Journal of Pharmacokinetics and Biopharmaceutics, Vol. 23, No. 1, 1995

Bayesian Design Criteria: Computation, Comparison, and Application to a Pharmacokinetic and a Pharmacodynamic Model

Yann Merlé[1,2] and France Mentré[1]

Received August 31, 1994—Final February 16, 1995

In this paper 3 criteria to design experiments for Bayesian estimation of the parameters of nonlinear models with respect to their parameters, when a prior distribution is available, are presented: the determinant of the Bayesian information matrix, the determinant of the pre-posterior covariance matrix, and the expected information provided by an experiment. A procedure to simplify the computation of these criteria is proposed in the case of continuous prior distributions and is compared with the criterion obtained from a linearization of the model about the mean of the prior distribution for the parameters. This procedure is applied to two models commonly encountered in the area of pharmacokinetics and pharmacodynamics: the one-compartment open model with bolus intravenous single-dose injection and the E_{max} model. They both involve two parameters. Additive as well as multiplicative gaussian measurement errors are considered with normal prior distributions. Various combinations of the variances of the prior distribution and of the measurement error are studied. Our attention is restricted to designs with limited numbers of measurements (1 or 2 measurements). This situation often occurs in practice when Bayesian estimation is performed. The optimal Bayesian designs that result vary with the variances of the parameter distribution and with the measurement error. The two-point optimal designs sometimes differ from the D-optimal designs for the mean of the prior distribution and may consist of replicating measurements. For the studied cases, the determinant of the Bayesian information matrix and its linearized form lead to the same optimal designs. In some cases, the pre-posterior covariance matrix can be far from its lower bound, namely, the inverse of the Bayesian information matrix, especially for the E_{max} model and a multiplicative measurement error. The expected information provided by the experiment and the determinant of the pre-posterior covariance matrix generally lead to the same designs except for the E_{max} model and the multiplicative measurement error. Results show that these criteria can be easily computed and that they could be incorporated in modules for designing experiments.

KEY WORDS: Bayesian designs; Bayesian estimation; prior distribution; pharmacokinetics; pharmacodynamics; E_{max} model; nonlinear models.

[1]INSERM U194, CHU Pitié-Salpêtrière, 91, Boulevard de l'Hôpital, 75634 Paris cedex 13 France.
[2]To whom correspondence should be addressed.

0090-466X/95/0200-0101$07.50/0

INTRODUCTION

Parameter estimation in pharmacokinetic or pharmacodynamic models from data from an experiment performed in a given individual is a commonly encountered procedure in the area of optimal drug dosage regimen. This estimation may require several measurements which raises ethical, practical, and economic problems. It has been shown that estimation procedures using complementary information yield more accurate estimates or reduce the number of samples required for a given target reliability (1). One source of such additional information is prior knowledge of the parameters in the population. As, for instance, the parameter distribution in the population when assuming that the parameters are random variates. When a structural model for the process being studied and an error model are defined, several methods are available to estimate the population's parameter distribution from a given set of measurements (2). A Bayesian approach may then be used to estimate the parameters for a new individual from only a few measurements. Many therapeutic applications of this approach have been proposed to individualize dosages of various drug regimens (3). The accuracy of the parameter estimation for a patient clearly depends on the design of the experiment (i.e., sampling times, administered doses) especially when the number of measurements is small.

Several approaches to optimizing experimental design for individual parameter estimation have been proposed (4–7). The principle of these methods is to optimize a criterion with respect to design variates. These design criteria either refer to the concept of Fisher information matrix or to that of the Shannon information. Most of the design criteria are scalar functions of the Fisher information matrix. It should be noted that, for nonlinear models, this matrix depends on the vector of parameters to be estimated. In that case, a nominal value of the parameter vector is chosen. When a prior parameter distribution is available this value is generally taken as the mean. For standard estimation, the D-optimality criterion, which is the determinant of the Fisher information matrix, is the most frequently used design criterion (8). Results concerning D-optimal designs for various pharmacokinetic models have already been given (9).

For Bayesian estimation, work on defining optimal designs has mainly concerned linear models with normal additive errors and normal prior distributions (4,10,11,12). The proposed optimality criteria are scalar functions of the Bayesian information matrix (13). The most well known is the determinant of this matrix which is called the Bayes D-optimality criterion. Few applications have been proposed for nonlinear models.

The other approach originates from Shannon's work (14). He defined a quantity called entropy that measures the uncertainty associated with

random phenomena. From this concept comes the notion of the entropy of a distribution, also called the Shannon information. For classical estimation the design criterion can be taken as the entropy of the distribution of the estimator. For Bayesian estimation and from the notion of entropy, Lindley (15) defined the information provided by an experiment as the difference between the entropy of the prior and the posterior distribution. The expectation of this quantity can be used as a Bayesian design criterion. It has already been used to optimize sampling times for Bayesian parameter estimation in therapeutics (16,17).

Unfortunately, for nonlinear models with continuous prior distributions these Bayesian criteria involve expectations that cannot be expressed analytically. This is surely an obstacle to their applications in designing optimal experiments. To simplify their computation, linearization of the model around the mean of the prior distribution has been proposed (18).

In this paper an approach for simplifying the computation of three Bayesian design criteria (the determinant of the Bayesian information matrix, the determinant of the pre-posterior covariance matrix, and the expected information provided by an experiment) for nonlinear models and continuous prior distributions is first proposed. The results obtained from this approach are compared with those provided by the linearization of the model about the mean of the prior distribution. This approach is then applied to design experiments for estimating the parameters of two models commonly encountered in the area of pharmacokinetics and pharmacodynamics: the one-compartment open model with a single bolus intravenous injection and the E_{max} model. The prior parameter distribution is assumed to be Gaussian. The measurement error is additive, normally distributed with either homoscedastic or an heteroscedastic variance. The optimal experiments obtained from the three Bayesian design criteria are compared with each other for various combinations of the variances of the parameters and of the measurement error. They were also compared to designs provided by standard approaches.

MODELS AND NOTATIONS

Let $y(\xi)$ be a vector of observations obtained after having conducted an experiment with a design ξ on a given individual. The predicted model vector is denoted $f(\xi, \theta)$ where θ is the individual parameter vector. The relationship between the observed vector and the vector of responses predicted by the model is assumed to be

$$y(\xi)=f(\xi, \theta)+\varepsilon \tag{1}$$

Where ε is a random error with zero mean and a variance-covariance matrix $\Omega(\theta, \xi)$. Let $X(\theta, \xi)$ be the matrix of derivatives vector of the model with respect to its parameters:

$$X(\theta, \xi) = \frac{\partial}{\partial \theta} f(\xi, \theta) \tag{2}$$

The distribution of the observations given the vector of parameters θ is denoted $p(y(\xi)|\theta)$ and is fully specified by the distribution of the measurement error. The prior distribution for the parameter is denoted $p(\theta)$. The posterior distribution $p(\theta|y(\xi))$ is estimated from the prior distribution and the distribution of the observations using Bayes theorem

$$p(\theta|y(\xi)) = \frac{p(y(\xi)|\theta))p(\theta)}{p(y(\xi))} \tag{3}$$

where

$$p(y(\xi)) = \int p(y(\xi)|\theta)p(\theta)\, d\theta \tag{4}$$

The mean and the variance-covariance matrix of the prior distribution are denoted μ and C, respectively.

The standard Fisher information matrix $M_F(\theta, \xi)$ is given by

$$M_F(\theta, \xi) = E_{Y|\theta}\left(\frac{\partial}{\partial \theta} \mathrm{Ln}(p(y(\xi)|\theta)) \frac{\partial}{\partial \theta'} \mathrm{Ln}(p(y(\xi)|\theta))\right) \tag{5}$$

If the error ε is Gaussian with constant variance Ω independent of the design and of the parameters then

$$M_F(\theta, \xi) = X(\theta, \xi)'\Omega^{-1}X(\theta, \xi) \tag{6}$$

For linear models, $X(\theta, \xi)$ is the design matrix and is independent of the parameters. In that case and if the error variance is constant

$$M_F(\xi) = X'(\xi)\Omega^{-1}X(\xi) \tag{7}$$

It should be noted that, from the Rao-Cramer inequality, the inverse of the Fisher information matrix is the lower bound of the covariance matrix of any unbiased estimator. The covariance matrix of the estimator reflects its accuracy. To increase the accuracy, design criteria based upon $M_F(\theta, \xi)$ were defined. The most widespread criterion is the D-optimality criterion. A D-optimal design is a design that maximizes $\mathrm{Det}(M_F(\theta, \xi))$.

DESIGN CRITERIA

Three Bayesian design criteria are described: the determinant of the Bayesian information matrix, the determinant of the pre-posterior covariance matrix, and the expected information provided by the experiment.

Determinant of the Bayesian Information Matrix

The Bayesian information matrix $N_F(\xi)$ is given by

$$N_F(\xi) = E_\theta\left[E_{Y|\theta}\left(\frac{\partial}{\partial\theta}\operatorname{Ln}(p(\theta|y))\,\frac{\partial}{\partial\theta'}\operatorname{Ln}(p(\theta|y))\right)\right] \tag{8}$$

or by

$$N_F(\xi) = E_\theta[M_F(\theta,\xi)] + E_\theta\left[\frac{\partial \operatorname{Ln}(p(\theta))}{\partial\theta}\,\frac{\partial \operatorname{Ln}(p(\theta))}{\partial\theta'}\right] \tag{9}$$

The first term of the previous expression is the expectation of the Fisher information matrix over the prior parameter distribution. For a Gaussian prior distribution with a covariance matrix C, the N_F criterion is given by

$$N_F(\xi) = E_\theta[M_F(\theta,\xi)] + C^{-1} \tag{10}$$

From the Rao-Cramer inequality for random parameters the inverse of this Bayesian information matrix is a lower bound of the expected covariance matrix of any unbiased estimator (19).

The expected variance–covariance matrix of the estimator of the random vector θ reaches its lower bound N_F^{-1} only if the posterior distribution is Gaussian which occurs, for example, for linear models with Gaussian additive noise, with constant variance, and a normal prior. The N_F matrix is then given by

$$N_F(\xi) = M_F(\xi) + C^{-1} \tag{11}$$

The Bayesian information matrix, for Bayesian estimation, is analogous to the Fisher information matrix in standard nonlinear regression. Similarly, optimal design criteria based upon $N_F(\xi)$ were defined to increase the accuracy of the estimator.

The design which maximizes the determinant of this matrix is known as Bayes D-optimal.

Determinant of the Pre-Posterior Covariance Matrix

The covariance matrix of the posterior distribution also called posterior covariance matrix reflects the accuracy of the Bayesian parameter estimator after a given experiment has been conducted. Its expectation, $V(\xi)$ is the

expected posterior covariance matrix over all the possible results of the experiments given the design ξ. This matrix is also called the pre-posterior covariance matrix.

Therefore the inverse of $V(\xi)$ can be viewed as a measure of the accuracy and can be used to define optimal design criterion. It is given by

$$V(\xi) = E_Y[\text{Var}(\theta|y(\xi))] \tag{12}$$

$$V(\xi) = E_\theta[E_{Y|\theta}(\text{Var}(\theta|y(\xi)))] \tag{13}$$

where $\text{Var}(\theta|y(\xi))$ is the covariance matrix of the posterior distribution. $V(\xi)$ can also be written as

$$V(\xi) = \int\int \text{Var}(\theta|y(\xi))p(y(\xi)|\theta)\, dy(\xi)p(\theta)\, d\theta \tag{14}$$

The design that maximizes the determinant of the inverse of this matrix is optimal for this design criterion. It should be noted that the pre-posterior covariance matrix is the expected covariance error of the mean a posteriori estimator. Therefore $N_F^{-1}(\xi)$ is a lower bound of $V(\xi)$.

Expected Information Provided by an Experiment

The information is defined from the notion of the entropy of a distribution, namely, the Shannon information. The entropy is a quantity that measures the uncertainty associated with random phenomena.

The entropy of a probability distribution is given by

$$H(p(\theta)) = -E_\theta(\text{Ln}(p(\theta))) \tag{15}$$

or by

$$H(p(\theta)) = -\int \text{Ln}(p(\theta))p(\theta)\, d\theta \tag{16}$$

The entropy of the p-dimensional gaussian distribution $N(\mu, C)$ is given by

$$H(p(\theta)) = \tfrac{1}{2}\,\text{Ln}[(2\pi)^p|C|] + \tfrac{1}{2}p \tag{17}$$

where $|C|$ is the determinant of C.

The information provided by an experiment is given by

$$I(\xi) = H(p(\theta)) - H(p(\theta|y(\xi))) \tag{18}$$

where $H(p(\theta))$ and $H(p(\theta|y(\xi))$ are the entropies of the prior and the posterior distributions, respectively. It quantifies the loss of uncertainty provided by the results of the experiment $y(\xi)$.

As for the previous criterion, the expected information provided by an experiment is therefore the expected value of $I(\xi)$ over all the results of the experiment $y(\xi)$

$$E_Y(I(\xi)) = E_Y[H(p(\theta)) - H(p(\theta|y(\xi)))] \tag{19}$$

$$E_Y(I(\xi)) = H(p(\theta)) - \iint H(p(\theta|y(\xi)))p(y(\xi)|\theta)p(\theta)\, dy(\xi)\, d\theta \tag{20}$$

A design that maximizes the expected information provided by an experiment is optimal for this design criterion. It corresponds to a design that minimizes the expected posterior uncertainty in the parameters after the experiment is conducted.

For linear models, a normal prior and normal additive measurement errors with a constant variance Eq. (18) does not depend on the observations but only on the experimental design. Therefore it is equal to its expectation and is given by

$$I(\xi) = -\tfrac{1}{2}\,\mathrm{Ln}|M_F(\xi) + C^{-1}| - \tfrac{1}{2}\,\mathrm{Ln}|C^{-1}| \tag{21}$$

So in this case, this criterion is equivalent to the determinant of the inverse of the Bayesian information matrix and to the determinant of the pre-posterior covariance matrix.

COMPUTATIONAL METHODS

For continuous prior and posterior distributions, the determinants of the Bayesian information matrix, the pre-posterior covariance matrix, and the expected information provided by an experiment have no analytical expression. The required integrals may be evaluated using stochastic techniques but these are time-consuming. The computation can, however, be simplified by discretization of the prior. This approach is compared with a first-order expansion of the model about the mean of the prior distribution.

Discretization of the Prior Distribution

An algorithm has been developed by Katz (20) to approximate a continuous univariate density function $p(\theta)$ on a closed interval $[a, b]$ by a discrete density distribution which consists of v vectors of parameters θ_k with associated probabilities δ_k.

The principle of the method is to minimize an appropriate measure of the difference between $p(\theta)$ and the discrete distribution. Instead of directly comparing the difference between these two functions, their respective cumulative distributions $F(\theta)$, for $p(\theta)$, and $G(\theta)$, for the discrete

distribution, are compared that leads to the L1 optimality criterion

$$\int_a^b |F(\theta) - G(\theta)|\, d\theta \tag{22}$$

which has to be minimized with respect to θ_k and δ_k and δ_k, $k = 1 \ldots v$.

An initial value for the first location (θ_1) is chosen, then the sequence of θ_k and δ_k, $k = 2 \ldots v$, is obtained from a simple iterative procedure which only requires to calculate $F(\theta_k)$ and $F^{-1}(\theta_k)$. The criterion is then computed by numerial integration (the Simpson method) from the sequence of θ_k and δ_k and is optimized with respect to θ_1 by using a one dimensional search. This method has been extended to multivariate continuous densities (21).

We use this approximation to discretize the prior distribution. With this approximation some of the expectations involved in the criterion become summations that can easily be evaluated.

The Bayesian information matrix given in Eq. (10) for a Gaussian prior can be approximated by

$$N_F(\xi) \approx \sum_{k=1}^{v} \delta_k M_F(\theta_k, \xi) + C^{-1} \tag{23}$$

Similarly, the first integral in Eq. (14) giving the pre-posterior variance–covariance matrix can be approximated by a discrete sum and the equation becomes

$$V(\xi) \approx \sum_{k=1}^{v} \delta_k \int \mathrm{Var}(\theta | y(\xi)) p(y(\xi) | \theta_k)\, dy(\xi) \tag{24}$$

The posterior distribution can be approximated by

$$p(\theta | y(\xi)) \approx \frac{p(y(\xi)|\theta) p(\theta)}{\sum_{k=1}^{v} p(y(\xi)|\theta_k)\delta_k} \tag{25}$$

Therefore $\mathrm{Var}(\theta | y(\xi))$ can be approximated by

$$\mathrm{Var}(\theta | y(\xi)) \approx \sum_{j=1}^{v} \theta_k^2 p(\theta_k | y(\xi)) - \left(\sum_{j=1}^{v} \theta_k p(\theta | y(\xi)) \right)^2 \tag{26}$$

The integral involved in Eq. (24) can be computed by stochastic simulation: Let $y_{i,k}(\xi)$ be samples from the distribution $p(y(\xi)|\theta_k)$. Given n samples for each k, the remaining integral involved in Eq. (24) can be approximated by a discrete sum and becomes

$$V(\xi) \approx \sum_{k=1}^{v} \delta_k \cdot \frac{1}{n} \sum_{i=1}^{n} \mathrm{Var}(\theta_k | y_{i,k}(\xi)) \tag{27}$$

where $\mathrm{Var}(\theta_k|y_{i,k}(\xi))$ is obtained using Eqs. (25) and (26) with $y_{i,k}(\xi)$ replacing $y(\xi)$.

Similarly the entropy of the prior distribution is approximated by

$$H(p(\theta)) \approx -\sum_{k=1}^{v} \delta_k \operatorname{Ln}(p(\theta_k)) \tag{28}$$

The posterior distribution can also be discretized. It has the same locations θ_k than the prior but the associated frequencies $\delta_k(y(\xi))$ are given by

$$\delta_k(y(\xi)) = \frac{\delta_k p(y(\xi)|\theta_k)}{\sum_{j=1}^{v} \delta_j p(y(\xi)|\theta_j)} \tag{29}$$

Then, the entropy of the posterior distribution is approximated by

$$H(p(\theta|y(\xi))) \approx -\sum_{k=1}^{v} \delta_k(y(\xi)) \operatorname{Ln}(p(\theta_k|y(\xi))) \tag{30}$$

The expectation of the information is approximated by

$$E_Y(I(\xi)) \approx H(p(\theta)) - \sum_{k=1}^{v} \delta_k \int H(p(\theta|y(\xi)))p((y(\xi))|\theta_k)\, dy(\xi) \tag{31}$$

The integral involved in this criterion can be approximated by stochastic simulation using the same technique as was used for the pre-posterior covariance matrix.

$$E_Y(I(\xi)) \approx H(p(\theta)) - \sum_{k=1}^{v} \delta_k \cdot \frac{1}{n} \sum_{i=1}^{n} H(p(\theta|y_{i,k}(\xi))) \tag{32}$$

where $H(p(\theta))$ and $H(p(\theta|y_{i,k}(\xi)))$ are obtained by using Eqs. (28) and (30), respectively.

Linearization of the Model

By expanding the model to the first order about the mean μ of the prior distribution an approximate linearized form of the model can be obtained and is given by

$$f(X, \theta) \approx f(X, \mu) + X(\xi, \mu)'(\theta - \mu) \tag{33}$$

The model linearization considerably simplifies the computation of these three criteria which become equivalent for a Gaussian prior distribution and

an additive error with a constant variance. In that case, the linearized form of the Bayesian information matrix is given by

$$N_F^{(L)}(\xi) = M_F(\mu, \xi) + C^{-1} \tag{34}$$

SIMULATIONS

We now compute the criteria we have just described in order to compare the criteria, the designs, and the approximations. The models, the prior distributions, the measurement error models, and the studied experimental designs are presented successively.

Pharmacokinetic and Pharmacodynamic Models

Our attention has been focused on one pharmacokinetic model: the one-compartment open model with single-dose bolus intravenous injection. Its expression is given by

$$f(t, \theta) = \frac{D}{V} \exp\left(-\frac{CL}{V} t\right) \tag{35}$$

where the design variable $\xi = t$ is the time of sampling and $\theta = (V, CL)$. This model involves two parameters: V, the apparent volume of distribution and CL the elimination clearance. D is the administered dose which was fixed to 1.

The pharmacodynamic model was assumed to be an E_{max} model. Its expression is given by

$$f(D, \theta) = \frac{E_{max} \cdot D}{D_{50} + D} \tag{36}$$

here the design variable $\xi = D$ is the dose x and $\theta = (E_{max}, D_{50})$. This model involves two parameters, E_{max} and D_{50}, which are the maximum drug-related effect and that value of the dose that causes 50% of the maximum effect, respectively.

Studied Experimental Designs

The design variates of interest were either the sampling times (for the pharmacokinetic model) or the doses (for the pharmacodynamic model). Our attention was restricted to designs with one or two measurements.

For designs including a single measurement, only the criteria developed in a Bayesian estimation context (namely, the N_F criterion, the expected information provided by an experiment and the pre-posterior covariance matrix) can be computed. These criteria were calculated for 20 sampling

times or doses included in the interval [0, 1] and [0, 20] for the one-compartment and the E_{max} models, respectively.

For designs including two measurements the determinant of the Bayesian information matrix after linearization, the determinant of the Bayesian information matrix, the determinant of the pre-posterior covariance matrix and the expected information provided by an experiment were computed. For the pharmacokinetic model 200 time pairs were explored between $t=0$ and $t=1$ with steps of 0.05 under the assumption $t_2 \geq t_1$. For the E_{max} model the same procedure was taken for doses in the interval [0, 20] using steps of 1.

Measurement Error Models

Two error models for observations y (concentrations or effects) performed at the design variate ξ (times or doses) were considered. For the homoscedastic model, it was assumed, as in Eq. (1), that these errors are additive and independently normally distributed with zero mean and a constant variance equal to

$$\text{var}(\varepsilon) = \Omega = \sigma^2 I \tag{37}$$

For the heteroscedastic model, it was assumed that

$$y = f(\xi, \theta)(1 + \varepsilon) \tag{38}$$

where ε is normally distributed with constant variance as in Eq. (37). However, this model was approximated as follows in order to be in the setting previously described of a constant error model: First a logarithmic transformation was applied; second, a first-order expression of $\text{Ln}(1+\varepsilon)$ about zero was done. Therefore it was assumed that

$$\text{Ln}(y(\xi)) \approx \text{Ln}(f(\xi, \theta)) + \varepsilon \tag{39}$$

where $\text{Ln}(y(\xi))$ was taken as the data and $\text{Ln}(f)$ as the model.

The variance of the error σ^2 was taken to be 0.0225. For the heteroscedastic error this corresponds to a coefficient of variation of 15%.

Prior Distributions

The prior distributions of the parameters were supposed to be Gaussian because this assumption is often made when estimating a prior. The expectation and the variance–covariance matrix of this prior was supposed to be known and the covariances between the parameters were assumed to be equal to zero.

For each model a nominal value for the hyperparameters was chosen. For the one-compartment open model with single-dose bolus intravenous

injection the mean value and the variance of the clearance were fixed at 0.5 and 0.01, respectively. For the volume of distribution, the mean and the variance were of 0.2 and 0.0016, respectively. For these two parameters this corresponds to a coefficient of variation of 20%. For the pharmacodynamic model the mean values and the variance for the E_{max} parameter were fixed to 1 and 0.09. The same values were chosen for the other parameter. For these two parameters this corresponds to a coefficient of variation of 30%.

The algorithm developed by Katz was used for discretizing the prior distributions. The number of locations for the discrete distributions was fixed at $15 \times 15 = 225$. For each location 100 vectors of observations y were simulated. The same simulated samples were used for the computation of every criterion.

The impact of altering (i) the ratio of the measurement error variance to the variances of the parameters and of altering (ii) the ratio of the variance of one parameter to the other was investigated on (a) the variations of the criteria with respect to the experimental designs; (b) the optimal designs.

Influence of the Ratio of the Intra- to the Interindividual Variability

The variances in the prior population were fixed to their nominal values. For the one-compartment open model five values of the measurement error variances were studied (0.005, 0.01, 0.05, 0.1, 0.5). For the E_{max} model, six values of the variance were investigated (0.0014, 0.0028, 0.00563, 0.0113, 0.045, 0.09).

Influence of the Ratio of the Variance of One Parameter to the Other

The variance of one parameter was fixed to its nominal value while the other was altered. For the E_{max} model, six values of the coefficient of variation of the other parameter were studied: 5, 10, 15, 20, 40, and 50%. The same procedure was applied to the other parameter. For the one-compartment open model four values were studied: 10, 15, 30, and 40%.

Analysis of the Results

For each value of the population parameters and combination of measurements, the design criteria were computed and compared. The criterion variations have been plotted with respect to the experimental designs.

Given a design criterion ϕ, either the determinant of N_F or the determinant of the pre-posterior covariance matrix, or the expected information provided by an experiment, a usual measure to compare a design ξ to the

optimal one ξ_ϕ is the ϕ-efficiency defined as (4)

$$Ef = \left(\frac{\phi(\xi_D)}{\phi(\xi_\phi)} \right)^{1/p} \tag{40}$$

where p is the number of parameters to estimate. This efficiency is defined from the ratio of the criteria for the design ξ and the optimal design ξ_ϕ, therefore it is always lower than 1. It is taken to the power $1/p$ in order to be normalized with respect to the number of parameters. The efficiency reflects the loss of information in the current design ξ instead of the optimal one. In our study we have evaluated the efficiency of the D-optimal design for the mean of the prior with respect to each design criterion. The gain of efficiency of a Bayesian optimal design instead of the D-optimal design, defined as

$$GEf = \frac{1 - Ef}{Ef} \tag{41}$$

was reported.

To examine if the pre-posterior covariance matrix $V(\xi)$ is close to its lower bound $N_F^{-1}(\xi)$, the ratio $\mathrm{Det}(V)/\mathrm{Det}(N_F^{-1})$ was also computed. This ratio quantifies the normality of the posterior distribution and therefore the nonlinearity of the model with respect to the design under study.

RESULTS

For the homoscedastic error, results on two-point and one-point designs are presented. For the heteroscedastic error only one-point designs were studied.

One-Compartment Open Model

The results on the Bayesian designs for this model are presented in Table I.

Homoscedastic Error

It was shown (9) that the two-sample D-optimal design in that case is $t_1 = 0$, $t_2 = CL/V$. Therefore, for the mean of the prior, the D-optimal design is $t_1 = 0$ and $t_2 = 0.4$. The loss of efficacy may be very large if the samples are not performed at optimal times. The optimal designs obtained from the determinant of the pre-posterior covariance matrix, the N_F criterion without or with linearization, and the expected information provided by an experiment are generally the same as the D-optimal designs for the mean of the prior distribution except when the ratio of the variance of the measurement error to the variance of the parameters is very large [$\mathrm{Var}(CL) = 0.01$,

Table I. Optimal Designs for the One-Compartment Open Model, Homoscedastic, Heteroscedastic Error, and Several Values of the Hyperparameters[a]

		Var(CL)[b]				Var(V)[b]				σ^b				
	Nominal	0.0025	0.00563	0.0225	0.04	0.0004	0.0009	0.0036	0.0064	0.005	0.01	0.05	0.1	0.5
Designs														
Two-point homo.	0–0.4	0–0.4	0–0.4	0–0.4	0–0.4	0–0.4	0–0.4	0–0.4	0–0.4	0–0.4	0.04	0–0.4	0–0.4	0–0[c]
One-point homo.	0	0	0	0	0	0	0	0	0	0	0	0	0	0
One-point hetero.	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ratio														
Two-point homo.	2.55	2.14	2.33	2.71	3.59	1.27	1.65	8.24	476.00	2.16	2.59	2.53	2.00	1.52
One-point homo.	1.89	1.85	1.89	2.00	2.71	1.06	1.28	5.45	281	1.53	1.85	1.76	1.71	1.5
One-point hetero.	3.4	2.10	2.63	6.40	9.50	13.00	1.66	43.00	47.5	20.00	7.31	1.80	1.33	1.00

[a] For each combination are given: the optimal designs, the ratio $\mathrm{Det}(V)/\mathrm{Det}(N_F^{-1})$ and, for two-point designs only, the GEf of the D-optimal designs.
[b] Other variances fixed to their nominal values.
[c] Except for $\mathrm{Det}(N_F^{(2)})$: 0–0.4.

$\mathrm{Var}(V)=0.0016$, $\sigma^2=0.5$]. In that case, the optimal Bayesian designs are designs with replicates: The two measurements must be repeated ideally at $t=0$ (except for the linearized form of the N_F criterion). For this case the gains of efficiency when performing the Bayesian optimal design instead of the D-optimal design are of 15, 11, and 7% for the N_F criterion, the determinant of the pre-posterior covariance matrix and the expected information, respectively. The variations of the expected information with respect to the sampling times are presented in Fig. 1 for the nominal values of the parameters. The variations of the other criteria are quite similar. The importance of an early measurement (ideally at time $t=0$) is shown. By contrast, the choice of the second sampling time seems less crucial. For the case where the measurements must be repeated twice at $t=0$ the variations of the four Bayesian design criteria are also very similar. For the optimal designs, the ratios of the determinant of the pre-posterior covariance matrix to the determinant of the inverse of the Fisher information matrix are within the interval [1.27, 476]. The highest value was found for large coefficients of variations of the volume of distribution.

For the one-point designs, the four Bayesian design criteria lead to the same result: The measurement has be be performed as soon as possible, theoretically at time $t=0$. The variations of these four criteria with respect

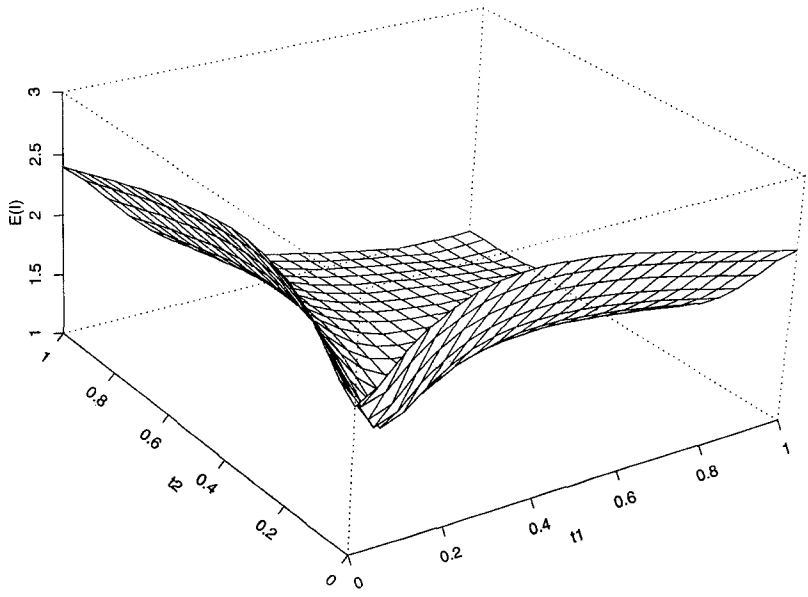

Fig. 1. One compartment open model, two-point designs, and homoscedastic error: Variations of $E(I)$ with respect to the two measurement times for the nominal values.

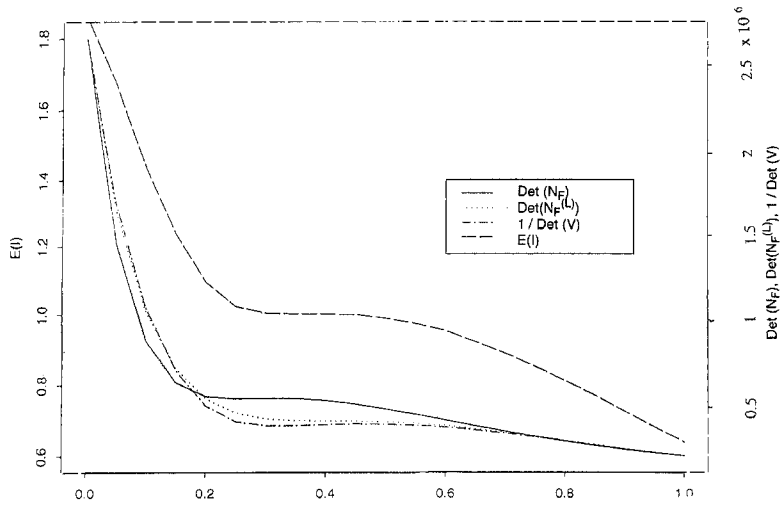

Fig. 2. One-compartment open model, one-point designs, and homoscedastic error: Variations of the four Bayesian design criteria with respect to the measurement time for the nominal values.

to the sampling time are shown in Fig. 2 for the nominal values of the hyperparameters. The efficacies of the designs decrease rapidly within the interval [0, 0.2]. The expected information and the inverse of the determinant of the pre-posterior covariance matrix vary very similarly. The pre-posterior covariance matrix is close to its lower bound for the optimal designs. That is not always true for nonoptimal designs particularly when the ratio of the variance of the measurement error to the variance of the parameters is very low [$\mathrm{Var}(CL)=0.01$, $\mathrm{Var}(V)=0.0016$, $\sigma^2=0.005$]. In that case the ratio of the two criteria reached 11 for sampling times within the interval [0.3, 0.4]. The variations of this ratio are shown in Fig. 3.

Heteroscedastic error

The optimal designs consist in measuring the concentrations as late as possible. It should be noted that for every studied case, the four criteria lead to the same design. The variations of the four Bayesian design criteria for the nominal values of the parameters are presented in Fig. 4. A comparison of Fig. 2 to Fig. 4 shows the importance of the error model on the criterion variations and on the optimal one-point design. The determinant of the pre-posterior covariance matrix is generally close to its lower bound except when the variance of the measurement error is low or for great values of the variances of the two parameters. It should be noted that because of the

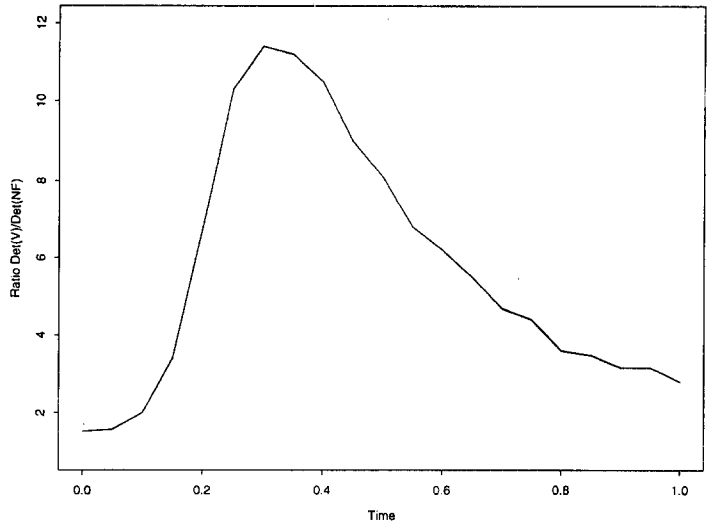

Fig. 3. One-compartment open model, one-point designs, and homoscedastic error: Variations of the ratio $\mathrm{Det}(V)/\mathrm{Det}(N_F^{-1})$ of the determinants of the pre-posterior covariance matrix and of the N_F criterion with respect to one measurement time for $\mathrm{Var}(CL)=0.01$, $\mathrm{Var}(V)=0.0016$, and $\sigma^2=0.005$.

stochastic simulation the determinant of the pre-posterior covariance matrix was found in some cases to be slightly lower than its lower bound.

E_{max} Model

Homoscedastic Error

The results are presented in Table II. It was shown that the D-optimal two-point design is theoretically $D1=D_{50}$ and $D2$ at an infinite value. In practice, the dose $D2$ has to be fixed to a large value, and, in that case, $D1$ is slightly lower than D_{50}. For the mean values of the parameters given the design constraints, the D-optimal design is $D1=1$, $D2=20$. The variations of the D-optimality criterion with respect to the two doses show that $D1$ must be equal or close to D_{50} otherwise the loss of efficacy can be large. The choice of $D2$ is less important provided this dose is large. For each studied combination, the four Bayesian design criteria generally lead to the same optimal designs: The same dose has to be repeated twice and must be as large as possible. In some cases the gain of efficacies when performing optimal Bayesian designs rather the D-optimal design for the mean parameters can be large. For instance, for $\mathrm{Var}(E_{max})=0.09$, $\mathrm{Var}(D_{50})=0.0025$ and $\sigma^2=0.0225$ the gains of efficacies are 41, 59, 41, and 19% for $\mathrm{Det}(N_F)$, $\mathrm{Det}(N_F^{(L)})$, $1/\mathrm{Det}(V)$, and $E(I)$, respectively.

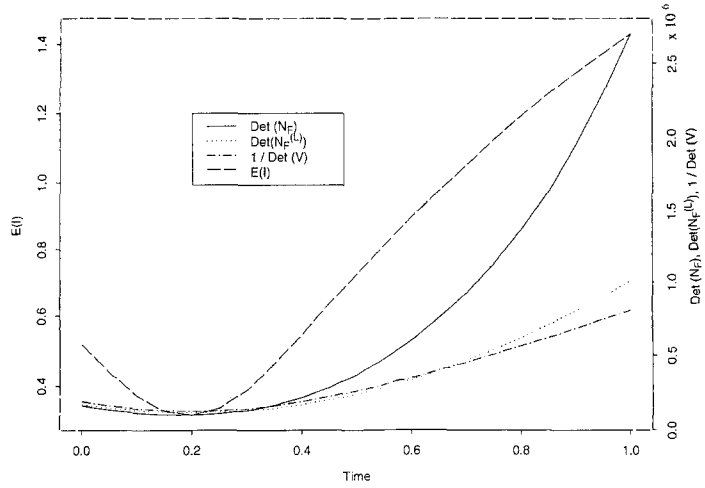

Fig. 4. One-compartment open model, one-point designs, and heteroscedastic error: Variations of the four Bayesian criteria for the nominal values.

In some cases the optimal designs are however different. Thus, the D-optimal design for the mean of the prior distribution has to be performed when (i) the ratio of the variance of the parameters to the variance of the measurement error is low: $\mathrm{Var}(E_{\max})$ and $\mathrm{Var}(D_{50})$ at their nominal values and σ^2 lower or equal to 0.00563; (ii) the ratio of the coefficient of variations of $E_{\max}$ to D_{50} is lower than 0.50. It should be noted that the dose must be equal to D_{50} and given twice when the previous ratio is equal to 1. The variations of the four Bayesian criteria with respect to second dose when the first one is equal to 20 are presented in Fig. 5.

The optimal one-point designs generally consist in giving the largest dose. In this example the dose must be greater than 5 otherwise the loss of efficacy can be important. The variations of the Bayesian criteria, with respect to the dose, for the nominal values of the hyperparameters are presented in Fig. 6. These criteria vary similarly. In these cases the determinant of the pre-posterior covariance matrix is always close to its lower bound. It should be noted that for low values of the variation coefficient of the $E_{\max}$ parameter (5%, 10%) the optimal designs obtained from the four Bayesian criteria consist in giving by contrast a dose equal to D_{50}.

Heteroscedastic Error

For the studied cases the determinant of the Bayesian information matrix and of its linearized form lead to the same optimal designs: The given dose should be as low as possible. Thus, the optimal dose found from

Table II. Optimal One-Point and Two-Point Designs for the E_{max} Model, Homoscedastic Error, and Several Values of the Hyperparameters[a]

		$Var(E_{max})$[b]					
	Nominal	0.0025	0.01	0.0225	0.04	0.16	0.25
Designs							
Two-point homo.	20–20[c]	1–1	1–20	1–20[d]	20–20	20–20	20–20
One-point homo.	20	1	1	20	20	20	20
Ratio							
Two-point homo.	0.90	0.96	0.96	1.00	0.92	1.00	0.90
One-point homo.	0.77	0.76	0.90	0.87	0.83	0.83	0.87
GEf							
$Det(N_F^{(L)})$	0	0.11	0	0.08	0.18	0.30	0.30
$Det(N_F)$	0.15	0.13	0	0.02	0.09	0.12	0.16
$1/Det(V)$	0.22	0.11	0	0	0.14	0.26	0.29
$E(I)$	0.09	0.31	0	0	0.09	0.10	0.07
		$Var(D_{50})$[b]					
	Nominal	0.0025	0.01	0.0225	0.04	0.16	0.25
Designs							
Two-point homo.	20–20[c]	20–20	20–20	20–20	20–20	20–20	1–20[c]
One-point homo.	20	20	20	20	20	20	20
Ratio							
Two-point homo.	0.90	0.93	0.94	1.07	0.92	1.66	1.25
One-point homo.	0.77	0.93	0.92	0.93	0.93	0.93	0.91
GEf							
$Det(N_F^L)$	0	0.47	0.42	0.41	0.37	0.12	0.02
$Det(N_F)$	0.15	0.45	0.42	0.58	0.33	0.14	0
$1/Det(V)$	0.22	0.47	0.42	0.41	0.35	0.05	0
$E(I)$	0.09	0.22	0.22	0.19	0.16	0.14	0
		σ^{2b}					
	Nominal	0.0014	0.00282	0.00563	0.0113	0.045	0.09
Designs							
Two-point homo.	20–20[c]	1–20	1–20	1–20[e]	20–20	20–20	20–20
One-point homo.	20	20	20	20	20	20	20
Ratio							
Two-point homo.	0.90	1.45	1.31	1.19	0.95	1.00	1.33
One-point homo.	0.77	0.93	0.94	0.91	0.99	0.75	0.66
GEf							
$Det(N_F^{(L)})$	0	0	0	0.02	0.14	0.28	0.25
$Det(N_F)$	0.15	0	0	0	0.14	0.22	0.21
$1/Det(V)$	0.22	0	0	0	0.12	0.25	0.34
$E(I)$	0.09	0	0	0	0.23	0.16	0.25

[a]For each combination are given: the optimal designs, the ratio $Det(V)/Det(N_F^{-1})$ and, for two-point designs only, the GEf of the D-optimal designs.
[b]Other variances fixed to their nominal values.
[c]Except for $Det(N_F^{(L)})$: 1–20.
[d]Except for $Det(N_F$ and $Det(N_F^{(L)})$: 20–20.
[e]Except for $Det(N_F^{(L)})$: 20–20.

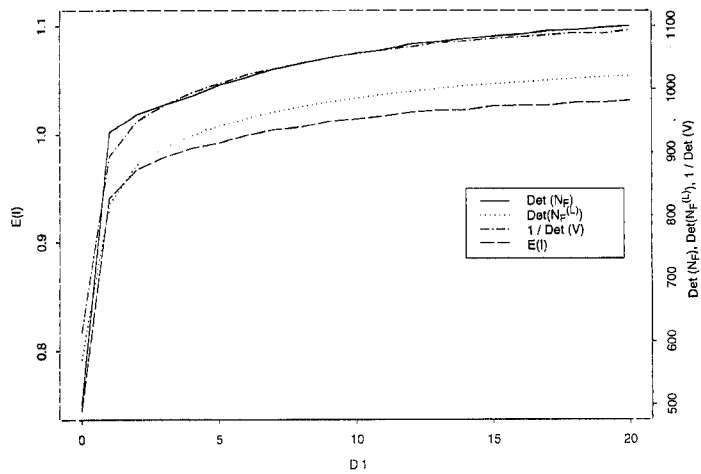

Fig. 5. E_{max} model, two-point designs, and homoscedastic error. Variations of the four Bayesian criteria with respect to *D2* for the nominal values when $D1 = 20$.

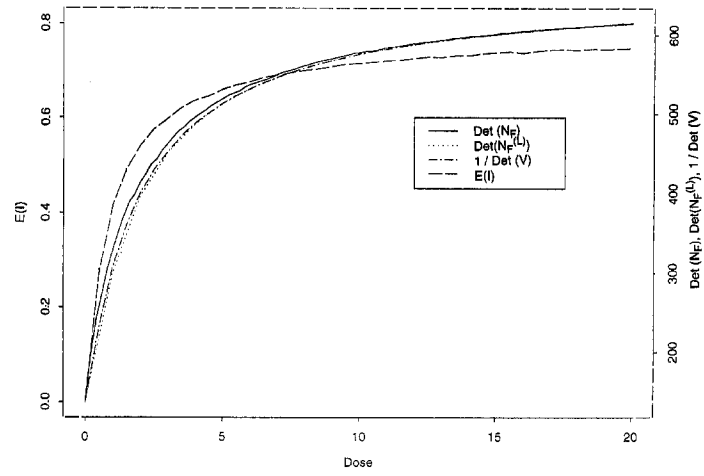

Fig. 6. E_{max} model, one-point designs, and homoscedastic error. Variations of the four Bayesian criteria with respect to the dose for the nominal values.

these criteria is equal to zero which has obviously no sense in practice. The optimal designs obtained from the expected information and the determinant of the pre-posterior covariance matrix generally consist in giving either the lowest dose or a dose equal to D_{50}. These two criteria cannot be computed for $D = 0$. It should be noted that for a large variance of the E_{max} parameter

(0.25) as well as for low values of σ^2 (0.0014 and 0.00282), these two criteria do not lead to the same optimal designs: $D=1$ for the expected information provided by an experiment and $D=20$ for the determinant of the pre-posterior covariance matrix. The variations of the four Bayesian criteria for the nominal values of the parameters are very similar to those presented for the homoscedastic case. The pre-posterior covariance matrix may be far from its lower bound especially for low values of the variances of the two parameters and of the variance of the measurement error.

DISCUSSION

In this paper a method for simplifying the computation of three Bayesian design criteria in case of nonlinear models with respect to their parameters has been proposed. This method has been applied assuming a Gaussian prior distribution and for two models often encountered in the area of pharmacokinetics and pharmacodynamics. This simplification procedure could be applied to other distributions, it is easy to perform and substantially decreases the time of computation especially for the expected information provided by an experiment as well as for the determinant of the pre-posterior covariance matrix. The approximation of the prior distribution to compute expectations could have been obtained by using a random variate generation technique. However, the sample required to obtain the same accuracy in the criterion computation might be larger than the value used ($15 \times 15 = 225$ locations) in our simulation procedure after discretization of the prior distribution. The number of locations chosen for the discrete prior distributions leads to a reasonable approximation of the continuous prior distributions. Thus, the values of the entropies of the prior distributions obtained from Eq. (28) were very close to those provided by the exact formula Eq. (17). The same remark can be made for the variance. Furthermore, with this number of locations, the discretization procedure is not time-consuming. A greater number of locations ($30 \times 30 = 900$) has been tried but it did not lead to a better approximation and took more time. However, the computation of the pre-posterior covariance matrix and of the expected information still requires stochastic simulations. In Eqs. (27) and (32) the value $n=100$ was chosen. It corresponds to a reasonable value: The variations of the criterion with respect to the design variates are regular and a higher value we have tried ($n=1000$) does not lead to a better approximation of the criterion and is obviously time-consuming.

Our results indicate that for every studied criterion, the value of the variance of the prior parameter distribution and of the measurement error may modify the optimal designs. The Bayesian two-point designs sometimes differ from the D-optimal design for the mean of the parameter and may

consist in a replication of the measurements. The gain of efficiency, when performing the Bayesian design, rather than the D-optimal design for the mean parameters may be large. The results provided by our approach have been also compared with those obtained from the linearization of the model about the mean of the prior distribution. This latter approach is to compute the determinant of the Bayesian information matrix for the mean of the prior distribution. In this case the computation of the criteria does not require stochastic simulations and is straightforward. In this study, our results show that the optimal designs obtained after linearization of the two models are mostly the same as those provided by the determinant of the Bayesian information matrix. In contrast, the optimal designs may differ from those obtained from the expected information provided by an experiment and from the determinant of the pre-posterior covariance matrix especially for the E_{max} model and a multiplicative measurement error. Furthermore our results indicate that the pre-posterior covariance matrix can be far from the lower bound. The optimal designs obtained from the expected information differ rarely from those provided from the pre-posterior variance–covariance matrix and only for the E_{max} model and the multiplicative measurement error.

General principles for Bayesian design of experiments for these two models can be drawn from this study. Thus, for the one-compartment open model and homoscedastic error, the two-point optimal design is the D-optimal design for the mean of the prior distribution except when the variance of the measurement error is large (the samples should be performed as soon as possible). In the case of one-point designs the sample must be performed as soon as possible. In contrast, for heteroscedastic error, the sample must be performed as late as possible. For the E_{max} model and homoscedastic error the two doses should be as large as possible except for a low variance of the measurement error (the D-optimal design for the mean parameters has to be performed) and when the variance of the E_{max} parameter is smaller than the one of the D_{50} parameter (the dose should be either equal to the mean D_{50} and repeated twice or the D-optimal designs should be performed). In the case of one-point designs the dose has to be as large as possible except when the variance of the E_{max} parameter is lower than those of the D_{50} parameter. For heteroscedastic error general principles are difficult to draw because the various Bayesian design criteria do not lead to the same optimal design and depend upon the values of the variances of the parameters and of the measurement error. Optimal experiments vary with respect to the model and prior assumptions so that the optimization of Bayesian design criteria should be performed in each case.

Bayesian forecasting programs often involve modules for designing experiments. The latter generally compute the D-optimal design for the

mean of the prior distribution. This approach is not appropriate for several reasons. Thus, the design obtained from this approach is optimal only if the individual parameters are equal to the mean of the prior distribution. In addition, the D-optimal design criterion has not been developed in a Bayesian estimation context but in a standard one; consequently it is not defined if the number of measurements is lower than the number of parameters to estimate. Furthermore, linearizing the model about the expectation of the Gaussian prior distribution does not take into account the shape of the latter. Bayesian design criteria should be employed instead because they take into account prior knowledge both for estimation and design procedures. It should be noted that the determinant of the pre-posterior covariance matrix and the expected information provided by an experiment could also be used in a standard context of estimation. In that case the design criteria are, respectively, the determinant of the variance–covariance matrix of the estimator (whose lower bound is the inverse of the Fisher information matrix) and the entropy of the distribution of this estimator. The studied Bayesian design criteria have rarely been used up to now for planning experiments in the area of pharmacodynamics and pharmacokinetics. The linearized form of the N_F criterion has been used in software to design optimal experiments (22). For Gaussian prior, the linearization procedure does not take into account the shape of the distribution. For that reason, discretization seems to be preferable to the linearization procedure.

Our results show that the determinant of the Bayesian information matrix as well as its linearized form are very easy to compute in case of a Gaussian prior distribution. However, for other continuous prior distributions, the Fisher information matrix involved in the general expression of the N_F criterion as well as the second term of Eq. (9) can be more difficult to calculate. In addition, our results show that the pre-posterior covariance matrix can be far from its lower bound N_F^{-1}. Furthermore the two other Bayesian design criteria involve less assumptions and their computation can be simplified by the discretization procedure. Therefore the expected information provided by an experiment and the determinant of the pre-posterior covariance matrix are more appropriate to design experiments. Nevertheless, the N_F optimal design could be chosen as an initial point in the algorithm used to optimize the two other Bayesian design criteria. In a context of parameter estimation, the expected information provided by an experiment can be more difficult to interpret than the pre-posterior covariance matrix which reflects the accuracy of the posterior distribution. However we would use the information criterion to design experiments that involve a more general measure of the uncertainty of a distribution. We applied this criterion to find the optimal two-point design to estimate the kinetics of iodine thyroid uptake (16).

In the present work, the optimal designs were found by performing a combinatorial computation of all the results. This latter approach can be easily used for designs with a small number of measurement (which is usually the case for Bayesian estimation) especially when the measurement times or the doses have to be chosen in a restricted set because of practical reasons. In the other cases optimization methods have to be employed. Since, the determinant of the Bayesian information matrix can be computed quickly, a global optimization may be performed. In constrast, for the determinant of the pre-posterior covariance matrix and the expected information, stochastic approximation methods seem to be more suitable for these criteria which involve expectation terms (23). Besides, with these algorithms, the criteria do not need to be evaluated at each iteration.

Other improvements could be performed. Thus, optimal designs obtained from the Bayesian design criteria could be individualized by taking into account the values of the covariates of the patient collected before the experiment. In that case optimal Bayesian designs may differ from one individual to the other (24). Such methods could be incorporated in modules of Bayesian forecasting programs for designing experiments. Such study as well as others conducted in the area of the population design (25) may contribute to better collect data.

REFERENCES

1. J. L. Steimer, A. Mallet, and F. Mentré. Estimating interindividual pharmacokinetic variability. In M. Rowland, L. B. Sheiner, and J. L. Steimer (eds.), *Variability in Drug Therapy: Description, Estimation and Control,* Raven Press, New York, 1985, pp. 65–111.
2. J. L. Steimer, M. E. Ebelin, and J. Van Bree. Pharmacokinetic and pharmacodynamic data and models in clinical trials. *Eur. J. Drug. Metab. Pharmacokin.* **18**:61–76 (1993).
3. A. H. Thomson and B. W. Whiting. Bayesian parameter estimation and population pharmacokinetics. *Clin. Pharmacokin.* **22**:447–467 (1992).
4. E. Walter and L. Pronzato. Qualitative and quantitative experiment design for phenomenological models—A survey. *Automatica* **26**:195–213 (1990).
5. V. V. Fedorov. *Theory of Optimal Experiments,* Academic Press, New York, 1972.
6. D. M. Steinberg and W. G. Hunter. Experimental design: Review and comment. *Technometrics* **26**:71–97 (1984).
7. A. C. Atkinson. Developments in the design of experiments. *Int. Statist. Rev.* **50**:161–177 (1982).
8. R. C. St John and N. R. Draper. D-Optimality for regression designs: A review. *Technometrics* **17**:15–23 (1975).
9. H. J. Rostami. Experimental design for pharmacokinetic modeling. *Drug Inform. J.* **24**:299–313 (1990).
10. I. Verdinelli. Advances in experimental design. In *Bayesian Statistics (4),* Oxford Unversity Press, Oxford, 1992, pp. 467–481.
11. H. N. Bandemer, N. Wolfgang, and J. Pilz. Once more: optimal experimental design for regression models. *Statistics* **18**:171–217 (1987).
12. I. Verdinelli and J. B. Kadane. Bayesian designs for maximizing information and outcome. *J. Am. Statist. Assoc.* **87**:510–515 (1992).

13. K. Chaloner. Optimal Bayesian experimental design for linear models. *Ann. Statist.* **12**:283–300 (1984).
14. C. E. Shannon. A mathematical theory of communication. *Bell System Tech. J.* **27**:379–423, 623–656 (1948).
15. D. V. Lindley. On the measure of information provided by an experiment. *Ann. Math. Statist.* **13**:986–1005 (1956).
16. Y. Merlé, F. Mentré, A. Mallet, and A. Aurengo. Designing an optimal experiment for Bayesian estimation: Application to the kinetics of iodine thyroid uptake. *Statist. Med.* **13**:185–196 (1994).
17. Y. Merlé, F. Mentré, A. Mallet, and A. Aurengo. Computer-assisted individual estimation of radioiodine uptake in Grave's disease. *Comput. Meth. Prog. Biomed.* **40**:33–41 (1993).
18. N. R. Draper and G. W. Hunter. The use of prior distributions in the design of experiments for parameter estimation in non-linear situation: Multiresponse case. *Biometrika* **54**:662–665 (1968).
19. J. Pilz. *Bayesian Estimation and Experimental Design in Linear Regression Models*, Wiley, New York, 1991.
20. D. Katz. Discrete approximations to continuous density functions that are L1 optimal. *Comput. Statist. Data Anal.* **1**:175–181 (1983).
21. D. Katz. Discrete approximation of multivariate densities with application to Bayesian estimation. *Comput. Statist. Data Anal.* **2**:27–36 (1984).
22. A. Iliadis, A. C. Brown, and M. L. Huggins. APIS: A software for model identification, simulation and dosage regimen calculations in clinical and experimental pharmacokinetics. *Comput. Meth. Prog. Biomed.* **38**:227–239 (1992).
23. L. Pronzato and L. E. Walter. Robust experiment design via stochastic approximation. *Math. Biosci.* **75**:103–120 (1985).
24. F. Mentré, A. Mallet, J. L. Steimer, and F. Lokiec. An application of the population pharmacokinetics to the clinical use of cyclosporine in bone marrow transplant patients. *Transpl. Proc.* **20**:466–470 (1988).
25. Y. Hashimoto and L. B. Sheiner. Designs for population pharmacodynamics: value of pharmacokinetic data and population analysis. *J. Pharmacokin. Biopharm.* **19**:333–353 (1991).