Biostatistics (2015), **0**, 0, *pp.* 1–35
doi:10.1093/biostatistics/MSMvsMeca-submitted-20150116

Dynamic versus marginal structural models for estimating the effect of HAART on CD4 in observational studies : application to the Aquitaine Cohort study and the Swiss HIV Cohort Study.

M. PRAGUE[1*], D. COMMENGES[2,3,4], J.M. GRAN[5], B. LEDERGERBER[6], J. YOUNG[7], H. FURRER[8], R. THIÉBAUT[2,3,4]

1 *Harvard T.H. Chan School of Public Health, Boston, USA*

2 *Univ. Bordeaux, ISPED, F-33000 Bordeaux, France*

3 *INSERM, Centre INSERM U897-Epidémiologie et Biostatistique, F-33000, Bordeaux, France*

4 *INRIA (SISTM) Centre Recherche Bordeaux Sud-Ouest, Univ. Bordeaux, Talence, France*

5 *Oslo Center for Biostatistics and Epidemiology, Department of Biostatistics, Univ. of Oslo, Norway*

6 *Division of Infectious Diseases and Hospital Epidemiology, Univ. Hospital Zurich, Univ. of Zurich, Zurich, Switzerland*

7 *Basel Institute for Clinical Epidemiology and Biostatistics, University Hospital Basel, Basel, Switzerland*

8 *Department of Infectious Diseases, Bern Univ. Hospital and Univ. of Bern, Switzerland*

mprague@hsph.harvard.edu

*To whom correspondence should be addressed.

Summary

Highly active antiretroviral therapy (HAART) has proved efficient in increasing CD4 counts in many randomized clinical trials. Because randomized trials have some limitations (e.g. short duration, highly selected subjects), it is still interesting to assess the effect of HAART using observational studies. This is however challenging because treatment is started preferentially in subjects with severe conditions, in particular in subjects with low CD4 counts. The difficulty of assessing treatment effect in observational studies is a general problem arising for other diseases which has been treated using Marginal Structural Models (MSM) relying on the counterfactual formalism. Another approach to causality is based on dynamical models. We present first discrete-time dynamic models, based on Linear Increment Models (LIM), which the simplest models consist in one difference equation for CD4 counts while a system of two difference equations allows modeling jointly CD4 counts and viral load. Then we further extend in continuous time with mechanistic models based on ordinary differential equations with random effects (ODE-NLME). Mechanistic models allow incorporating biological knowledge when available which leads to increased power for detecting a treatment effect. Inference in ODE-NLME however is challenging from a numerical point of view, and requests specific methods and softwares. LIM are a valuable intermediary option in terms of consistency, precision and complexity. The different approaches are compared using a simulation study and then applied to two HIV cohorts studies (the ANRS CO3 Aquitaine Cohort and the Swiss HIV Cohort Study).

Key words: Causality; Dynamic models; Dynamic treatment regimen; HAART; Linear Increment Models (LIM); Marginal Structural Models (MSM); Mechanistic models; Non Linear Mixed Effect Models (NLME); Observational study; ODE-NLME; Ordinary Differential Equation (ODE), Treatment effect.

1. Introduction

Assessing the effect of a treatment in observational studies is useful because randomized clinical trials have often short durations and often include highly selected subjects. This is however challenging because most of the time the treatment regime is a dynamic process, and covariates history of a subject up to time t may influence treatment given after t, and may also influence the outcome of interest, which induces a time-dependent confounding (Psaty *and others*, 1999). For instance, one may wish to assess the effect of antiretroviral therapy in HIV infected subjects. As CD4+ T-lymphocytes (CD4, in short) are the main target cells of the HIV virus, it is possible to assess the effect of a treatment on the blood concentration of these cells: CD4 counts are measurements of this concentration. In observational studies however, the decision to start an antiretroviral therapy may depend on CD4 counts as well as on other covariates. In this setting, it has been demonstrated that conventional regressions lead to biased estimates of the treatment effect, typically underestimating it and possibly (wrongly) indicating a negative effect. This phenomenon can be called "confounding by indication" (Walker, 1996).

Robins (1994) introduced the structural nested models which can be applied to this problem but difficulties appear in the presence of time-dependent treatment and confounders, see Robins (2000) for theoretical explanations. G-computation (Westreich *and others*, 2012) and inverse probability weighted (Robins *and others*, 2000) marginal structural models (MSM) are alternative approaches for treating this problem. As estimators based on G-computation require more parametric modeling assumptions (Young *and others*, 2011), we will focus on the latter. Thus, when referring to MSM later in this paper we generally think of MSM based on inverse probability weighting. The weights are inverse-probability-of-treatment and are determined in order to eliminate the bias due to the dependance of treatment attribution on covariates linked to the outcome, and are obtained through a model of treatment attribution. The parameters are estimated through a weighted generalized estimating equation (GEE). The approach has been

applied by Hernán *and others* (2002) and by Cole *and others* (2005) for estimating the effect of zidovudine and of highly active antiretroviral therapy (HAART) on CD4 count. Cole *and others* (2007), Sterne *and others* (2005) and Cole *and others* (2010) used it for estimating the effect of HAART on viral load and on AIDS or death. The estimation of treatment effect comes from the contrast between the expected outcome with and without treatment.

Whereas time dynamics are often ignored in causal modeling such as MSM, an alternative approach is to use dynamic models. A dynamic approach to causality has been pioneered by Granger (1969), Schweder (1970), Aalen (1987), and further developed by Didelez (2008); Commenges and Gégout-Petit (2009); Gégout-Petit and Commenges (2010); Eichler and Didelez (2010). Assumptions needed for a causal interpretations of dynamic models have been presented in Arjas and Parner (2004) and Commenges and Gégout-Petit (2015). Dynamical models in discrete time, and in particular Linear Increment models (LIM), have been proposed by Diggle *and others* (2007) and Hoff *and others* (2015) ; Aalen *and others* (2012) have suggested that such models can be useful for studying HAART effect on CD4 counts or viral load. Discrete-time models however may not be completely satisfactory because the processes of interest most often live in continuous time. Systems of differential equations in continuous time can also be used to model the interaction between HIV and CD4 cells populations. Models based on differential equations, called mechanistic, considerably helped in understanding some important features of the infection: see Perelson (2002) for a review. First, they were used to estimate biological rates in HIV infection: Ho *and others* (1995) uncovered the high turnover rates of virus and HIV infected CD4 cells. More interestingly in our setting, it is possible to model the treatment effect from a biological perspective. Introducing random effects allows analyzing a sample of subjects with different parameter values without increasing too much the number of parameters (Wu, 2005; Guedj *and others*, 2007; Lavielle *and others*, 2011). Refined models taking into account pharmacokinetics or pharmacodynamic aspects have also been proposed for treatment monitoring (Prague *and others*,

2012). In principle stochastic differential equations could be considered, but they are difficult to handle statistically as well as numerically, and methods are restricted for the moment to rather simple systems: see for instance Delattre *and others* (2013); so we restrict to models based on Ordinary Differential Equations (ODE) systems. Fitting ODE-based models with random effects (ODE-NLME) is numerically challenging because of the need to solve the ODE system and to compute numerical integrals which appear in the likelihood. Up to now, mechanistic models have been used to analyze data from clinical trials. Using mechanistic models to estimate the effect of HAART based on data of large observational cohorts is a possibility, that to our knowledge has never been attempted.

The aim of this paper is to compare the dynamic models in discrete (LIM) and continuous time (ODE-NLME) to the MSM in the problem of assessing the causal effect of a treatment on a marker in observational studies. Specifically, we aim at estimating HAART effect in HIV infected patients. We present several possible models and compare them using simulations and real data. In Section 2, we present the statistical models: the naive model, the MSM models, the dynamic models and a mechanistic model. In Section 3, we compare the results of these models in a simulation study, where the data are generated from a complex mechanistic model. Section 4 presents the application of these different models on the data of two cohorts of HIV infected patients: the Swiss HIV Cohort Study (SHCS) and the ANRS CO3 Aquitaine cohort. Section 5 concludes.

2. Modeling the treatment effect in observational studies

2.1 *Notations and the naive model*

We denote the value of a physiological marker Y for subject i at time t by Y_t^i. The vector Y_t represents the biomarkers for all subjects at time t. The value of a treatment given at time t for subject i is denoted by A_t^i. For sake of simplicity, we only model two treatment states: $A(t) = 0$

when treatment is not given, and $A(t) = 1$ when treatment is given at time t, and we assume that once initiated, the treatment is not interrupted; generalization to different treatment levels is possible. If treatment is started at time t then $A_t^i = 1$ and $A_{t-1}^i = 0$. In our application Y_t^i is the CD4 counts, A_t^i is treatment (HAART) attribution which is binary, and V_t^i may be the viral load or other biological processes linked to the infection. We use overbars to represent histories of the processes: for instance $\bar{A}_t^i = (A_0^i, A_1^i, \ldots, A_t^i)$. We denote by $\text{cum}(\bar{A}_t^i)$ the cumulative time under treatment until time t for subject i. Since A is binary we can write $\text{cum}(\bar{A}_t) = \sum_{j=1}^{t} A_j^i$

In the absence of confounding by indication, a regression of Y_t on the history of treatment would give the effect of treatment on the marker. The simplest model would be to regress Y_t on $\text{cum}(\bar{A}_{t-1})$. It has been noted however that a piecewise linear regression model allowing a change of treatment effect after one year was better suited (Cole *and others*, 2005). Thus, the naive model that we consider is our **Model 1**:

$$\text{E}(Y_t|\bar{V}_t, \bar{A}_t) = \beta_0 + \beta_1 \text{cum}(\bar{A}_{t-1}) + \beta_2 \text{cumlag}(\bar{A}_{t-1}). \tag{2.1}$$

where $\text{cumlag}(\bar{A}_t)$ is the cumulative time under treatment up to time t minus one year: $\text{cumlag}(\bar{A}_t) = max(0, \text{cum}(\bar{A}_t) - 1)$. The β's are estimated by conventional GEE because it is very likely that the Y_t's are positively correlated, and because we are interested in the population average (Hubbard *and others*, 2010). A working correlation structure has to be chosen and the choice may have an impact, not on the bias, but on the efficiency of the estimators (Liang and Zeger, 1986). However, in presence of time-varying covariates, the independence model should be chosen, otherwise results could be seriously biased, as shown by Pepe and Anderson (1994). Thus an identity working correlation matrix is used for all GEE models.

2.2 *Marginal structural models (MSM)*

As suggested in the introduction, in observational studies there is often an indication bias using naive models: treatment is given to subjects with low CD4 counts, hence treated subjects tend

to have low CD4 counts. The true value of parameter β_1 in Model 1 cannot be interpreted as the causal effect. The MSM have been designed to estimate the causal effect of a treatment in the case of dynamic treatment regimes. It is assumed that to each particular value $\bar{a}_t$ of $\bar{A}_t$ (or treatment regime), a potential outcome $\bar{Y}_t^i(\bar{a})$ is associated for subject i. This means that if subject i had (possibly contrary to the fact) treatment trajectory $\bar{a}_t$, his outcome would be $\bar{Y}_t^i(\bar{a})$. A model is postulated to describe how the potential outcomes vary as a function of the different treatment regimes. Then it has been shown that the parameters of this model, called causal parameters, can be estimated with a suitably weighted GEE. The weights are inverse-probability-of-treatment (IPT); extension to the case with censoring has also been developed (Cole *and others*, 2005; Cole and Hernán, 2008) but this is not be used in this paper. The weights are generally estimated using logistic models for each time and are the product over time of weights deriving from a model of treatment attribution. The causal parameters can be estimated consistently if all the confounders (factors influencing both the outcome of interest and treatment attribution) have been taken into account. Hernán *and others* (2002) and Cole *and others* (2005) proposed benchmark models also adjusting for confounders such as time (t) and baseline value of the biomarker (Y_0) in the regression. The model for potential outcomes that we consider is the same for our **Models 2** and **Model 3**; it is:

$$\mathrm{E}(Y_t(\bar{a})|\bar{V}_t) = \beta_0 + \beta_1 \mathrm{cum}(\bar{A}_{t-1}) + \beta_2 \mathrm{cumlag}(\bar{A}_{t-1}) + \beta_3 t + \beta_4 Y_0. \tag{2.2}$$

We use the stabilized weights defined as :

$$SW(t) = \prod_{k=1}^{t} \frac{Pr(A_k = 1|\bar{A}_k, L_0)}{Pr(A_k = 1|\bar{A}_k, \bar{L}_k)},$$

where the probabilities are predicted for every subject from a logistic regression depending on a subset of baseline covariates denoted L_0 and their history up to time k denoted $\bar{L}_k$ ($L_0 \in \bar{L}_k$). The MSM correction compared to Model 1 depends on the correctness of the IPT model. So, we tried two models of treatment attribution. We define the subsets L for treatment attribution

model in **Model 2** as baseline and time-varying CD4 count in class (< 200, $[200; 400]$, > 400) only. We extend this list for **Models 3** to the main potential confounders: viral load in categories (< 401, $401 - 10000$ and > 10000), and an indicator of undetectable viral load.

In Models 1, 2 and 3, the effect of treatment on CD4 counts during the first year is given by β_1 and the effect after one year of treatment is given by $\beta_1 + \beta_2$. The impact of treatment on CD4 counts after an infinite time is infinite: it tends to $+\infty$ or $-\infty$ according to the sign of $\beta_1 + \beta_2$.

2.3 *Discrete-time dynamical models (LIM)*

When using dynamical models, we regress the change in the marker of interest $Z_t^i = Y_t^i - Y_{t-1}^i$. This fits well with a causal thinking which considers that the change of a process depends on its present, and possibly past, state. The independence assumption for the Z_t^i's is much more acceptable than for the Y_t^i's. However, there remains an inter-subject variability that we model by a random effect b_i, assumed normally distributed with zero expectation. This leads to a model for Z_t^i which is not marginal, but conditional on the b_i; this is **Model 4**:

$$Z_t^i = \beta_0 + \beta_1 A_{t-1}^i + \beta_2 A_{t-2}^i + b_i + \varepsilon_t^i, \tag{2.3}$$

where the ε_t^i's are iid normally distributed variables with zero expectation. This model can be fitted easily since this is a linear mixed-effects model. In order to account for non equally spaced measurement of biomarkers, we extend the notation to obtain a model which has approximately the same meaning: it is natural to think that the change is be proportional to the time elapsed between two observations, that we note $\Delta_t^i = c_t^i - c_{t-1}^i$, where c_t^i is the t^{th} calendar time of observation since baseline measure of subject i. This extended model is obtained by redefining the increment as $Z_t^i = \frac{Y_t^i - Y_{t-1}^i}{\Delta_t^i}$. In Models 4, the effect of treatment on the CD4 counts during the first year is approximated by $\beta_1 \bar{\Delta}_t$, and the effect after one year of treatment by $(\beta_1 + \beta_2)\bar{\Delta}_t$, where $\bar{\Delta}_t$ is the mean of all the Δ_t^i. As in previous models the value of Y tends to infinity as time increases.

Many deterministic dynamical models have equilibrium points; similarly many stochastic dynamical models tend toward a stationary process: this property fits very well with the behavior of biological systems since concentrations of many molecules or cells have a tendency to return around the same value, a property called homeostasis. The difference equation of the type $Y_t - Y_{t-1} = \gamma_0 + \gamma_1 Y_{t-1} + \varepsilon_t$ corresponds to an autoregressive model of order one, noted AR(1): $Y_t = \gamma_0 + \gamma' Y_{t-1} + \varepsilon_t$ with $\gamma' = (\gamma_1 + 1)$. It is well known that is $|\gamma'| < 1$ this process converges toward a stationary process (in discrete time) with expectation $\mathrm{E}(Y_t) = -\frac{\gamma_0}{1-\gamma'} = -\frac{\gamma_0}{\gamma_1}$; this is always defined unless $\gamma_1 = 0$, as is the case in Model 4 which does not have a finite stationary expectation. The condition amounts to $-2 < \gamma_1 < 0$ and to get a positive stationary expectation we must have $\beta_0 > 0$. When using a model which has this convergence property, it may not be necessary to have a two-slope model, so we define **Model 5** as:

$$Z_t^i = \beta_0 + \beta_1 A_{t-1}^i + \beta_2 Y_{t-1}^i + b_i + \varepsilon_t^i. \tag{2.4}$$

If $-2 < \beta_2 < 0$ and $\beta_0 + \beta_1 > 0$, Y tends to a stationary process with expectation $-\frac{\beta_0+\beta_1}{\beta_2}$ for treated patients.

A more realistic modeling of CD4 counts is to take viral load into account. Here, we make a step toward mechanistic models because we know that the virus concentration and the CD4 concentration are two inter-related processes. Thus, we propose **Model 6** based on a system of two difference equations:

$$\left\{ \begin{array}{rcl} Z_t^i & = & \beta_0 + \beta_1 A_{t-1}^i + \beta_2 Y_{t-1}^i + \beta_3 V_{t-1}^i + b_i + \varepsilon_{it}^1, \\ W_t^i & = & \alpha_0 + \alpha_1 A_{t-1}^i + \alpha_2 Y_{t-1}^i + \alpha_3 V_{t-1}^i + d_i + \varepsilon_{it}^2. \end{array} \right. \tag{2.5}$$

where $W_t^i = \frac{V_t^i - V_{t-1}^i}{\Delta_t^i}$, with V_t^i the viral load at time t, d_i and b_i are normally distributed independent random effects, and ε_{it}^1 and ε_{it}^2 are normally iid error variables with zero expectation. Assessing the long term treatment effect in Model 5 and 6 is possible by analytically computing the the stationary expectations of CD4 counts and vial load with ($A^i = 1$) or without ($A^i = 0$) treatment. However, one year and subsequent years increase of CD4, as well as the long term

change, are more easily computed by solving the difference equations numerically. The standard deviation of the of CD4 increase could be computed using predictions from the mean value of baseline CD4 and viral load by simulating parameters values from their estimated asymptotic distribution. However, for testing whether the treatment has an effect, it is easier and more powerful to test the hypotheses $\beta_1 = 0$ and $\alpha_1 = 0$, by Wald tests for instance.

2.4 *Continuous Dynamical models, Mechanistic Models (ODE-NLME)*

In reality, biomarkers processes live in continuous time. A natural extension of the dynamical models in discrete time leads to models based on differential equations. We use the "target cells model" that proved to provide a good fit and to have good prediction abilities (Prague *and others*, 2013). The combination of the target cells model, a model for inter-individual variability of the parameters, and an observation model specifies our **Model 7**.

2.4.1 *Biological system* We know that only infected cells can produce viruses (V), so we should distinguish between uninfected (T) and infected cells (T^*). Moreover, it has been shown that not all CD4 T-cells can be infected with the same probability, so we can distinguish between quiescent cells (Q) and target cells (T). The instantaneous change of concentrations of these populations at time t, for all real value of $t > 0$, is given by the ODE system:

$$\left\{\begin{array}{lcl} \frac{\mathrm{d}Q_t^i}{\mathrm{d}t} & = & \lambda^i + \rho^i T_t^i - \alpha^i Q_t^i - \mu_{Q^i}^i Q_t^i, \\ \frac{\mathrm{d}T_t^i}{\mathrm{d}t} & = & \alpha^i Q_t^i - \gamma^i T_t^i V_t^i - \rho^i T_t^i - \mu_T^i T_t^i, \\ \frac{\mathrm{d}T^{*i}_t)}{\mathrm{d}t} & = & \gamma^i T_t^i V_t^i - \mu_{T^*}^i T^{*i}_t, \\ \frac{\mathrm{d}V_t^i}{\mathrm{d}t} & = & \pi^i T^{*i}_t - \mu_V^i V_t^i. \end{array}\right. \tag{2.6}$$

The system is graphically represented in Figure 1a. Here, the parameters have biological meanings: λ is the production rate of new CD4 cells, the μ's are death rates, α and ρ are transition rates between quiescent and activated cells, π is the rate of production of virions by infected cells, and γ is the infectivity parameter. The model assumes that the rate of infection of target T cells is γV_t. It has been proved that a patient-by-patient approach is not powerful and allows estimating only

a very limited number of parameters. A population approach obtained by introducing random effects (Huang *and others*, 2006; Guedj *and others*, 2007; Lavielle *and others*, 2011), allows estimating more parameters.

2.4.2 *Inter-individual variability* The model for inter-individual variability of the parameters is a mixed-effect model and includes possibly time-dependent explanatory variables (such as the treatment) and random effects. In this application, based on Prague *and others* (2012), normal random effects u^i_λ and $u^i_{\mu_{T^*}}$ is put on the two log-transformed parameters (denoted with a tilde) $\tilde{\lambda}^i = \tilde{\lambda}_0 + u^i_\lambda$ and $\tilde{\mu}^i_{T^*} = \tilde{\mu}_{T^*0} + u^i_{\mu_{T^*}}$. Biologically, the causal effect of treatment can be modeled as an effect on the infectivity parameter γ. The parameter γ consequently depends on t through A_t:

$$\tilde{\gamma}^i(t) = \tilde{\gamma}_0 + \beta A^i_t, \tag{2.7}$$

where we expect $\beta < 0$, so that biologically the treatment decreases the infectivity of the virus.

2.4.3 *Observation model* One important consequence of using continuous time models is that we must distinguish between the biological system which lives in continuous time and observations which are made at discrete time. We only have observations of total CD4 counts and viral loads at discrete times t_{ij}, $Q^i_{t_{ij}} + T^i_{t_{ij}} + {T^*}^i_{t_{ij}}$ and $V^i_j = V^i_{t_{ij}}$ respectively. We also need to model the measurement error. To make an additive model for measurement error acceptable, we use 4th-root transformation for CD4 and a $\log_{10}$ transformation for the viral load respectively. Thus, the observation model is:

$$(Y^i_j)^{1/4} = [Q^i_{t_{ij}} + T^i_{t_{ij}} + {T^*}^i_{t_{ij}}]^{1/4} + \varepsilon^1_{ij} \quad ; \quad \log_{10} V_{ij} = V^i_{t_{ij}} + \varepsilon^2_{ij}, \tag{2.8}$$

where ε^1_{ij} and ε^1_{ij} are measurement errors, independently normally distributed.

In this Model 7, inference is much more complex and computationally demanding than in discrete-time models. Inference is based on a penalized maximum likelihood approach, and a

special Newton-like algorithm has been implemented in the NIMROD program (Prague *and others*, 2013). As in Models 5 and 6, assessing the long-term treatment effect in Model 7 is possible by analytically computing the equilibrium point. On the other hand, one year and subsequent years increase of CD4 after treatment initiation can be computed by solving the ODE system for given values of the random effects. The marginal effect can be computed as the mean of the individual effects in the population. The infectivity parameter gives an indicator of the effect of treatment, and a Wald test can be used to test the no-effect hypothesis "$\beta = 0$" .

3. Simulation study

3.1 *Design of the simulation*

We perform simulations to investigate the properties of the models and methods described above concerning the estimation of causal treatment effect and the long-term prediction of biomarkers. In order to simulate biomarkers trajectories in HIV cohort data, Adams *and others* (2005) extended the work of Callaway and Perelson (2002) and developed a tool to mimic HIV progression in an individual. They describe a "typical" mechanistic model to be employed for further discussion of mathematical and statistical methodology. This model is much more complex than the one described in Equation 2.6 for Model 7: it includes two co-circulating populations of target cells and immune effectors (such as cytotoxic T-lymphocytes) population, see Figure 1b and Table 1 for a description and the values of parameters chosen for the simulations. Biomarkers trajectories for given values of the parameters are numerically computed by an ODE solver; we use the R package deSolve (Soetaert *and others*, 2010). All parameters have inter-individual variability modeled by drawing them from a normal law (with mean values listed in Table 1 and variances chosen to obtain a variation coefficient of 50%). Thus, the simulated patients have various biomarkers trajectories because they have different parameters values.

Patients' parameters are related with steady state values for biomarkers without treatment;

as there is no analytic solution for this system, this steady state may be found by running the solver long enough. Random effects are simulated so that the steady state baseline distribution of CD4 counts and viral load is consistent with the baseline values distribution found in Aquitaine cohort and SHCS dataset (see Table 3). An observation model is also simulated with visits every 3 months (90 days). We generate observations according to the same model as (2.8); the standard deviations of the measurement errors are chosen approximately equal to those estimated from real data $\sigma_{VL} = 0.6$ and $\sigma_{CD4} = 0.1$ respectively. Viral load is artificially censored at the level of 50 copies/mL. Treatment assignment is done by simulating a CD4 count assessment at every visit (every 3 months) and by fixing a probability of treatment attribution depending on the observed CD4 count. We took empirical probabilities from the Aquitaine cohort and SHCS dataset: treatment is attributed in 2%, 28% or 47% of the cases if CD4 count is > 400, $[400, 200]$ or < 200. Neither confounder nor drop-out is considered. We simulate n=200 and n=1500 patients to look at small and large sample properties. Table 4 gives a general description of both simulated and real data sets: no major inconsistency in descriptive statistics values appears between simulated and real real cohort data sets.

3.2 *Evaluating the true treatment effect in the simulation*

It is possible to get an analytic value of the treatment effect in the simulated model. For computing it, we build two datasets without adding noise on the biomarkers. The first dataset (D1) consists in the true value (without noise) of the biomarkers trajectories according to the observed schedule of treatment for each simulated patient. The second dataset (D2) is a counterfactual dataset where we simulate patients' biomarkers trajectories without noise if no treatment is attributed. We define the "mean causal effect in treated patients", which is the real increase of CD4 in patients before 1 year, between years 1 and 2 and after an infinite time when treatment was initiated (from D1) compared to their counterfactual value without treatment (from D2). We expect a 350

cells increase of CD4 after 1 year, a 12 cells increase between 1 and 2 years and an overall increase of 370 CD4 cells after treatment initiation. Of note, this increase is probably in the upper range of clinical reality. However, as instance, Anthony *and others* (2003) found similar mean values (respectively an increase of 206 CD4 after one year and 326 CD4 2 years). We use those values as benchmark for comparison of methods.

3.3 *Practical concerns*

For Models 1-3 which are fitted by ordinary or weighted GEE, we used the R software and the function *geeglm* in the package *geepack* (Halekoh *and others*, 2006). Independence working correlation matrices are used. For models 2 and 3, weights are computed with the function *ipwtm* in the package *ipw* (van der Wal and Geskus, 2011) but could have been implemented from scratch. Models 4-6 are fitted using non linear mixed effect models package *lme4* (Bates *and others*, 2014), particularly the function *lmer*. In Model 7, in order to avoid identifiability issues, we use a priori knowledge on mechanistic parameters: the priors are set according to past estimates of the biological parameters in the literature (Prague *and others*, 2012) and are given in Table 2. Mechanistic models are fitted using NIMROD (Prague *and others*, 2013).

3.4 *Results of the simulation*

We run the analysis of Models 1-7 proposed in Section 2; results are presented in Table 5 for the simulated data set n=200 and n=1500. All results and conclusions are similar in small and large sample. First, we notice that the naive Model 1 largely underestimates the treatment effect. This is corrected by the MSM Models 2 and 3. Figure 3 shows the box-plot of the weights for Model 3 estimated by the treatment attribution model using a logistic regression; we found values close to the theoretical probabilities of treatment attribution. Models 2 and 3 give similar results because the true model of weight is only based on CD4 count. Thus in the simulation, the model for

treatment attribution is no better specified in Model 3 than in Model 2. Model 4 also gives good estimates of the mean causal effect in treated patients both for the first year and subsequent years. However, we can notice that long-term increase of CD4 is infinite in Models 1 to 4. Models 5-7 have an equilibrium point which makes it possible to consider the long-term causal effect of treatment. All dynamic models give a correct estimate of the long-term effect of the treatment. However, the misspecification of the models can induce biases. Actually, the initial increase in CD4 during the first year is not correctly caught by Model 5. Models 6 and 7 which both incorporate the dynamics of viral load give a correct estimation of the increase of CD4 count. Moreover, both models allow a qualitative interpretation of the pathway of action of treatment with a mediation through the decrease of the viral load. Model 7 allows a more detailed mechanistic interpretation of the infection and provides estimates of important biological parameters. Table 6 shows the *a posteriori* means and standard deviations for the log-transformed biological parameters. To some extent, these values can be compared to the generating values in Table 1. For instance, the viral clearance is 13 virions/day and is estimated at around 7 virions/day with our misspecified model. Even if all models found a significant effect of treatment on CD4 counts in the first year, the Z-statistics for the no-effect hypothesis are larger or much larger for dynamical models (LIM and ODE-NLME models) than for the GEE-based models, indicating more power to reject the null hypothesis.

4. Real data

4.1 *The Aquitaine cohort and the Swiss cohort*

Real data analysis is based on datasets from two large cohorts: the ANRS CO3 Aquitaine cohort (Thiébaut *and others*, 2000) and the Swiss HIV Cohort Study (SHCS) (Schoeni-Affolter *and others*, 2010). We compare the estimates of the effect of HAART on CD4 counts obtained by the naive, the MSM and the dynamic models.

For the Aquitaine cohort, the raw dataset includes 4541 patients and a total of 110,663 observations. For the SHCS, we use the same dataset as previous methodological work (Sterne *and others*, 2005) which includes 2161 patients and a total of 77,838 observations. Similarly to Cole *and others* (2005) we take a subsample of patients who are alive, HIV positive, yet untreated and under follow-up in April 1996 when HAART became available. All patients taking ARV in mono- or bi-therapy instead of HAART are excluded. Once a patient is on any therapy, we assume he or she remains on it. For each of them, the follow-up begins with the first visit after April 1996 and ends with 1) the last visit at which he or she was seen alive, 2) the last visit before patient discontinued the study, or 3) April 2005, whichever comes first. Patients with at least 2 observations are included. Table 7 show the patient selection flow chart. In Sterne *and others* (2005), last observation carried forward analysis (LOCF) was used to analyze SHCS to account for missing visits. However, Cook *and others* (2004) showed that this approach is not optimal and may lead to biases. Thus, we prefer to make the assumption that data are MCAR. Thus, we delete all observations where viral load or CD4 count were missing. Finally, we have two datasets of approximately the same size with a total of 1591 patients (19,597 observations) for the Aquitaine Cohort and 1726 patients (15,158 observations) for the SHCS. Figure 2 describes trajectories for the biomarkers in the two studies and Table 4 gives basic descriptive statistics.

4.2 *Results from the cohort data*

Table 8 displays the results we obtained for the treatment effect on CD4 counts. Computation times were less than 1 minute on a classical laptop for Models 1 to 6. Fitting Model 7 took about 10 hours of parallel computing over 100 cores. The naive Model 1, not correcting for treatment attribution, indicates a small and non-significant increase of CD4 for SHCS cohort, and a significant negative effect for the Aquitaine Cohort; this illustrates the need for correcting for treatment attribution. Models 2 and 3 have different weights but show rather similar results, probably because

treatment initiation was mainly driven by observed CD4 count.Both models yield a significant increase of CD4 counts for one year of treatment and after one year. The one-year increase however is much smaller for the Aquitaine cohort than for the SCHS. This is potentially due to the fact that the overall percentage of treatment is very different between the two cohorts (34% in SHCS vs. 64% in the Aquitaine cohort) and inverse probability weighting techniques are known to be highly sensitive to the models for treatment attribution. The dynamical models (LIM and ODE-NLME), especially Models 6 and 7, show a better consistency between the two cohorts.

Model 6 is interesting because it dissociates the effect of the treatment on CD4 and on viral load. We see that treatment has a non significant effect on CD4 change for the Aquitaine cohort (Z-stat=1.53); although this effect is significant for the SHCS (Z-stat=5.0) the absolute values of the Z-statistics are much higher for the effect on the viral load. This is consistent with the type of action of antiretroviral treatments: the increase of the CD4 count is essentially mediated by the decrease in viral load, which is the direct effect of antiretroviral treatments. Such biological knowledge is incorporated in Model 7 where the treatment acts on the infectivity parameter. In view of the Z-statistics obtained by a Wald test of the hypothesis $\beta = 0$ in 2.7, the power obtained in Model 7 is very high. Such high values are rarely encountered in statistical analysis. To confirm the very low p-values that can be found in this model, we performed a log likelihood ratio test. We ran the NIMROD analysis again fixing $\beta = 0$ and obtained log-likelihood under the null hypothesis. We found: $LL = -2 * -13836 + 2 * -12561 = 2550$ for the SHCS and LL=-2*-25211+2*-24579=1264 for the Aquitaine cohort, also yielding to a very low p-value. Finally, Model 7 gives an insight into the value of the biological birth and death rates of cells during the infection (Table 9). The estimates from the two datasets are rather consistent in the sense that they have the same order of magnitude, although a formal comparison would show that several parameters are different, potentially due to different characteristics of the patients in the two cohorts.

Finally, a simple way to look at these results and to compare them, is to consider the mean evolution of CD4 along time. Figure 4 represents the predicted CD4 count with Models 1, 3, 6 and 7 for treated patients starting at baseline with 365 CD4 cells/mm^3 and a viral load of 4.4 $\log_{10}$ cp/mL (which are approximately the mean values at treatment initiation in the cohorts). For Models 1 to 3, these curves are deterministic, which is not the case for Models 4 to 7 that have random effects. For these latter models, we computed the mean predicted curves depending on the value of the random effect, which have to be set to values compatible with the baseline condition for values of biomarkers. In order to set them, in both case, we computed the equilibrium point of the system without treatment and solved the system of equations. This was possible since we had two random effects and two conditions equations. In Figure 4, it is clear that the naive Model 1 badly under-estimates the effect. Both Models 1 and 3 are highly unstable, whereas Model 6 or the mechanistic Model 7 are more consistent between the two studies and have equilibrium points. Model 6 however is way faster and easier to run than Model 7.

5. Conclusion

We described and compared a naive model, two MSM and four dynamic models (2 LIM and 2 ODE-NLME) for estimating the treatment effect of HAART on CD4 in observational studies. We tried such models on simulated data and on the real data of the Swiss HIV Cohort Study and of the Aquitaine cohort study. The naive regression model (Model 1) strongly underestimated the effect of the treatment. The MSM models (Models 2 and 3) corrected this misleading result but failed to reach significance or were unstable among studies. In contrast, the discrete dynamic models LIM (Model 4, 5 and 6) gave rather good estimates and appear to have a higher power. Models 6 and 7, which jointly model CD4 and viral load gave the most consistent results, with a richer interpretation since they take into account that the effect of HAART on CD4 is mediated by viral load. However, these models may sometimes be too rigid, as was the case for Model

5 which wrongly caught the dynamics. The mechanistic model ODE-NLME (Model 7) directly incorporates biological knowledge. This leads to a more powerful test for the parameter of interest. Moreover, in the ODE-NLME approach, we distinguish the system living in continuous time and observations taken at discrete times. One of the advantages of this distinction is that we would be able to use the data as they are collected and thus a greater amount of information: with discrete-time models we must have approximately equally-spaced observations, which rarely occurs in real observational studies. Moreover, ODE-NLME may take into account more data since the likelihood may be computed if either the CD4 count or the viral load is observed. Thus, less data are omitted which leads to both increased power and less biases due to selection. However, results may be sensitive to misspecification even if the simulations in this paper suggest that misspecification (Target cell model versus Adams *and others* (2005) model) does not necessarily prevent to obtain a reasonable and significant estimate of treatment effect. All the discrete-time models can be easily fitted with classical softwares. The drawback of ODE-NLME is that it is numerically challenging and requires special software running on cluster computers.

More generally, MSM provide consistent estimates but lack precision. Adding more information, as we did with dynamic modeling, improves precision at the risk of increased bias; we encounter here the well known bias-variance trade off issue. This result is consistent with findings in Fitzmaurice *and others* (2008) showing that the most parametric a model is the most sensitive it is to bias. In other words, the MSM is fairly robust to bias from misspecification of a model, but may be an impractical method in many applications because of poor power that results in uninformative and wide confidence intervals (Hernán *and others*, 2006). On the other hand, dynamical models provide more efficient estimates, but require a better specification of the dynamical mechanism which may be challenging. Although mechanistic models are already used in HIV, infection diseases and oncology, another limitation is that we do not always have sufficient biological knowledge for constructing a mechanistic model for any application. Alto-

gether, this study showed that a dynamical approach for estimating the treatment effect on large observational cohorts is possible, with either discrete or continuous-time models, and may have better efficiency than marginal structural models.

6. Acknowledgement

This work of M.P. was supported by the National Institutes of Health grant R37 AI 51164. The authors would like to thank the investigators of the Aquitaine Cohort and the Swiss HIV Cohort Study. Parallel computing was used thanks to the computing facilities MCIA (Mésocentre de Calcul Intensif Aquitain) of the Université de Bordeaux and of the Université de Pau et des Pays de l'Adour.

References

Aalen, O.O. (1987). Dynamic modelling and causality. *Scandinavian Actuarial Journal* **1987**(3-4), 177–190.

Aalen, O.O., Røysland, K., Gran, J.M. and Ledergerber, B. (2012). Causality, mediation and time: a dynamic viewpoint. *JRSS-A* **175**(4), 831–861.

Adams, B.M., Banks, H.T., Davidian, M., Kwon, H.D., Tran, H.T., Wynne, S.N. and Rosenberg, E.S. (2005). HIV dynamics: modeling, data analysis, and optimal treatment protocols. *Journal of Computational and Applied Mathematics* **184**(1), 10–49.

Anthony, K.B., Yoder, C., Metcalf, J.A., DerSimonian, R., Orenstein, J.M., Stevens, R.A., Falloon, J., Polis, M.A., Lane, H.C. and Sereti, I. (2003). Incomplete CD4 T cell recovery in HIV-1 infection after 12 months of highly active antiretroviral therapy is associated with ongoing increased CD4 T cell activation and turnover. *Journal of acquired immune deficiency syndromes* **33**(2), 125–133.

Arjas, E. and Parner, J. (2004). Causal reasoning from longitudinal data*. *Scandinavian Journal of Statistics* **31**(2), 171–187.

Bates, D., Mächler, M., Bolker, B. and Walker, S. (2014). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **55**, 1–9.

Callaway, D.S. and Perelson, A.S. (2002). HIV-1 infection and low steady state viral loads. *Bulletin of Mathematical Biology* **64**(1), 29–64.

Cole, S.R. and Hernán, M.A. (2008). Constructing inverse probability weights for marginal structural models. *American journal of epidemiology* **168**(6), 656–664.

Cole, S.R., Hernán, M.A., Anastos, K., Jamieson, B.D. and Robins, J.M. (2007). Determining the effect of highly active antiretroviral therapy on changes in human immunodeficiency virus type 1 RNA viral load using a marginal structural left-censored mean model. *American journal of epidemiology* **166**(2), 219–227.

Cole, S.R., Hernán, M.A., Margolick, J.B., Cohen, M.H. and Robins, J.M. (2005). Marginal structural models for estimating the effect of highly active antiretroviral therapy initiation on CD4 cell count. *American journal of epidemiology* **162**(5), 471–478.

Cole, S.R., Jacobson, L.P., Tien, P.C., Kingsley, L., Chmiel, J.S. and Anastos, K. (2010). Using marginal structural measurement-error models to estimate the long-term effect of antiretroviral therapy on incident AIDS or death. *Am. j. epi.* **171**(1), 113–122.

Commenges, D. and Gégout-Petit, A. (2009). A general dynamical statistical model with causal interpretation. *JRSS-B* **71**(3), 719–736.

Commenges, D. and Gégout-Petit, A. (2015). The stochastic system approach for estimating dynamic treatments effect. *Lifetime Data Analysis* **submitted**, –.

Cook, R.J., Zeng, L. and Yi, G.Y. (2004). Marginal analysis of incomplete longitudinal binary data: a cautionary note on LOCF imputation. *Biometrics* **60**(3), 820–828.

Delattre, M, Genon-Catalot, V. and Samson, A. (2013). Maximum likelihood estimation for stochastic differential equations with random effects. *Scan J Stat* **40**(2), 322–343.

Didelez, V. (2008). Graphical models for marked point processes based on local independence. *JRSS-B* **70**(1), 245–264.

Diggle, P., Farewell, D. and Henderson, R. (2007). Analysis of longitudinal data with drop-out: objectives, assumptions and a proposal. *JRSS-C* **56**(5), 499–550.

Eichler, M. and Didelez, V. (2010). On granger causality and the effect of interventions in time series. *Lifetime data analysis* **16**(1), 3–32.

Fitzmaurice, G., Davidian, M., Verbeke, G. and Molenberghs, G. (2008). Longitudinal data analysis. *(Book)* **Chapter 23**, p 584.

Gégout-Petit, A. and Commenges, D. (2010). A general definition of influence between stochastic processes. *Lifetime data analysis* **16**(1), 33–44.

Granger, C.W.J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society* **37**(3), 424–438.

Guedj, J., Thiébaut, R. and Commenges, D. (2007). Maximum likelihood estimation in dynamical models of HIV. *Biometrics* **63**(4), 1198–1206.

Halekoh, U., Højsgaard, S. and Yan, J. (2006). The R package geepack for generalized estimating equations. *Journal of Statistical Software* **15**(2), 1–11.

Hernán, M.A., Lanoy, E., Costagliola, D. and Robins, J.M. (2006). Comparison of dynamic treatment regimes via inverse probability weighting. *Basic & clinical pharmacology & toxicology* **98**(3), 237–242.

Hernán, M. a, Brumback, B.A. and Robins, J.M. (2002). Estimating the causal effect of zidovudine on CD4 count with a marginal structural model for repeated measures. *Statistics in medicine* **21**(12), 1689–709.

Ho, D.D., Neumann, A.U., Perelson, A.S., Chen, W., Leonard, J.M., Markowitz, M. *and others*. (1995). Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* **373**(6510), 123–126.

Hoff, R., Gran, J.M. and Farewell, D. (2015). Farewell's linear increments model for missing data: The FLIM package. *The R Journal* **-**(-), To be published.

Huang, Y., Liu, D. and Wu, H. (2006). Hierarchical Bayesian methods for estimation of parameters in a longitudinal HIV dynamic system. *Biometrics* **62**(2), 413–423.

Hubbard, A.E., Ahern, J., Fleischer, N.L., Van der Laan, M., Lippman, S.A., Jewell, N., Bruckner, T. and Satariano, W.A. (2010). To GEE or not to GEE: comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. *Epidemiology* **21**(4), 467–474.

Lavielle, M., Samson, A., Karina Fermin, A. and Mentré, F. (2011). Maximum likelihood estimation of long-term HIV dynamic models and antiviral response. *Biometrics* **67**(1), 250–259.

Liang, K. and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**(1), 13–22.

Pepe, M.S. and Anderson, G.L. (1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics-Simulation and Computation* **23**(4), 939–951.

Perelson, A.S. (2002). Modelling viral and immune system dynamics. *Nature* **2**(1), 28–36.

Prague, M., Commenges, D., Drylewicz, J. and Thiébaut, R. (2012). Treatment monitoring of hiv-infected patients based on mechanistic models. *Biometrics* **68**, 902–911.

Prague, M., Commenges, D., Guedj, J., Drylewicz, J. and Thiébaut, R. (2013*a*). NIMROD: a program for inference via a normal approximation of the posterior in models with random effects based on ordinary differential equations. *Comp Meth Prog* **111**(2), 447–458.

Prague, M., Commenges, D. and Thiébaut, R. (2013*b*). Dynamical models of biomarkers and clinical progression for personalized medicine: The HIV context. *Advanced drug delivery reviews* **65**, 954–965.

Psaty, B.M., Koepsell, T.D., Lin, D., Weiss, N.S., Siscovick, D.S., Rosendaal, F.R., Pahor, M. and Furberg, C.D. (1999). Assessment and control for confounding by indication in observational studies. *American journal of geriatrics society* **47**, 749–754.

Robins, J.M. (1994). Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics-Theory and methods* **23**(8), 2379–2412.

Robins, J.M. (2000). Marginal structural models versus structural nested models as tools for causal inference. *Stat. models in epi.* **116**(1986), 95–133.

Robins, J.M., Hernán, M.A. and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11**(5), 550–60.

Schoeni-Affolter, F., Ledergerber, B., Rickenbach, M., Rudin, C., Günthard, H.F., Telenti, A., Furrer, H., Yerly, S. and Francioli, P. (2010). Cohort profile: the swiss HIV cohort study. *International journal of epidemiology* **39**(5), 1179.

Schweder, T. (1970). Composable markov processes. *J Appl. Probab.* **7**(2), 400–410.

Soetaert, K., Petzoldt, T. and Setzer, R.W. (2010). Solving differential equations in R: Package deSolve. *Journal of Statistical Software* **33**(9), 1–25.

Sterne, J.A.C., Hernán, M.A., Ledergerber, B., Tilling, K., Weber, R., Sendi, P., Rickenbach, M., Robins, J.M. and Egger, M. (2005). Long-term effectiveness of potent antiretroviral therapy in preventing AIDS and death: a prospective cohort study. *The Lancet* **366**(9483), 378–384.

Thiébaut, R., Morlat, P., Jacqmin-Gadda, H., Neau, D., Mercié, P., Dabis, F., Chêne, G. and the Groupe d'Epidmiologie du SIDA en Aquitaine (GECSA). (2000). Clinical progression of HIV-1 infection according to the viral response during the first year of antiretroviral treatment. *Aids* **14**(8), 971–978.

van der Wal, W.M. and Geskus, R.B. (2011). Ipw: an R package for inverse probability weighting. *Journal of Statistical Software* **43**(i13), 1–23.

Walker, A.M. (1996). Confounding by indication. *Epidemiology* **7**(4), 335–336.

Westreich, D., Cole, S.R., Young, J.G., Palella, F., Tien, P.C., Kingsley, L., Gange, S.J. and Hernán, M.A. (2012). The parametric g-formula to estimate the effect of highly active antiretroviral therapy on incident AIDS or death. *Stat. med.* **31**(18), 2000–2009.

Wu, H. (2005). Statistical methods for HIV dynamic studies in AIDS clinical trials. *Statistical Methods in Medical Research* **14**(2), 171–192.

Young, J.G., Cain, L.E., Robins, J.M., O'Reilly, E.J. and Hernán, M.A. (2011). Comparative effectiveness of dynamic treatment regimes: an application of the parametric g-formula. *Statistics in biosciences* **3**(1), 119–143.

7. Tables and Figures

Table 1: Parameter values used for data simulation from the Adams *and others* (2005) model presented in Figure 1b

			Parameter Value for each population (X)†			
Name	Description	Units	Type 1	Type 2	Effectors	Virus
λ_X	Natural production rate	$\frac{cells}{mL.day}$	5000	31.98	1.0	-
(1-ϵ_X)	Treatment efficacy	no unit	50%	83%	-	-
d_X	Natural death rate	$\frac{1}{day}$	0.01	0.01	0.25	-
δ_X	Infection-induced death rate	$\frac{1}{day}$	0.7	0.7	0.1	-
ρ_X	Number of virions infecting a cell	$\frac{virions}{cells}$	1	1	-	-
m_X	Immune-induced clearance rate	$\frac{mL}{cells.day}$	1×10^{-5}	1×10^{-5}	-	-
k_X	Infection rate	$\frac{mL}{virions.day}$	8×10^{-7}	1×10^{-4}	-	-
c	Virions natural death rate	$\frac{1}{day}$	-	-	-	13
N_T	Virions production per infected cells	$\frac{virions}{cells}$	-	-	-	100
K_b	Saturation constant cells birth	$\frac{cells}{mL}$	-	-	100	-
K_d	Saturation constant cells death	$\frac{cells}{mL}$	-	-	500	-
b_E	Infection-induced birth rate for E cells	$\frac{1}{day}$	-	-	0.3	-
†For each simulated patient, every parameter got a random effect leading to 50% coefficient of variation						

Table 2: Prior means and standard deviations for normal a priori distributions used for estimation of mechanistic parameters in Model 7. Reference and explanation for these choices can be found in Prague *and others* (2012).

Parameter	Mean	sd.
λ	2.55	1.90
μ_{T^*}	-0.05	0.68
μ_Q	-9.00	1.00
α	-4.00	2.00
ρ	-4.34	1.38
μ_T	-2.59	0.34
γ	-5.76	4.02
π	4.04	2.66
μ_V	2.83	0.68

Table 3: Baseline distribution of CD4 counts and viral load in Aquitaine cohort and SHCS dataset combined

	Viral load $copies/mL$			
CD4 count $cells/mm^3$	< 500	$[500 - 10000]$	> 10000	P(treatment attribution)
< 200	0%	1%	11%	0.46
$[200 - 400]$	1%	8%	18%	0.28
> 400	5%	28%	28%	0.13

Table 4: Data description for illustrations : Average viral load, CD4 counts and percentage of treatment attribution in the population are displayed for simulated data and real data from the Aquitaine cohort and the SHCS. Statistics displayed are mean [Q1;Q3].

	Simulated dataset		Aquitaine Cohort	SHCS
n	200	1500	1591	1726
CD4 count				
Baseline	428 [266 ; 545]	420 [253 ; 530]	471 [298 ; 612]	536 [357 ; 670]
Follow-up untreated	594 [485 ; 675]	588 [478 ; 656]	625 [440 ; 762]	543 [363 ; 675]
Follow-up treated	627 [417 ; 837]	606 [405 ; 801]	492 [315 ; 638]	507 [300 ; 660]
Viral Load				
Baseline	3.9 [3.3 ; 4.6]	4 [3.4 ; 4.7]	4.2 [3.6 ; 4.8]	4.0 [3.4 ; 4.6]
Follow-up untreated	3.5 [2.9 ; 4.2]	3.7 [3.1 ; 4.4]	3.3 [2.3 ; 4.2]	3.8 [3.1 ; 4.5]
Follow-up treated	2.6 [1.7 ; 3.2]	2.6 [1.7 ; 3.2]	2.7 [1.7 ; 3.6]	3.2 [2.4 ; 4.1]
% censored (baseline,untreated, treated)	(3%;4%;40%)	(2%;3%;38%)	(7%;22%;48%)	(10%;15%;57%)
Treatment attribution				
Time (day)	412 [1 ; 631]	377 [91 ; 451]	727 [1 ; 1281]	548 [183 ; 752]
% treated	69%	65%	64%	34%

Table 5: Estimated treatment effect on CD4 counts from simulated data: Model 1: Naive regression; Model 2: MSM with simple weights; Model 3: MSM with more complete weights; Model 4: simple dynamic model; 5: autoregressive model; Model 6: bivarariate dynamic model; Model 7: mechanistic model.

		Simulated Dataset with Adams *and others* (2005) model					
		n=200			n=1500		
Model	β treatment■	Effect	Sd.	Z-stat†	Effect	Sd.	Z-stat†
Model 1	<1 yr	136	29	4.68	172	11	16.34
	>1 yr	-11	43	-0.38	-12	16	-1.13
	∞	$-\infty$	-	-	$-\infty$	-	-
Model 2	<1 yr	320	31	10.44	322	11	28.71
	>1 yr	-15	46	-0.49	-8	17	-0.72
	∞	$-\infty$	-	-	$-\infty$	-	-
Model 3	<1 yr	327	31	10.64	325	11	28.81
	>1 yr	-14	46	-0.45	-7	17	-0.62
	∞	$-\infty$	-	-	$-\infty$	-	-
Model 4	<1 yr	362	17	21.60	378	6	61.35
	>1 yr	8	24	0.33	7	9	0.8
	∞	$+\infty$	-	-	$+\infty$	-	-
Model 5	<1 yr	133	-	-	136	-	-
	>1 yr	84	-	-	86	-	-
	∞	359	-	-	370	-	-
	Param.	149	5	31.24	154	2	89.36
Model 6	<1 yr	325	-	-	334	-	-
	>1 yr	31	-	-	34	-	-
	∞ CD4	360	-	-	371	-	-
	∞ VL	-1.9	-	-	-2	-	-
	Param. CD4	600	21	28.42	630	8	82.22
	Param. VL	-7	0	-40.86	-7	0	-120.51
Model 7	<1 yr	312	-	-	304	-	-
	>1 yr	2	-	-	4	-	-
	∞ CD4	308	-	-	306	-	-
	∞ VL	-5.6	-	-	-4.98	-	-
	Param. γ	-1.12	0.014	-79.3	-1.03	0.003	-295.6

†Estimates for treatment effect (β) are significant at level 10% if the absolute value of Z-stat is greater than 1.64 and significant at level 5% if the absolute value of Z-stat is greater than 1.96.
■ To be compared with mean treatment effect in treated for (<1 year; >1 year; ∞): benchmarks values are (350;12;370) for these simulations.

Table 6: Estimates for mechanistic parameters in log transform with model 7 on the simulated data.

	Simulated Dataset with Adams *and others* (2005) model			
	n=200		n=1500	
Parameter	Mean	sd.	Mean	sd.
λ	1.99	0.05	1.73	0.01
μ_{T^*}	-2.95	0.06	-3.08	0.01
μ_Q	-9.04	0.99	-9.10	0.99
α	1.06	1.08	1.19	0.51
ρ	2.17	1.06	2.61	0.67
μ_T	-3.02	0.33	-3.12	0.32
γ	-3.52	0.33	-3.45	0.32
π	-1.92	0.57	-1.98	0.59
μ_V	1.38	0.57	1.42	0.59
σ_λ	0.46	0.026	0.65	0.007
σ_{μ_T*}	0.61	0.031	0.55	0.006
σ_{CV}	0.82	0.007	0.88	0.003
σ_{CD4}	0.19	0.001	0.20	0.001

Table 7: Real data selection flowchart for Aquitaine cohort and SHCS.

	SHCS		Aquitaine Cohort	
	n pat.	n obs.	n pat.	n obs.
Raw dataset	2161	77838	4541	110663
No LOCF values	2161	17307	4541	110663
Date range Apr 97 - Apr 05	2124	17050	3727	59517
Only HAART (no ARV)	2124	17050	3567	49514
Only naive patients at baseline	2066	16237	1792	20288
Nb observation / patient $\geqslant 2$	1726	15897	1591	20087
MCAR assumption	1726	15158	1591	19597
Selection Ratio	80%	20%	35%	18%

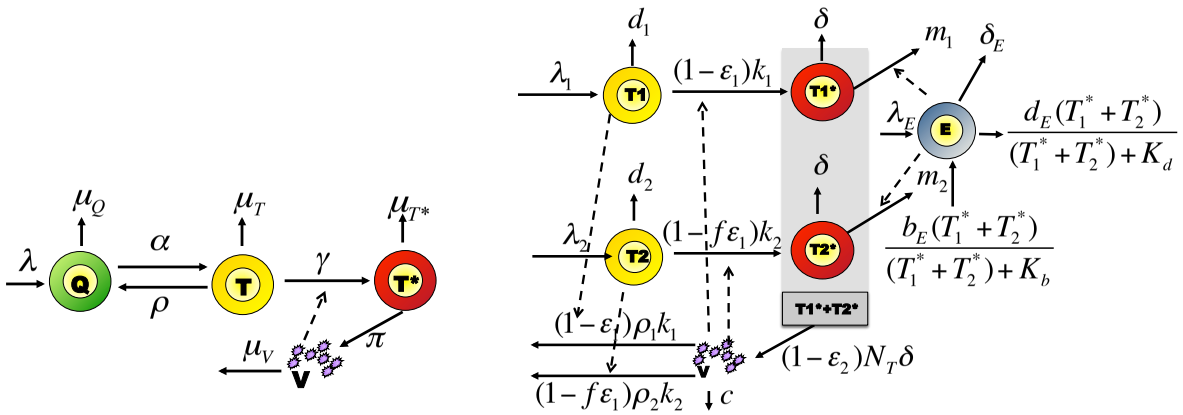
Table 8: Estimated treatment effect on CD4 counts from real data of the Aquitaine cohort and SHCS: Model 1: Naive regression; Model 2: MSM with simple weights; Model 3: MSM with more complete weights; Model 4: simple dynamic model; 5: autoregressive model; Model 6: bivarariate dynamic model; Model 7: mechanistic model.

		Real Dataset observational studies					
		SHCS			Aquitaine Cohort		
Model	β treatment	Effect	Sd.	Z-stat†	Effect	Sd.	Z-stat†
Model 1	< 1 yr	6	16	0.34	-94	12	-7.55
	> 1 yr	30	6	5.42	30	3	9.75
	∞	$+\infty$	-	-	$+\infty$	-	-
Model 2	< 1 yr	206	18	11.47	59	15	3.87
	> 1 yr	54	8	6.67	41	5	8.03
	∞	$+\infty$	-	-	$+\infty$	-	-
Model 3	< 1 yr	208	18	11.31	36	20	1.87
	> 1 yr	50	9	5.79	53	5	9.62
	∞	$+\infty$	-	-	$+\infty$	-	-
Model 4	< 1 yr	189	11	17.33	109	9	12.03
	> 1 yr	73	16	4.54	55	13	4.28
	∞	$+\infty$	-	-	$+\infty$	-	-
Model 5	< 1 yr	26	-	-	45	-	-
	> 1 yr	14	-	-	19	-	-
	∞	55	-	-	79	-	-
	Param.	60	4	16.04	14	3	4.12
Model 6	< 1 yr	73	-	-	92	-	-
	> 1 yr	26	-	-	16	-	-
	∞ CD4	104	-	-	111	-	-
	∞ VL	-2.0	-	-	-2.3	-	-
	Param. CD4	80	16	5	28	18	1.53
	Param. VL	-3.29	0.09	-38.4	-3.19	0.1	-30.55
Model 7	< 1 yr	104	-	-	71	-	-
	> 1 yr	18	-	-	9	-	-
	∞ CD4	127	-	-	86	-	-
	∞ VL	-4.09	-	-	-3.14	-	-
	Param. γ	-1.73	0.05	-34.79	-0.89	0.01	-85.77

†Estimates for treatment effect (β) are significant at level 10% if the Z-stat is greater than 1.64 and significant at level 5% if the Z-stat is greater than 1.96.

Table 9: Estimates for mechanistic parameters in log transformation with model 7 on the SHCS and the Aquitaine cohort datasets.

	Real Dataset observational studies			
	SHCS		Aquitaine Cohort	
Parameter	Mean	sd.	Mean	sd.
λ	0.61	0.04	0.94	0.05
μ_{T^*}	-2.74	0.04	-2.88	0.01
μ_Q	-5.42	0.06	-5.39	0.08
α	-5.63	0.09	-5.28	0.09
ρ	-1.35	0.60	-2.22	0.42
μ_T	-3.02	0.28	-2.91	0.29
γ	-2.39	0.54	-3.41	0.35
π	2.80	0.65	2.11	0.66
μ_V	2.82	0.66	3.07	0.65
σ_λ	0.63	0.011	0.67	0.006
σ_{μ_T*}	0.73	0.013	0.71	0.006
σ_{CV}	0.87	0.004	1.18	0.005
σ_{CD4}	0.43	0.001	0.50	0.001



(a) Target cell model

(b) Adams *and others* (2005) model.

Fig. 1: Mechanistic models for HIV dynamics

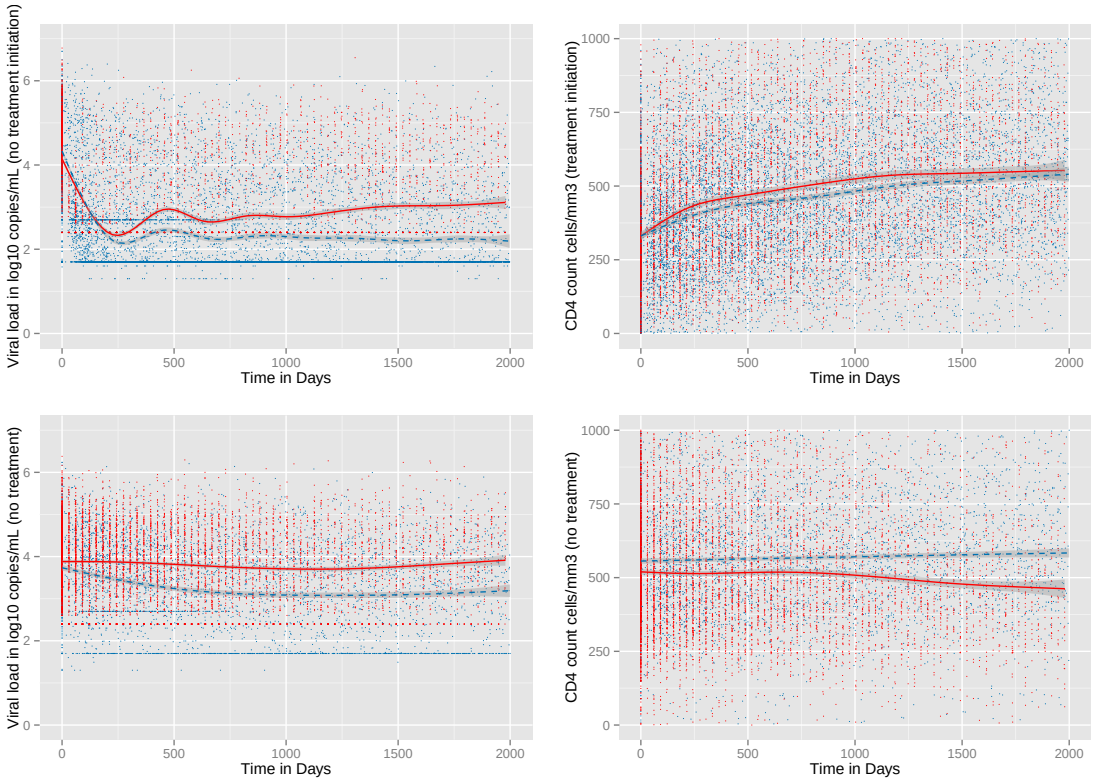


Fig. 2: Scatterplot and smoothed mean trajectories of viral load and CD4 counts for patients without treatment and after treatment initiation in the Aquitaine cohort (dashed blue) and the SHCS (plain red). Artifact horizontal lines for viral load result from points in the scatterplot due to detection threshold (mainly 250 copies/mL for SHCS and 50 or 500 copies/mL for the Aquitaine Cohort).

[*Received April, 2014; revised ???; accepted for publication ???*]

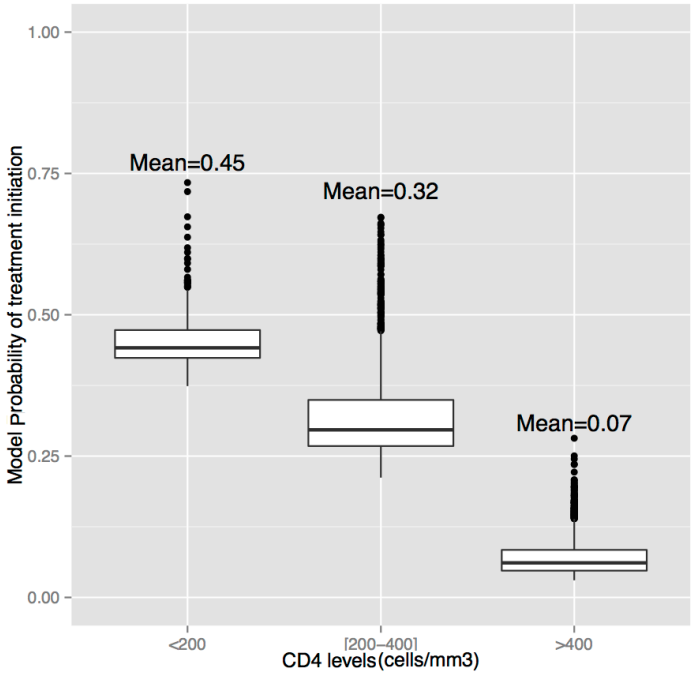


Fig. 3: Probability of treatment attribution predicted by the weighting model for MSM only depending on the CD4 counts. Theoretical values for these probabilities are 47% in the group < 200, 28% in the group $[200; 400]$ and 2% in the group > 400.

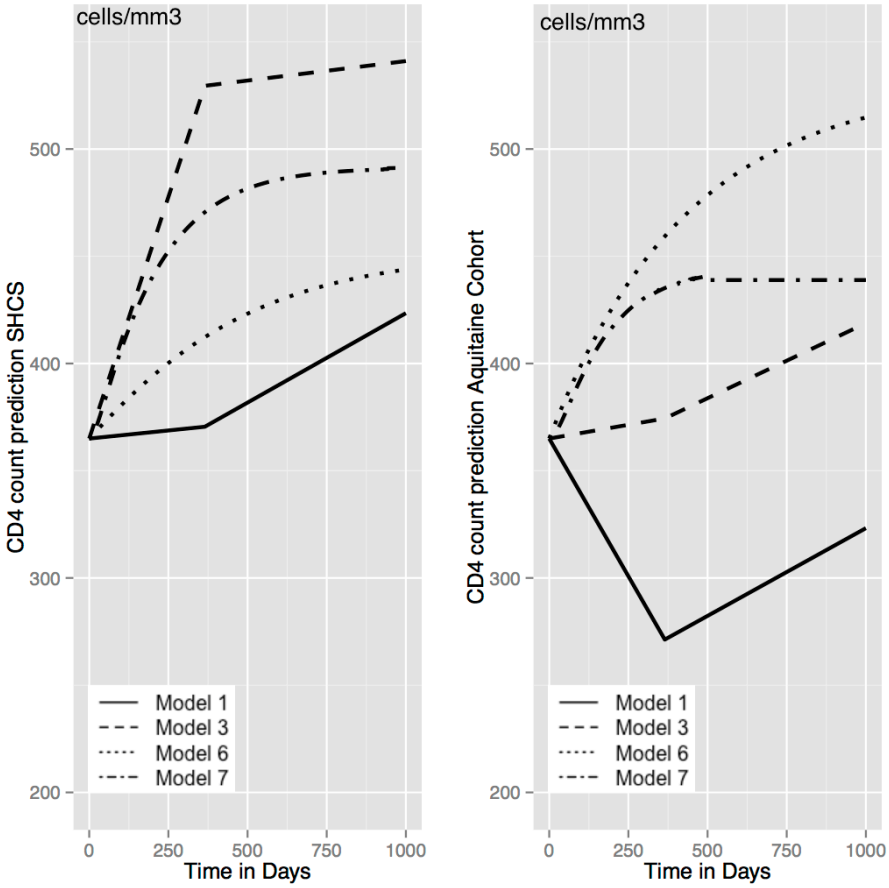


Fig. 4: Mean evolution of CD4 predicted by Model 1 (plain line, simple regression), Model 3 (dashed line, MSM), Model 6 (dotted line, linear incremental system) and Model 7 (dashed-dotted line, mechanistic model) for treated patients starting with 325 CD4 cells/mL and a viral load of 3.9 log10 copies/mm^3: (left) estimates from the SHCS data (right) estimates from the aquitaine cohort.