Web-based Supplementary Materials for Maximum likelihood estimation of long term HIV dynamic models and antiviral response by

Marc Lavielle
INRIA Saclay, France

Adeline Samson
CNRS UMR8145, Université Paris Descartes, France

Ana Karina Fermin
Université de Nanterre, France

France Mentré
INSERM U738 and Université Paris-Diderot, France

Web Appendix A

The steady state values of model $\mathcal{M}_B$are given by:

$$\begin{aligned} T_{NI}(0) &= \frac{\mu_I \mu_V}{p\,\gamma}, \qquad (1) \\ T_I(0) &= \frac{\mu_V V_I(0)}{p}, \\ V_I(0) &= \frac{\lambda - \mu_{NI} T_{NI}(0)}{\gamma\, T_{NI}(0)}, \qquad (2) \\ V_{NI}(0) &= 0 \end{aligned}$$

The steady state values of model $\mathcal{M}_Q$are given by:

$$\begin{aligned} T_Q(0) &= \frac{\lambda + \rho\, T_{NI}(0)}{\alpha_Q + \mu_Q} \\ T_{NI}(0) &= \frac{\mu_V\, \mu_{NI}}{\gamma\, p} \\ T_I(0) &= \frac{\mu_V\, V_I(0)}{p} \\ V_I(0) &= \frac{\alpha_Q\, T_Q(0)}{\gamma\, T_{NI}(0)} - \frac{\rho + \mu_{NI}}{\gamma} \\ V_{NI}(0) &= 0 \end{aligned} \tag{3}$$

The steady state values of model $\mathcal{M}_L$are given by:

$$\begin{aligned} T_{NI}(0) &= \frac{\mu_A \mu_V (\alpha_L + \mu_L)}{\gamma p (\alpha_L + \pi \mu_L)} \\ T_A(0) &= \frac{\mu_V V_I(0)}{p} \\ T_L(0) &= \frac{(1-\pi)\gamma\, T_{NI}(0) V_I(0)}{\alpha_L + \mu_L} \\ V_I(0) &= \frac{\lambda - \mu_{NI} T_{NI}(0)}{\gamma\, T_{NI}(0)} \\ V_{NI}(0) &= 0. \end{aligned} \tag{4}$$

Web Appendix B

Algebraic conditions of parameters identifiability of the three dynamics models can not be found in a closed form. Thus, we decided to examine the identifiability problem by simulation, *i.e.* by performing some sensitivity analysis. We selected a set of parameter $\theta^A = (\theta_1^A, \theta_2^A, \ldots, \theta_{11}^A)$ (some typical values for a "responder") and computed the predicted (log) viral load $f_{\theta^A}(t)$ and the predicted CD4 cells count $g_{\theta^A}(t)$ obtained with this set of parameters. Then, for $k = 1, 2, \ldots, 11$,

1. we modified the kth component of θ^A by setting $\tilde{\theta}_k^A = 0.7 \times \theta_k^A$,
2. we created a new set of parameters $\tilde{\theta}^A$ by modifying θ^A "as less as possible" and such that $f_{\tilde{\theta}^A}$ and $g_{\tilde{\theta}^A}$ are "as close as possible" to f_{θ^A} and g_{θ^A}.

The table displays θ^A and 3 different set of parameters $\tilde{\theta}^A$ that give the same predicted viral loads and CD4 counts:

λ	2.5	2.5	2.5	2.5
γ	0.00225	**0.001575**	0.00225	0.00225
π	0.46	**0.46853**	0.46	0.46
α_L	1.5e-005	1.5e-005	1.5e-005	1.5e-005
p	620	**629.25**	**434**	620
μ_{NI}	0.008	0.008	0.008	0.008
μ_L	0.008	0.008	0.008	0.008
μ_A	0.3	**0.30156**	0.3	0.3
μ_V	30	30	**21.006**	30
η_{PI}	0.95	**0.9081**	0.95	**0.665**
η_{RTI}	0.92	0.92	0.92	**0.988**

We repeated this analysis with another set of parameters θ^B (some typical values for a "non responder"). The table below displays θ^B and 3 different set of parameters $\tilde{\theta}^B$ that give the same predicted viral loads and CD4 counts:

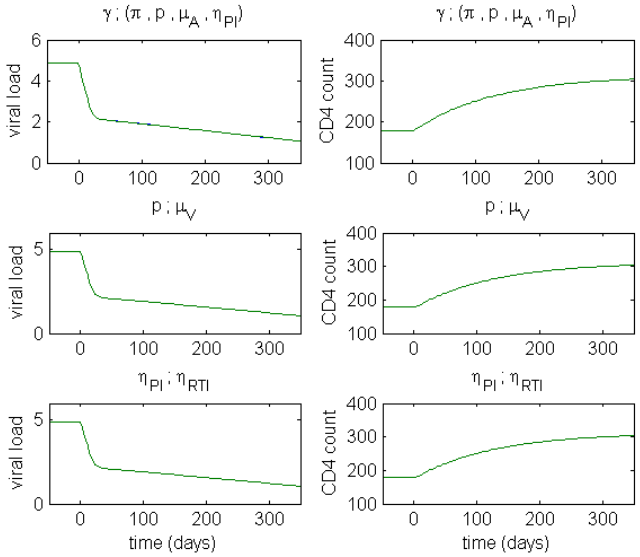
λ	2.5	2.5	2.5	2.5
γ	0.00225	**0.001575**	0.00225	0.00225
π	0.46	**0.46644**	0.46	0.46
α_L	1.5e-005	1.5e-005	1.5e-005	1.5e-005
p	620	**724.91**	**434**	620
μ_{NI}	0.008	0.008	0.008	0.008
μ_L	0.008	0.008	0.008	0.008
μ_A	0.3	**0.34932**	0.3	0.3
μ_V	30	30	**20.997**	42.849
η_{PI}	0.3	**0.16081**	0.3	**0.2**
η_{RTI}	0.3	0.3	0.3	**0.3798**

In other words,

- γ, π, p, μ_A and η_{PI} are not identifiable simultaneously,
- p and μ_V are not identifiable simultaneously,
- η_{PI} and η_{RTI} are not identifiable simultaneously.

Figure 1 displays f_{θ^A} and g_{θ^A} in blue and $f_{\tilde{\theta}^A}$ and $g_{\tilde{\theta}^A}$ in green. Figure 2 displays f_{θ^B} and g_{θ^B} in blue and $f_{\tilde{\theta}^B}$ and $g_{\tilde{\theta}^B}$ in green.

Figure 1: Predicted viral load and CD4 cell count profiles obtained with different set of parameters (responder profiles)



This sensitivity analysis only considers the problem of fitting the structural model for a given subject. This method is not completely appropriate for a population approach where the distribution of the individual parameters is also part of the (statistical) model. In other word, if a parameter is not algebraically (or structurally) identifiable, it can be statistically identifiable. Consider the following "toy example":

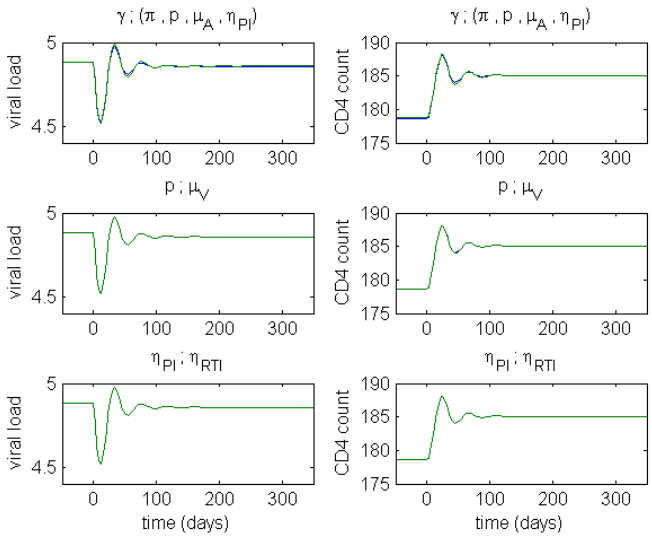
$$y_{ij} = A_i t_{ij} + B_i t_{ij} + \varepsilon_{ij}$$

This model is clearly not algebraically identifiable. Assuming now that A_i and B_i are random variables, statistical identifiability of the population parameters depends on the population distribution of A_i and B_i. We see in Figure 3 the convergence of the SAEM algorithm for different initializations and different distributions. The population parameters A and B are not identifiable when A_i and B_i are normally distributed. They are identifiable when A_i and B_i are defined using a logit transformation.

In our viral kinetic model, even if only the product $(1-\eta_{PI})(1-\eta_{RTI})$ is structurally identifiable, both η_{PI} and η_{RTI} are statistically identifiable.

In summary, identifiability of non linear mixed effects models does not reduce to algebraic identifiability and pracival identifiability should also be

Figure 2: Predicted viral load and CD4 cell count profiles obtained with different set of parameters (non responder profiles)



addressed. As a practical diagnostic tool, we propose to use the Fisher Information Matrix for detecting some over-parametrization in the model. We are aware that it is not completely satisfactory but from our experience, we know for certain that very large s.e. (or NaN) indicate some issue in the parametrization. Unfortunately, the reverse is not necessarily true.

Figure 3: Convergence of SAEM with different initial values and different population distributions. Left: normal distribution ($A + B$ is identifiable) , Right: logit transformation (both A and B are identifiable).