Web-based Supplementary Materials for "Dynamic models for estimating the effect of HAART on CD4 in observational studies: application to the Aquitaine Cohort study and the Swiss HIV Cohort Study" by M. Prague, D. Commenges, B. Ledergerber, J. Young, H. Furrer and R. Thiébaut.

A - Additional information on the simulated data analysis

1 - Generation of simulated data

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). 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 to reach stabilisation to an equilibrium defined by: 1) difference between two consecutive values of CD4 is lower than 10^{-3}; 2) two consecutive value of log10 viral load are lower than 10^{-5}). Random effects are simulated so that the steady state baseline distributions of CD4 counts and viral load are approximately consistent with the baseline values distributions found in Aquitaine cohort and SHCS dataset (see Table 1).

[Table 1 about here.]

2 - Estimation of weights for treatment attribution in the simulated data

Figure 1 shows the box-plot of the weights for Model 3 estimated by the treatment model using a logistic regression; we found values close to the theoretical probabilities of treatment attribution.

[Figure 1 about here.]

3 - Estimation of mechanistic parameters in the simulated data

Table 2 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 in the manuscript. For instance, the viral clearance is 13 virions/day and is estimated at around 7 virions/day with our misspecified model.

[Table 2 about here.]

4 - Are these simulations giving a competitive advantage to Model 7?

The first question is: 'Are the Adams et al. model and Model 7 different enough?'. The model 7 (Target cells model) and simulation model (the Adams et al. (2005)) are not the same and there is no mapping between the two. Even in the most dramatic case, where all the parameters in the Adams et al. model are equal to zero except $\lambda_1, d_1, (1-\epsilon_1)k_1, \delta, (1-\epsilon_2)N_T\delta, c$ (which is not the case in the simulations - See Table 1 in the main manuscript for the used values), the structure of the dynamical model is different. Actually, the simulation model will have three-compartments compared to four-compartments for the model 7. Drylewicz *and others* (2010) showed that the dynamics of these two mechanistic models is very different because the three-compartment model allows a delay between the creation of the CD4 and its infection/production of virus. We acknowledge that if, in this same extreme case, α and ρ were put to 0 in the Model 7, this would make the two dynamics models very close. However, Table 2 in this Web-Supplementary Material shows that in our estimations $\alpha > 2.88$ day^{-1} and $\rho > 8.76$ day^{-1}. Thus we believe that these models are different enough not to give a major advantage to dynamical approaches compared to MSM.

The second question is: 'Are the marginal models appropriate for catching the dynamics produced by the simulation model?'. In this simulation, we know that the marginal model is not correctly specified. However, in reality we have the same assumption and the model for data generation is probably way more complicated that the Adams et al. model. Thus

we do not think that the specification of the marginal model is worse in the simulation than in the real dataset.

B - Additional information on real data analysis

1 - Sample selection for real data

For the Aquitaine cohort, the raw data set included 4541 patients and a total of 110,663 observations. For the SHCS, we used the same data set as previous methodological work (Sterne *and others*, 2005; Gran *and others*, 2013) which included 2161 patients and a total of 77,838 observations. Similarly to Cole *and others* (2005) we took a sub-sample of patients who were 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 were excluded. Once a patient was on any therapy, we assumed he or she remained 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 2004, whichever comes first. Patients with at least 2 observations were included. Table 3 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 preferred to make the assumption that data are MCAR, and we deleted all observations where viral load or CD4 count were missing. Finally, we had two data sets 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.

[Table 3 about here.]

[Figure 2 about here.]

2 - Model 2-3 IPW analysis

The one-year increase of CD4 is much smaller for the Aquitaine cohort than for the SCHS (for Model 3, 36 CD4 versus 208 CD4). Even if Figure 2 shows slightly different trajectories, the difference seems to be less than 75 CD4. This result of Model 2-3 is potentially due to different weights distributions between these two cohorts (see Figure 3 for Model 3). In particular, the large weights are only driven by patients with high viral load (> 10000) and low CD4 count (< 200) who remain untreated. It includes only 5 observations in SHCS dataset and 48 observations in the Aquitaine Cohort. In the Aquitaine cohort the probability to be treated is overall very high (64%); thus, non-treated patients with high probability of being treated have even higher weights. In the Aquitaine cohort there is a large number of large weights denoting a possible practical violations of the experimental treatment assumption Cole and Hernán (2008).

[Figure 3 about here.]

3 - Model 7 mechanistic modeling

The apparent power of the Wald test in Model 7 seems striking; we looked at a confirmation with likelihood ratio test. We ran the NIMROD analysis again fixing $\beta = 0$ and obtained log-likelihood under the null hypothesis. The log-Likelihood test ratio gives the following statistics: $LL - ratio = -2 * -13836 + 2 * -12561 = 2550$ for the SHCS and LL-ratio=-2*-25211+2*-24579=1264 for the Aquitaine cohort, also yielding to a very low p-value.

The estimates of the parameters in the mechanistic Model 7 are given Table 4.

[Table 4 about here.]

4 - Impact of MCAR assumption for missing data

The correction for treatment is obviously important since treatment was started mainly in view of the CD4 counts. Censoring may also depend on the outcome (CD4 counts) but this

dependance is probably weak because the subjects are regularly followed-up in these cohorts: in the Aquitaine cohort lost of follow-up represents only 5.6% and 9.2% in the SHCS. To convince ourselves that censoring is not a major issue in this dataset, we performed a IPTW-IPCW analysis of the real dataset for Model 3. With the same notations as in the article and C the censoring indicator for lost of follow up or death, we have :

$$SW_T(t) = \prod_{k=1}^{t} \frac{Pr(A_k = 1|\bar{A}_{k-1}, L_0)}{Pr(A_k = 1|\bar{A}_{k-1}, \bar{L}_k)},$$
$$SW_C(t) = \prod_{k=1}^{t} \frac{Pr(C_k = 1|\bar{C}_{k-1}, L_0)}{Pr(C_k = 1|\bar{C}_{k-1}, \bar{L}_{k-1})},$$
$$SW(t) = SW_T(t) \times SW_C(t).$$

The probabilities at time k of being missing are predicted for every subject from logistic regressions depending on $\bar{L}_k$ ($L_0 \in \bar{L}_k$). We defined the subsets L for treatment model and censoring model as baseline and time-varying CD4 count in class (< 200, $[200; 400]$, > 400), viral load in categories (< 401, $401 - 10000$ and > 10000), and an indicator of undetectable viral load. This is the same model as in Cole *and others* (2005). Table 4 shows an extremely good concordance of the results with and without IPCW. This is partly explained by the fact that loss of follow-up is almost non-informative which gives IPC weights close to 1 (see Figure 4 for their distribution). Moreover regarding dynamical model, the MAR assumption does not request extra modeling because the estimation is likelihood-based.

[Table 5 about here.]

[Figure 4 about here.]

C - Technical details for fitting the models

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 were used. For models 2 and 3, weights were computed with the

function *ipwtm* in the package *ipw* (van der Wal and Geskus, 2011). Models 4-6 were fitted using the non linear mixed-effect models package *lme4* (Bates *and others*, 2014), particularly the function *lmer*. Standard errors (SE) are estimated with a sandwich approach. However, adjusting for uncertainty in the estimation of weights would lead to greater SE and thus does not impact the conclusion in this article. The simulated data analyzed in Section 3 and a R program implementing models 1-6 are available with this paper at the Biometrics website on Wiley Online Library. Programs to estimate the parameters with model 7 are available on a dedicated website: http://www.isped.u-bordeaux.fr/NIMROD.

References

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.

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

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., 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.

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.

Drylewicz, Julia, Guedj, Jérémie, Commenges, Daniel and Thiébaut, R. (2010). Modeling the dynamics of biomarkers during primary hiv infection taking into account the uncertainty of infection date. *The Annals of Applied Statistics* **4**(4), 1847–1870.

Gran, J.M., Roysland, K., Ledergerber, B., Young, J. and Aalen, O. (2013).

Analysing time-dependent confounding by imputation of counterfactual covariate trajectories: Application to data from the swiss HIV cohort study. *Statistics in medecine* **-**(-), submitted.

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

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.

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.

Received February 2015. *Revised ??? ???. Accepted ??? ???.*

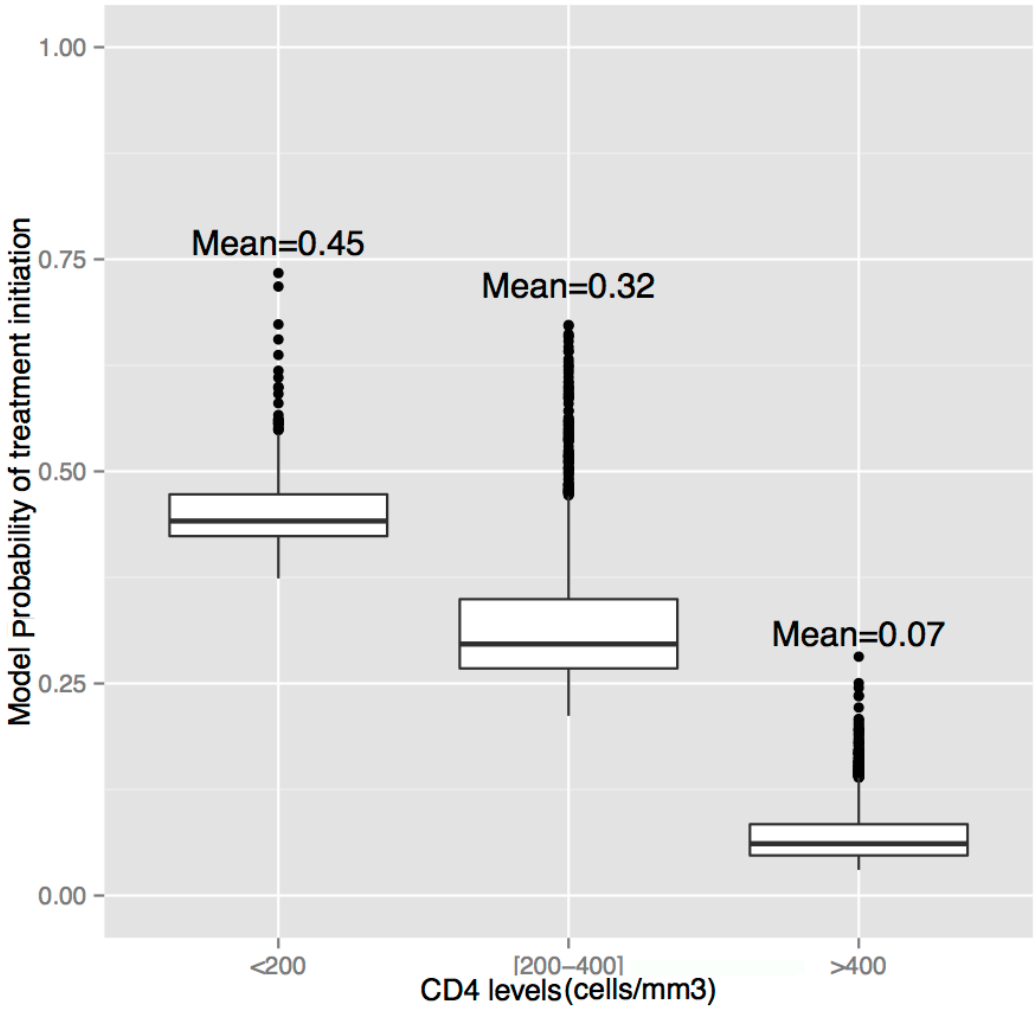

Figure 1: Probability of treatment attribution predicted by the treatment model only depending on the CD4 counts for the MSM (Model 2). Theoretical values for these probabilities are 47% in the group < 200, 28% in the group $[200; 400]$ and 2% in the group > 400.

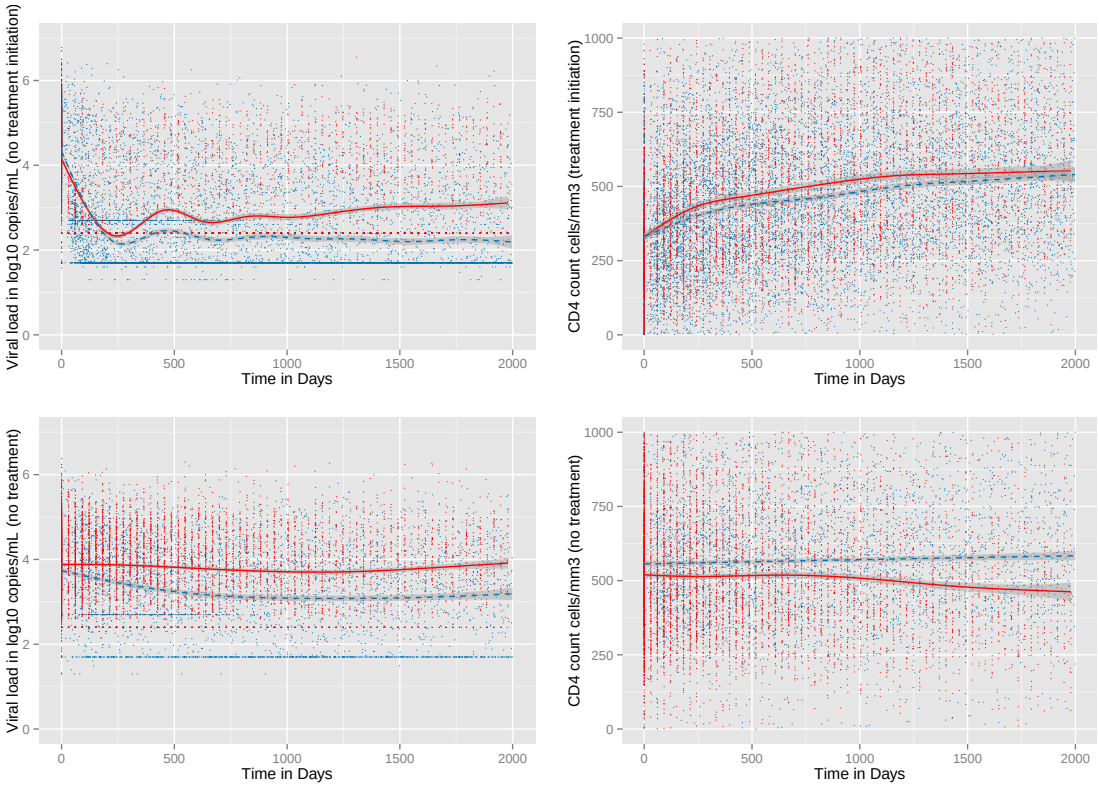

Figure 2: Scatterplot and smoothed mean trajectories of viral load and CD4 counts for patients without treatment and after treatment initiation in the Aquitaine cohort (dashed blue) and the SHCS (plain red). Artifact horizontal lines for viral load result from points in the scatterplot due to detection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

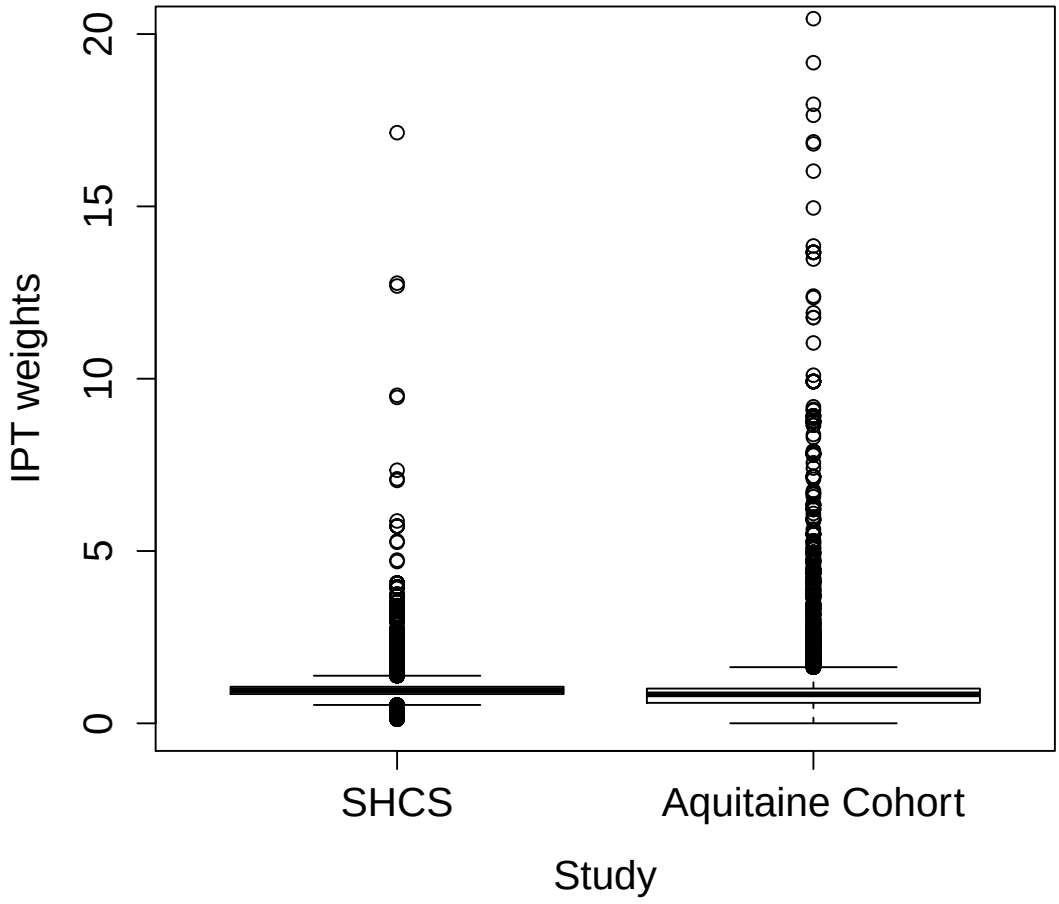

Figure 3: Distributions of weights for inverse probability of treatment in the Aquitaine cohort and the SHCS cohort. The plot had been truncated for weights > 20. Weights > 20 are only found in the Aquitaine cohort.

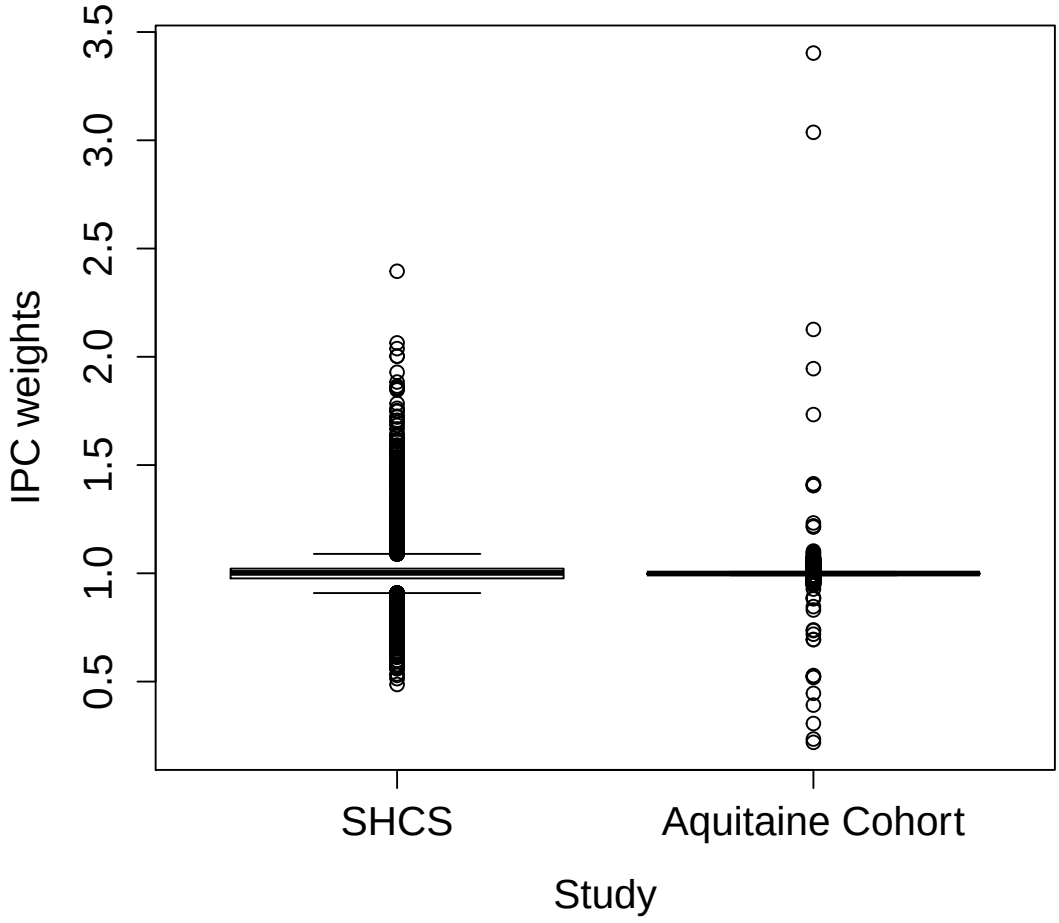

Figure 4: Values of weights for Inverse Probability of Censoring.

Table 1: Baseline distributions of CD4 counts and viral load in Aquitaine cohort and SHCS data set 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 2: 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
$\sigma_{\mu_{T^*}}$	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 3: 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 observations / patient $\geqslant 2$	1726	15897	1591	20087
MCAR assumption	1726	15158	1591	19597
Selection Ratio	80%	20%	35%	18%

Table 4: Estimates for mechanistic parameters in log-transform with Model 7 on the SHCS and the Aquitaine cohort data sets.

	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
$\sigma_{\mu_{T^*}}$	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

		Real Dataset observational studies					
		SHCS			Aquitaine Cohort		
Model	β treatment	Effect	Sd.	Z-stat†	Effect	Sd.	Z-stat†
Model 3	< 1 yr	208	18	11.31	36	20	1.87
IPTW	> 1 yr	50	9	5.79	53	5	9.62
	∞	$+\infty$	-	-	$+\infty$	-	-
Model 3	< 1 yr	207	18	11.26	36	19	1.86
IPTW + IPCW	> 1 yr	51	8	6.15	53	5	9.78
	∞	$+\infty$	-	-	$+\infty$	-	-

†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.