Design evaluation and optimization for models of hepatitis C viral dynamics.

Jérémie Guedj, Caroline Bazzoli, Avidan Neumann, France Mentré

Design evaluation and optimization for models of hepatitis C viral dynamics

Jeremie Guedj[a,b,*], Caroline Bazzoli[c], Avidan U. Neuman[b,d], France Mentré[c]

[a] *Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos 87545, USA*
[b] *The Everard and Mina Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan 52900, Israel*
[c] *UMR 738, INSERM and University Paris Diderot, 75018 Paris, France*
[d] *Santa Fe Institute, Santa Fe NM 87501, USA*

Abstract

Mathematical modeling of hepatitis C viral (HCV) kinetics is widely used for understanding viral pathogenesis and predicting treatment outcome. The standard model is based on a system of five non-linear ordinary differential equations (ODE) that describe both viral kinetics and changes in drug concentration after treatment initiation. In such complex models parameter estimation is challenging and requires frequent sampling measurements on each individual. By borrowing information between study subjects, non-linear mixed effect models can deal with sparser sampling from each individual. However, the search for optimal designs in this context has been limited by the numerical difficulty of evaluating the Fisher information matrix (FIM). Using the software PFIM, we show that a linearization of the statistical model avoids most of the computational burden, while providing a good approximation to the FIM. We then compare the expected precision of the parameters that can be expected using five study designs from the literature. We illustrate the usefulness of rationalizing data sampling by showing that, for a given level of precision, optimal design could reduce the total number of measurements by up 50%. Our approach can be used by a statistician or a clinician aiming at designing an HCV viral kinetics study.

Keywords: Design Evaluation, Design Optimization, Hepatitis C, Non-linear mixed effect models, Ordinary Differential Equations, Viral kinetics

1. Introduction

Chronic hepatitis C virus (HCV) infection is one of the most common causes of chronic liver disease, with as many as 170 million people infected worldwide (1). HCV is considered as eradicated when a sustained virological response (SVR), defined by the persistent absence of serum HCV RNA for 6 months after therapy, is achieved. The current standard of care combines weekly injections of pegylated interferon (Peg-IFN) and daily oral ribavirin and leads to an SVR rate of 50% in HCV genotype 1 patients (2), which is the largest group of patients in western countries.

By mathematically describing early HCV RNA (viral load) decay after initiation of IFN-based antiviral therapy, crucial parameters of the *in vivo* viral kinetics have been estimated, such as the rates of production and clearance of free virus, and the rate of loss of infected cells (3). Furthermore, by suggesting mechanisms of action for IFN and ribavirin, mathematical modeling has provided a means for evaluating and optimizing treatment strategies (4). The standard approach is based on five non-linear ordinary differential equations (ODE) that describe the interaction between the virus and its target cells, and the changes in antiviral drug effectiveness with varying levels of Peg-IFN and ribavirin (5).

*Corresponding author
Email address: `jeremie.guedj@gmail.com` (Jeremie Guedj)

A first step when dealing with such a complex model is to ensure that model parameters can be uniquely identified, e.g., that there is a one-to-one map between the model outputs and the parameters. The identifiability study of viral dynamics models has attracted a great deal of attention from mathematicians (6; 7) and has resulted in important theoretical results, such as the identification of parameters that cannot be uniquely determined from data and the minimal number of measurements required for parameter estimation (8; 9). However, these results only provide the minimal amount of information that needs to be collected in order to identify the parameters under perfect conditions. In reality, more data are needed to obtain reliable estimates of model parameters that circumvent measurements error and/or model misspecification. To address this issue, the notion of practical identifiability has been introduced and refers to the ability to provide reliable parameter estimates (10–12). When the parameters are estimated using maximum likelihood approach, the practical identifiability boils down to the ability to get "satisfactory" standard errors of the parameter estimates. Hence, the practical identifiability is a relative notion, that will depend on the information available, both qualitatively and quantitatively. In nonlinear contexts, such as viral dynamics models, whether a parameter is practically identifiable may highly depend on the data, the schedule of the sampling measurements and on the number of patients, when the parameters are estimated in a population framework, i.e., using nonlinear mixed effect models (13).

In order to study and optimize the practical identifiability, it is natural to evaluate the Fisher Information Matrix (FIM). Indeed the Cramer-Rao bound states that maximum likelihood estimated parameters are asymptotically efficient, with an asymptotic variance given by the inverse of the FIM. Nevertheless, in nonlinear mixed effect models, the FIM has no closed form, and its computation involves solving a multi-dimensional integral. The level of complexity and the computational time burden are still increased by the complexity of the nonlinear ODE model used in viral dynamics models. We have proposed an approach that advantageously uses the structure of ODEs to compute "exactly" (more rigorously, as precisely as wanted) the FIM (11); nevertheless, this approach involves greedy computations that makes its use complicated for design evaluation and impossible for design optimization. Another approach developed by Mentré *et al.* is to use a linearization of the statistical model around the expectation of the random effects (14). This approximation provides analytical values for the FIM elements and thus avoids the computational burden. Relevance of this approximation has been shown in models with single and multiple responses but only for models with analytical solutions (15). An application to multiple responses with nonlinear ODEs was published but without evaluation of the approximation (16). Here, we show the relevance of this approximation in the context of HCV viral dynamics models when two responses (Peg-IFN & HCV RNA) modeled by nonlinear ODEs are observed. Using this approximation we compare the practical identifiability that can be expected using five study designs of the literature, and we propose several optimal designs according to the number of sampling measurements allowed in each patient.

2. Methods

2.1. Viral dynamics modeling

The viral dynamics model considers target cells, T, productively infected cells, I and viral particles, V. Target cells are produced at a rate s and die at a rate d. Cells become infected with *de novo* infection rate b. After infection, these cells are lost with rate δ. In the absence of treatment, virus is produced by infected cells at a rate p and cleared at a rate c. The time origin, $t = 0$, corresponds to the time of treatment initiation. We assume that treatment is initiated sufficiently long after the acute infection so that the system is in equilibrium, i.e., that viral production is balanced by viral clearance, and $\frac{dV}{dt} = \frac{dT}{dt} = \frac{dI}{dt} = 0$.

Peg-IFN acts mainly by reducing the rate of virion production per infected cell, with effectiveness ϵ, therefore the viral production under treatment is $p(1-\epsilon)$, where $0 < \epsilon < 1$. Consequently, treatment initiation leads to an initial rapid decline of viral load with rate c and magnitude $\log_{10}(1-\epsilon)$. After this initial drop, the viral load enters a new slower mode of decline that reflects the progressive loss rate of infected cells, δ. However, with weekly administration of Peg-IFN, drug levels can drop rapidly and become too low to maintain a high degree of suppression of viral load. Consequently, a new round of cell infection may occur and give rise to increases in viral load towards the end of the dosing interval. These rebounds can be captured by assuming an E-max relationship between drug effectiveness, $\epsilon(t)$, and the drug concentration, $C(t)$, such that $\epsilon(t) = \frac{C(t)^n}{C(t)^n + EC_{50}^n}$, where EC_{50} is the drug concentration in blood at which the drug is 50% effective, and n is the Hill coefficient, a parameter that determines how steeply the treatment effectiveness changes with variations in drug concentration (5). The changes in Peg-IFN concentrations, $C(t)$, can be described by a first-order absorption and elimination model (5). In the long-run, weekly viral load levels in good responders will decline linearly with rate δ (Figure 1, A and B).

Although ribavirin showed only little antiviral efficacy in monotherapy (17), the combination of IFN or Peg-IFN with ribavirin dramatically improves SVR rates, particularly in patients that respond poorly to interferon-based treatment (18). The mechanisms of action of ribavirin are poorly understood, although one predominant mechanism is thought to be mutagenesis, whereby the proportion of non-infectious virus released is increased (19). This can be modeled by assuming that ribavirin reduces the *de novo* infection rate with effectiveness η (e.g., infection under treatment becomes $(1-\eta)b$, where $0 < \eta < 1$) (20). Since ribavirin is administrated on a daily basis, its effectiveness can be considered constant, although a slow accumulation of ribavirin has been observed in plasma (19).

Furthermore, the HCV viral dynamics model with Peg-IFN and ribavirin treatment is based on a set of five nonlinear ordinary differential equations (more details on the modeling aspects can be found in (21)):

$$
\begin{cases}
\frac{dX}{dt} = -k_a X + e(t) \\
\frac{dA}{dt} = k_a X - k_e A \\
\frac{dT}{dt} = s - (1-\eta) bVT - dT \\
\frac{dI}{dt} = (1-\eta) bVT - \delta I \\
\frac{dV}{dt} = \left(1 - \frac{C(t)^n}{C(t)^n + EC_{50}^n}\right) pI - cV
\end{cases}
\tag{1}
$$

for $t = [0, 7, 14, 21, ...]$, $e(t) = FD$ where D is the drug dose and F is the bioavailability $(0 \leq F \leq 1)$. $X(t)$ and $A(t)$ are the amount of drug at the injection site and in the blood, respectively. The concentration of drug in the blood is given by $C(t) = A(t)/V_d$ where V_d is the drug's volume of distribution.

Since only the viral load and the Peg-IFN levels are measured, the parameters p, s, d, V_d, F cannot be identified (8; 9). Therefore, without loss of generality, we fix these parameters to the following values: $p = 100, s = 20000, d = 0.001$ and $F = 1$. Although η and b are theoretically identifiable, the lack of information on the dynamics of T and I make them poorly identifiable (9). Hence, unless it is specified otherwise, we will consider $b = 10^{-7}$ and $\eta = 0$.

2.2. Statistical Model

To ensure that parameters take on positive values, let us define $[\log(EC_{50}), \log(n), \log(\delta), \log(c), \log(k_a), \log(k_e), \log($ the vector of dimension $p = 7$ of the log-transformed parameters to be estimated. The between-patients variability is taken into account by considering that each individual parameter value, β_i, is composed of a fixed part (the fixed effects parameter β) and a random part b_i, assumed to be Gaussian with a diagonal variance-covariance matrix $\Omega = diag(\omega_1^2, ..., \omega_p^2)$:

$$
\begin{cases}
\beta_i = \beta + b_i \\
b_i \sim \mathcal{N}(0, \Omega)
\end{cases}
\tag{2}
$$

The observations are of concentrations of viral load (HCV RNA) and Peg-IFN in the blood plasma. Both markers are observed at the same time points. We denote by $d_i = \{t_{ij}\}_{j=1,...,n_i}$ the elementary design of the measurement times j for patient i, and $\mathcal{D} = \{d_i\}_{i=1,...,N}$ the design of the study (for all patients). Because of measurement errors, random fluctuations, and model limitations, the observations are not the exact output predicted by the mathematical model. We ensure a Gaussian framework and homoskedasticity of the error measurements by working on the log-transformed values of the viral load data. We summarize all sources of deviations by introducing for each observation an independent and additive white noise process, ϵ_k, with variances σ_k^2. We set $\sigma = (\sigma_1, \sigma_2)^T$, where $k = 1$ indicates HCV

RNA and $k = 2$ indicates Peg-IFN. Furthermore, the model for the observations is:

$$\begin{cases} Y_{ij1} = \log_{10}(V((t_{ij}, \beta_i)) + \epsilon_{ij1}, \ j = 1, ..., n_i \\ Y_{ij2} = C(t_{ij}, \beta_i) + \epsilon_{ij2}, \ j = 1, ..., n_i \end{cases} \tag{3}$$

As usual, for all N patients, $\epsilon_i | b_i, \ i = 1, .., N$ are assumed to be independent from one individual to the other and from one response to the other. Lastly, for each individual, ϵ_i and b_i are also independent.

2.3. Fisher information matrix

Let ψ be the vector of parameters to be estimated, such as $\psi^T = \left(\beta^T, \omega_1^2, ..., \omega_p^2, \sigma^T\right)$, and let λ be the vector of variance terms, such as $\lambda^T = \left(\omega_1^2, ..., \omega_p^2, \sigma^T\right)$, such that $\psi^T = \left(\beta^T, \lambda^T\right)$. Let $l_i(\psi; Y_i)$ be the observed likelihood for individual i:

$$l_i(\psi; Y_i) = \int_{\mathbb{R}^p} \prod_{j=1,\dots,n_i} \frac{1}{\sigma_1\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{Y_{ij1} - \log_{10}(V(t_{ij}, \beta_i))}{\sigma_1}\right)^2\right] \times \frac{1}{\sigma_2\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{Y_{ij2} - C(t_{ij}, \beta_i)}{\sigma_2}\right)^2\right] dp(b_i). \tag{4}$$

The log-likelihood is $L_i(\psi; Y_i) = \log(l_i(\psi; Y_i))$, and the global observed log-likelihood is $L(\psi; Y) = \sum_{i=1,\dots,N} L_i(\psi; Y_i)$ by independence between patients. The elementary Fisher Information Matrix (FIM) for patient i is defined as

$$M_F(\psi; d_i) = E_\psi \left(-\frac{\partial^2 L_i(\psi; Y_i)}{\partial\psi\partial\psi^T}\right), \tag{5}$$

and the FIM for the whole sample is defined by

$$M_F(\psi; \mathcal{D}) = E_\psi \left(-\frac{\partial^2 L(\psi; Y)}{\partial\psi\partial\psi^T}\right). \tag{6}$$

By independence between patients, the FIM for the whole sample is simply $M_F(\psi; \mathcal{D}) = \sum_{i=1,\dots,N} M_F(\psi; d_i)$. If all patients have the same design d, then the FIM for the whole sample is $M_F(\psi; \mathcal{D}) = N \times M_F(\psi; d)$.

2.4. The PFIM software

Approximation of the FIM.

By making a first-order approximation around the expectation of the random effects in the likelihood expression (Eq. (4)) and using the assumptions made in section 2.3, Mentré *et al.* showed that an analytical approximation of the FIM can be obtained (14). The approximated FIM is block diagonal with a block for the fixed effects parameters and a block for the random effects parameters. A major advantage to this approximation is avoiding the multiple integrals involved in the exact computation of the FIM (Eq. (6)). This approximation of the FIM has been proposed in some free software, including

PFIM, PopED and WinPopt. A comparison of the different codes can be found at http://www.page-meeting.org/?abstract=1179. Here we focus on PFIM, which has the advantages of being written in R, and of using the Federov-Wynn algorithm for design optimization. This software has recently been expanded to deal with multiple response models (PFIM 3.0), including those defined as ODE models (15; 16). The software is available at www.pfim.biostat.fr and the code used in this article is available from the authors upon request.

Comparisons of study designs.

The comparison of the designs is based on their efficiencies. The efficiency criterion, $\phi(\mathcal{D})$, is defined by the determinant of $M_F^{-1}(\psi;\mathcal{D})$ normalized by the number of parameters to estimate, namely $\phi(\mathcal{D}) = |M_F(\psi;\mathcal{D})|^{\frac{1}{dim(\psi)}}$. Then the efficiency of a design $\mathcal{D}_1$ may be compared to that of $\mathcal{D}_2$ by computing the ratio of the efficiency criteria, Eff$(\mathcal{D}_1,\mathcal{D}_2)$, i.e., Eff$(\mathcal{D}_1,\mathcal{D}_2) = \phi(\mathcal{D}_1)/\phi(\mathcal{D}_2)$. If Eff$(\mathcal{D}_1,\mathcal{D}_2) > 1$, then $\mathcal{D}_1$ is more efficient than $\mathcal{D}_2$ and Eff$(\mathcal{D}_1,\mathcal{D}_2)$ is the mean gain in precision using $\mathcal{D}_1$ instead of $\mathcal{D}_2$.

Design Optimization.

For a given value ψ of the parameters, a population design, $\mathcal{D}$, is D-optimal, denoted by $\mathcal{D}_D$, if it minimizes the determinant of the inverse of the FIM:

$$\mathcal{D}_D = \underset{\mathcal{D}}{Argmin} \frac{1}{|M_F(\psi;\mathcal{D})|} \tag{7}$$

When the total number of sampling measurements is fixed and there is only a finite set of admissible times, the Fedorov-Wynn algorithm (22; 23) implemented in PFIM can be used to optimize designs based on the criterion $\phi(\mathcal{D})$. This algorithm provides an optimal design described by a number of elementary designs with the corresponding proportion of subjects in each elementary design.

3. Design comparisons

3.1. Designs studied and parameter set-up

We study the practical identifiability that can be expected with five designs used in seminal clinical trials aimed at describing viral dynamics with Peg-IFN (5; 24–27) (Table 1). These designs assume that both viral load and Peg-IFN concentrations are measured at each sampling time. Each design uses a different strategy. For example, in $\mathcal{D}_1$, there are only 8 measurements per patient, spread over four weeks, with sample measurements one day after each injection, close to viral load troughs (and drug peaks). On the other hand, $\mathcal{D}_2 - \mathcal{D}_5$ focus their samplings on the first week of treatment and on injection times that correspond to viral load peaks and drug troughs.

Moreover, in order to study how parameter precisions depend on the values of the fixed effects, we studied the SE obtained for four possible parameter set-ups A-D for the fixed effects (Table 2). These four set-ups are meant to cover the different median values that could occur in a clinical study of Genotype 1

Table 1: Design used in five studies of viral dynamics. M is the number of sampling points for each marker and each individual

Design name	Authors	Measurement times (days)	M
$\mathcal{D}_1$	Zeuzem et al. (2005) (24)	$\{0, 1, 4, 7, 8, 15, 22, 29\}$	8
$\mathcal{D}_2$	Sherman et al. (2005)(25)	$\{0, 0.25, 0.5, 1, 2, 3, 7, 10, 14, 28\}$	10
$\mathcal{D}_3$	Herrmann et al. (2003)(26)	$\{0, 0.25, 0.5, 1, 2, 3, 4, 7, 10, 14, 21, 28\}$	12
$\mathcal{D}_4$	Zeuzem et al. (2001)(27)	$\{0, 0.040, 0.080, 0.12, 0.20, 0.33, 1, 2, 3, 4, 7, 14, 21, 28\}$	14
$\mathcal{D}_5$	Talal et al. (2006)(5)	$\{0, 0.25, 0.5, 1, 2, 3, 5, 6, 7, 7.25, 7.50, 8, 9, 14, 15, 16\}$	16

Table 2: Values of the parameters used in the simulations (on their natural scale), according to four different patterns of viral decline in the population (Fig 2). The dose is the same for all the patients (180 μg). All the random effects have a standard deviation $\omega = 0.5$ and the error measurements are such that $\sigma_1 = \sigma_2 = 0.2$.

Patterns of viral kinetics	EC_{50} ($\mu g L^{-1}$)	n	δ (d^{-1})	c (d^{-1})	k_a (d^{-1})	k_e (d^{-1})	V_d (L)
A. Good responders with low inter-dose variation	0.12	2	0.2	7	0.8	0.15	100
B. Good responders with high inter-dose variation	0.20	4	0.2	7	1.4	0.24	100
C. Poor responders with low inter-dose variation	0.30	1	0.1	7	1.4	0.24	100
D. Poor responders with high inter-dose variation	0.30	4	0.1	7	1.4	0.46	100

patients: good responders with low or large inter-dose variations (A, B), and poor responders with low or large inter-dose variation (C, D) (Figure 1). Large inter-dose variations (B, D) are typically observed with Peg-IFN-α2b treatment (5; 28), whereas low variations (A, C) are more likely to be observed with Peg-IFN-α2a treatment (characterized by a slower absorption rate but also a slower clearance rate)(28; 29). Since drug concentration per se was not found to be associated with long-term response (5; 29), good and poor responders differ primarily in their pharmacodynamics (EC_{50}, n) and in their loss rate of infected cells, δ.

3.2. Comparison between the approximated standard errors given by PFIM and the empirical standard errors

We first investigate whether the approximate standard errors SE_{PFIM} are a good approximation of the empirical standard errors SE_{EMP}. To address this issue, we simulated, for each design $\mathcal{D}_1$-$\mathcal{D}_5$, 500 data set replications of $N = 30$ patients with parameter set-up A. For each data set, parameters were estimated using MONOLIX version 2.4 (http://software.monolix.org), a software program based on a Stochastic Approximation Expectation-Approximation (SAEM) algorithm (30). Parameters initial values used with the SAEM algorithm were the true values of the parameters. The empirical SE were obtained by computing the standard deviation of the 500 estimates. Figure 2 displays, for each design and each parameter, the difference $\Delta(SE) = SE_{PFIM} - SE_{EMP}$, which is a measure of the error made in the estimation of the confidence interval amplitude. We found that PFIM provides, in this challenging context, a very satisfactory approximation of the empirical SE, with $\Delta(SE) < 0.01$ for all fixed effects and all designs, except in three cases: EC_{50} with $\mathcal{D}_1$ and $\mathcal{D}_3$, and c with $\mathcal{D}_1$. Moreover, the gain in computation speed is considerable: for each design, PFIM takes only about 1 minute, as compared to > 5 days of computation needed to get the empirical standard error (on an Intel Core 2 Duo 2.5 GHz).

The fact that the linear approximation seems to work less well for EC_{50}, especially for the poorest design $\mathcal{D}_1$, was already observed in another context (15). Indeed, EC_{50} bridges the pharmacokinetics and the pharmacodynamics and, hence, is associated with a high degree of nonlinearity (31), which has been shown to hamper the quality of the first-order approximation (32; 33). Of note, $\Delta(SE) > 0.1$ for c in $\mathcal{D}_1$ results from its high degree of correlation with EC_{50} in this design, where SE for these parameters are the largest.

3.3. Design evaluation

We compared the SE for the designs $\mathcal{D}_1$-$\mathcal{D}_5$ using parameter set-ups A-D. Using the δ-method, the SEs of the estimators of the log-transformed fixed effects β are asymptotically equal to the relative error RE that can be expected for the parameters on their natural scale. Hence the results given in Table 3 can be read in the following way (first line): with design $\mathcal{D}_1$, a patient sample size $N = 30$, and parameter set-up A, the expected relative error RE for δ is $RE(\delta) = SE(\delta)/\delta \simeq SE(\log(\delta)) = 0.10$. Hence, the confidence interval (at the level of 95%) that can be expected for δ is $CI_{95\%}(\delta) = [0.16; 0.24]$. Not surprisingly, the RE tend to improve as the number of sampling points increases. However since the model is nonlinear, the relevance of a sampling point depends on the sensitivity of the biological and statistical model at this particular timepoint and for this particular parameter set-up. For instance, the most intensive design, $\mathcal{D}_5$, improves the RE of $\{EC_{50}, n, c\}$ compared to $\mathcal{D}_1$-$\mathcal{D}_4$, but degrades the RE of δ for good responders (A and B). Indeed, $\mathcal{D}_5$ concentrates most of the sampling points in the first week of treatment with no sampling point at week 3 or week 4. Given the mathematical properties of the biological model, the large number of early sampling points allows for a good estimation of $\{EC_{50}, n, c\}$. However, since the effects of the loss rate of infected cells, δ, are mainly seen in the long-term viral decline, $\mathcal{D}_5$ is not the most appropriate for estimating δ. This result is less pronounced for the parameter set-ups C and D, for which patients respond poorly to therapy and have a large viral rebound from days 2-3. Indeed, the kinetics of the rebound depend on the infected cell turnover and, hence, substantial information on δ can be gained in this case within the first week of treatment.

In general, a poor response (design B and D, with less than $1 \log_{10}$ reduction at the end of the first week) leads to a poor estimation of the viral parameters. As expected, the study of the FIM reveals that the correlation between the estimators of EC_{50} and n is important and increases with EC_{50}. For design $\mathcal{D}_1$, good responders with low (resp. high) inter-dose variation exhibited a correlation equal to 0.25 (resp. 0.58) compared to 0.70 (resp. 0.77) in poor responders with low (resp.) inter-dose variation. This is also true for the correlation between c and δ that increases from -0.12 (resp. -0.08) to -0.71 (resp. -0.49). The study of the eigenvectors of $M_F(\psi; \mathcal{D})$ reveals that the difficulty in practical identifiability stems from the ratio δ/c. In practice, parameter set-up associated with poor responders (i.e. a low δ) would make the estimation of both parameters difficult. Since δ is known to have greater importance and to be more variable than c (3), it should be recommended to fix c. For instance, when fixing c and using $\mathcal{D}_1$ with parameter set-up C, we found $RE(\delta) \simeq 0.12$ as compared to 0.20 when estimating both parameters. This leads to a confidence interval $CI_{95\%}(\delta) \simeq 0.10 \pm 0.024$ as compared 0.10 ± 0.04 when both parameters are included in the estimation procedure.

For the pharmacokinetic parameters, all the profiles except the parameter set-up D show comparable levels of precision. Indeed, this set of parameters is characterized by a rapid degradation of Peg-IFN, k_e, and a low second order sensitivity n, and only limited information can be gained after day 3. Analyzing the variances of the random effects is less intuitive but follows the same trend already described for the fixed effects parameters. The SE of the error measurements are roughly proportional to the number of sampling points (34), but our results indicate that the sampling times and the parameter set-up choices can have substantial effects.

It must be noted that the interpretation of the viral dynamics is seriously limited by the lack of practical identifiability for the parameters related to the host dynamics. Indeed, if the *de novo* infectivity, b, is included in the set of parameters to estimate, we found $RE(\beta) > 20$ in all cases (not shown). Moreover, including b seriously impaired the estimation of the other parameters, particularly the loss of infected cells, δ, and resulted in a correlation larger than 0.8 in all cases, with $RE(\delta) > 0.18$. This result reflects that the loss rate of infected cells and the infectivity rate share similar effects on the viral load. Yet, mathematical simulations of the model show that the effects of these two parameters on the host dynamics are different, and thus observation of T and/or I would help to solve this indetermination. Similarly, we studied the case in which ribavirin is effective by reducing the ability of virus to infect new cells, with 90% effectiveness ($\eta = 0.90$). We found that $RE(\eta) > 3$ and $CI_{95\%}(\eta) \simeq [0.30; 1.0]$ for all cases: thus, the ribavirin effectiveness is not practically identifiable. The non-identifiability of the ribavirin effect has already been discussed in empirical sensitivity studies (3; 19). However studying the FIM enables a rigorous assessment.

Table 3: Expected SE for a sample size of 30 patients, according to five popular designs and four different sets of parameter values, corresponding to four patterns of viral kinetics during Peg-IFN based therapy.

		Fixed effects							Random effects							Error measurements		$\Phi(\mathcal{D})$
	M	$\log(EC_{50})$	$\log(n)$	$\log(\delta)$	$\log(c)$	$\log(k_a)$	$\log(k_e)$	$\log(V_d)$	$\omega^2_{\log(EC_{50})}$	$\omega^2_{\log(n)}$	$\omega^2_{\log(\delta)}$	$\omega^2_{\log(c)}$	$\omega^2_{\log(k_a)}$	$\omega^2_{\log(k_e)}$	$\omega^2_{\log(V_d)}$	σ_1	σ_2	
A. Good responders with low inter-dose variation																		
$\mathcal{D}_1$	8	0.20	0.12	0.10	0.13	0.12	0.10	0.10	0.19	0.080	0.070	0.094	0.10	0.080	0.075	0.017	0.013	162.1
$\mathcal{D}_2$	10	0.16	0.10	0.095	0.11	0.13	0.11	0.10	0.15	0.072	0.068	0.082	0.10	0.081	0.073	0.011	0.010	184.4
$\mathcal{D}_3$	12	0.16	0.10	0.094	0.11	0.12	0.10	0.10	0.15	0.072	0.068	0.082	0.10	0.077	0.071	0.0091	0.0084	197.7
$\mathcal{D}_4$	14	0.16	0.11	0.094	0.10	0.12	0.10	0.10	0.15	0.076	0.068	0.075	0.099	0.077	0.071	0.0082	0.0076	203.1
$\mathcal{D}_5$	16	0.14	0.10	0.10	0.11	0.11	0.10	0.10	0.13	0.072	0.074	0.085	0.084	0.080	0.072	0.0075	0.0070	213.1
B. Good responders with high inter-dose variation																		
$\mathcal{D}_1$	8	0.12	0.10	0.097	0.11	0.11	0.10	0.10	0.095	0.074	0.071	0.081	0.086	0.077	0.075	0.014	0.010	200.1
$\mathcal{D}_2$	10	0.11	0.096	0.096	0.10	0.12	0.10	0.097	0.085	0.069	0.070	0.076	0.10	0.073	0.070	0.011	0.0093	219.1
$\mathcal{D}_3$	12	0.10	0.096	0.095	0.10	0.11	0.097	0.095	0.083	0.069	0.069	0.075	0.092	0.071	0.069	0.0096	0.0083	235.6
$\mathcal{D}_4$	14	0.10	0.098	0.095	0.098	0.11	0.098	0.096	0.085	0.071	0.069	0.072	0.090	0.071	0.069	0.0085	0.0076	242.2
$\mathcal{D}_5$	16	0.10	0.095	0.099	0.10	0.10	0.097	0.095	0.078	0.068	0.074	0.076	0.080	0.071	0.069	0.0077	0.0070	254.9
C. Poor responders with low inter-dose variation																		
$\mathcal{D}_1$	8	0.25	0.19	0.20	0.20	0.12	0.11	0.11	0.25	0.25	0.15	0.15	0.17	0.089	0.081	0.011	0.011	103.1
$\mathcal{D}_2$	10	0.20	0.14	0.18	0.17	0.13	0.10	0.10	0.20	0.11	014	0.13	0.12	0.076	0.071	0.0098	0.0095	128.6
$\mathcal{D}_3$	12	0.20	0.14	0.18	0.17	0.13	0.10	0.097	0.19	0.10	0.13	0.13	0.11	0.073	0.070	0.0086	0.0085	138.7
$\mathcal{D}_4$	14	0.19	0.16	0.19	0.18	0.13	0.10	0.098	0.16	0.12	0.13	0.13	0.12	0.077	0.070	0.0078	0.0077	141.8
$\mathcal{D}_5$	16	0.19	0.12	0.18	0.17	0.11	0.10	0.098	0.19	0.091	0.13	0.13	0.094	0.075	0.072	0.0072	0.0071	153.4
D. Poor responders with high inter-dose variation																		
$\mathcal{D}_1$	8	0.19	0.28	0.14	0.15	0.16	0.12	0.13	0.12	0.26	0.11	0.097	0.099	0.084	0.084	0.013	0.010	122.4
$\mathcal{D}_2$	10	0.12	0.12	0.14	0.15	0.15	0.12	0.11	0.094	0.084	0.10	0.10	0.11	0.086	0.078	0.010	0.0091	161.9
$\mathcal{D}_3$	12	0.11	0.11	0.13	0.14	0.14	0.12	0.11	0.084	0.079	0.10	0.10	0.10	0.087	0.077	0.0093	0.0081	179.9
$\mathcal{D}_4$	14	0.11	0.11	0.14	0.14	0.14	0.12	0.11	0.082	0.078	0.10	0.10	0.10	0.083	0.078	0.0083	0.0074	185.1
$\mathcal{D}_5$	16	0.10	0.10	0.11	0.12	0.11	0.10	0.10	0.079	0.073	0.088	0.091	0.089	0.077	0.073	0.0077	0.0069	220.9

4. Design optimization

Computational limitations prevent extensively searching for an optimal study design over continuous sampling ranges in the first month of treatment. Nevertheless, study design can be optimized over a smaller grid consisting of all the potential sampling points described in $\mathcal{D}_1$-$\mathcal{D}_5$. Two constraints were imposed. First, both markers were measured at the same sampling time points, including $t = 0$. Second, the total number of samples for each marker was $N * M = 240$, where N is the number of patients and M is the number of sampling times for each marker; this value corresponds to the one used in the most parsimonious design, $\mathcal{D}_1$, described above ($N = 30$, $M = 8$).

Table 4 displays the optimal designs according to the maximal number of sampling times M for each marker and each patient. For instance, if $M = 3$, then the number of patients that can be enrolled in the study is $N = 80$. To take advantage of the population framework, the optimal design can be composed of different elementary designs where all the patients do not necessarily have the same number of measurements or the same schedule for the sample measurements. Predicting the best balance between the number of patients, N, and the number of samples per patient, M, is not straightforward. In other words, what should be preferred: rich information within each individual and poor information on between-patient variability, or *vice-versa* ?

Interestingly, the design efficiency, as measured by $\Phi(\mathcal{D})$, varies substantially according to this choice. We found the best efficiency when $M = 4$ (and hence $N = 60$ enrolled patients). This design outperforms any other designs considered above. Indeed, while $\mathcal{D}_5$ used twice as much information ($N * M = 30 * 16 = 480$ samples per marker), its efficiency is equal to 213.1 (Table 3) compared to 230.2 for our optimized design, i.e., a mean gain in precision of 1.05 while the raw amount of information to be collected is reduced by 50%. By using the population framework and the possibility of subgroups with different elementary designs, the fixed effects parameters are substantially improved. For instance, this optimal design gives $RE(\delta) = 0.057$, versus $RE(\delta) \geq 0.095$ in all other contexts in Table 3. Hence, using this design, δ can be estimated with very good precision ($CI_{95\%}(\delta) = [0.18; 0.22]$). As expected, estimating the variances of the random effects is also strongly improved since there are $N = 48$ patients compared to $N = 30$ in A-D. However, since the number of points per patient is only equal to 4, the SE for the error terms is elevated, compared to previous cases. For instance, $SE(\sigma_1) = 0.024$, compared to $SE(\sigma_1) = 0.0075$ obtained with the same parameter set-up A and the design $\mathcal{D}_5$.

Although an intuitive approach could lead us to focus on early time points to estimate the initial viral kinetics, these results show that designs focused on the first week of treatment can generate redundant information and are not optimal. Furthermore, it would be preferable to enroll more patients and to complete long-term follow-up, rather than to conduct intensive sampling with fewer patients. Although this can imply a slight disadvantage in the estimation of treatment effectiveness, it is compensated by a more accurate estimation of the loss rate of infected cells δ, which is the most critical parameter for predicting treatment outcome (35). This is even more evident if the pharmacokinetic parameters are not constant over time, but could change over time as found in (5; 29).

5. Conclusion

We have proposed a methodology for rationalizing the measurement sampling times in models of HCV dynamics. Although some limitations for EC_{50} in sparse designs have been pointed out, this methodology has been shown to provide good and rapid approximations of the standard errors in this challenging context.

We have shown that an appropriate choice of the sampling times and the elementary designs allows for a substantial reduction of the information needed to estimate the parameters. In particular, we showed that, as compared to intensive designs used in previous studies, the number of sampling times could be reduced by up to 50%. Furthermore, parameters could be obtained with excellent precision with only four measurements of each marker. Interestingly, these data should not be focused on only the very first days that follow treatment initiation, but rather should be balanced throughout the entire follow-up period. Indeed, long-term sampling points (third and fourth week) provide substantial information, particularly for investigating the viral parameters that could be responsible for a poor response to treatment.

Although population approaches advantageously use the between-patient variability to optimize inference, we showed that the measurement of only the viral load and the drug level is not sufficient for practically estimating all the parameters of the current model, in particular those related to the dynamics of the host and to the antiviral effectiveness of ribavirin (11). This feature highlights more generally that in the absence of data on the host dynamics, the understanding of HCV viral kinetics provided by modeling approaches remains speculative.

There is a paradox in the approach based on the study of the FIM. It is assumed that the values of the parameters are known, whereas the aim of the study is ultimately to estimate these parameters. Hence, it could be argued that the figures found here are valid only for the specific parameter set-up that was chosen. While it is true that the precise values of the parameters are unknown, their order of magnitude, in general, is guaranteed. Indeed, these parameters have a biological and physiological meaning and, hence, they must take on values within certain ranges. We have proposed to partially circumvent this issue by showing that the main conclusions were consistent when changing the values of the fixed effects parameters. Another approach is to use *a priori* distributions for the parameters (36; 37), but a major drawback of Bayesian approaches is the computational cost, since the expectation of the posterior variance-covariance matrix of the parameters must be (numerically) evaluated.

This work constitutes a first step towards improved rationalization of data sampling designs for HCV viral kinetics studies, but several extensions of this work are necessary. First, the information brought by data under the limit of detection (lod), whose proportion is increased with the development of new, highly potent, direct-acting antiviral agents, needs to be rigorously taken into account (38; 39). Another natural extension of this work is to take into account the constraint of the financial cost of patient

Table 4: Optimal designs according to the maximum number of samples M allowed per patient for a fixed number of $N * M = 240$ sample measurements for each marker, similarly to design $\mathcal{D}_1$ (Table 1). The parameter values correspond to the set-up A (Table 2).

		Optimal design	Fixed effects							$\Phi(\mathcal{D})$
M	N	{ (sampling times),number of patients per group }	$\log(EC_{50})$	$\log(n)$	$\log(\delta)$	$\log(c)$	$\log(k_a)$	$\log(k_e)$	$\log(V_d)$	
3	80	$(0,7,9),3$ $(0,10,28),11$ $(0,1,28),16$ $(0,4,29),19$ $(0,1,4),31$	0.21	0.12	0.081	0.096	0.14	0.11	0.084	193.2
4	60	$(0,1,4,28),38$ $(0,1,10,28),22$	0.17	0.090	0.070	0.090	0.12	0.083	0.080	230.2
5	48	$(0,1,4,16,28),14$ $(0,1,7,10,29),34$	0.14	0.061	0.057	0.075	0.087	0.068	0.061	224.0
6	40	$(0,1,4,7,16,28),40$	0.15	0.095	0.084	0.10	0.011	0.095	0.090	208.3
7	34	$(0,0.040,1,4,7,9,29),4$ $(0,1,4,7,9,28,29),10$ $(0,0.040,1,4,7,16,28),20$	0.15	0.070	0.065	0.081	0.094	0.075	0.070	193.0

studies. Although increasing the number of patients improves the identifiability of the parameters, it often increases the cost of the study. Recently, a new development of PFIM integrates the financial cost of study designs into the design optimization procedure but this has not been yet implemented in the available version of PFIM (40). Lastly, predicting treatment outcome is done at the individual level, and the precision at both the population and the individual levels are important. New approaches need to be developed to study the appropriate trade-off between these two levels of prediction.

Acknowledgments

We thank L. Murillo, A. Smith and A.S. Perelson (LANL) for helpful comments; we are also grateful to both anonymous reviewers and the AE for very insightful remarks. This work was supported by the French Ministry of Foreign Affairs, the Israel Science Foundation (ISF 939/2008), and the US Department of Energy under contract RR06555 and P20-RR18754.

References

[1] Shepard C, Finelli L, Alter M. Global epidemiology of hepatitis C virus infection. *Lancet Infectious Diseases* 2005; **5**(9):558–567.

[2] Awad K T amd Thorlund, Hauser G, Mabrouk M, Gluud C. Peginterferon alpha-2a is associated with higher sustained virological response than peginterferon alfa-2b in chronic hepatitis c: a systematic review of randomized trials. *Hepatology* 2010; **51**(4):1176–1184.

[3] Neumann AU, Lam NP, Dahari H, Gretch DR, Wiley TE, Layden TJ, Perelson AS. Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon-alpha therapy. *Science* 1998; **282**:103–107.

[4] Mihm U, Herrmann E, Sarrazin C, Zeuzem S. Review article: predicting response in hepatitis C virus therapy. *Alimentary Pharmacology & therapeutics* 2006; **23**(8):1043–54.

[5] Talal A, Ribeiro R, Powers K, Grace M, Cullen C, Hussain M, Markatou M, Perelson A. Pharmacodynamics of PEG-IFN Differentiate HIV/HCV Coinfected Sustained Virological Responders From Nonresponders. *Hepatology* 2006; **43**:943–953.

[6] Miao H, Xia X, Perelson A, Wu H. On identifiability of nonlinear ODE models with applications in viral dynamics. *SIAM Review* 2010; .

[7] Saccomani M. An effective automatic procedure for testing parameter identifiability of hiv/aids models. *Bulletin of Mathematical Biology* 2010; :DOI 10.1007/s11 538–010–9588–2.

[8] Xia X, Moog CH. Identifiability of nonlinear systems with applications to HIV/AIDS models. *IEEE transactions on automatic control* 2003; **48**:330–336.

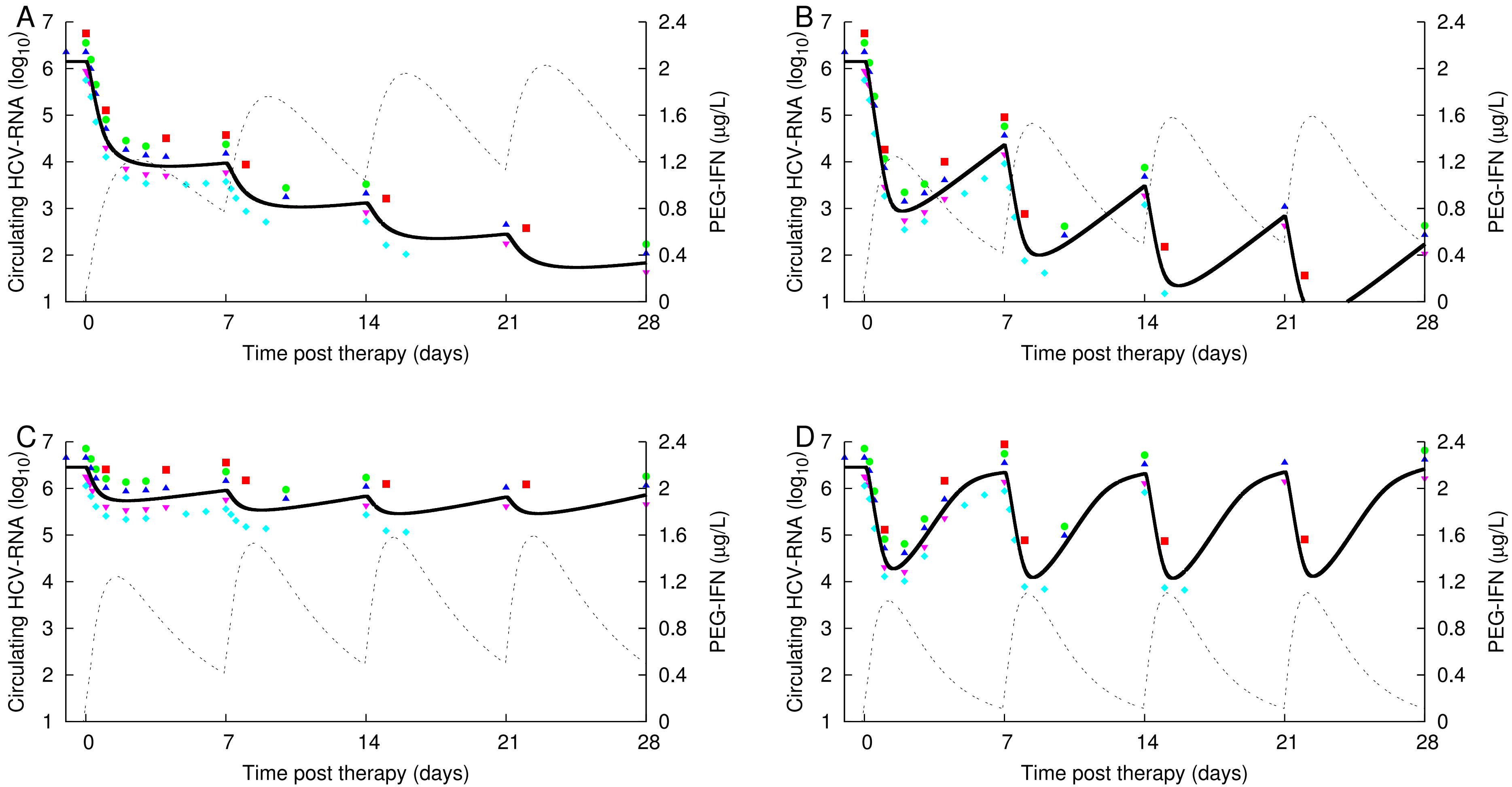

FIGURE 1

Simulation of the median viral dynamics (black) and Peg-IFN profile (dashed) according to the four parameter set-ups given in Table 2

A) Good responders with low inter-dose variations. **B**) Good responders with high inter-dose variations. **C**) Poor responders with low inter-dose variations. **D**) Poor responders with high inter-dose variations. For each case, we indicate the data points where the viral load is sampled according to the five designs given in Table 1: ■ is $\mathcal{D}_1$, ● is $\mathcal{D}_2$, ▲ is $\mathcal{D}_3$▼ is $\mathcal{D}_4$,◆ is $\mathcal{D}_5$.

For the sake of readibility, the symbols are drawn above and under the curve and are not indicated for the Peg-IFN profiles since the measurements are done at the same time point.

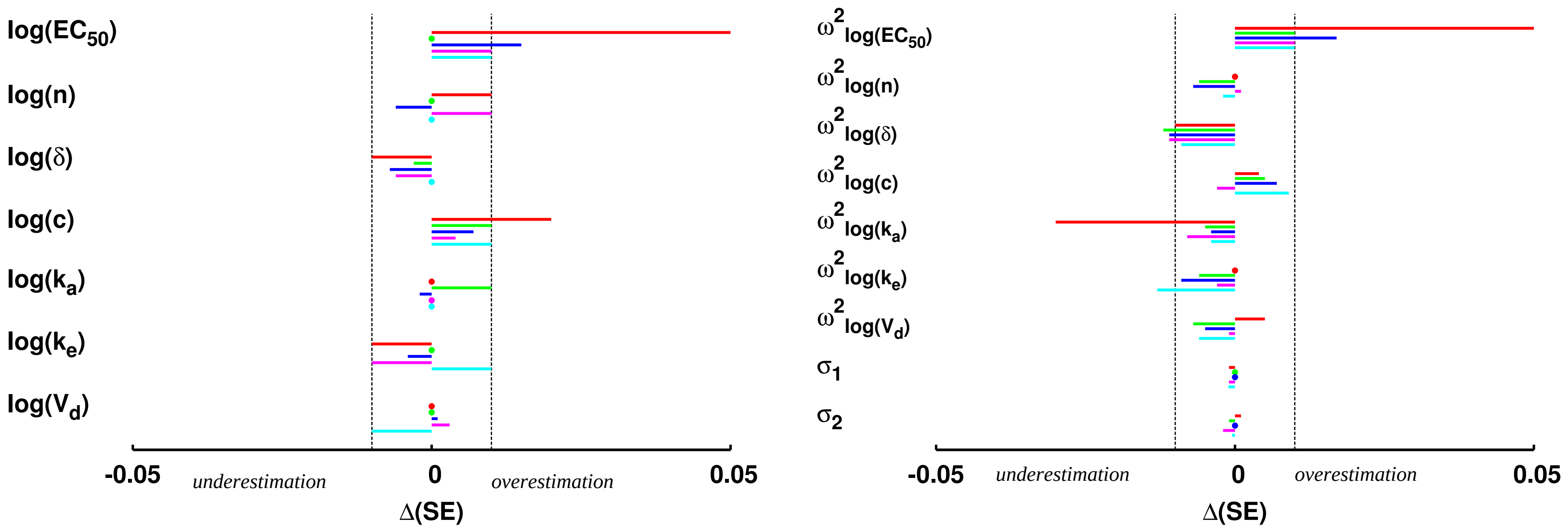

FIGURE 2

Difference between the approximated and the empirical standard errors according to the study design. The difference is denoted by $\Delta(SE) = SE_{PFIM} - SE_{EMP}$ and is displayed separately for the fixed effects parameters (left) and the variance parameters (right). The vertical lines define the area where $-0.01 < \Delta(SE) < 0.01$ and the symbol "$\bullet$" indicates that $\Delta(SE) = 0$ (up to three significant digits); the code color for the study designs is similar to Figure 1.

[9] Wu H, Zhu H, Miao H, Perelson A. Parameter Identifiability and Estimation of HIV/AIDS Dynamic Models. *Bulletin of Mathematical Biology* 2008; **70**(3):785–799.

[10] Han C, Chaloner K. Design of population studies of hiv dynamics. *Deterministic and stochastic models of AIDS epidemics and HIV infections with intervention*, Tan W, Wu H (eds.). World Scientific Pub Co Inc, 2005.

[11] Guedj J, Thiebaut R, Commenges D. Practical Identifiability of HIV Dynamics Models. *Bulletin of Mathematical Biology* 2007; **69**(8):2493–2513.

[12] Miao H, Dykes C, Demeter L, Cavenaugh J, Park S, Perelson A, Wu H. Modeling and Estimation of kinetics parameters and replicative fitness of HIV-1 from flow-cytometry-based growth competition experiments. *Bulletin of Mathematical Biology* 2008; **70**(6):1749–1771.

[13] Lavielle M, Samson A, Fermin A, Mentré. Maximum likelihood estimation of long-term hiv dynamic models and antiviral response. *Biometrics* ; .

[14] Mentré F, Mallet A, Baccar D. Optimal design in random-effects regression models. *Biometrika* 1997; **84**:429–442.

[15] Bazzoli C, Retout S, Mentré F. Fisher information matrix for nonlinear mixed effects multiple response models: evaluation and comparison by simulation to results from FO, FOCE and SAEM algorithms . *Statistics in Medicine* 2009; **28**(14):1940–1956.

[16] Bazzoli C, Retout S, Mentré F. Design evaluation and optimisation in multiple response nonlinear mixed effect models: PFIM 3.0. *computer methods and programs in biomedicine* 2010; **98**:55–65.

[17] Pawlotsky J, Dahari H, Neumann A, Hezode C, Germanidis G, Lonjon I, Castera L, Dhumeaux D. Antiviral action of ribavirin in chronic Hepatitis C. *Gastroenterology* 2004; **126**:703–714.

[18] Layden-Almer J, Ribeiro R, Wiley T, Perelson A, Layden T. Viral dynamics and response differences in HCV-infected African American and white patients treated with IFN and ribavirin. *Hepatology* 2003; **37**(6):1343–1350.

[19] Dixit N, Layden-Almer J, Layden T. Modelling how ribavirin improves interferon response rates in hepatitis C virus infection. *Nature* 2005; **432**:922–924.

[20] Herrmann E, Neumann A, Schmidt J, Zeuzem S. Hepatitis C virus kinetics. *Antiviral Therapy* 2000; **5**(2):85–90.

[21] Guedj J, Rong L, Dahari H, Perelson A. A perspective on modeling hepatitis C virus infection. *Journal of Viral Hepatitis* 2010; .

[22] Atkinson A, Donev A. *Optimum Experimental Designs*. Oxford University Press, 1992.

[23] Retout S, Comets E, Samson A, Mentré F. Design in nonlinear mixed effects models: optimization using the Fedorov-Wynn algorithm and power of the Wald test for binary covariates. *Statistics in Medicine* 2007; **26**(28):5162–79.

[24] Zeuzem S, Pawlotsky J, Lukasiewicz E, von Wagner M, Goulis I, Lurie Y, Gianfranco E, Vrolijk J, Esteban J, Hezode C, *et al.*. International, multicenter, randomized, controlled study comparing dynamically individualized versus standard treatment in patients with chronic hepatitis C. *Journal of Hepatology* 2005; **43**(2):250–257.

[25] Sherman K, Shire N, Rouster S, Peters M, James Koziel M, Chung R, Horn P. Viral kinetics in hepatitis C or hepatitis C/human immunodeficiency virus-infected patients. *Gastroenterology* 2005; **128**(2):313–27.

[26] Herrmann E. Effect of ribavirin on hepatitis C viral kinetics in patients treated with pegylated interferon. *Hepatology* 2003; **37**(6):1351–1358.

[27] Zeuzem S, Herrmann E, Lee J, Fricke J, Neumann A, Modi M, Colucci G, Roth W. Viral kinetics in patients with chronic hepatitis c treated with standard or peginterferon α2a. *Gastroenterology* 2001; **120**(6):1438–1447.

[28] Bruno R, Sacchi P, Ciappina V, Zochetti C, Patruno S, Maiochhi L, Filice G. Viral dynamics and pharmacokinetics of peginterferon alpha-2a and peginterferon alpha-2b in naive patients with chronic hepatitis C: a randomized, controlled study. *Antiviral Therapy* 2004; **9**:491–498.

[29] Dahari D, Affonso de Araujo E, Haagmans B, Layden J, Cotler S, Barone A, Neumann AU. Pharmacodynamics of peg-ifn alpha-2a in hiv/hcv co-infected patients: Implications for treatment outcomes. *Journal of Hepatology* 2010; .

[30] Kuhn E, Lavielle M. Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistic & Data Analysis* 2005; **49**:1020–1038.

[31] Bates D, Watts D. Relative curvature measures of nonlinearity. *Journal of the Royal Statistical Society. Series B (Methodological)* 1980; :1–25.

[32] Tan W, Wu H. *Deterministic and stochastic models of AIDS epidemics and HIV infections with intervention*. World Scientific Singapore:, 2005.

[33] Jones B, Wang J. Constructing optimal designs for fitting pharmacokinetic models. *Statistics and Computing* 1999; **9**(3):209–218.

[34] Retout S, Mentré F, Bruno R. Fisher information matrix for non-linear mixed-effects models: evaluation and application for optimal design of enoxaparin population pharmacokinetics. *Statistics in Medicine* 2002; **21**(18):2623–2639.

[35] Neumann A, Layden T, Reddy K, Levi-Drummer R, Poulakos 2nd J. The 2nd phase slope of HCV decline is highly predictive of sustained virologic response following consensus IFN treatment for chronic hepatitis C and is determined by genotype but not dose. *Hepatology* 2000; **32**:356A.

[36] Han C, Chaloner K, Perelson A. Bayesian analysis of a population HIV dynamic model. *Case studies in Bayesian statistics* 2002; **6**:223–237.

[37] Han C, Chaloner K. Bayesian experimental design for nonlinear mixed-effects models with application to HIV dynamic. *Biometrics* 2004; **60**:25–33.

[38] Thiébaut R, Guedj J, Jacqmin-Gadda H, Chêne G, Trimoulet P, Neau D, Commenges D. Estimation of dynamical model parameters taking into account undetectable marker values. *BMC Medical Research Methodology* 2006; **6**(1):1–10.

[39] Guedj J, Neumann A. Understanding hepatitis C viral dynamics with direct-acting antiviral agens due to the interplay between intracellular replication and cellular infection dynamics. *Journal of Theoretical Biology* ; **267**(3):330–340.

[40] Retout S, Comets E, Thus C, Mentré F. Design optimization in nonlinear mixed effects models using cost functions: application to a joint model of infliximab and methotrexate pharmacokinetics. *Communications in Statistics-Theory and Methods* 2009; **38**(18):3351–3368.