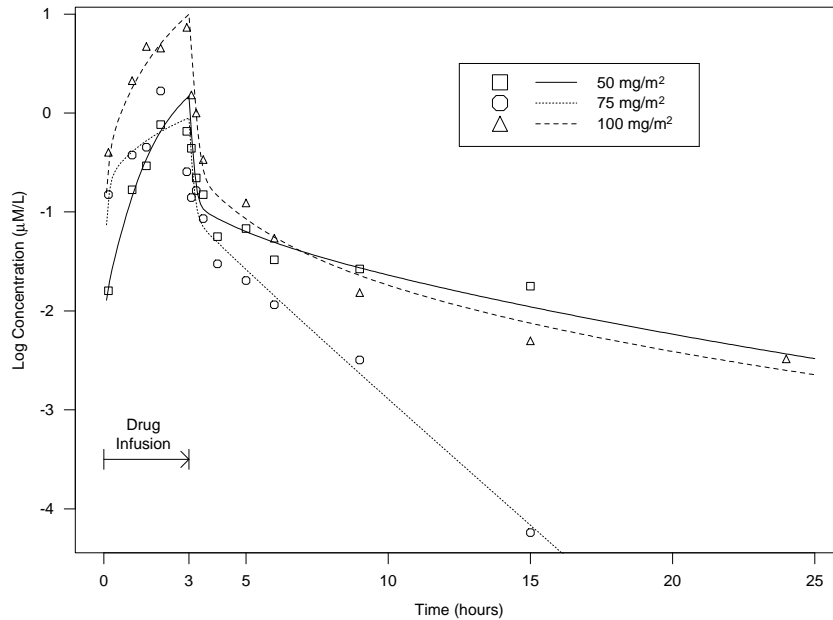


Fig. 2. Log-concentrations and fitted curves for selected patients receiving three-hour infusions of paclitaxel.

4.1. A Three-Compartment Model

To capture the complex pharmacokinetics of paclitaxel, we rely on a three-compartment model with nonlinear Michaelis-Menten kinetics. A diagram of the model is shown in Fig. 3. This model was first proposed by Kearns et al. (1995) for ovarian cancer patients receiving three-hour paclitaxel infusions at a higher dosage. Compartment 1 represents the plasma, in which the drug concentration is measured. Compartments 2 and 3 have no exact anatomical interpretation.

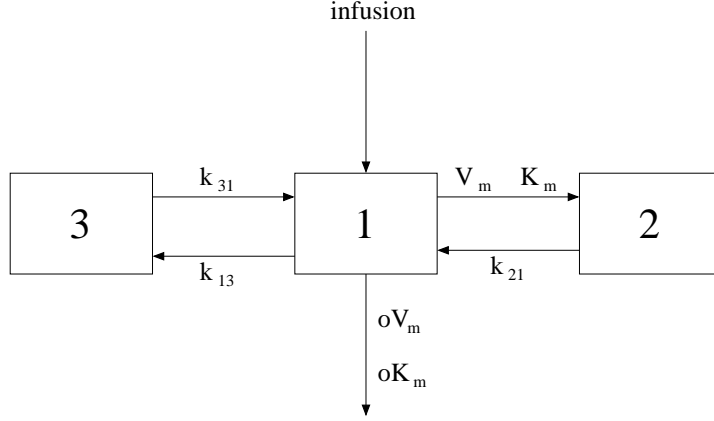


Fig. 3. Structure of the three-compartment model, used to describe the disposition of paclitaxel.

The model is defined by a system of ordinary differential equations (ODEs). Let $X_j(t)$ denote the level of the drug in compartment j at time t , and $R(t, x)$ be the infusion rate at time t and dosage x . The ODE system is given by

$$\begin{aligned} \frac{dX_1(t)}{dt} &= - \left\{ \frac{V_c V_m}{V_c K_m + X_1(t)} + k_{13} + \frac{V_c oV_m}{V_c oK_m + X_1(t)} \right\} X_1(t) + \\ &\quad + k_{21} X_2(t) + k_{31} X_3(t) + R(t, x) \\ \frac{dX_2(t)}{dt} &= \left\{ \frac{V_c V_m}{V_c K_m + X_1(t)} \right\} X_1(t) - k_{21} X_2(t) \\ \frac{dX_3(t)}{dt} &= k_{13} X_1(t) - k_{31} X_3(t) \end{aligned} \quad (12)$$

with initial conditions $X_1(0) = X_2(0) = X_3(0) = 0$. In our example, the infusion rate is $R(t, x) = x$ for $t \leq 3$ hours, and $R(t, x) = 0$ otherwise. The parameters of the model are: V_c , the apparent volume of the central compartment; V_m and oV_m , the maximum velocity of distribution and elimination, respectively; K_m and oK_m , the concentrations associated with one half the maximum velocities of distribution and elimination, respectively; and $k_{jj'}$, the rate constant of transit from compartment j to j' . The parameter vector $\theta_i = (\log V_{c_i}, \log V_{m_i}, \log K_{m_i}, \log k_{21_i}, \log k_{13_i}, \log k_{31_i}, \log oV_{m_i}, \log oK_{m_i})$ collects all $p = 8$ log-PK parameters for patient i .

Given $X_1(t)$, the theoretical drug concentration in the plasma, $C_1(t)$, is:

$$C_1(t) = X_1(t)/V_c. \quad (13)$$

Table 1. Prior and posterior means and 95% probability intervals for the population parameters μ . (Values are reported on the original scale.)

Population Parameter	Prior		Posterior	
	Mean	95% Interval	Mean	95% Interval
V_c	4.00	(2.26, 7.08)	3.61	(1.98, 6.47)
V_m	10.20	(3.98, 26.1)	9.93	(9.08, 10.8)
K_m	0.32	(0.20, 0.52)	0.18	(0.01, 5.54)
k_{21}	0.68	(0.58, 0.79)	0.63	(0.13, 2.96)
k_{13}	2.20	(1.55, 3.11)	2.10	(1.37, 3.34)
k_{31}	0.65	(0.55, 0.77)	0.64	(0.20, 1.89)
oV_m	18.80	(5.80, 60.9)	18.65	(18.1, 19.2)
oK_m	5.50	(2.77, 10.9)	5.77	(3.30, 10.3)

This defines the mean function in our nonlinear regression model (1). That is, given the parameters θ_i and dosage x_i , we set $f(\theta_i, t, x_i) = \log\{C_1(t)\}$ for all sampling times $t = \{t_{ij}\}_{j=1}^{n_i}$. Unfortunately, the solution to (12) is not available in closed form, so the ODE system must be solved numerically. This is done using the *Numerical Recipes* routine STIFBS (Press et al., 1992), which relies on a modified Bulirsch-Stoer method for stiff systems. We use this computationally intensive numerical solution only for posterior predictive simulation for a new patient $i = N + 1$. To evaluate the likelihood function $p(y_i|\theta_i, \sigma^2)$ for $i = 1, \dots, N$, we rely on a multivariate normal approximation.

To complete the model specification, we select hyperparameters for the priors in (3). For the observation variance σ^2 , we set $a_0 = 2.0$ and $b_0 = 0.2$. The prior mean for μ is obtained from the literature: we set m_0 equal to the log values reported in Kearns et al. (1995, p. 19). Following Gelman et al. (1996), we assume that the components of μ are *a priori* independent, and select a coefficient of variation of .2 for all components of μ . We fix $c_0 = 1$ and $d_0 = 10$, which leads to a prior covariance matrix of $A_0 = .04 \text{diag}(m_{01}^2, \dots, m_{08}^2)$, where m_{0k} is the k -th element of m_0 .

4.2. Posterior Inference

We use the MCMC scheme described in Section 2.1 to simulate from the joint posterior distribution based on the $N = 10$ patients. The parameters μ , Σ , and σ^2 are sampled directly from their full conditionals. Updating the patient-specific PK parameters θ_i , $i = 1, \dots, N$, is more difficult because of the computation-intensive evaluation of $C_1(t)$ which involves solving the ODE system (12). To evaluate the likelihood function, $p(y_i|\theta_i, \sigma^2)$, we use the normal approximation described in Section 2.1. The MLE and information matrix for each patient were obtained using the ADAPT II software package of D'Argenio and Schumitzky (1997).

To obtain the results reported in this section, we discarded the first 1000 realizations of the simulated chain and collected the next $M = 10,000$ simulated samples. Convergence diagnostics proposed by Geweke (1991) indicate that the chain converged for all parameters. Table 1 shows the prior and posterior means and 95% probability intervals for the population PK parameters μ . The prior values were obtained analytically, while the posterior values were obtained from the MCMC output.

Table 1 shows only slight differences between the prior and posterior means. This is because the individual MLEs, $\hat{\theta}_i$, are generally in close agreement with the prior mean m_0 . A notable change,

however, occurs for the parameter K_m . The increase in uncertainty about this parameter is due to two patients with extremely high MLEs for K_m . As part of the analysis, we also explored posterior sensitivity to changes in A_0 . We considered two alternative values: $A_0 = \tau \text{diag}(m_{01}^2, \dots, m_{08}^2)$, with $\tau = .01$ and $.16$. These changes slightly affected the widths of the posterior intervals for μ , but did not change the posterior means.

4.3. Choosing Sampling Times

Given the historical data Y , we seek to determine the optimal sampling times $d = (d_1, \dots, d_n)$ for the next patient $i = N + 1$. We assume that the new patient receives a three-hour infusion at an hourly rate of $x = 50 \text{ mg/m}^2$, and consider possible sampling times between 0 and 25 hours. To simplify notation we will in the following write θ for the patient-specific parameters θ_{N+1} of the new patient. We specify a utility function of the form (9), and assume that the patient leaves the hospital at $\delta = 7$ hours from the start of the infusion. The target parameters $\phi(\theta)$ are described below.

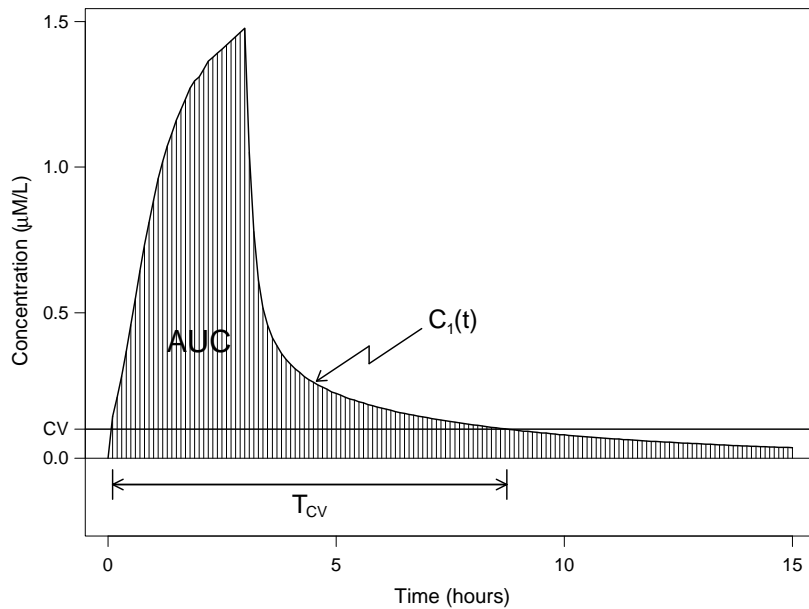


Fig. 4. Fitted concentration-time curve, $C_1(t)$, and the two target parameters: $\phi(\theta) = AUC$, area under the curve; and $\phi(\theta) = T_{CV}$, time above the critical value.

Fig. 4 shows a fitted concentration-time curve, $C_1(t)$, for one of the patients, along with our two target parameters: area under the curve (AUC) and time above critical value (T_{CV}). Following Huizing et al. (1995) and Gianni et al. (1995), we consider two critical values in our analysis: $CV = 0.10$ and $0.05 \mu M/L$. Since $C_1(t)$ can only be evaluated pointwise, we must compute $\phi(\theta)$ numerically. We obtain AUC by summing the area from 0 to 25 hours, computed using a trapezoidal rule, and the area from 25 hours to infinity, computed by log-linear extrapolation of $C_1(24.9)$ and $C_1(25)$. T_{CV} is computed by interpolating the fitted log-concentrations.

Table 2. Optimal designs and expected posterior precisions for $\alpha = 0$ and all $\phi(\theta)$. (Designs are given in hours.)

Number of Samples (n)	AUC		$T_{.10}$		$T_{.05}$	
	d^*	\hat{U}^*	d^*	\hat{U}^*	d^*	\hat{U}^*
0		0.23		0.03		0.01
1	(3)	0.53	(8)	0.08	(25)	0.05
2	(3,25)	0.90	(4,25)	0.15	(24,25)	0.07
3	(1,3,25)	0.99	(3,5,25)	0.21	(1,20,25)	0.09

Before choosing the optimal designs, we compute prior summaries of the next patient’s target parameters. The summaries are based on a sample of size $M_p = 1000$ from the next patient’s prior, $p(\theta)$, which is generated from the MCMC output. The first row of Table 2 shows the prior precision of AUC, $T_{.10}$ and $T_{.05}$. These values can be compared with the expected posterior precisions for $n = 1, 2, 3$ to assess the expected amount of information obtained from the experiment.

Next, we consider grid-based methods to find the optimal design for $n = 1, 2$ and 3 sampling times. We consider each value of n separately, and report the optimal designs for all n and target parameters. First, we consider the case $n = 1$. We use an equally-spaced grid, with possible designs at 30 minutes, 60 minutes, \dots , up to 25 hours. For each design d , we simulate $M_y = 100$ observations, and reweight the $M_p = 1000$ prior samples as in (10) to compute the posterior precision of $\phi(\theta)$. Repeating this procedure for M_y future observations, we obtain an estimate for the expected posterior precision ($\alpha = 0$). The optimal designs d^* and corresponding expected utilities $\hat{U}^* = \hat{U}(d^*)$ are reported in row two of Table 2.

For $n = 2$, we consider a design grid with points every hour from 1 to 25 hours, imposing the constraint $d_1 < d_2$. We use $M_p = 1000$ prior samples and $M_y = 100$ future observations, and repeat the above procedure. The optimal designs are reported in row three of Table 2. Fig. 5(a) shows the fitted expected precision surface ($\alpha = 0$) for AUC, with $n = 2$ sampling times. The surface was obtained by smoothing the simulated expected utilities using the GAM function in Splus. A similar procedure is repeated for $n = 3$ sampling times with an hourly-spaced grid, and we report the results in row four of Table 2.

We implemented the Metropolis-Hastings algorithm described in Section 3.2, with a power transformation $h_J(d) \propto \{U(d)\}^J$ to obtain a more peaked criterion surface. After experimenting with different values of the parameter J , we settled on $J = 10$. This value produced designs clustered tightly enough to accurately estimate the optimal design for $\phi(\theta) = AUC$ and $n = 2$. As proposal distribution $g(\cdot)$, we used a bivariate spherical Gaussian density centered at the current design with a standard deviation of 2 hours. Fig. 5(b) shows a scatterplot of the $M_d = 1000$ sampled designs from this algorithm. Note how the majority of the designs cluster near the point $d^* = (3, 25)$. This confirms our grid-based result that the optimal two-point design includes one sample at the peak and another at 25 hours.

We also ran the Metropolis-Hastings algorithm for $\phi(\theta) = AUC$ and $n = 3$. We increase the power parameter to $J = 25$ to provide a criterion surface that is sufficiently peaked. Again, we find that the optimal design based on the $M_d = 2500$ sampled designs is similar to the optimal design from the grid-based algorithm, with the first sample one hour after the start of infusion, the second at the peak, and the third as late as possible. The method can be extended to $n = 4$ and higher.

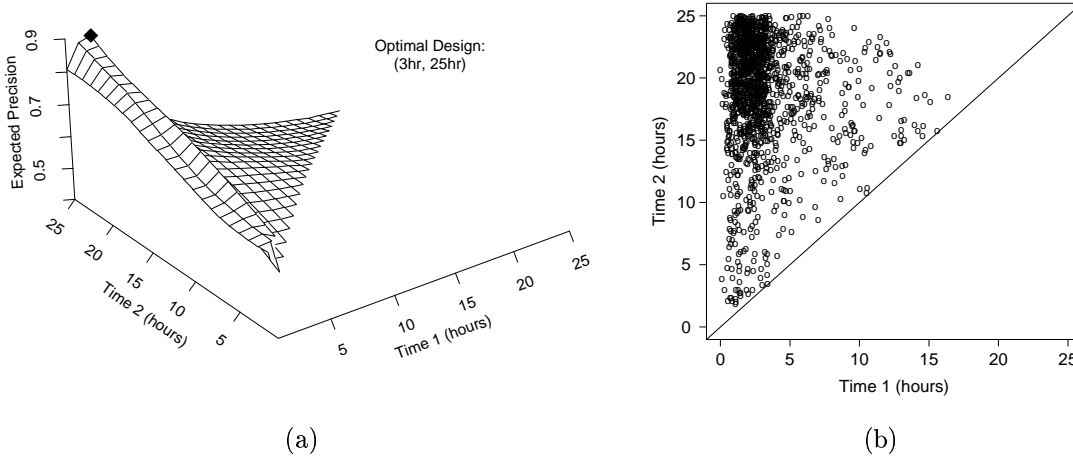


Fig. 5. Optimal designs for $n = 2$ sampling times with $\phi(\theta) = AUC$ and $\alpha = 0$. (a): Fitted expected precision surface from a grid-based scheme. (b): Simulated designs from a Metropolis-Hastings scheme with $J = 10$.

We now incorporate sampling costs by allowing $\alpha > 0$. Table 3 shows the optimal designs and corresponding expected utilities for AUC with $n = 1, 2, 3$ sampling times, and six different values of the cost parameter α . The values d^* and \hat{U}^* reported in the table were obtained by subtracting off the sampling cost from the fitted expected precision surface.

For one-point designs ($n = 1$), the optimal sampling time of $d^* = 3$ hours remains unchanged as we increase α , since no costs are incurred before the patient leaves the hospital. For two-point designs ($n = 2$), a small change in the cost coefficient quickly drives the optimal second sampling time from 25 to 7 hours. After reaching that point, the optimal design remains fixed with one sample at the end of the infusion, and another before the patient leaves the hospital. It is interesting to note that the two-point design has higher expected utility than the one-point design, regardless of the cost coefficient. For three-point schemes, we see a similar pattern to that of $n = 2$, with the final time quickly moving from 25 to 7 hours, while the earlier times remain unchanged. As with $n = 2$, the optimal design always includes one sample at the peak, and another as late as possible, subject to cost constraints.

5. Conclusions

We have explored Bayesian decision-theoretic solutions to the problem of choosing optimal sampling times in PK studies. The proposed methods are simulation based and, therefore, allow a wide variety of probability models and utility functions. We considered two methods for maximizing expected utility: a grid-based algorithm, and a Metropolis scheme based on an augmented probability model.

The advantage of the grid-based scheme is that it provides an estimate of the entire expected utility surface. Having the full surface allows the practitioner to assess the relative efficiency of various suboptimal designs. This can be important in clinical trials when precisely-timed samples are not always possible.

The advantage of the Metropolis algorithm is that it focuses sampling on designs with high

Table 3. Optimal designs and expected utilities for different values of α : $\phi(\theta) = AUC$. (Designs are given in hours.)

Cost Coeff. α	$n = 1$		$n = 2$		$n = 3$	
	d^*	\hat{U}^*	d^*	\hat{U}^*	d^*	\hat{U}^*
.00000	(3)	0.53	(3,25)	0.90	(1,3,25)	0.99
.00002	(3)	0.53	(3,20)	0.82	(1,3,21)	0.89
.00004	(3)	0.53	(3,10)	0.74	(1,3,14)	0.83
.00006	(3)	0.53	(3,7)	0.70	(1,3,9)	0.79
.00008	(3)	0.53	(3,7)	0.70	(1,3,7)	0.75
.00010	(3)	0.53	(3,7)	0.70	(1,3,7)	0.75

expected utility. A disadvantage of the method is that sampling costs cannot be ignored during simulation, as was possible with the grid-based scheme. Neglecting costs would focus the sampling on regions away from the optimal design. This forces us to choose the cost coefficient α before implementing the simulation. A similar problem arises if the investigator wishes to learn about multiple target parameters simultaneously. In this case, a grid-based method might be preferable to Metropolis, as the optimal designs for one parameter may not coincide with those for another.

For this application, $n = 3$ is the highest dimension for which a grid-based method can be implemented within reasonable computation time. For example, the results reported in Table 2 for AUC and $n = 3$ required about 30 hours of CPU time on a 433 MHz DEC Personal Workstation, using code written in C. More than three sampling times could be accommodated by using lower dimensional parameterizations of (fixed) multiple sampling times: for example, first sampling time, number of samples, and gap between sampling times on an equally spaced grid of times.

6. Acknowledgements

This research was partially supported by National Cancer Institute, grant number CA75981; the National Cancer Institute for the Cancer & Leukemia Group B (CALGB) Statistical Office, grant number CA33601; National Science Foundation, grant number DMS-9704934. We gratefully acknowledge the CALGB for providing the data. We wish to thank Nick Polson, two anonymous referees, and the Associate Editor for their valuable comments and suggestions.

References

- Atkinson, A., K. Chaloner, A. Herzberg, and J. Juritz (1993). Optimum experimental designs for properties of a compartmental model. *Biometrics* 49, 325–337.
- Beatty, D. and W. Piegorsch (1997). Optimal statistical design for toxicokinetic studies. *Statistical Methods in Medical Research* 6, 359–376.
- Berry, D. (1993). A case for Bayesianism in clinical trials. *Statistics in Medicine* 12, 1377–1404. with discussion.
- Berry, D. and D. Stangl (1996). Bayesian methods in health related research. In D. Berry and D. Stangl (Eds.), *Bayesian Biostatistics*, pp. 3–66. New York: Marcel Dekker.

- Bielza, C., P. Müller, and D. R. Insua (1999). Monte Carlo methods for decision analysis with applications to influence diagrams. *Management Science* 45(7), 995–1007.
- Bois, F., T. Smith, A. Gelman, H. Chang, and A. Smith (1999). Optimal design for a study of Butadiene toxicokinetics in humans. *Toxicological Sciences* 49, 213–224.
- Carlin, B., J. Kadane, and A. Gelfand (1998). Approaches for optimal sequential decision analysis in clinical trials. *Biometrics* 54, 964–975.
- Chaloner, K. and I. Verdinelli (1995). Bayesian experimental design: A review. *Statistical Science* 10, 273–304.
- D’Argenio, D. and A. Schumitzky (1997). *ADAPT II User’s Guide: Pharmacokinetic/Pharmacodynamic Systems Analysis Software*. Los Angeles, California: Biomedical Simulations Resource.
- D’Argenio, D. Z. (1990). Incorporating prior parameter uncertainty in the design of sampling schedules for pharmacokinetic parameter estimation experiments. *Mathematical Biosciences* 99, 105–118.
- DeGroot, M. (1970). *Optimal Statistical Decisions*. New York: McGraw Hill.
- Egorin, M., A. Forrest, C. Belani, M. Ratain, J. Abrams, and D. van Echo (1988). A limited sampling strategy for Cyclophosphamide pharmacokinetics. *Cancer Research* 49, 3129–3133.
- Erkanli, A., R. Soyer, and A. Angold (1998). Optimal Bayesian two-phase designs for prevalence estimation. *Journal of Statistical Planning and Inference* 66, 171–191.
- Gelfand, A. and A. Smith (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85, 398–409.
- Gelman, A., F. Bois, and J. Jiang (1996). Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *Journal of the American Statistical Association* 91, 1400–1412.
- Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* 57, 1317–1339.
- Geweke, J. (1991). MCMC diagnostics. In J. Bernardo, M. DeGroot, D. Lindley, and A. Smith (Eds.), *Bayesian Statistics 4*. Oxford: Oxford University Press.
- Gianni, L., C. Kearns, A. Giani, G. Capri, L. Vigano, A. Locatelli, G. Bonadonna, and M. Egorin (1995). Nonlinear pharmacokinetics and metabolism of paclitaxel and its pharmacokinetic/pharmacodynamic relationships in humans. *Journal of Clinical Oncology* 13, 180–190.
- Gibaldi, M. and D. Perrier (1982). *Pharmacokinetics* (2nd ed.). New York: Marcel Dekker, Inc.
- Gilks, W., S. Richardson, and D. Spiegelhalter (1995). *Markov Chain Monte Carlo in Practice*. Chapman & Hall.

- Huizing, M., L. van Warerdam, H. Rosing, W. ten Bokkel Huinink, M. Stewart, H. Pinedo, and J. Beijnen (1995). Limited sampling strategies for investigating paclitaxel pharmacokinetics in patients receiving 175 mg/m^2 as a 3-hour infusion. *Clinical Drug Investigation* 9, 244–255.
- Kearns, C., L. Gianni, and M. Egorin (1995). Paclitaxel pharmacokinetics and pharmacodynamics. *Seminars in Oncology* 22, 16–23.
- Kuo, L., R. Soyer, and F. Wang (1999). Optimal design for quantal bioassay via Monte Carlo methods. In J. Bernardo, J. Berger, A. Dawid, and A. Smith (Eds.), *Bayesian Statistics 6*, pp. 795–802. Oxford: Oxford University Press.
- Lee, H. (2000). Model selection for neural network classification. Technical Report 00-18, Institute of Statistics and Decision Sciences, Duke University.
- Merlé, Y. and F. Mentré (1995). Bayesian design criteria: Computation, comparison, and application to a pharmacokinetic and a pharmacodynamic model. *Pharmacokinetics and Biopharmaceutics* 23, 101–125.
- Merlé, Y., F. Mentré, A. Mallet, and A. H. Aurengo (1994). Designing an optimal experiment for Bayesian estimation: Application to the kinetics of iodine thyroid uptake. *Statistics in Medicine* 13, 185–196.
- Müller, P. (1999). Simulation based optimal design. In J. Bernardo, J. Berger, A. Dawid, and A. Smith (Eds.), *Bayesian Statistics 6*, pp. 459–474. Oxford: Oxford University Press.
- Müller, P. and G. Parmigiani (1995). Optimal design via curve fitting of Monte Carlo experiments. *Journal of the American Statistical Association* 90, 1322–1330.
- Palmer, J. and P. Müller (1998). Bayesian optimal design in population models of hematologic data. *Statistics in Medicine* 18, 1613–1622.
- Press, W., S. Teukolsky, W. Vetterling, and B. Flannery (1992). *Numerical Recipes in C* (2nd ed.). Cambridge, MA: Cambridge Press.
- Ratain, M., A. Staibus, and R. Schilsky (1988). Limited sampling models for Amonafide (NCS 308847) pharmacokinetics. *Cancer Research* 48, 4127–4130.
- Ratain, M. and N. Vogelzang (1987). Limited sampling model for Vinblastine pharmacokinetics. *Cancer Treatment Reports* 71, 935–939.
- Spiegelhalter, D., L. Freedman, and M. Parmar (1994). Bayesian approaches to randomised trials. *Journal of the Royal Statistical Society, Series A* 157, 357–416.
- Sun, D., R. Tsutakawa, and W. W. Lu (1996). Experiment for quantal responses: What’s promised and what’s delivered. *Journal of Statistical Planning and Inference* 52, 289–306.
- Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics* 22, 1701–1762.

Venook, A., M. Egorin, G. Rosner, T. Brown, T. Jahan, G. Batist, R. Hohl, D. Budman, M. Ratain, C. Kearns, and R. Schilsky (1998). A Phase I trial of paclitaxel in patients with hepatic dysfunction: Cancer and Leukemia Group B 9264. *Journal of Clinical Oncology* 16, 1811–1819.

Wakefield, J. (1996). Bayesian individualization via sampling-based methods. *Pharmacokinetics and Biopharmaceutics* 24(1), 103–131.