Published in final edited form as:

J Stat Plan Inference. 2008 January 1; 138(1): 105–113. doi:10.1016/j.jspi.2007.05.019.

Bayesian Experimental Design for Long-Term Longitudinal HIV Dynamic Studies

Yangxin Huang^a and Hulin Wu^b

^aDepartment of Epidemiology & Biostatistics, College of Public Health MDC 56, University of South Florida Tampa FL 33612, U.S.A., hwu@bst.rochester.edu

^bDepartment of Biostatistics & Computational Biology, School of Medicine and Dentistry, University of Rochester, Rochester, NY 14642, U.S.A., yhuang@hsc.usf.edu

Abstract

The study of HIV dynamics is one of the most important developments in recent AIDS research for understanding the pathogenesis of HIV-1 infection and antiviral treatment strategies. Currently a large number of AIDS clinical trials on HIV dynamics are in development worldwide. However, many design issues that arise from AIDS clinical trials have not been addressed. In this paper, we use a simulation-based approach to deal with design problems in Bayesian hierarchical nonlinear (mixed-effects) models. The underlying model characterizes the long-term viral dynamics with antiretroviral treatment where we directly incorporate drug susceptibility and exposure into a function of treatment efficacy. The Bayesian design method is investigated under the framework of hierarchical Bayesian (mixed-effects) models. We compare a finite number of feasible candidate designs numerically, which are currently used in AIDS clinical trials from different perspectives, and provide guidance on how a design might be chosen in practice.

Keywords

AIDS; Clinical trial simulation; HIV dynamics; Longitudinal data; mixed-effects models; Optimal design

1. Introduction

Within the HIV/AIDS research community, there has been substantial interest in investigating HIV dynamics as a mechanism for understanding the pathogenesis of HIV-1 infection and for assessing the potency of antiviral therapies. Analysis of the dynamics of HIV-1 infection in response to drug therapy has elucidated crucial properties of viral dynamics. Hierarchical nonlinear mixed-effects models are often used to quantify individual heterogeneity among subjects, for example, in studies of pharmacokinetic, pharmacodynamic and viral dynamics (Huang et al., 2006; Sheiner and Beal, 1980; Wakefield, 1996; Wu and Ding, 1999). Viral dynamics can be used to evaluate the virological response to antiviral therapies more efficiently and in a more timely manner. This knowledge will accelerate the development of new antiretroviral drugs and speed up the life-saving process in fighting this fatal disease. Although a great number of HIV

© 2007 Elsevier B.V. All rights reserved.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

dynamic studies have been developed in the last decade the design of these studies has not been investigated carefully.

Marschner (1998) studied the design problems for viral dynamics under a linear mixed-effects model framework, but his assumptions are too idealistic and do not reflect methods that can be readily implemented in practice. Wu and Ding (2002) and Han and Chaloner (2004) used computer simulations to evaluate a finite number of designs for viral dynamic studies under the setting of nonlinear mixed-effects models. Wu and Ding (2002) considered different markers and different designs for assessing the potency of antiviral therapies in AIDS clinical trials. Han and Chaloner (2004) investigated the Bayesian experimental design where the design issue is formulated as a decision of choosing the values of a predictor, while minimizing the preposterior risk. These approaches assumed that once data are available, they are analyzed in a Bayesian paradigm. Unfortunately, Marschner (1998), Wu and Ding (2002) and Han and Chaloner (2004) have only addressed the design issues for short-term viral dynamic studies, and their methods are not applicable to long-term AIDS clinical studies, particularly where viral loads rebound for some individuals.

In this paper, we consider a model to characterize long-term viral dynamics. We combine the Bayesian analytic methods and mixed-effects modeling approach to evaluate a finite number of plausible candidate designs, which are currently used and proposed in AIDS clinical trials. The design issues for viral dynamic studies generally include (1) how frequently HIV RNA should be measured and how the design time points should be located; (2) how to balance the sample size (the number of subjects) and the number of measurements per subject; and (3) whether the measurement time points should be the same for all subjects. This paper will address some of these issues. We propose a simulation-based method for obtaining the Bayesian optimal design. Markov chain Monte Carlo (MCMC) simulation techniques are applied to the nonlinear mixed-effects model and to evaluate different design schemes. Following the study by Stroud et al. (2001), we assumed that the same prior distribution will be used both for design and for estimation and prediction.

The technical details on Bayesian analysis of hierarchical nonlinear mixed-effects models can be found in Wakefield et al. (1994) and Wakefield (1996). Han et al. (2002) and Huang et al. (2006) studied Bayesian approaches in analyzing the short-term and long-term viral dynamic models, respectively. We employ the model and estimation approach proposed in Huang et al. (2006) to address study design issues in this article.

The paper is organized as follows. In Section 2, we briefly describe the viral dynamic model and Bayesian modeling approach for hierarchical nonlinear mixed-effects models. Sampling schedules are described and the design criterion is introduced in Section 3. We compare different sampling schedules via intensive simulations and present the results in Section 4. Finally, the paper concludes with discussion in Section 5.

2. HIV dynamic models and statistical methods

Details of the HIV dynamic models and statistical methods are described in Huang et al. (2006) and Wu et al. (2005). For completeness, a brief summary of the models and methods is given as follows.

2.1. Drug efficacy models

As Molla et al. (1996) suggested, the phenotype marker, median inhibitory concentration (IC_{50}), can be used to quantify agent-specific drug susceptibility. We use the following model to approximate the within-host changes over time in IC_{50} ,

$$IC_{50}(t) = \begin{cases} I_0 + \frac{I_r - I_0}{t_r} t & \text{for } 0 < t < t_r, \\ I_r & \text{for } t \geq t_r, \end{cases} \quad (1)$$

where I_0 and I_r are respective values of $IC_{50}(t)$ at baseline and at time point t_r , where resistant mutations dominate. In our study, t_r is the time of virological failure which is observed from clinical studies.

Poor adherence to a treatment regimen is one of the major causes of treatment failure (Ickovics and Meisler, 1997). The following model is used to represent adherence for a time interval $T_k < t \leq T_{k+1}$,

$$A(t) = \begin{cases} 1 & \text{if all does are taken in } (T_k, T_{k+1}], \\ R_k & \text{for } 100 R_k \% \text{ does are taken in } (T_k, T_{k+1}], \end{cases} \quad (2)$$

where $0 \leq R_k < 1$ with R_k indicating adherence rate during the interval (T_k, T_{k+1}) denotes the adherence evaluation time at the k th clinical visit.

In most viral dynamic studies, investigators assumed that either drug efficacy was constant over treatment time (Perelson and Nelson, 1999; Wu and Ding, 1999) or antiviral regimens had perfect effect in blocking viral replication (Ho et al., 1995; Perelson et al., 1996). However, the drug efficacy may change as concentrations of antiretroviral drugs and other factors (*e.g.* drug resistance) vary during treatment. We employ the following modified E_{max} model (Sheiner, 1985) to represent the timevarying drug efficacy for two antiretroviral agents within a class,

$$\gamma(t) = \frac{C_1 A_1(t) / IC_{50}^1(t) + C_2 A_2(t) / IC_{50}^2(t)}{\phi + C_1 A_1(t) / IC_{50}^1(t) + C_2 A_2(t) / IC_{50}^2(t)} = \frac{IQ_1(t) A_1(t) + IQ_2(t) A_2(t)}{\phi + IQ_1(t) A_1(t) + IQ_2(t) A_2(t)}, \quad (3)$$

where C_1 and C_2 represent the respective drug concentrations or any of the PK parameters such as C_{trough} , C_{12h} , C_{max} or AUC_{0-12h} , $IC_{50}^1(t)$ and $IC_{50}^2(t)$ indicate the median inhibitory concentrations of the two drugs, $A_1(t)$ and $A_2(t)$ are adherence profiles of the two drugs as measured by pill counts. Parameter ϕ can be regarded as a conversion factor between *in vitro* and *in vivo* IC_{50} s and will be estimated from the data $IQ_i(t) = C_i / IC_{50}^i(t)$ is the inhibitory quotient (IQ) (Kempf et al., 2001) and $\gamma(t)$ ranges from 0 to 1.

2.2. Antiviral response model

We consider a simplified HIV dynamic model with antiviral treatment as follows (Huang et al., 2005),

$$\begin{aligned} \frac{dT}{dt}(t) &= \lambda - d_T T(t) - [1 - \gamma(t)] k T(t) V(t), \\ \frac{dT^*}{dt}(t) &= [1 - \gamma(t)] k T(t) V(t) - \delta T^*(t), \\ \frac{dV}{dt}(t) &= N \delta T^*(t) - c V(t), \end{aligned} \quad (4)$$

where the three differential equations represent three compartments: target uninfected cells (T), infected cells (T^*) and free virions (V). Parameter λ represents the rate at which new T cells are generated from sources within the body, such as the thymus, d_T is the death rate of T cells, k is the infection rate without treatment, δ is the death rate of infected cells, N is the

number of new virions produced from each infected cell during its life-time and c is the clearance rate of free virions. The time-varying parameter $\gamma(t)$ is the antiviral drug efficacy at treatment time t .

If we assume that the system of equations (4) is in a steady-state before initiating antiretroviral treatment, then it is easy to show that the initial conditions for the system are

$$T_0 = \frac{c}{kN}, T_0^* = \frac{cV_0}{\delta N}, V_0 = \frac{\lambda N}{c} - \frac{d_T}{k} \quad (5)$$

If the regimen is not 100% effective (not perfect inhibition) the system of ordinary differential equations cannot be solved analytically. The solutions to (4) then have to be evaluated numerically. In the estimation procedure, we only need to evaluate the difference between observed data and numerical solutions of $V(t)$. So there is no need for an explicit solution of equation (4).

2.3. Bayesian nonlinear mixed-effects model

A number of studies investigated various statistical methods, including Bayesian approaches, to fit viral dynamic models and to predict virological responses using short-term viral load data (Ho et al., 1995; Perelson et al. 1996; Wu et al., 1998; Wu and Ding, 1999; Han et al. 2002). Huang et al. (2006) extended the existing methods to model long-term HIV dynamics of virological response. In this paper, we focus on design issues evaluated under the setting of a hierarchical Bayesian nonlinear model.

We denote the number of subjects by n and the number of measurements on the i th subject by m_i . For notational convenience, let $\boldsymbol{\mu} = (\ln \phi, \ln c, \ln \delta, \ln \lambda, \ln d_T, \ln N, \ln k)^T$, $\boldsymbol{\theta}_i = \{\ln \phi_i, \ln c_i, \ln \delta_i, \ln \lambda_i, \ln d_{Ti}, \ln N_i, \ln k_i\}^T$, $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_i, i = 1, \dots, n\}$, $\boldsymbol{\Theta}_{\{i\}} = \{\boldsymbol{\theta}_l, l \neq i\}$ and $\mathbf{Y} = \{y_{ij}, i = 1, \dots, n; j = 1, \dots, m_i\}$. Let $f_{ij}(\boldsymbol{\theta}_i, t_j) = \log_{10}(V_i(\boldsymbol{\theta}_i, t_j))$ where $V_i(\boldsymbol{\theta}_i, t_j)$ denotes the numerical solution of the differential equations (4) for the i th subject at time t_j . Let $y_{ij}(t)$ and $e_i(t_j)$ denote the repeated measurements of common logarithmic viral load and a measurement error with mean zero, respectively. The Bayesian nonlinear mixed-effects model can be written as the following three stages (Huang et al., 2006; Davidian and Giltinan, 1995).

Stage 1. Withinsubject variation:

$$\mathbf{y}_i = \mathbf{f}_i(\boldsymbol{\theta}_i) + \mathbf{e}_i, \quad \mathbf{e}_i | \sigma^2, \boldsymbol{\theta}_i \sim \mathcal{N}\left(0, \sigma^2 \mathbf{I}_{m_i}\right) \quad (6)$$

where $\mathbf{y}_i = (y_{i1}(t_1), \dots, y_{im_i}(t_{m_i}))^T$, $\mathbf{f}_i(\boldsymbol{\theta}_i) = (f_{i1}(\boldsymbol{\theta}_i, t_1), \dots, f_{im_i}(\boldsymbol{\theta}_i, t_{m_i}))^T$, $\mathbf{e}_i = (e_i(t_1), \dots, e_i(t_{m_i}))^T$.

Stage 2. Between subject variation:

$$\boldsymbol{\theta}_i = \boldsymbol{\mu} + \mathbf{b}_i, \quad [\mathbf{b}_i | \Sigma] \sim \mathcal{N}(0, \Sigma) \quad (7)$$

Stage 3. Hyperprior distributions:

$$\sigma^{-2} \sim \text{Ga}(a, b), \quad \boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\eta}, \Lambda), \quad \Sigma^{-1} \sim \text{Wi}(\Omega, \nu) \quad (8)$$

where the mutually independent Gamma (Ga), Normal ($\mathcal{N}$) and Wishart (Wi) prior distributions are chosen to facilitate computations (Davidian, et al., 1995). The hyperparameters $a, b, \boldsymbol{\eta}, \Lambda, \Omega$, and ν were determined from previous studies and literature (Ho, et

al., 1995; Perelson, et al., 1996; Perelson and Nelson, 1999; Nowak and May, 2000). See Huang et al. (2006) for a detailed discussion of the Bayesian modeling approach, including the choice of the hyper-parameters, and Wakefield (1996) for the implementation of the Markov chain Monte Carlo (MCMC) procedures.

3. Sampling schedules and design criteria

In a general nonlinear regression problem, an optimal design of sampling schedules can be obtained by optimizing a selected criterion which measures the goodness of the design. These criteria, which are usually based on the Fisher information matrix, include D-optimal, A-optimal, E-optimal and G-optimal criteria (Atkinson and Donev, 1992). Among them, D-optimality (determinant of the information matrix) is the most popular one and is widely used in practice (Atkinson et al., 1993). The optimal design points are derived under the assumed parameter values. Therefore, they may not be optimal if different parameter values are assumed. To solve this problem, other design strategies such as a Bayesian design and a robust design have been developed (Atkinson and Donev, 1992; Atkinson et al., 1993; Pronzato and Walter, 1988; Chaloner and Verdinelli, 1995).

Wu and Ding (2002) investigated several design sampling schedules within a period of eight weeks for a short-term HIV dynamic study. Following the idea in Wu and Ding (2002) for a choice of design, Han and Chaloner (2004) examined eight sampling schedules within a one week period. However, treatment evaluation in AIDS clinical trials usually require a much longer period than those scheduled in the above two articles. In other words, those designs of sampling schedules may not be feasible in practice, particularly, in AIDS clinical trials. In this article, we follow a similar approach to specify a number of practically feasible candidate designs for a long-term period (twenty-four weeks) and to evaluate their performance.

There exist some practical limitations in the design of viral dynamic studies. One limitation is that the total volume of blood drawn (number of samples) for viral load measurements for each individual is limited within a time period. Secondly, patients usually find it convenient to visit a clinic on the same day of the week. Thus we can not design too many points within a week. Due to these limitations and other practical considerations for AIDS clinical trials, we propose and examine the sampling schedules summarized in Table 1 and Fig. 1.

Schedule 1 is the original schedule used in an AIDS clinical trial developed by the AIDS Clinical Trials Group (ACTG) (Protocol A5055) See Acosta et al. (2004) and Wu et al. (2005) for details of this study in which the planned design took nine measurements from each of 44 patients within a 24 week period. This schedule was designed to take all measurements at weekly clinical visits and has four measurements (Days 0, 7, 14, 28) within the first month after treatment initiation. Compared with Schedule 1, Schedule 2 adds two visits, Days 3 and 21, in the first month, but drops two later visits (Days 84 and 140). Schedule 3 reduces the visits for individual patients from Schedule 1 and only has six weekly visits, which reflect standard sampling schedule currently used in many AIDS clinical trials (Weeks 0, 4, 8, 12, 16, 24). Schedule 4 is designed to take the same number of measurements as Schedule 3, but moves the two later samples (Days 56 and 112) to the first month (Days 7 and 14). Notice that by comparing Schedules 1 versus 2 and Schedules 3 versus 4, we can evaluate whether the early samples can improve the performance based on the proposed criterion. Moreover, comparing Schedules 1 versus 4 and Schedules 2 versus 3 allows us to evaluate the two different design strategies, one is frequent sampling per subject but with fewer subjects and another is sparse sampling per subject but with more subjects (keeping the total number of measurements fixed). Another reason we choose these schedules for evaluation is that these schedules are proposed and currently used in many

AIDS clinical trials, such as ACTG 398 (Hammer et al., 2002) and ACTG 388 (Fischl et al., 2003) For these chosen candidate designs (schedules), both population and individual dynamic parameters can be estimated using a Bayesian nonlinear mixed-effects modeling approach (Huang et al., 2005; Wakefield, 1996).

Following Han and Chaloner (2004), a quadratic loss function is used for the estimation of μ , the population mean of the random effects in the mean function: $L(\mu, \mathbf{a}) = (\mu - \mathbf{a})^T (\mu - \mathbf{a})$. The Bayes estimator of μ , is $E(\mu|Y)$ which is the mean of the posterior distribution under the prior distribution (8). This results in the posterior risk $\sum_k V(\mu_k|Y)$ where μ_k 's are the components of μ . Let G_ξ denote the marginal distribution of Y when the prior distribution is taken as (8) and the design is ξ ($\xi \in \Xi$). For given data Y , we define the loss to be

$\sum_k V(\mu_k|Y)$. Then the goal is to find a design with $\xi = \arg \min_{\xi \in \Xi} E^{G_\xi} \left\{ \sum_k V(\mu_k|Y) \right\}$, among a finite collection of candidate designs (schedules).

As shown in Han and Chaloner (2004), the sum of the posterior variances of the fixed effects is finite with a probability of 1 in the marginal distribution of data. When the posterior variances exist Markov chain Monte Carlo (MCMC) method can be used to estimate them.

4. Design evaluation via Monte Carlo simulations

4.1. Simulation method

Based on the results from A5055 study (Acosta et al. 2004; Wu et al, 2005, Huang et al., 2006), we designed simulation experiments to evaluate the aforementioned sampling schedules and compared their performance in terms of design criteria introduced in Section 3. We considered the Bayesian nonlinear mixed-effects model (6)–(8) but for illustration purposes and for computational convenience we proposed to only estimate the three parameters $\ln \phi$, $\ln \delta$, and $\ln d_T$ which are identifiable in our model and assumed that the other four parameters ($\ln c$, $\ln \lambda$, $\ln N$, $\ln k$) are given as constants whose values are (1.1, 4.6, 6.9, -9.0), respectively. Based on the discussion in Section 2, the prior distribution for μ , = $(\ln \phi, \ln \delta, \ln d_T)^T$ was assumed to be $\mathcal{N}(\eta, \Lambda)$ with Λ being a diagonal matrix. The prior distribution which has the form (8) was constructed based on several studies available and details of the construction of the prior distribution are given in Huang et al (2006) Thus the values of hyperparameters were chosen as follows.

$$a=4.5, \quad b=9.0, \quad \nu=5.0, \quad \eta=(3.0, -1.0, -2.5)^T, \\ \Lambda=\text{diag}(1000.0, 1000.0, 1000.0), \quad \Omega=\text{diag}(2.5, 2.5, 2.5).$$

Note that noninformative priors were chosen for $\ln \phi$, $\ln \delta$ and $\ln d_T$.

To simulate studies for different design schedules, we assumed 20 subjects for Schedules 1 and 2 with each subject having 9 measurements. Thus, the total number of viral load measurements is 180 for each simulated study. For Schedules 3 and 4, we assumed 30 subjects for each simulated study and each subject has 6 measurements with a total number of viral load measurements of 180 for each study. Thus all the simulated studies have the same number of total measurements (approximately the same cost).

The true individual dynamic parameters, $\ln \phi_i$, $\ln \delta_i$ and $\ln d_{Ti}$, were generated by $\ln \phi_i = \ln \phi + b_{1i}$, $\ln \delta_i = \ln \delta + b_{2i}$ and $\ln d_{Ti} = \ln d_T + b_{3i}$ where $\ln \phi = 3.0$, $\ln \delta = -1.0$ and $\ln d_T = -2.5$

are the true values of population parameters b_{ji} ($j = 1, 2, 3$) are random effects following a normal distribution with mean 0 and standard deviation of 0.2.

Based on generated true individual parameters and four known constant parameters as well as the data (C_{12h} , IC_{50} and $A(t)$) which were taken from A5055 study the observations y_{ij} (the common logarithm of total viral load) were generated by perturbing the solution of the differential equations (4) with a within-subject measurement error, i.e., $y_{ij} = \log_{10}(V_{ij}) + e_i$, where V_{ij} is the numerical solution of viral load to the differential equations (4) for the i th subject at time t_j . It is assumed that the within-subject measurement error e_i is normally distributed with $N(0, 0.1^2)$.

We simulated the dataset for the four schedules introduced in Section 3, respectively, and 100 replications were used for each schedule (the simulation sample size of 100 runs was selected by considering both simulation error and computational limitation). For each replication sample, one long chain was a run with 20000 iterations. After an initial number of 5000 burn-in iterations, every 2nd MCMC sample was retained from the next 15000 iterations. Thus, 7500 MCMC samples were used to calculate the posterior variance of $\ln \phi$, $\ln \delta$ and $\ln d_T$. We implemented our Bayesian approach using the MCMC procedure consisting of a series of Gibbs sampling and Metropolis-Hastings algorithms (Huang et al. 2006). The convergence of the MCMC algorithms was carefully monitored (Wakefield, 1996).

4.2. Simulation results

Table 2 summarizes the results of estimated preposterior risk and relative risk ($\widehat{RR}$), and Fig. 2 displays the estimated preposterior risk of the three parameters for the four sampling schedules. The term “relative risk” refers to the ratio between a preposterior risk and the minimum preposterior risk over the four schedules. From Fig. 2 we can see that Schedule 4 is the best in terms of the estimated preposterior risk for all three parameter among the four proposed schedules although the performance of the four schedules is quite similar for estimated preposterior risk of the parameter $\ln(d_T)$. For $\widehat{E}\{V(\ln \phi|Y)\}$ and $\widehat{E}\{V(\ln \delta|Y)\}$, Schedule 2 is the next best. Schedules 1 and 3 perform similarly for the estimation of $\ln(\phi)$ but Schedule 1 is better than Schedule 3 in estimating $\ln(\delta)$. Thus, we can conclude that the sparse schedule (Schedule 4) with more frequent early measurements and more subjects provides the most information in estimating unknown dynamic parameters in long-term HIV dynamic studies, among the schedules we evaluated.

5 Discussion

Bayesian design is an exciting and fast-developing area of research. The Bayesian methodology has much to offer in guiding experimental design, where prior information which has always been used can be more systematically applied in the choice of experiment sampling schedules, explanatory factors sample size and model. In this article, a pragmatic Bayesian approach for optimal design, which minimizes the estimated preposterior risk for nonlinear mixed-effects models, is investigated. As pointed out in Han and Chaloner (2004), it is unclear what other approaches could be used to choose designs for nonlinear mixed-effects models. In addition, as stated previously, the design of experiments is traditionally formulated as an optimization problem, such as applying various alphabetical optimality criteria, under the assumption of a proposed model. This kind of optimal design may result in infeasible designs for implementation in practice, especially for clinical trial designs. In theory, although the exact calculation of these optimality criteria could be used, the Fisher information matrix is itself a matrix of integrals, whose integrands are functions of integrals all of which must be calculated numerically. We notice that a major difficulty with nonlinear

mixed-effects models is the lack of analytic expression of the likelihood function, and hence a related problem for design is the lack of analytic expression of the expected Fisher information matrix. To facilitate the computation of the information matrix and overcome this difficulty for design, Han and Chaloner (2005) and Mentré et al. (1997) suggest the use of a firstorder approximation of the mean function to the model for design of an optimal experiment.

In many study design problems, there only exists a finite number of feasible candidate designs. By taking the advantage of powerful computers, we can evaluate the finite number of designs from many different perspectives. We believe that computer simulation is a very useful tool for designing clinical trials. In this paper, we use intensive computer simulations, combined with the MCMC modeling approach, to evaluate the posterior risk over the four candidate designs. Our simulation results reveal that Schedule 4 is clearly superior to the other schedules. Our conclusion is that early measurements are more informative than later measurements, and a sparse sampling schedule with more subjects is better than a frequent sampling schedule with fewer subjects. These results are consistent with those from short-term HIV dynamic studies (Marschner, 1998; Wu and Ding, 2002; Han and Chaloner, 2004). This is also intuitively understandable since early HIV dynamics data provide more information for the dynamic parameters and a greater nonlinear curvature of viral load trajectories often occurs in an earlier stage of antiviral treatment in most patients, requiring more frequent measurements during this early stage. The sparse schedule with more subjects (more replicates at the same design points) is also favored since this is consistent with the results from the design studies with nonlinear regression and nonlinear mixed-effects models (Mentré et al., 1997; Pukelsheim, 1993).

It should be noted that the prior distributions for data analysis have been constructed using results from short-term viral dynamic studies in the scientific literature. The conclusion regarding the risks and their ordering could potentially change when a different set of prior distributions is used. It has been assumed in this paper that taking 9 measurements within 24 week period on each of the 20 subjects would incur the same cost as taking 6 measurements on each of thirty subjects. This may be an oversimplification since recruiting more participants will typically involve more effort and cost.

We identify some limitations of the experimental designs in our model. We only have the IC_{50} data at baseline and failure time as is available in most AIDS clinical trial studies such as A5055 (Acosta, et al., 2004) and ACTG 398 (Hammer et al. 2002). We extrapolated the data linearly to the whole treatment period in our modeling. The linear extrapolation is the best approximation that we can get from the sparse IC_{50} data (Huang et al., 2006). In addition, we assumed that the time point of the development of drug resistant mutations is known. The linear assumption might have some influence on the estimation results since the IC_{50} values might have jumped to a higher level earlier prior to the failure time when we obtained the sample for drug resistance testing. But we expect that this assumption had little effect on the prediction of virological response. Another problem is that as a measurement of adherence, pill counts may not reflect actual adherence profiles for individual patients. The validity of the adherence measurement would affect our results for trial designs in practice. More accurate measurements for adherence such as electronic monitoring devices (e.g. MEMs caps) may improve data quality and accuracy of design. Nevertheless, since our results are obtained from a simulationbased approach, to deal with design problems these limitations would not offset the major findings from our modeling approach and the results can be used to provide guidance on how a design might be chosen in practice although further improvement may be warranted.

The basic tool for investigating model uncertainty/assumptions is the sensitivity analysis. That is we simply make reasonable modifications to the assumptions in question, recompute the posterior quantities of interest, and see whether they have changed in a way that significantly affects the resulting interpretations or conclusions. If the results are robust against the suspected assumptions, we can report the results with confidence and our conclusions will be solid. However, if the results are sensitive to the assumptions, we would then communicate the sensitivity results and interpret the results with caution (Gamerman, 1997). In order to examine the dependence of dynamic parameter estimates/posterior quantities on the prior distributions and initial values, we carried out a sensitivity analysis. We summarize the sensitivity analysis results as follows: (i) The posterior quantities were not sensitive to both the priors and/or the initial values and the results were reasonable and robust. (ii) When different priors and/or different initials were used the results were similar to those presented in this paper. See (Huang et al. 2006) for more detailed discussions on sensitivity analysis.

We also like to point out that some of the results from our simulations are based on a selected set of identifiable parameters in our model and clinical data. We are conducting studies to further investigate how our results would change if other identifiable combinations of parameters in our model are used.

Acknowledgments

Authors are extremely grateful to the editor and two reviewers for their insightful comments and suggestions that led to a marked improvement of the article. This work was supported in part by NIH research grants R01 AI052765 and R01 AI055290 (H. Wu).

References

- Acosta EP, Wu H, Walawander A, Eron J, Pettinelli C, Yu S, Neath D, Ferguson E, Saah AJ, Kuritzkes DR, Gerber JG, the Adult ACTG 5055 Protocol Team. Comparison of two indinavir/ritonavir regimens in treatment-experienced HIV-infected individuals. *Journal of Acquired Immune Deficiency Syndromes*. 2004; 37:1358–1366. [PubMed: 15483465]
- Atkinson, AC.; Donev, AN. *Optimum Experimental Designs*. Clarendon Press; Oxford: 1992.
- Atkinson AC, Chaloner K, Herzberg AM, Juritz J. Optimum experimental designs for properties of a computational model. *Biometrics*. 1993; 49:325–337. [PubMed: 8369370]
- Chaloner K, Verdinelli I. Bayesian experimental design: a review. *Statistical Science*. 1995; 10:273–304.
- Davidian, M.; Giltinan, DM. *Nonlinear models for repeated measurement data*. Chapman & Hall; London: 1995.
- Fischl MA, Ribbaudo HJ. A randomized trial of 2 different 4-drug antiretroviral regimens versus a 3-drug regimen in advanced human immunodeficiency virus disease. *Journal of Infectious Diseases*. 2003; 188:625–634. [PubMed: 12934177]
- Gamerman, D. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall; London: 1997.
- Hammer SM, Vaida F, et al. Dual vs single protease inhibitor therapy following antiretroviral treatment failure: a random trial. *Journal of the American Medical Association*. 2002; 288:169–180. [PubMed: 12095381]
- Han, C.; Chaloner, K.; Perelson, AS. Bayesian analysis of a population HIV dynamic model. In: Gatsoiquiry, C.; Kass, RE.; Carriquiry, A.; Gelman, A.; Higdon, D.; Pauler, DK.; Verdinellinis, I., editors. *Case Studies in Bayesian Statistics*. Vol. 6. Springer-Verlag; New York: 2002. p. 223–237.
- Han C, Chaloner K. Bayesian experimental design for nonlinear mixedeffects models with application to HIV dynamics. *Biometrics*. 2004; 60:25–33. [PubMed: 15032770]

- Han, C.; Chaloner, K. Design of population studies of HIV dynamics. In: Tan, W-Y.; Wu, H., editors. *Deterministic and Stochastic Models of AIDS Epidemics and HIV Infections with Intervention*. World Scientific; London: 2005. p. 525-548.
- Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature*. 1995; 373:123–126. [PubMed: 7816094]
- Huang Y, Liu D, Wu H. Hierarchical Bayesian methods for estimation of parameters in a longitudinal HIV dynamic system. *Biometrics*. 2006; 62 in press.
- Ickovics JR, Meisler AW. Adherence in AIDS clinical trial: a framework for clinical research and clinical care. *Journal of Clinical Epidemiology*. 1997; 50:385–391. [PubMed: 9179096]
- Kempf, DJ.; Hsu, A.; Jiang, P., et al. Response to ritonavir intensification in indinavir recipients is highly correlated with virtual inhibitory quotient; 8th Conference on Retroviruses and Opportunistic Infections; Chicago, IL, USA. 2001; Abstract 523
- Marschner IC. Design of HIV dynamics studies. *Statistics in Medicine*. 1998; 17:2421–2434. [PubMed: 9819837]
- Mentré F, Mallet A, Baccar D. Optimal design in random effects regression models. *Biometrika*. 1997; 84:429–442.
- Molla A, Korneyeva M, et al. Ordered accumulation of mutations in HIV protease confers resistance to ritonavir. *Nature Medicine*. 1996; 2:760–66.
- Nowak, MA.; May, RM. *Virus dynamics mathematical principles of immunology and virology*. Oxford University Press; Oxford: 2000.
- Pronzato L, Walter E. Robust experiment design via maximin optimization. *Mathematical Biosciences*. 1988; 89:161–176.
- Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. HIV-1 dynamics *in vivo*: virion clearance rate, infected cell life-span, and viral generation time. *Science*. 1996; 271:1582–1586. [PubMed: 8599114]
- Perelson AS, Nelson PW. Mathematical analysis of HIV1 dynamics *in vivo*. *SIAM Review*. 1999; 41:3–44.
- Pukelsheim, F. *Optimal design of experiments*. Wiley; New York: 1993.
- Sheiner LB, Beal SL. Evaluation of methods for estimating population pharmacokinetic parameters. I. Michaelis-Menten model: Routine clinical pharmacokinetic data. *Journal of Pharmacokinetics and Biopharmaceutics*. 1980; 8:553–571.
- Sheiner, LB. Modeling pharmacodynamics parametric and nonparametric approaches. In: Rowland, M., et al., editors. *Variability in Drug Therapy: Description, Estimation, and Control*. Raven Press; New York: 1985. p. 139-152.
- Stroud JR, Müller P, Rosner GL. Optimal sampling times in population pharmacokinetic studies. *Applied Statistics*. 2001; 50:345–359.
- Wakefield JC, Smith AFM, Racine-Poon A, Gelfand AE. Bayesian analysis of linear and nonlinear population models using the Gibbs sampler. *Applied Statistics*. 1994; 43:201–21.
- Wakefield JC. The Bayesian analysis to population Pharmacokinetic models. *Journal of the American Statistical Association*. 1996; 91:62–75.
- Wu H, Ding AA, de Gruttola V. Estimation of HIV dynamic parameters. *Statistics in Medicine*. 1998; 17:2463–2485. [PubMed: 9819839]
- Wu H, Ding AA. Population HIV-1 dynamics *in vivo*: applicable models and inferential tools for virological data from AIDS clinical trials. *Biometrics*. 1999; 55:410–418. [PubMed: 11318194]
- Wu H, Ding AA. Design of viral dynamic studies for efficiently assessing potency of anti-HIV therapies in AIDS clinical trials. *Biometrical Journal*. 2002; 44:175–196.
- Wu H, Huang Y, Acosta EP, Rosenkranz SL, Kuritzkes DR, Eron JJ, Perelson AS, Gerber JG. Modeling long-term HIV dynamics and antiretroviral response: effects of drug potency pharmacokinetics, adherence and drug resistance. *Journal of Acquired Immune Deficiency Syndromes*. 2005; 39:272–283. [PubMed: 15980686]

Figure 1.
The four sampling schedules for HIV dynamic experiments.

Figure 2. Comparisons of estimated preposterior risk of $\ln \phi \ln \delta$ and $\ln d_T$ for the four sampling schedules.

Table 1

The four sampling schedules for HIV dynamic experiments

Schedule 1	0	7	14	28	56	84	112	140	168
Schedule 2	0	3	7	14	21	28	56	112	168
Schedule 3	0	28	56	84	112	168			
Schedule 4	0	7	14	28	84	168			

Table 2

Estimates for $E\{V(\ln \phi|Y)\}$, $E\{V(\ln \delta|Y)\}$, $E\{V(\ln d_T|Y)\}$ and estimated relative risks ($\widehat{RR}$) which is the ratio between a preposterior risk and the minimum preposterior risk

	$\hat{E}\{V(\ln \phi Y)\}$	$\widehat{RR}$	$\hat{E}\{V(\ln \delta Y)\}$	$\widehat{RR}$	$\hat{E}\{V(\ln d_T Y)\}$	$\widehat{RR}$
Schedule 1	0.04600	1.591	0.00613	2.609	0.01170	1.194
Schedule 2	0.03868	1.338	0.00413	1.757	0.01121	1.143
Schedule 3	0.04476	1.652	0.01207	5.136	0.01085	1.107
Schedule 4	0.02891	1.000	0.00235	1.000	0.00980	1.000