TOXICOLOGICAL SCIENCES **49,** 213–224 (1999)

Optimal Design for a Study of Butadiene Toxicokinetics in Humans

Frédéric Y. Bois,*[,1] Thomas J. Smith,† Andrew Gelman,‡ Ho-Yuan Chang,† and Andrew E. Smith†

Lawrence Berkeley National Laboratory, Berkeley, California; †Harvard School of Public Health, Boston, Massachusetts; and ‡Statistics Department, Columbia University, New York, New York

Received September 4, 1998; accepted February 28, 1999

The derivation of the optimal design for an upcoming toxicokinetic study of butadiene in humans is presented. The specific goal of the planned study is to obtain a precise estimate of butadiene metabolic clearance for each study subject, together with a good characterization of its population variance. We used a two-compartment toxicokinetic model, imbedded in a hierarchical population model of variability, in conjunction with a preliminary set of butadiene kinetic data in humans, as a basis for design optimization. Optimization was performed using Monte Carlo simulations. Candidate designs differed in the number and timing of exhaled air samples to be collected. Simulations indicated that only 10 air samples should be necessary to obtain a coefficient of variation of 15% for the estimated clearance rate, if the timing of those samples is properly chosen. Optimal sampling times were found to closely bracket the end of exposure. This efficient design will allow the recruitment of more subjects in the study, in particular to match prescribed levels of accuracy in the estimate of the population variance of the butadiene metabolic rate constant. The techniques presented here have general applicability to the design of human and animal toxicology studies.

Key Words: **butadiene population toxicokinetics; human inhalation experiments; Markov chain Monte Carlo simulations; optimal experimental design.**

An individual's risk from exposure to a metabolically activated or detoxified agent is defined by the intensity and duration of internal exposure in combination with physiologic and genetic factors that control metabolic enzyme activity. Individuals with the highest exposure intensity *and* the highest rate of activation and/or the slowest deactivation rates will have the highest levels of the active metabolite, and over time the highest risk. Consequently, the average risk per unit of exposure for epidemiologic studies will depend in part on the population's distribution of metabolic phenotypes. Thus, knowledge of human metabolic rates can be critical to the estimation of risk from occupational and environmental exposures, and for extrapolating risk from one population to another. Unfortunately, in most cases, these data are unavailable. Animal bioassay data are commonly used to estimate the quantitative risks of chemical exposures and fill the gaps in human data. Animal risks can be extrapolated to humans using a variety of toxicologic approaches, most recently physiologic-pharmacokinetic (PBPK) modeling (National Research Council (NRC), 1994; Ramsey and Andersen, 1984; Reitz *et al*., 1988). The common approach is to "scale-up" rodent models to humans, making some allowance for species differences in metabolic rates on the basis of *in vitro* data (when available). However, because of differences in metabolism, there can be considerable interspecies variation in potency in animal bioassays, which prevents simple extrapolations to humans. Without human data, it is impossible to determine whether humans are similar to the most sensitive species, or not. Consequently, it is important to obtain human metabolic data to characterize human risks of exposure for prevalent hazards.

Timed sets of samples of exhaled breath, urine, or blood during and after a measured exposure are useful to estimate human metabolic rates. These data can potentially be obtained by 3 strategies: evaluation of environmentally or occupationally exposed subjects, or laboratory exposures of volunteers. The first 2 approaches are conceptually possible, but highly impractical. In the field, it is rarely possible to obtain a temporal control of exposure, or to collect precisely timed series of biological samples. Properly executed laboratory studies of brief, low-level exposures can produce high quality data to estimate apparent metabolic rates and intersubject variability. These studies must use protocols approved by a human-subjects review committee to insure that risks for the subjects are minimal. Study approval is possible if several conditions are met: (1) environmental and occupational exposures to the agent are common, (2) the lifetime risk of the lab exposure is less than one per million using standard, most-sensitive species risk assessment procedures (California Environmental Protection Agency Air Resources Board, 1992), and (3) volunteers are fully informed that the exposure may produce a small increase in their lifetime risk of a health effect.

Our primary goal in this work was to define an efficient testing protocol for assessing 1,3-butadiene metabolic clearance in human volunteers.

[1] To whom correspondence should be sent at B3E–INSERM U444, Faculté de Médecine St. Antoine, 27 rue Chaligny, 75012 Paris, France. Fax: 011 33 1 44 73 84 62. E-mail: fbois@diana.lbl.gov.

1,3-Butadiene as a Model Compound

Environmental and occupational exposures to butadiene are common, because it is a component of urban air pollution (0.001 ppm) (Mullins, 1990; U.S. Environmental Protection Agency, 1989), cigarette smoke (2 ppm) (Brunnemann *et al*., 1990), and gasoline vapors (0.01 ppm) (Rappaport *et al*., 1987), while occupational exposures range from 0.01 ppm to 300 ppm (Fajen *et al*., 1990). The cumulative exposures range from 8.8 ppm × h (*i.e*., 8.8 ppm for 1 h or 4.4 ppm for 2 h etc.) per year for urban air pollution to 2,000 ppm × h per year for a 1 ppm average occupational exposure.

Considerable data have been accumulated to suggest that butadiene is an animal carcinogen, and may be a human carcinogen. The U.S. EPA classified 1,3-butadiene as a "probable human carcinogen" (Group 2B) on the basis of 2 rodent inhalation studies, where the rodent exposures ranged from 18,000 to 750,000 ppm × h accumulated in daily exposures over a year or more (U.S. Environmental Protection Agency, 1985). The International Agency for Research on Cancer (IARC, 1992) has given 1,3-butadiene a rating of 2A, which is "probably carcinogenic in humans." This determination was based on both animal and epidemiologic studies. Recently, IARC has re-examined the epidemiologic evidence on butadiene and concluded that it was not possible to separate potential effects of styrene and other exposures from those of butadiene, so the classification was left at 2A. The U.S. Occupational Safety and Health Administration (OSHA) recently reduced the allowable daily occupational exposure to 2.0 ppm for 8 h (16 ppm × h) for butadiene. There is considerable interest in the human risks of butadiene exposures.

The California EPA (1992) performed a population risk assessment assuming a continuous 1 ppm exposure for 70 years (6.1×10^5 ppm × h) based on tumor rates for mice, which are the most sensitive species. The lifetime risk of cancer ranged from 1.4×10^{-4} to 0.8. However, if the rat tumor rates are used instead of the mouse, the lifetime risks are orders of magnitude lower because of differences in rat and mouse metabolism of butadiene. Based on *in vitro* metabolism measurements on mouse, rat, and human tissue samples, it appears that humans may be more like rats than mice (Bond *et al*., 1996; Csanady *et al*., 1992; Johanson and Filser, 1996). Thus, it is important to measure human metabolic rates to assess human risks.

Optimizing Design of Volunteer Exposures to Butadiene

The purpose of the study reported here was to help develop a human testing protocol that would maximize the precision of estimates for butadiene metabolic rate and its population variance, while minimizing the number of breath samples taken (and specifying their optimal timing). As we sought to expose the subjects to a minimal risk of toxicity, exposure concentration and length were not optimized but set, according to criteria of risk and feasibility. Preliminary experiments in Dr. Chang's laboratory, using an approved human subjects protocol, used inhalation exposures of 5.0 ppm for 2 h, 10 ppm × h, equivalent to about one day of occupational exposure allowed under OSHA rules. These data, whose collection is described below, were used as a training set for our analysis. Taking advantage of improved analytical sensitivity, the exposure protocol for the planned experiments has been set to 2.0 ppm for 20 min (i.e., 0.67 ppm × h). We were left with defining the number and timing of breath samples for these future experiments. To that effect, we present a new method for experimental design optimization, based on Monte Carlo simulations. Our method has the advantage of following naturally from the Markov chain Monte Carlo techniques applicable to population pharmacokinetic models (Bois *et al*., 1996a,b; Gelman *et al*., 1996; Wakefield, 1996).

MATERIALS AND METHODS

Preliminary data. Eight human volunteers were recruited and tested in Dr. Chang's laboratory at the National Cheng Kung University, College of Medicine, in Taiwan, Taiwan. The tests were conducted, under informed consent, with an Institutional Review Board-approved human subjects protocol. The consent form clearly indicated that 1,3-butadiene is a suspected human carcinogen and that exposure to it may cause a small increase in the subject's lifetime risk of cancer. An inhalation exposure of 5 ppm for 2 h, (i.e.,10 ppm × hr, equivalent to about one day of occupational exposure allowed under OSHA rules), was chosen. That exposure was the minimum that could be precisely measured, and was well below Taiwan's allowable occupational exposure of 10 ppm per 8-h work day for a working lifetime. The exposure was generated using a permeation tube, dynamic calibration apparatus (Metronics Assoc. Inc., Palo Alto, CA) that is designed to produce parts-per-million concentrations. In this apparatus, a liquefied calibration gas slowly permeates through the walls of a Teflon tube, which is sealed with double stainless steel balls at either end (O'Keefe and Ortman, 1966). The operating temperature is tightly controlled, so the permeation tube releases gas at a fixed low rate, and can be used as a primary standard by periodically weighing the tube. After testing the minute ventilation rate of each subject before experiment, the default flow rate was set at 11.50 L/min, which included the flow rate of the permeation apparatus (0.18 L/min) and a dilution flow rate of 11.32 L/min. Three permeation tubes were used in the permeation chamber of the standard gas generator and each generated 42.5 μg/min at the flow rate of 0.18 L/min. Together with the dilution flow this gives a final concentration of 5.0 ppm of butadiene. This gas flowed into a reservoir Tedlar® bag, which provided sufficient capacity to meet the cyclic demand of the subject's breathing. To accommodate differences in the basic ventilation rate of each subject, a bypass was also included to release excess gas flow before it entered the reservoir bag. The flow rates out of the standard gas generator and for the dilution system are critical determinants of the accuracy of the butadiene concentration generated; therefore, the flow rates were calibrated and measured before and after each exposure experiment. Additionally, before and after each exposure, an air sample was collected from the exposure system to measure, by gas chromatography (GC) analysis, the exact concentration of butadiene being produced. This is a very safe exposure system, which has little risk of overexposure from errors in dilution or accidental releases.

Subjects were evaluated by a physician to verify their health and absence of medical problems that might affect their metabolism, and to verify that they understood the possible cancer risk. Body weight (*Bw*) was recorded (Table 1). No specific dietary information was collected at the time of testing. Subjects were seated and exposed by facemask. A one-way valve in the mask allowed the subjects to draw in breathing air with butadiene and to exhale into either a hood or a 15 L. Tedlar® breath-collection bag. At most, 41 timed one-min breath samples were collected: 11 during the 2-h exposure (every 10 min); and

TABLE 1
Physiological Characteristics Measured for Each of the Eight Human Subjects Studied in Preliminary Experiments

Subject ID	Body weight (Bw, kg)	Minute ventilation rate (K_{in}, L/min)	Blood/air partition coefficient (P_{ba})
A	71	8.4	1.5
B	68	7.1	1.8
C	60	6.0	0.9
D	50	4.5	1.3
E	67.5	7.8	1.4
F	70	7.8	1.3
G	64	7.6	1.1
H	48	5.0	1.3

30 during the 57-min post-exposure period (every min for the first 10 min, every 2 min for the next 20 min, then every 3 min for the remaining 27 min). Some subjects had less than 41 samples collected (a few sample times were missed), the lowest number of samples collected was 32. To avoid residual contamination, different masks were used during exposure and after exposure. Minute ventilation (K_{in}) was estimated for each subject by measuring the volume of breath collected in the breath samples (Table 1). Replicate determinations of K_{in} had a coefficient of variation (CV) of 2%. Immediately after collection, the breath was drawn from the bag through a tributyl catechol (TBC, to prevent self-polymerization) treated charcoal tube at 100 ml/min. Butadiene adsorbed on the charcoal tubes was desorbed with heptane and analyzed by GC using a flame ionization detector and a 50 m, 0.32 mm, OD fused silica porous layer open tubular column previously coated with Al_2O_3/KCl. The limit of detection was 0.01 ppm in a 10 L breath sample and the coefficient of variation (mean/SD) was 5.6% for replicate samples.

For each subject, the butadiene blood/air partition coefficient, P_{ba}, was measured (Table 1) with a method developed by Dr. Chang. A measured volume of venous blood (approximately 20 ml) was added to a flask with a septum, two side arms, and a valve in the bottom. A known amount of pure butadiene was added through a septum, and the flask was equilibrated with occasional shaking for 30 min at 37°C. Approximately 90% of the blood was removed and the volume determined. The remaining butadiene was flushed from the flask with nitrogen, into a TBC-treated charcoal tube, under a flow of 50 ml/min for 40 min. The butadiene collected was measured by the GC method described above. The partition coefficient was calculated from the mass of butadiene added (M_{added}), the measured mass of butadiene flushed from the air and residual blood in the flask (M_{a+r}), the volume of the flask (V_{flask}), and the volumes of blood added (VB_{added}) and removed (VB_{rm}). The mass of butadiene in the blood removed is $M_b = M_{added} - M_{a+r}$, and the volume of air in the flask is $V_{air} = V_{flask} - VB_{added}$. The formula used is:

$$P_{ba} = \frac{M_b/VB_{rm}}{\left(M_{a+r} - M_b\left(\frac{VB_{added}}{VB_{rm}} - 1\right)\right)/V_{air}}$$

Replicate determinations of P_{ba} had a CV of 16%. No compensation was attempted for variations in blood lipids associated with meals consumed before the testing was performed. However, the normal Taiwanese diet is a low fat diet, so the change in blood lipids is likely to have been small.

Toxicokinetic/statistical model. A standard two-compartment toxicokinetic model was selected to describe the above data and conduct the design optimization. Preliminary work (not reported) had shown that a two-compartment model in which metabolism took place in the peripheral compartment led to the same estimates of metabolic rate; that model was therefore not considered here. In addition, a one-compartment model was unable to correctly describe the data (this was assessed using the statistical techniques described below). We chose the simplest compartmental model able to correctly describe the data at hand. A discussion of the possible physiological meaning of the 2 compartments is not in order, *a priori*, and will be presented in the Results section, on the basis of the posterior parameter values obtained. The model adopted here assumes that metabolic elimination takes place in the central compartment (Fig. 1). It corresponds to the following set of differential equations:

$$\begin{cases} \dfrac{\partial Q_{central}}{\partial t} = K_{in}C_{inhaled} + \dfrac{K_{cp}}{P_{pc}}\dfrac{Q_{periph}}{V_p} - \left(K_{cp} + \dfrac{K_{in}}{P_{ca}}\right)\dfrac{Q_{central}}{V_c} - K_{met}Q_{central} \\ \dfrac{\partial Q_{periph}}{\partial t} = K_{cp}\dfrac{Q_{central}}{V_c} - \dfrac{K_{cp}}{P_{pc}}\dfrac{Q_{periph}}{V_p} \\ \dfrac{\partial Q_{metab}}{\partial t} = K_{met}Q_{central} \end{cases} \quad (1)$$

where $Q_{central}$ and Q_{periph} are the quantities of butadiene in the central and peripheral compartments, respectively (in mmoles). $C_{inhaled}$ is the inhalation concentration (in mmoles/L); K_{in} is the minute ventilation rate (in L/min); K_{cp} is the rate constant for distribution from the central to the peripheral compartment (in L/min); P_{pc} is the peripheral over central partition coefficient (dimensionless); V_c and V_p are the volumes (in L) of the central and peripheral compartments, respectively; P_{ca} is the central over air partition coefficient (dimensionless); K_{met} is the butadiene metabolic rate constant (in min^{-1}), which here is the parameter of primary interest. The concentration of butadiene in exhaled air (in mmol/L) was given by the algebraic equation:

$$C_{exhaled} = \frac{0.7 \cdot Q_{central}}{P_a \cdot V_c} + 0.3 \cdot C_{inhaled} \quad (2)$$

which assumes a physiological pulmonary dead space of 30%.

Central volume was scaled to body weight, Bw (in kg), and its scaling coefficient, sc_V_c, was the parameter actually estimated:

$$V_c = sc\_V_c \times Bw \quad (3)$$

It can be noticed, in Equation 1 that P_{pc} and V_p are not separately identifiable, and therefore their product was defined as one parameter, $P_{pc}V_p$.

The above equations were coded for the MCSim simulation software and

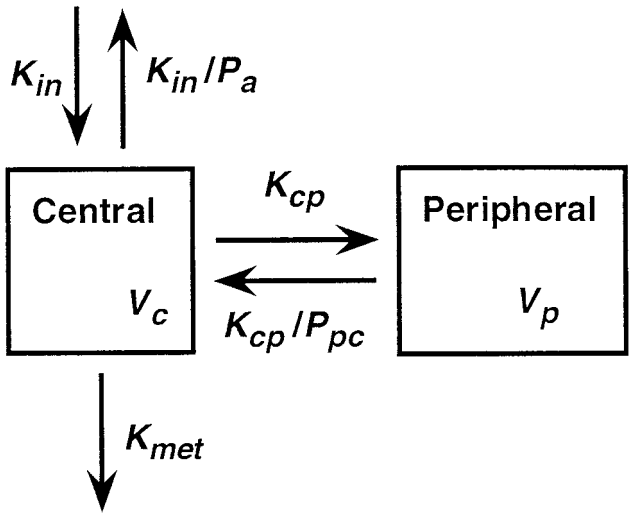

FIG. 1. Schematic representation of the toxicokinetic 2-compartment model used. The model parameters are first order rate constants, partition coefficients and volumes (see text).

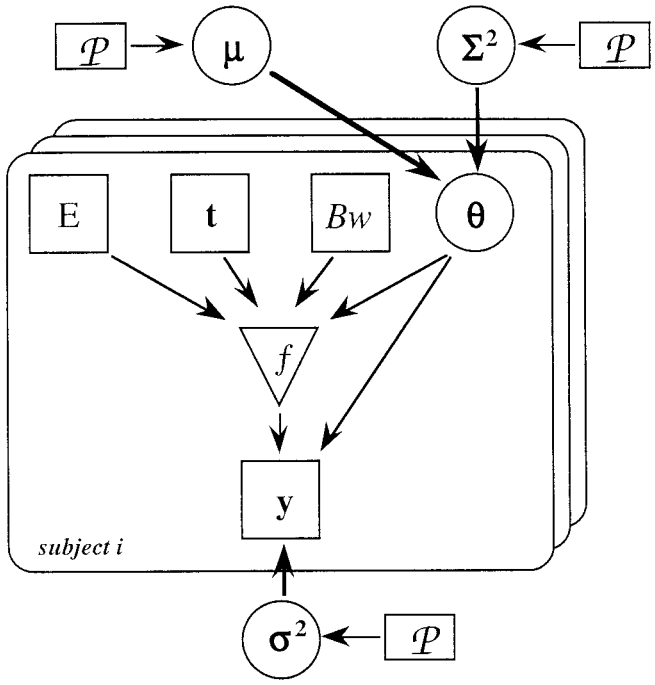

FIG. 2. Graph of the hierarchical statistical model describing dependencies between groups of variables. Symbols are: P, prior distributions; $\boldsymbol{\mu}$: mean population parameters; $\boldsymbol{\Sigma}^2$, variances of the parameters in the population; **E**, butadiene exposure concentrations; **t**, experimental sampling times; $\boldsymbol{\theta}$, unknown physiological parameters; Bw, body weight, exactly measured; f, PBPK model; **y**, measured butadiene concentration (in exhaled air) and measured values of K_{in} and P_{ba}; $\boldsymbol{\sigma}^2$, variances of the experimental measurements. Square nodes represent variables of known value; circular nodes represent unknown variables; the triangle represents the deterministic physiological model.

solved with the Lsodes integrator (Bois and Maszle, 1997; Maszle and Bois, 1993).

The 2-compartment toxicokinetic model was imbedded in a hierarchical population model (Fig. 2) to describe the various levels of variability present in the data, according to a population toxicokinetic approach (Bois *et al*., 1996b). At the individual level, for each subject, the data (**y**) consisted of exhaled air concentrations of butadiene, blood to air partition coefficients, (P_{ba}, used to estimate P_{ca}), and minute ventilation rate, K_{in}, all measured experimentally with uncertainty. Body weight, Bw, relatively precisely measured, was considered as a covariate. The expected values of the exhaled air concentrations are a function (f) of exposure level (E), time (**t**), a set of physiological parameters of *a priori* unknown values (θ, which includes sc_V_c, K_{cp}, $P_{pc}V_p$, K_{met}, K_{in}, and P_{ca}), and Bw. E, **t**, $\boldsymbol{\theta}$, and Bw were subject-specific. The function f was the toxicokinetic model described above. The data collected were also affected by measurement and modeling errors, which were assumed, as usual, to be independent and log-normally distributed, with mean of zero and variance $\boldsymbol{\sigma}^2$ (on the log scale). The variance vector $\boldsymbol{\sigma}^2$ had three components: $\sigma^2{}_1$ for butadiene exhaled air concentrations, $\sigma^2{}_2$ for blood over air partition coefficients, and $\sigma^2{}_3$ for minute ventilation rates.

At the population level, the individual parameters, $\boldsymbol{\theta}$, were assumed to be distributed normally (in log-space) around a population mean $\boldsymbol{\mu}$, with population variance $\boldsymbol{\Sigma}^2$.

The hierarchical population model is symbolically illustrated by Figure 2. Three types of nodes are featured in the figure: square nodes represent variables for which the values are known by observation, such as **y** or Bw, or were fixed by us, such as E and **t**, or the priors on $\boldsymbol{\sigma}^2$, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}^2$; circular nodes represent variables with unknown values, such as $\boldsymbol{\theta}$, $\boldsymbol{\sigma}^2$, $\boldsymbol{\mu}$, or $\boldsymbol{\Sigma}^2$; the triangle represents the deterministic compartmental model f.

An arrow between two nodes indicates a direct statistical dependence between the variables of those nodes.

The model f was non-physiological and we had no prior information on its parameters $\boldsymbol{\theta}$. Therefore, uniform (i.e., non informative) prior distributions were used for the population means $\boldsymbol{\mu}$. For the population variances $\boldsymbol{\Sigma}^2$, inverse-gamma priors were used; inverse-gamma priors are standard reference priors for unknown variances (Bernardo and Smith, 1994). Their shape parameters were set to 1, and their scales were chosen to correspond to 20% CV for all parameters, except for K_{met}. CVs of about 20% have been found for distribution parameters in other studies (Bois *et al*., 1996a; Gelman *et al*., 1996). For K_{met}, the scale corresponded to 100% CV, because we expected *a priori* the population variability for this parameter to be higher than for the others, and a factor 2 variability is commensurate with values found in the literature on variability of metabolic parameters (Bois *et al*., 1996b). For $\sigma^2{}_1$, we used the standard non-informative prior distribution $P(\sigma^2{}_1) \sim \sigma^2{}_1$. The experimental variances $\sigma^2{}_2$ and $\sigma^2{}_3$ were known *a priori* (see Preliminary Data section, above) and set to the corresponding values.

From Bayes' theorem, the joint posterior distribution of the parameters to estimate, $P(\boldsymbol{\theta}, \boldsymbol{\sigma}^2, \boldsymbol{\mu}, \Sigma^2 \mid \mathbf{y}, Bw, E, \mathbf{t})$ is proportional to the likelihood of the data multiplied by the parameters' priors:

$$P(\boldsymbol{\theta}, \boldsymbol{\sigma}^2, \boldsymbol{\mu}, \boldsymbol{\Sigma}^2 \mid \mathbf{y}, Bw, E, \mathbf{t}) \sim P(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\sigma}^2, Bw, E, \mathbf{t}) \cdot P(\theta \mid \boldsymbol{\mu}, \Sigma^2) \cdot P(\boldsymbol{\sigma}^2) \cdot P(\boldsymbol{\mu}) \cdot P(\boldsymbol{\Sigma}^2). \quad (4)$$

where:

$P(\boldsymbol{\theta}, \boldsymbol{\sigma}^2, \boldsymbol{\mu}, \boldsymbol{\Sigma}^2 \mid \mathbf{y}, Bw, E, \mathbf{t})$ is the joint posterior distribution,

$P(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\sigma}^2, Bw, E, \mathbf{t})$ is the likelihood of the data,

$P(\boldsymbol{\theta} \mid \boldsymbol{\mu}, \Sigma^2)$ is the conditional distribution of $\boldsymbol{\theta}$, given $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}^2$,

$P(\boldsymbol{\sigma}^2)$, $P(\boldsymbol{\mu})$, and $P(\boldsymbol{\Sigma}^2)$ are the prior distributions (as specified above) for $\boldsymbol{\sigma}^2$, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}^2$, respectively.

The likelihood term was given by the lognormal measurement model:

$$\log(\mathbf{y}) \sim \mathcal{N}(\log f(\boldsymbol{\theta}, Bw, E, \mathbf{t}), \boldsymbol{\sigma}^2) \quad (5)$$

Current standard practice in Bayesian statistics is to summarize a complicated high-dimensional posterior distribution (such as that given by Eq. 4) by random draws from the conditional distribution of each individual model parameter, given the other parameters (Smith and Roberts, 1993). This can be shown equivalent, at equilibrium, to random draws from their joint posterior distribution (Smith and Roberts, 1993). By conditional independence arguments, the conditional distributions of individual model components are simpler than Equation 4:

$$P(\boldsymbol{\theta} \mid \boldsymbol{\sigma}^2, \boldsymbol{\mu}, \boldsymbol{\Sigma}^2, \mathbf{y}, Bw, E, \mathbf{t}) \sim P(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\sigma}^2, Bw, E, \mathbf{t}) \cdot P(\boldsymbol{\theta} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}^2) \quad (6)$$

$$P(\boldsymbol{\mu} \mid \boldsymbol{\theta}, \boldsymbol{\sigma}^2, \boldsymbol{\Sigma}^2, \mathbf{y}, Bw, E, \mathbf{t}) \sim P(\boldsymbol{\theta} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}^2) \cdot P(\boldsymbol{\mu}) \quad (7)$$

$$P(\boldsymbol{\Sigma}^2 \mid \boldsymbol{\theta}, \boldsymbol{\sigma}^2, \boldsymbol{\mu}, \mathbf{y}, Bw, E, \mathbf{t}) \sim P(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\Sigma}^2) \cdot P(\boldsymbol{\Sigma}^2) \quad (8)$$

$$P(\boldsymbol{\sigma}^2 \mid \boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\Sigma}^2, \mathbf{y}, Bw, E, \mathbf{t}) \sim P(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\sigma}^2, Bw, E, \mathbf{t}) \cdot P(\boldsymbol{\sigma}^2) \quad (9)$$

Each term on the right can be numerically evaluated, given the distributional assumptions listed above. We used Metropolis-Hasting sampling to update each parameter in its turn (using Eqs. 6–9), therefore performing a random walk through the posterior distribution. This iterative sampling belongs to a class of Markov chain Monte Carlo (MCMC) techniques that has recently received much interest (Gelfand and Smith, 1990; Gelman, 1992; Gelman and Rubin, 1996; Gilks *et al*., 1996; Wakefield *et al*., 1994). Two independent Markov chain Monte Carlo runs were performed for each computation. Convergence was monitored using the method of Gelman and Rubin (1992). Once a posterior sample of parameter values is obtained, further simulations can then be performed to compute, under specified conditions, posterior distributions of estimands of interest (e.g., a total quantity of metabolites formed).

Optimal design determination. The Harvard School of Public Health Institutional Review Board has approved, for our planned study, a protocol similar to the one used for preliminary experiments. The planned exposure level was reduced to 2 ppm for 20 min (instead of the 5 ppm for a 2-h exposure, used here). This exposure represents an inhalation dose similar to that received during everyday life from exposure to urban air pollution, passive cigarette smoke, gasoline vapors, and automobile exhaust. Subjects will be paid a nominal amount (equal to 6 h of a laboratory technician's salary) to compensate for the discomfort of the procedure, time required, and any travel expenses. The study has been approved by the Institutional Review Board because (1) the planned exposures will be in the range of everyday ambient inhalation doses, (2) they will be below the allowable occupational exposures, (3) the subjects will be well informed about the carcinogenic potential of 1,3-butadiene and its resulting risk (less than $1/10^6$ based on California EPA methodology, extrapolating from the mouse, the most sensitive species), and (4) data about human metabolism of 1,3-butadiene are critically needed for population risk assessments of ambient and occupational exposures.

We were faced with the problem of defining the number and timing of breath samples for the planned experiments. The method used to optimize that part of the design is based on decision theory, as presented for example by Müller and Parmigiani (1995): the goal of collecting and analyzing data is to make decisions. Such decisions may range from addressing an estimation problem under squared error loss (the usual least-square estimation) to recommending a certain treatment in clinical settings. Each decision has a payoff, and any rational decision-maker should take the decision leading to maximum payoff. That payoff depends on some set of unknown quantities, $\boldsymbol{\theta}$, taking values in a set of possibilities Θ and having *a priori* distribution Π. In our case, one can think of the payoff as "being right about the value of K_{met}," of the decision as "publishing that K_{met} is equal to 0.1," and of $\boldsymbol{\theta}$ as being the true value of K_{met}. In scientific research, experiments are typically conducted to gain additional knowledge of $\boldsymbol{\theta}$. The usual sequence of actions in an experiment is:

1. Choose a design e among all possible experiments;
2. Conduct the experiment and observe data $\mathbf{y}'$**.** The data are assumed to have as probability model (*i.e.*, error model) $f_e(\mathbf{y}'/\boldsymbol{\theta})$;
3. Make a final decision d, knowing $\mathbf{y}'$**,** to report $\boldsymbol{\theta}_r$ as value for $\boldsymbol{\theta}$, so as to maximize the payoff $U(e, d, \boldsymbol{\theta})$. The payoff can be, for example, inversely proportional to the square of the difference between $\boldsymbol{\theta}_r$ and $\boldsymbol{\theta}$.

In addition, recruiting subjects and collecting data involve costs, which may depend on the design, as well as on $\boldsymbol{\theta}$ or $\mathbf{y}'$. The cost function is $C(e, \mathbf{y}', \boldsymbol{\theta})$. From a theoretical point of view, consistent with the expected utility principle (Bernardo and Smith, 1994), the optimal experiment is that which maximizes the function:

$$U(e) = \int_{y,\Theta} [U(e, d, \boldsymbol{\theta}) - C(e, \mathbf{y}', \boldsymbol{\theta})] dP_e(\boldsymbol{\theta}, \mathbf{y}') \tag{10}$$

where P_e is the joint distribution of $\mathbf{y}'$ and $\boldsymbol{\theta}$. The interpretation of $U(e)$ is quite simple: it is the expected value, over all possible parameter values and data outcomes (*i.e.* future observations) of the net utility (cost subtracted) of performing experiment e.

Cost depends on the number of recruited subjects and on the number of measurements made. We chose here to neglect costs because we wanted to focus on the scientific question of precisely identifying K_{met}. As a utility function we chose:

$$U(e, d, \boldsymbol{\theta}) = -(\log K_{met} - E[\log K_{met} | \boldsymbol{\theta}])^2 \tag{11}$$

where $E(\cdot)$ denotes mathematical expectation. This corresponds to measuring utility by the opposite of the variance of the parameter of interest, K_{met} (i.e., the smaller the variance of K_{met} the greater the utility of the design). We evaluated the expected utility $U(e)$ in Equation 10 by averaging over the distribution of $\boldsymbol{\theta}$ given $\mathbf{y}'$ and averaging over the distribution of $\mathbf{y}'$ given e and $\boldsymbol{\theta}$. We did the averaging by simulation (randomly sampling from the distributions and forming the sample averages). The analysis of the preliminary data by MCMC simulations provided us with a set of parameter vectors sampled from their joint posterior distribution. Those vectors were used to simulate datasets for the range of possible future experiments. Using each dataset, we then obtained an updated posterior variance for K_{met}. Finally the average K_{met} variance, over all datasets (according to Eq. 10), was used to select the most informative experiment, *i.e.* the one leading to the smallest estimation variance.

More precisely, N parameter sets ($\boldsymbol{\theta}'$) (corresponding to "individuals") were generated by sampling one random parameter vector from $\mathcal{N}(\mu, \Sigma^2)$, for each of N pairs of μ and Σ^2 obtained by the MCMC sampling. Half of N $\boldsymbol{\theta}'$ vectors were then used as input to the model to create as many predicted datasets ($\mathbf{y}'$). Each dataset included the same number M of data points, obtained along a grid of time values ($t_j, j = 1 \ldots M$). Simulated experimental noise was added to the predictions data as specified by the measurement error model (Eq. 5) and $N/2$ estimates of experimental variance, obtained by MCMC sampling.

Even along a small grid of design points, the number of possible combinations can be enormous (e.g., there are 1,099,511,627,776 possibilities with 40 points). To avoid searching the entire design space, our algorithm proceeded heuristically, along 2 directions: forward or backward (Atkinson and Donev, 1992). In the "forward" mode, the algorithm started with no points in the design, *i.e.* from the prior knowledge. The criterion variance s^2 (variance of K_{met}) was directly obtained from $\boldsymbol{\theta}'$ since K_{met} was included in $\boldsymbol{\theta}'$ (if the variance of a model prediction were of interest, the model should first be used to compute those predictions for each θ'). At the next step, each possible position, j, of a design point along the time-grid was examined for its potential at reducing the criterion variance. For each design point j and each dataset y', the variance estimate was computed by importance reweighting (Bernardo and Smith, 1994), given the corresponding datum y'_j. The reweighted estimate of variance for any variable ϕ (model parameter or prediction) is given by:

$$s^2 = \sum_i^N r_i (\varphi_i - \bar{\varphi})^2 \tag{12}$$

where the average is:

$$\bar{\varphi} = \sum_i^N r_i \varphi_i \tag{13}$$

the importance weights are:

$$r_i = L_i / \sum_i^N L_i \tag{14}$$

where L_i is the likelihood of the datum y'_j given θ'_i, under the lognormal measurement model (Eq. 5). The variance estimates obtained were averaged over all $N/2$ datasets, according to Equation 10. The point giving the lowest variance for K_{met} was selected and definitely included in the design, closing that step. The algorithm then went on to determine the best time point for a second position and so on, iteratively, until all candidate design points were included. In the "backward" mode, all time points were included at start and the algorithm removed sequentially the least informative points, until the null design was reached.

The MCSim software (Bois and Maszle, 1997; Maszle and Bois, 1993), has been extended to perform the above calculations and was used throughout.

RESULTS

Adjustment of the Population Model to Butadiene Data

To parameterize the butadiene population pharmacokinetic model, two Markov chains of 50,000 iterations each were run.

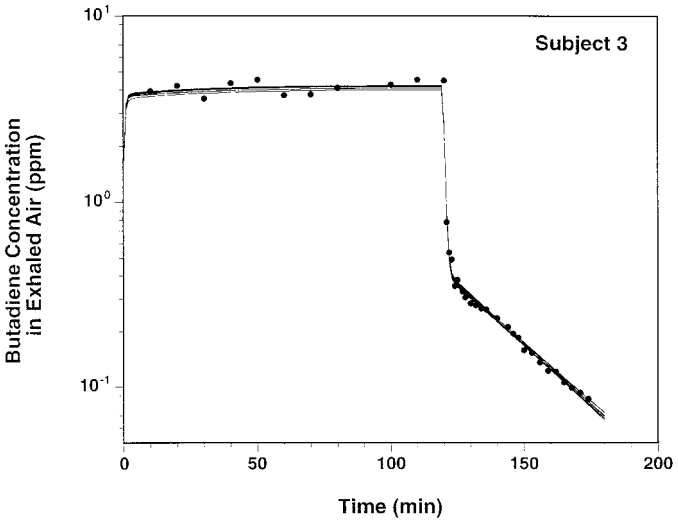

FIG. 3. Evolution in time of butadiene concentration in exhaled air for subject 3 (preliminary experiments). Exposure was to 5 ppm for 2 h. The solid lines correspond to the maximum posterior probability fit and to 10 other fits from randomly-sampled parameter sets. The data are represented by circles.

After 10,000 iterations, the trajectory of each parameter started to oscillate randomly around a mean value, and these oscillations had stabilized to a stationary distribution. The Gelman-Rubin convergence diagnostic less than 1.01 for all parameters, meaning that we would have expected to achieve at most 1% reduction in variance by continuing the simulations. One in 40 of the last 40,000 simulations of 2 chains were recorded, yielding 2,000 sets of parameter values from which the inferences and predictions presented in the following were made.

Figure 3 presents, for an individual (subject A), the model-predicted time course of butadiene exhaled air concentration, overlaid with the measured values. Predictions were made with the parameter set of highest posterior density, and with 10 parameter sets randomly sampled from their joint posterior. The fit was excellent, and the same was observed for all subjects. The residuals appeared evenly spread in log-space, indicating that the log-normal error model was reasonable and that little modeling error was present (observed and predicted data values differ on average by only 6%, and at most by 25%). Visual inspection of the residuals did not show any evidence of autocorrelation between them. In this extended design, data sampling was thorough and left little room for model uncertainty. A two-compartment model appeared sufficient to describe those data. The posterior geometric mean of the measurement SD for butadiene concentrations, σ_1, was 1.077 ($\pm$ 0.0036), corresponding to a CV of 7.7%. That estimate was close to the reported assay error of 5.6%.

Posterior Parameter Distributions

The joint distribution of all population and individual parameters was obtained in output of the MCMC simulations.

TABLE 2

Summary Statistics for the Posterior Distributions of the Population Geometric Means, μ, and Geometric Standard Deviations, Σ (See Fig. 2 and Corresponding Text), Derived from Analysis of the Preliminary Experiments

Model parameter[a]	μ[b]	Σ[b]
sc_V_c	0.0813 (1.09) [0.0693, 0.0983]	1.17 (1.05) [1.09, 1.33]
K_{cp}	1.21 (1.08) [1.03, 1.42]	1.17 (1.05) [1.09, 1.30]
$P_{pc}V_p$	31.5 (1.08) [27.4, 36.6]	1.15 (1.04) [1.09, 1.27]
K_{met}	0.239 (1.20) [0.165, 0.350]	1.63 (1.14) [1.35, 2.26]
K_{in}	0.373 (1.06) [0.330, 0.423]	1.18 (1.05) [1.11, 1.33]
P_{ca}	1.24 (1.10) [1.03, 1.51]	1.25 (1.07) [1.13, 1.46]

[a] Units: see text.

[b] Summary statistics given: geometric mean (geometric standard deviation) [2.5th percentile, 97.5th percentile].

This allowed consideration of marginal distributions (distributions of the parameters considered individually), but also of correlations of any order. Summary statistics for the population parameters means, μ, and geometric standard deviations, Σ, are presented in Table 2. The volume of the central compartment represented only 8% of the body weight on average. This was close to volume of blood in the body. Distribution to the peripheral compartment was fast (corresponding to a half-life of 0.6 minutes). It is difficult to comment on the product $P_{pc}V_p$ because of its composite nature; given its value (about 30), the peripheral compartment might correspond to the extra-vascular space, with large volume, partition coefficient lower than 1 and a fast exchange rate with blood. The metabolic elimination rate constant corresponded to a half-life of about 3 min. Variability across the 8 individuals was estimated to be 17% CV for sc_V_c and K_{cp}, 15% for $P_{pc}V_p$, and 50% for K_{met}. Table 3 presents the posterior distribution of K_{met} for each subject (distributions for the other parameters are not presented). These K_{met} values correspond to clearances of about 1.4 L/min (95% confidence interval: 0.45 – 4.1). For all model parameters, individual

TABLE 3

Summary Statistics of the Posterior Distributions for Individual K_{met} Values, in Min^{-1} (Determined from the Preliminary Experiments)

Subject	Geometric mean	Geometric SD	2.5th percentile	97.5th percentile
A	0.284	1.12	0.223	0.350
B	0.200	1.13	0.154	0.252
C	0.198	1.26	0.121	0.292
D	0.249	1.13	0.194	0.317
E	0.206	1.16	0.153	0.270
F	0.194	1.22	0.125	0.267
G	0.270	1.14	0.204	0.346
H	0.273	1.12	0.219	0.340

estimates were clustered around the corresponding population values, without "outliers". Uncertainties (in terms of CV) amounted to 2% to 25% approximately, most of the CVs being close to 10%. Parameter estimates were therefore quite precise with such an extended design. There were however, very large correlations between parameter estimates for any given individual (correlation coefficients ranging in absolute value from 0.52 to 0.94 were found between sc_V_c, K_{cp}, $P_{pc}V_p$, and P_{ca}). However, K_{met}, estimates were not strongly correlated to other parameters (correlation coefficients ranging from −0.07 to 0.32).

*Optimal Design of the Butadi*ene Experiments

Our interest resided in optimizing the design of butadiene exposure experiments. The optimization criterion chosen was minimization of the expected posterior variance of the parameter K_{met} (in log space), for a random individual. Butadiene exposure length and concentration were set to 20 minutes, and 2 ppm, respectively. These choices were dictated by safety reasons (subjects have no benefits in being exposed to butadiene) and analytical considerations (resulting concentrations in exhaled air must be detectable).

A prior sample of 1,000 parameter vectors, describing "future" individuals, was obtained by sampling from 1,000 pairs of μ and Σ population parameters drawn by MCMC sampling (see above). Body weights were randomly drawn from a lognormal distribution with geometric mean 70 kg and geometric standard deviation 1.2, truncated at ±2 SDs. This prior sample represents our knowledge of the population before the new (coming) experiments.

We discretized experimental time along a grid of 20 possible values. Time points were at 2, 5, 10, 15, 19, 21, 22, 24, 26, 28, 32, 38, 44, 50, 56, 62, 68, 80, 92, and 104 minutes after start of exposure. The final time was set with regard to the detection limit of the analytical method. Predictive datasets were obtained using half of the prior parameter vectors. Each of these datasets represents a possible response of new individuals in the future study. This allowed us to assess whether the optimal design was stable across subjects. Experimental noise was simulated according to a lognormal distribution, using a geometric SD of 1.077, as estimated by MCMC sampling (see above).

Given the stochastic noise present in the optimization results, we first wanted to assess their stability with respect to prior sample size. Obviously high sample sizes are preferable, but the computational burden increases in proportion. A first set of optimizations was performed, recruiting prior samples of either 250, 500, or 1000 parameter sets. Figure 4 presents the evolution of the estimation SD (square root of the variance) in log-space for K_{met} as locally optimal points were progressively included in the design. We present the SD, rather than the variance, for easier interpretation of the figure. The results were similar for sample sizes of 500 and 1000; with 250 vectors the SD was consistently underestimated (in comparison to estimates obtained with 500 or 1000 parameter sets). Note that the starting SD is quite high, close in fact to the population SD given in Table 2, since it represents uncertainty about the K_{met} of an individual not yet observed. The lowest SD achievable with such an experiment appeared to be 0.10 (which corresponds to a CV of 10%) (Fig. 4), close to what was obtained in the preliminary experiments. After adding 10 sampling times, 90% of the potential reduction in SD was achieved (corresponding to a CV of 15%). The corresponding results for backward optimization are similar (data not shown). The results appeared reasonably stable and a sample size of 500 was chosen for all subsequent simulations. Note that the average value of K_{met} did not change in those simulations, since no new experimental data were brought in, but only model-simulated "data".

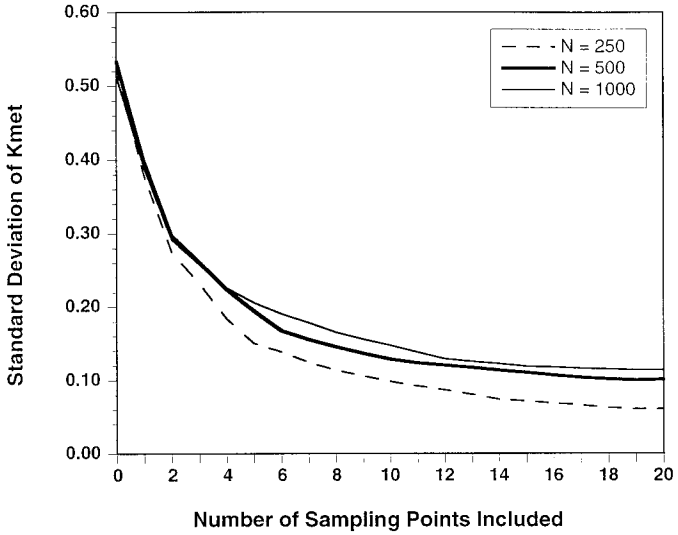

FIG. 4. Evolution of K_{met} estimation SD (in log space), for a random individual, as locally optimal sampling time points are progressively included in the design (forward optimization). Results are given for 3 prior sample sizes.

Given that a 10-point design was nearly optimal, the question remained as to which time points should be selected. The answer is given by the consideration of plots such as Figure 5. The expected estimation variance associated with each design point is presented at each step of the forward optimization procedure (*i.e.* when progressively including locally optimal points). When no point had yet been included (step 1), the 6th point (time 21 min, just after end of exposure) lead to the lowest estimation variance (about 0.15). That point was therefore chosen and included in the proposed design (Note that the next best points at step 1 were at later times). At step 2, interestingly, early time points were clearly preferred. At step 3, the last point lead to the lowest variance and was included. Table 4 gives the order of inclusion of the various time points in the first 10 steps. From consideration of the above, we propose for the coming experiments to sample exhaled air at 2, 5, 10, 15, 19, 21, 22, 28, 38, and 104 min after the start of

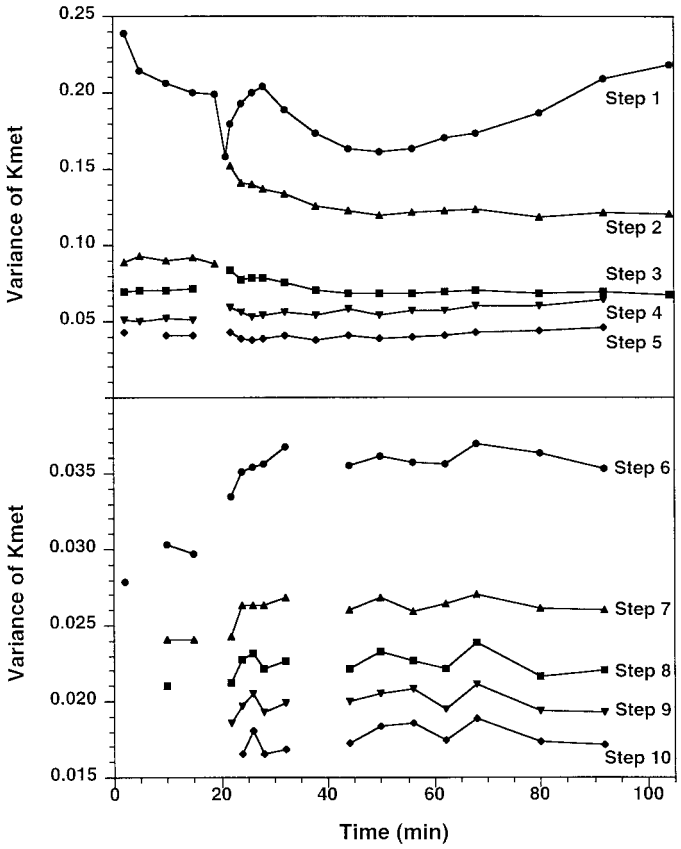

FIG. 5. Expected estimation variance associated with each design time point, for the first 10 steps of the forward optimization algorithm (at each step, the optimal time point is included in the proposed design). Gaps correspond to already included points.

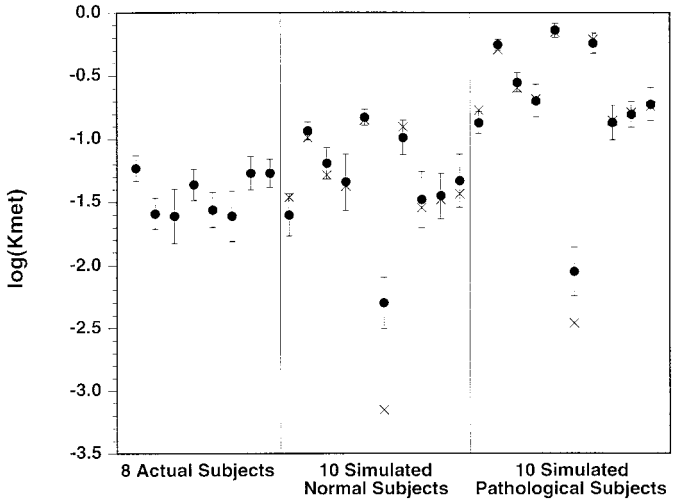

FIG. 6. Estimates (circles) of $\log_e(K_{met})$ obtained by MCMC sampling, for the 8 actual subjects of preliminary experiments (40-point sampling schedule), and simulated subjects (10-points optimal sampling schedule). All subjects were included in the same population model. Data for simulated normal subjects were drawn using estimates of population parameters for actual subjects (i.e., they mimic actual subjects' data). Data for simulated pathological subjects were drawn using a doubled value of K_{met}. True values of K_{met} are indicated by crosses. Error bars represent estimation SDs. An optimal 10-point sampling schedule leads to estimation SDs comparable to those obtained using the 40-point sampling schedule.

exposure. This design, which should lead to an estimation CV of 15%, is about evenly distributed in time; the times bracketing the end of exposure are definitely needed and sampling during exposure should be very informative.

The proposed sampling schedule was tested by simulating experimental data at the specified time points and fitting the toxicokinetic model using MCMC sampling. The aim of these simulations was to check if, under realistic conditions, the estimation SDs of K_{met} for future subjects would indeed be low (or at least close to those obtained with a dense sampling schedule in preliminary experiments). Data for several "normal subjects" were simulated as above, as well as data for "pathological subjects," with metabolic rates twice as high as for the normal subjects. All simulated subjects were introduced in the population model, together with the 8 subjects of the preliminary experiments. Figure 6 presents for the 8 actual subjects, 10 "normal" and 10 "pathological" simulated subjects, the resulting K_{met} estimates. The optimal 10-point sampling schedule lead to estimation SDs comparable to those obtained using the preliminary 40-point sampling schedule (which lead to CVs of about 15% for K_{met}). Note that K_{met} for "pathological subjects" (fast metabolizers) had lower SD than for "normal subjects" (low metabolizers). For the simulated subjects the figure also indicates the position of the "true" value of K_{met} (the value used to simulate the subjects). The estimate was usually close to the true value, except for the 2 subjects whose K_{met} is far lower than those for the rest of the population (a condition which, we just mentioned, leads to a low precision of estimates).

Figure 7 presents expected sample-based estimates of the CV of the population variance of K_{met} (in log-space) together with an analytical approximation: Given the distributional assumptions of our population model, if the K_{met} values of individuals were known perfectly, the posterior density of the

TABLE 4
Order of Inclusion of Candidate Time Points by Forward Optimization of Planned Butadiene Experiments

Step	Included time point (min)	Resulting variance	Resulting SD
1	21	0.158	0.397
2	19	0.0878	0.296
3	104	0.0679	0.261
4	5	0.0501	0.224
5	38	0.0375	0.194
6	2	0.0279	0.167
7	15	0.0241	0.155
8	10	0.0211	0.145
9	22	0.0186	0.136
10	28	0.0165	0.128

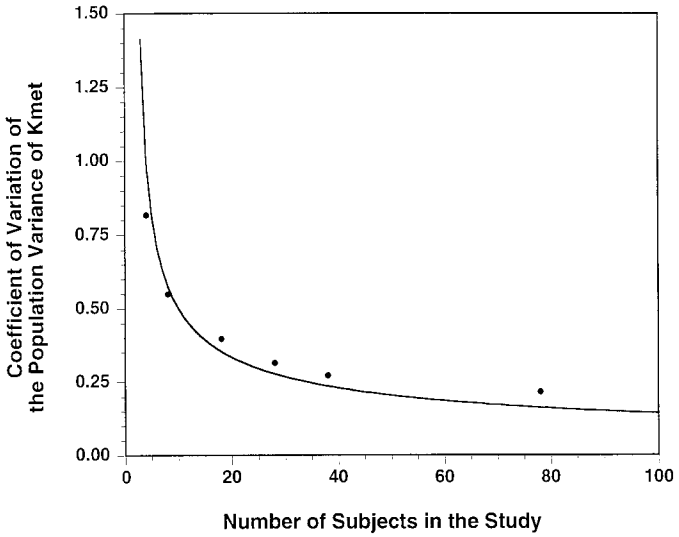

FIG. 7. Relationship between the planned study sample size and the CV of the population variance K_{met} (in log-space). The circles are estimates obtained by MCMC sampling of simulated experiments; the solid line is an analytical approximation.

population variance of K_{met} (in log-space) would follow an inverse-gamma distribution with a CV of $1/\sqrt{n/2+1}$, n being the number of subjects recruited. A "1" appears in the denominator because we are setting to 1 the shape parameter of the prior distributions for population variances (corresponding to vague prior distributions). In fact, the subjects' K_{met} are affected by uncertainty, but the approximation is reasonable within the range presented, even if slightly underestimating. According to the figure, with 40 subjects a CV of 25% should be achieved. One hundred subjects would bring this value down to 20%. The CV of the corresponding SD for K_{met}, in log-space, would be about half of that value, and hence close to 10%.

DISCUSSION

In the introductory section, it was noted that knowledge of human metabolic rates is critical to the estimation of risk from occupational and environmental exposures, and for extrapolating risk from one population to another. We are interested in more than just an estimate of the average human metabolic rate. There is evidence for butadiene that both formation and deactivation of epoxy metabolites are mediated by enzymes whose activity is determined by the individual's genotype (Carriere *et al*., 1996; Hassett *et al*., 1994). Thus, the average risk per unit-of-exposure for epidemiologic studies of petrochemical workers will depend on the population's distribution of metabolic genotypes.

It is not straightforward to obtain *in vivo* human metabolic data for both ethical and practical reasons. An alternative approach could be *in vitro* studies of enzyme activity in tissue slices, microsomes, or short-term cell cultures. These are suitable for performing comparisons when differences are large, such as cross-species differences in microsomal and cytosolic enzyme activities (Csanady *et al*., 1992). However, as noted by Guengerich (1996), all *in vitro* approaches have limitations in replicating *in vivo* metabolism. Metabolism depends on both the intrinsic reaction rate(s) of the enzyme(s) involved and on the amount of enzyme protein. Humans have a variety of intrinsic, genetically-determined enzyme rate capabilities. An individual's basic genetic capability is further modified by concurrent intake of alcohol, drugs, or foods, etc. The occurrence of such modifications in a population is difficult to evaluate by *in vitro* tests. A population-based collection of tissue samples would be necessary to characterize this variation, which is impractical and would not provide good estimates of the total rates. Since we wish to estimate the distribution of human metabolic rates to project population risks from exposure, it seems most appropriate to conduct efficient, short-duration laboratory exposures on a relatively large number of subjects. It is important that human testing be as efficient as possible, so that the risk to the volunteers is minimized. This means that the smallest possible exposure is used in the tests, the smallest number of subjects are tested, and the smallest number of sample time points is used that will efficiently estimate individual metabolic rates. Resolution of these considerations depends on the pharmacokinetics of the agent, measurement capabilities, and statistical factors.

The Bayesian approach with Monte Carlo simulation is an ideal means to optimize a measurement strategy to assess any one of the parameters of a pharmacokinetic model of human exposure to an environmental agent. Pilot data, with oversampling of time points during and after exposure, provided the input information for both the prior estimates of the model parameters and a test of the model's predictive capability. In particular, this work brings together for the first time, to our knowledge, a Bayesian population toxicokinetic analysis, and a simulation-based approach to optimal design. The population approach enables an easy summarizing of the population, the straightforward generation of simulated individuals, and the assessment of the behavior of precision in variance estimates (Bois *et al*., 1996a,b; Fanning *et al*., 1997; Gelman *et al*., 1996; Jonsson *et al*., 1997; Wakefield *et al*., 1994). It is easy in this framework to account for the uncertainty about measured covariates, such as minute ventilation or blood to air partition coefficient. Their measurements were treated as data, with known measurement SDs. The fits obtained with a 2-compartment model are excellent (among the best we have seen). Advantage of the 2-compartment model are obvious: it is simple, scientifically economical, quickly computed, and gives estimates of total metabolic clearance. This 2-compartment model does not describe a terminal elimination phase which would be controlled by the slow release of butadiene from the fat, but this phase would start beyond the measurements analyzed here. By not explicitly considering butadiene redistribu-

tion to body fat, we may have slightly overestimated metabolism. However, this potential bias is expected to be small because blood flow to the fat represents only a few percent of cardiac output and there was no observable slow-compartment contribution to the washout. Given a relative blood flow of 6:1 for the liver versus the fat, and a high extraction for butadiene, the metabolic rate might be at most overestimated by about 17%. Arguably, retention in fat could bear more on the total amount butadiene metabolized, independently of metabolic rate, and should be considered if cancer risk assessment, for example, was the objective of the analysis. We purposefully limited ourselves to the estimation of metabolic clearance, given the data at hand, which was our research objective. We also did not need to extrapolate to other species, other routes of exposure, or another dose range. Would that be the case, a PBPK model could be used for optimal design determination. Some approaches require linearized models (Mentré *et al*., 1997), in contrast, our method would still be applicable in the case of a complex nonlinear model.

Classical approaches to design tend to be the exclusive realm of biostatisticians and specialized software (Beatty and Piegorsch, 1997). We hope to make the technique more widely accessible, in particular to toxicologists interested in modeling. Our method is strongly model-based rather than databased. Model-based approaches (Atkinson *et al*., 1993; Burstein *et al*., 1997; D'Argenio, 1981, 1990; Kashuba *et al*., 1996; Palmer and Müller, 1998) are very flexible, because they allow the predictive design of experiments; preliminary data are needed to define a reasonable model, but it is then possible to optimize the design of any experiment susceptible to be simulated by the model. On the other hand, data-based approaches (Mager and Göller, 1998; Mahmood, 1998; Pai *et al*., 1996; Tse and Nedelman, 1996) trim down already-performed experiments through bootstrap or similar techniques; they are less sensitive to modeling errors, but also less powerful. Note that not only times, but also doses, or exposure length could be optimized by our method. In the case presented here, those variables were set based on safety considerations. Note also that the reduction in variance of a "useful prediction" could have been chosen as optimization criterion, instead of the variance of a crucial model parameter (K_{met}). Such a prediction, relevant to cancer risk, could be the amount of metabolites formed during the last h of an 8-h exposure. We could also have computed a global criterion, such as Shannon information (Merlé and Mentré, 1995; Polson, 1992). The choice of criterion should be dictated by the ultimate goal of the planned studies. In addition, in this analysis the balancing of costs and benefits was done informally. We did not need to formally put costs and benefits on a common utility scale. A complete cost function could be specified if formal cost minimization or more sophisticated decision making was requested (Wakefield and Racine-Poon, 1995). Cost functions could be used in particular if the simultaneously sampling optimization of several output variables (e.g., exhaled air, venous blood, and urine) was sought. A valuable aspect of optimal design might actually be the opportunity it gives to think about sophisticated utility functions: considering, for example, balancing the cost of the planned experiments, the expected benefits to society, and the potential risk incurred to the study subjects…

The forward and backward optimizations used here are heuristic approaches and are not guaranteed to give the very best results. A global optimization on all possible time point combinations would be required for that, but the resolution of such large problems is still an open question, and in any case very difficult (Atkinson and Donev, 1992). Forward and backward optimizations gave the same results in our problem, which reassures us that the results are reasonable and close to optimal. One of the drawbacks of our approach is the noise induced by the stochastic nature of the algorithm. Consequently, points offering an almost equal reduction of variance at a given step can be chosen quite randomly. The path leading to an optimal design might be missed in some cases. However, the size of the sample used can always be increased as computing power expands. In fact, there should be a set of near optimal designs to choose from, and the path to them does not need to be unique. It might actually be better to think in terms of informative design "regions," as the examination of variance profiles as in Figure 5 reveals. In the case studied, the most informative region for the elimination rate appears to be the discontinuity at the end of exposure. The early exposure period and the final portion of the elimination phase come next. Of those 3 periods, only the last one is usually considered as useful, on the basis of simple sensitivity analysis. Our approach takes into account the entire statistical framework used to derive estimates of the metabolic elimination rate constant (in particular the strong estimation covariances) and leads to somewhat different results. It should be noted that the form of the error distribution is of importance in determining the optimal design, since points determined with precision will tend to be preferred. We used a log-normal error model here, which fits the preliminary data very well. It is possible that a refined model, with a constant error term added (implying a detection limit), would be slightly better suited to the prediction of the low-concentration, short-exposure experiments envisioned. Using alternative error distributions can easily be accommodated by our approach.

An important finding from these simulations is the description of the diminishing returns for increasing the number of sampling time points. As expected, large increases in numbers are needed to reduce the coefficient of variation of the estimated metabolic rate constant below 10%. This value is close to what is expected given the estimation covariances and practical limits to measurements. However, until this study, the quantitative relationship between number of samples and precision was not known, because it is not determined by the simple $1/\sqrt{n}$ expression. The data presented here showed that relatively small numbers of data points, about 10, can provide precise estimates (*i.e*., with a 13% CV) of the apparent first

order metabolic rate for butadiene oxidation. Actually, additional simulations (data not shown) indicated that if higher precision were needed, a design with two 10 min exposure periods and 30 sampling times can achieve an estimation CV 9%.

Analysis of the pilot data gave an indication of the variability among individuals to be expected in butadiene metabolism for an ethnic Chinese group. This variability amounts to a 50% CV. Broader populations including other ethnic backgrounds may be expected to show more variation as a result of differences in enzyme genotypes. With genotypic data, it becomes possible to define the characteristic population rate parameter for a genotype, and to determine if rates are significantly different between different genotypes or ethnic groups (Jonsson *et al*., 1997). At this point, it will be useful to take advantage of PBPK modeling to compensate for differences in body size, body fat, inhalation rates, and blood partition coefficients.

It is tempting to take advantage of the proposed reduction in cost of individual experiments to include more subjects in the study, with the aim of increasing the precision of the population estimate of K_{met}. Our results show that, with up to a hundred subjects in the study, a 20% CV can be obtained on the population variance of K_{met} (this is a 10% CV for the estimation SD). Recruitment of about 50 subjects would yield a CV already close to those values. Again, given the limit to the precision of individual K_{met} estimates, there should be a lower limit to the precision achievable for the estimate of population's variability. We did not formally optimize the design with the objective of precision of population variance, but instead decoupled the optimization of design for individual-specific K_{met} and their population variance. This decoupling should have little implications here. However, a global approach would be more elegant, and we need to explore ways to implement the full population approach in simulation-based design optimization.

Many issues in designs routinely used in toxicology (air-chamber or gavage experiments) remain to be investigated. It would be interesting to see if some aspects of the optimal design for butadiene carry for other compounds, or to ask, for example, whether exhaled air sampling can replace blood-sampling altogether. Many experimenters are performing many similar experiments in animals and humans and all should benefit from the availability of a simple, almost automatic method for design optimization.

ACKNOWLEDGMENTS

This work has been primarily supported by Grant ES07581 from the National Institutes of Health. F. Bois is also supported by grant R 824755–01–0 from the U.S. Environmental Protection Agency and by the Association pour la Recherche contre le Cancer. T. Smith also is partially supported by the National Institute for Environmental Health Sciences research grant ES07586 and the center grant ES000002. We thank the reviewers for their thorough reviews and very helpful comments.

REFERENCES

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

Atkinson, A. C., and Donev, A. N. (1992). *Optimum Experimental Designs*. Oxford, New-York.

Beatty, D. A., and Piegorsch, W. W. (1997). Optimal statistical design for toxicokinetic studies. *Stat. Methods Med. Res.* **6**, 359–376.

Bernardo, J. M., and Smith, A. F. M. (1994). *Bayesian Theory*. Wiley, New York.

Bois, F. Y., Gelman, A., Jiang, J., Maszle, D., Zeise, L., and Alexeef, G. (1996b). Population toxicokinetics of tetrachloroethylene. *Arch. Toxicol.* **70**, 347–355.

Bois, F., Jackson, E., Pekari, K., and Smith, M. (1996a). Population toxicokinetics of benzene. *Environ. Health Perspect.* **104**(suppl. 6), 1405–1411.

Bois, F. Y., and Maszle, D. (1997). MCSim: a simulation program. *J. Stat. Software* **2**, http://www.stat.ucla.edu/journals/jss/v02/i09.

Bond, J. A., Himmelstein, M. W., and Medinsky, M. A. (1996). The use of toxicologic data in mechanistic risk assessment: 1,3-butadiene as a case study. *Int. Arch. Occup. Environ. Health* **68**, 415–420.

Brunnemann, K. D., Kagan, M. R., Cox, J. E., and Hoffmann, D. (1990). Analysis of 1,3-butadiene and other selected gas-phase components in cigarette mainstream and sidestream smoke by gas chromatography-mass selective detection. *Carcinogenesis* **11**, 1863–1868.

Burstein, A. H., Gal, P., and Forrest, A. (1997). Evaluation of a sparse sampling strategy for determining vancomycin pharmacokinetics in preterm neonates: application of optimal sampling theory. *The Ann. Pharmacotherapy* **31**, 980–983.

California Environmental Protection Agency Air Resources Board (1992). *Proposed Identification of 1,3-Butadiene as a Toxic Air Contaminant*. Stationery Source Division, Sacramento, CA.

Carriere, V., Berthou, F., Baird, S., Belloc, C., Beaune, P., and de Waziers, I. (1996). Human cytochrome P450 2E1 (CYP2E1): from genotype to phenotype. *Pharmacogenetics* **6**, 203–211.

Csanady, G. A., Guengerich, F. P., and Bond, J. A. (1992). Comparison of the biotransformation of 1,3-butadiene and its metabolite, butadiene monepoxide, by hepatic and pulmonary tissues from humans, rats, and mice. *Carcinogenesis* **13**, 1143–1153.

D'Argenio, D. Z. (1981). Optimal sampling times for pharmacokinetic experiments. *J. Pharmacok. Biopharm.* **9**, 739–756.

D'Argenio, D. Z. (1990). Incorporating prior parameter uncertainty in the design of sampling schedules for pharmacokinetic parameter estimation experiments. *Math. Biosci.* **99**, 105–118.

Fajen, J. M., Roberts, D. R., Ungers, L. J. and Krishnan, E. R. (1990). Occupational exposure of workers to 1,3 butadiene. *Environ. Health Perspect.* **86**, 11–18.

Fanning, E., Bois, F. Y., Rothman, N., Bechtold, B., Li, G., Hayes, R., and Smith, M. (1997). Population toxicokinetics of benzene and its metabolites. *Toxicol. Appl. Pharmacol.*(suppl.), 164.

Gelfand, A. E., and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* **85**, 398–409.

Gelman, A. (1992). Iterative and non-iterative simulation algorithms. *Comput. Sci. Stat.* **24**, 433–438.

Gelman, A., Bois, F., and Jiang, J. (1996). Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *J. Am. Stat. Assoc.* **91**, 1400–1412.

Gelman, A., and Rubin, D. B. (1992). Inference from iterative simulation, using multiple sequences (with discussion). *Stat. Sci.* **7**, 457–511.

Gelman, A., and Rubin, D. B. (1996). Markov chain Monte Carlo methods in biostatistics. *Stat. Methods Med. Res.* **5**, 339–355.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.

Guengerich, F. P. (1996). *In vitro* techniques for studying drug metabolism. *J. Pharmacok. Biopharm.* **24**, 521–533.

Hassett, C., Aicher, L., Sidhu, J. S., and Omiecinski, C. J. (1994). Human microsomal epoxide hydrolase: genetic polymorphism and functional expression *in vitro* of amino acid variants. *Hum. Mol. Genet.* **3**, 421–428.

International Agency for Research on Cancer (IARC) (1992). *Occupational Exposures to Mists and Vapours from Strong Inorganic Acids and Other Industrial Chemicals*. International Agency for Research on Cancer, Lyon, France.

Johanson, G. and Filser, J. G. (1996). PBPK model for butadiene metabolism to epoxides: Quantitative species differences in metabolism. *Toxicology* **113**, 40–47.

Jonsson, F., Johanson, G., and Bois, F. Y. (1997). A new approach to estimate population variability in target dose based on genetic polymorphism data and PBTK modeling. *Toxicol. Appl. Pharmacol.*(suppl.), 34.

Kashuba, A. D. M., Ballow, C. H., and Forrest, A. (1996). Development and evaluation of a Bayesian pharmacokinetic estimator and optimal, sparse sampling strategies for ceftazidime. *Antimicrob. Agents Chemother.* **40**, 1860–1865.

Mager, H., and Göller, G. (1998). Resampling methods in sparse sampling situations in preclinical pharmacokinetic studies. *J. Pharm. Sci.* **87**, 372–378.

Mahmood, I. (1998). A limited sampling approach in pharmacokinetic studies: experience with the antiepilepsy drug tiagabine. *Br.J. Clin. Pharmacol.* **38**, 324–330.

Maszle, D., and Bois, F. Y. (1993). *Program MCSim—User Manual*. Available from the authors, or at ftp://sparky.berkeley.edu/pub/mcsim and http://sparky.berkeley.edu/users/fbois.

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

Merlé, Y., and Mentré, F. (1995). Bayesian design criteria: computation, comparison, and application to a pharmacokinetic and a pharmacodynamic model. *J. Pharmacokinet. Biopharm.* **23**, 101–125.

Müller, P., and Parmigiani, G. (1995). Optimal design via curve fitting of Monte Carlo experiments. *J. Am. Stat. Assoc.* **90**, 1322–1330.

Mullins, J. A. (1990). Industrial emissions of 1,3-butadiene. *Environ. Health Perspect.* **86**, 9–10.

National Research Council (NRC) (1994). *Science and Judgement in Risk Assessment*. National Academy Press, Washington, DC.

O'Keefe, A. E., and Ortman, G. O. (1966). Primary standards for trace gas analysis. *Anal. Chem.* **38**, 760–763.

Pai, S. M., Fettner, S. H., Hajian, G., Cayen, M. N., and Batra, V. K. (1996). Characterization of AUCs from sparsely sampled populations in toxicology studies. *Pharm. Res.* **13**, 1283–1290.

Palmer, J. L., and Müller, P. (1998). Bayesian optimal design in population models for haematologic data. *Stat. Med.* **17**, 1613–1622.

Polson, N. G. (1992). On the expected amount of information from a nonlinear model. *J. Roy. Statist. Soc. B* **54**, 889–895.

Ramsey, J. C., and Andersen, M. (1984). A physiologically based description of the inhalation pharmacokinetics of styrene in rats and humans. *Toxicol. Appl. Pharmacol.* **73**, 159–175.

Rappaport, S. M., Selvin, S., and Waters, M. A. (1987). Exposures to hydrocarbon components of gasoline in the petroleum industry. *Appl. Ind. Hyg.* **2**, 148–154.

Reitz, R. H., MacDougal, J. N., Himmelstein, M. W., Nolan, R. J., and Schumann, A. M. (1988). Physiologically based pharmacokinetic modeling with methylchloroform: implications for interspecies, high dose/low dose, and dose route extrapolations. *Toxicol. Appl. Pharmacol.* **95**, 185–199.

Smith, A. F. M., and Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *J. Roy. Statist. Soc. B* **55**, 3–23.

Tse, F. L. S., and Nedelman, J. R. (1996). Serial versus sparse sampling in toxicokinetic studies. *Pharm. Res.* **13**, 1105–1108.

U.S. Environmental Protection Agency (1985). *Mutagenticity and Carcinogenicity Assessment of 1,3-Butadiene, Publication No. EPA/600/8–85/004F*. Office of Health Effects and Environmental Assessment, Washington, DC.

U.S. Environmental Protection Agency (1989). *Revised Draft Final Report: Locating and Estimating Air Emissions for Sources of 1,3-Butadiene*. Noncriteria Pollutant Programs Branch, Air Quality Management Division, Research Triangle Park, NC.

Wakefield, J. C. (1996). The Bayesian analysis of population pharmacokinetic models. *J. Am. Stat. Assoc.* **91**, 62–75.

Wakefield, J. C., and Racine-Poon, A. (1995). An application of Bayesian population pharmacokinetic pharmacodynamic models to dose recommendation. *Stat. Med.* **14**, 971–986.

Wakefield, J. C., Smith, A. F. M., Racine-Poon, A., and Gelfand, A. E. (1994). Bayesian analysis of linear and non-linear population models using the Gibbs sampler. *Appl. Statist. – J. Roy. Statist. Soc. C* **43**, 201–221.