Web-based Supplementary Material for "Dynamic Predictions and Prospective Accuracy in Joint Models for Longitudinal and Time-to-Event Data"

Dimitris Rizopoulos

Department of Biostatistics, Erasmus Medical Center, PO Box 2040, 3000 CA Rotterdam, the Netherlands

1. Web AIDS Data Set

1.1 *Descriptives*

We present three illustrative descriptive plots for the AIDS data set. In particular, Web Figures 1 and 2 depict the $\text{CD4}^{1/4}$ cell count subject-specific longitudinal trajectories and smooth average longitudinal evolutions, and Web Figure 3 the Kaplan-Meier estimate of the survival functions of the two treatment arms. We observe that both groups of patients show similar variability in their longitudinal profiles. From the Kaplan-Meier estimate in Web Figure 3 it seems that the ddC group has slightly higher survival than the ddI group after the six month of follow-up. From the 467 patients, 188 have died by the end of follow-up corresponding to about 60% censoring. In total we have 1405 recorded CD4 cell count measurements, with an average of three measurements per patient (standard deviation 1.1 measurements).

1.2 *Joint Model Specification & Results*

The joint model fitted to the AIDS data set has the following form:

- Event Process

$$\begin{aligned} h_i(t \mid \mathcal{M}_i(t), w_i) &= h_0(t) \exp\{\gamma^\top w_i + \alpha m_i(t)\} \\ &= h_0(t) \exp\{\gamma_1 \texttt{ddI}_i + \gamma_2 \texttt{failure}_i + \gamma_3 \texttt{prevOI}_i + \\ &\quad \gamma_4 \texttt{male}_i + \alpha m_i(t)\}, \end{aligned}$$

email: d.rizopoulos@erasmusmc.nl

where the baseline risk function is assumed piecewise constant:

$$h_0(t) = \sum_{q=1}^{7} \xi_q I(v_{q-1} < t \leq v_q),$$

with the internal knots $v_1 < \cdots < v_6$ placed at equally spaced percentiles of the observed event times, $v_0 = 0$, v_7 is taken larger than the largest observed time, and ξ_q denotes the value of the hazard in the interval $(v_{q-1}, v_q]$, `ddI` is the dummy variable for treatment group ddI, `failure` is the dummy variable for AZT failure, `prevOI` is the dummy variable for previous opportunistic infections, and `male` is the dummy variable for males.

- Longitudinal Process

$$\begin{aligned} y_i(t) &= m_i(t) + \varepsilon_i(t) \\ &= (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})t + \beta_2 \texttt{ddI}_i + \beta_3 \texttt{failure}_i + \beta_4 \texttt{prevOI}_i + \\ &\quad \beta_5 \texttt{male}_i + \beta_6 (t \times \texttt{failure}_i) + \beta_7 (t \times \texttt{prevOI}_i) + \\ &\quad \beta_8 (t \times \texttt{male}_i) + \beta_9 (t \times \texttt{ddI}_i) + \varepsilon_i(t), \;\; \varepsilon_i(t) \sim \mathcal{N}(0, \sigma^2), \end{aligned}$$

where $y_i(t)$ denotes the $\text{CD4}^{1/4}$ cell count for subject i at time t, $A \times B$ denotes the interaction between A and B, and the random effects $\{b_{i0}, b_{i1}\}$ are assumed bivariate normal with mean 0 and covariance matrix D.

The estimated parameters and standard errors for the joint model fitted to the AIDS data set are presented in Web Table 1.

1.3 *Comparison with Naive Cox Model*

To illustrate the virtues of the joint modelling approach versus a simplistic analysis that uses the baseline $\text{CD4}^{1/4}$ cell count measurement in a Cox model (and the same baseline covariates as in the relative risk submodel presented above), we compare the conditional survival probabilities of Patient 68 who is a male belonging to the ddI group, who had showed AZT failure, had a previous opportunistic infection, he provided four CD4 cell count measurements (192, 78, 33, and 11 cells/mm^3) and he was lost to follow-up at 14.17 months, and Patient 185 who is a male belonging to the ddI group, who

had showed AZT failure, had a previous opportunistic infection, he provided three CD4 cell count measurements (0, 60, and 70 cells/mm^3) and he was lost to follow-up at 13.37 months. We note that these two patients exhibit sharply decreasing and sharply increasing CD4 cell count profiles, with considerable difference between the baseline and last CD4 cell counts. For the joint model $\pi(u \mid t)$ was calculated using the procedure described in Section 3, whereas for the Cox model using ratios of the Breslow estimates of the survival function. The results are depicted in Web Figures 4 and 5 for Patients 68 and 185, respectively. We clearly observe the under-performance of the simplistic Cox analysis that does not acknowledge for the changes in the CD4 cell count of the two patients.

2. Web Simulation Study

2.1 *Design*

We have performed a simulation study to investigate the finite sample performance of the two estimators for $\pi_i(u \mid t)$, given by Equations (9) and (11) in the main text. In particular, we compare the two estimators with the 'gold standard' estimate of $\pi_i(u \mid t)$,

$$S_i\big\{u \mid \mathcal{M}_i(u, b_i)\big\} \Big/ S_i\big\{t \mid \mathcal{M}_i(t, b_i)\big\},$$

that is based on the true (i.e., simulated) values for the random effects, and the true value for the parameters. The design of the simulation study is almost entirely motivated by the joint model fitted in the AIDS data set, presented in Web Section 1.2. The only difference is that instead of a piecewise constant baseline risk function we opted for a Weibull baseline risk $h_0(t) = \phi\xi t^{\xi-1}$, with $\phi = 0.067$ and $\xi = 0.762$. The censoring distribution was taken exponential with mean 50, such that we achieve the same percentage of censoring (i.e., about 60%) as for the AIDS data set. Under this setting we simulated 200 data sets.

For each simulated data set we fitted the joint model with exactly the same design as the one we used to simulate from, but assuming a piecewise constant baseline risk function with internal knots placed at equally spaced percentiles of the observed event times. Following for each data set we picked randomly 20 subjects, and for each of them we esti-

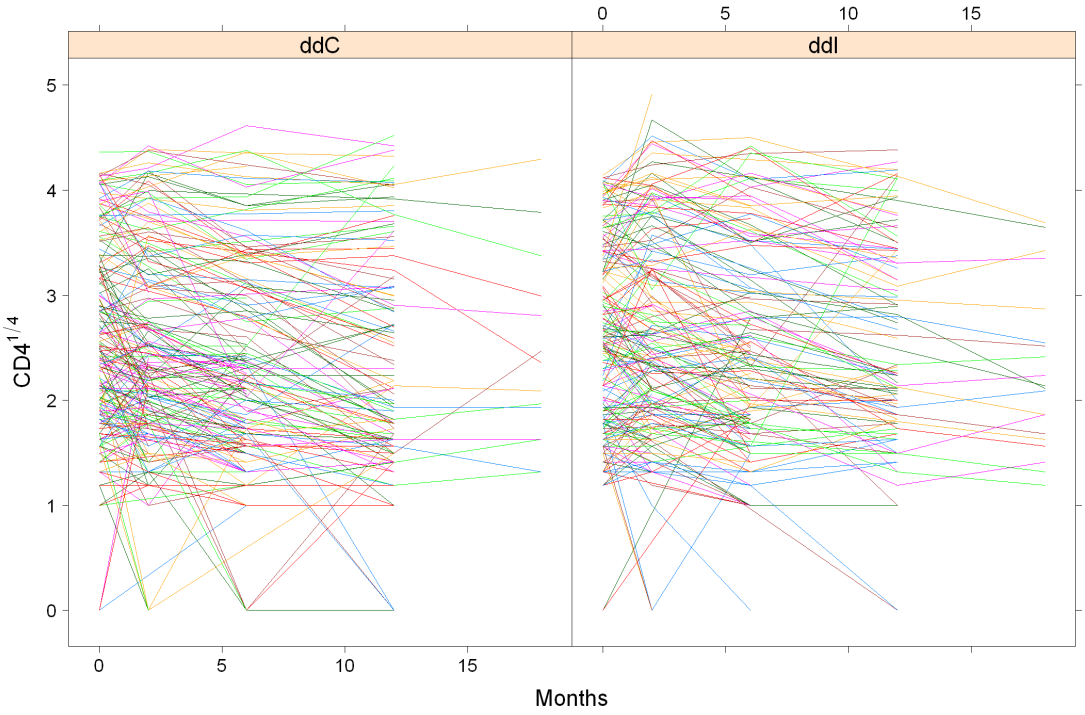
mated $\pi_i(t+\Delta t \mid t)$ using both the Empirical Bayes estimator (9), the Monte Carlo estimator (11), and the gold standard estimator $S_i\{t+\Delta t \mid \mathcal{M}_i(t+\Delta t, b_i)\}/S_i\{t \mid \mathcal{M}_i(t, b_i)\}$, for $t = 2$, 6, 12, and 18, and various values for Δt. The different follow-up times t have been chosen to reflect an increasing number of repeated measurements per subject n_i, which is expected to result in a decreasing degree of shrinkage in the estimation of the random effects. This procedure yielded about 45000 estimates of $\pi_i(u \mid t)$ for each estimator. For each simulated data set and for estimator (11) we used the median $\hat{\pi}_i(u \mid t) = \text{median}\{\pi_i^{(l)}(u \mid t), l = 1, \ldots, L\}$ over $L = 200$ Monte Carlo samples.

2.2 *Results*

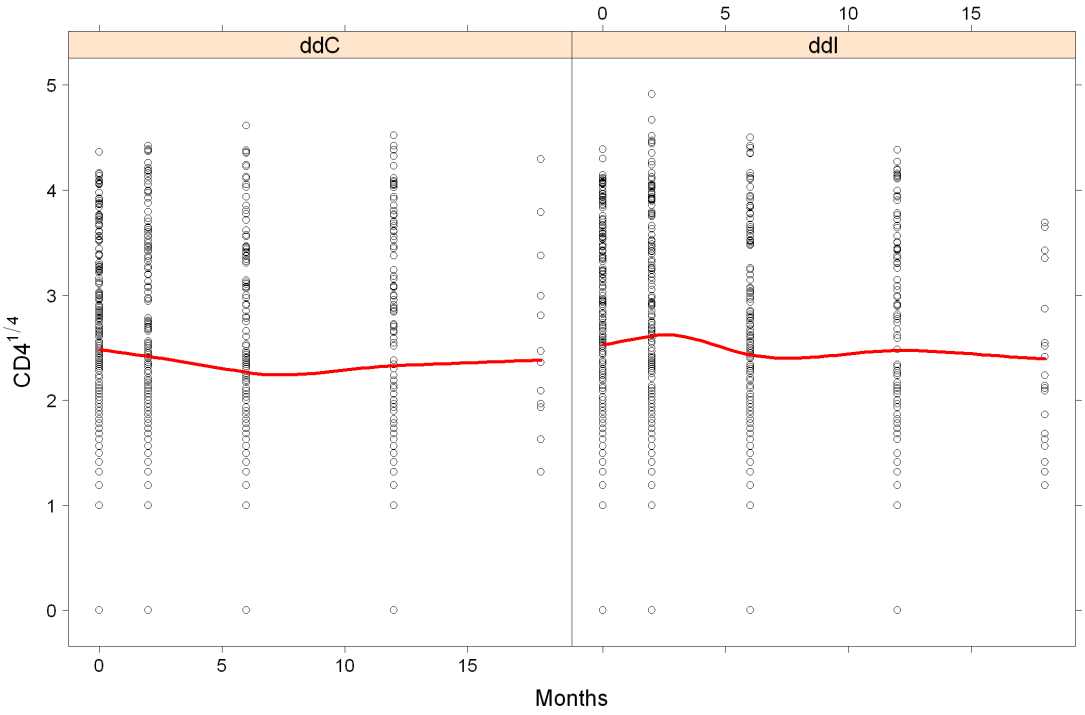
Since we are primarily interested in agreement between the gold standard estimator and the two estimators (9) and (11), we compared the 45000 estimates of $\pi_i(u \mid t)$ using Bland & Altman plots. Moreover, since the survival probabilities are constrained in the $[0, 1]$ interval, we expect more variability around 0.5 while less near the boundaries. Therefore, in order to stabilize the variance we present all results in the $\log\{x/(1-x)\}$ scale. Web Figures 6 and 7 depict the comparison of the two estimators over the whole follow-up period. For each plot the limits of agreement were constructed as plus/minus two times the sample standard deviation of the differences between the two estimators and the gold standard. For $t = 2$, 6, 12, and 18 months, and $\Delta t = 2$, 4, and 8 months, the estimated bias and limits of agreement are presented in Web Table 3.

Before discussing the results we should note that we should always expect some degree of shrinkage in both estimators, since this is an inherent characteristic of random-effects models. Focusing on the whole follow-up period in Web Figures 6 and 7, we observe that for both estimators the average bias is negligible, and that the expected number of comparisons (i.e., about 95%) is located well within the limits of agreement. The performance of the two estimators is almost indistinguishable, with the Monte Carlo estimator performing slightly better (i.e., less estimates outside the limits of agreement). Following, we concentrate on different values for Δt for fixed t. From Web Table 3 it is evident that both the bias and the variability in the differences between both estimators and the gold standard increase analogously with Δt – again, no considerable differences

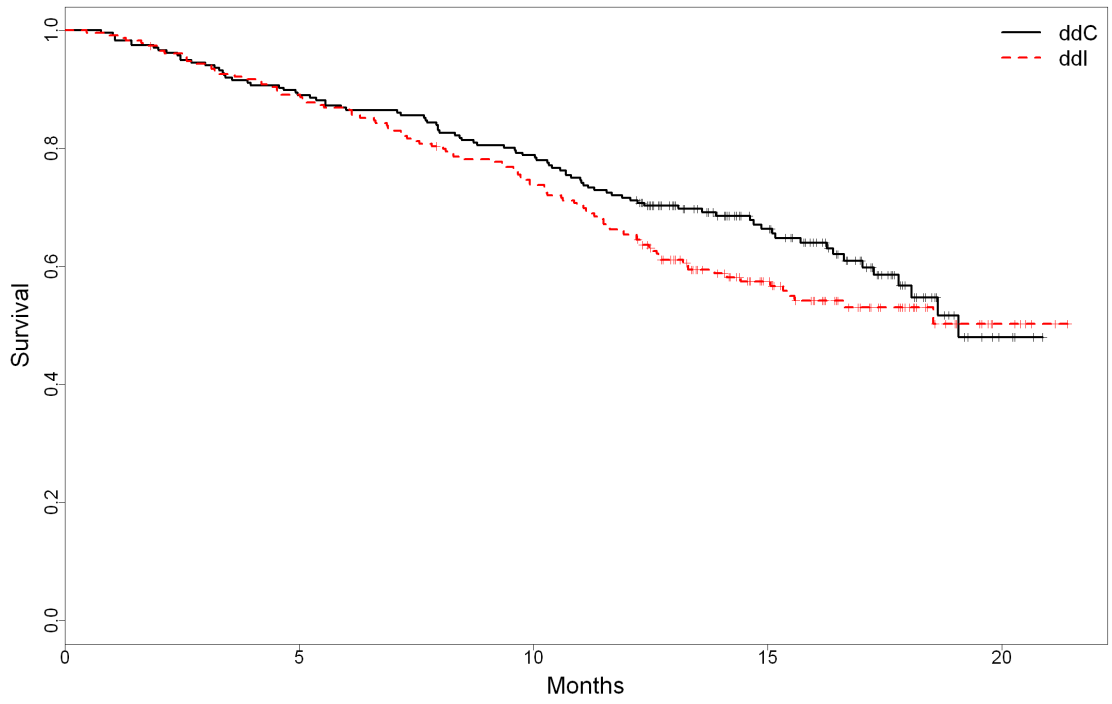
are observed between the behaviour of the two estimators. Taking into account the shape of the survival function, this behaviour is explained by the fact that, small differences between the true random effect values and their estimates are expected to translate to greater differences between the corresponding survival curves as time progresses (i.e., for greater Δt). Finally, we focus on increasing follow-up time t for the same time window Δt. Observing the results from Web Table 3, a counterintuitive relation emerges, namely, the bias decreases from $t = 2$ months to $t = 6$, but then increase again at $t = 12$ to the same levels of $t = 2$, and even more at $t = 18$. To understand why is this happening, we need to note that there is a 'competition' for information between the longitudinal and event time processes. More specifically, as time progresses, we collect more measurements for each subject and therefore we decrease the degree of shrinkage. However, for later stages of the follow-up, and especially for settings in which many events occur early (as in our simulation set-up), there are fewer patients left in the study, which means that the estimated hazard at these time points will have greater variability. The performance of both estimators (9) and (11) versus the gold standard (which uses the true values for both parameters and random effects) will depend on the combined performance of the MLEs and the random-effects estimates, and therefore will be affected when either of the two is less accurately estimated.



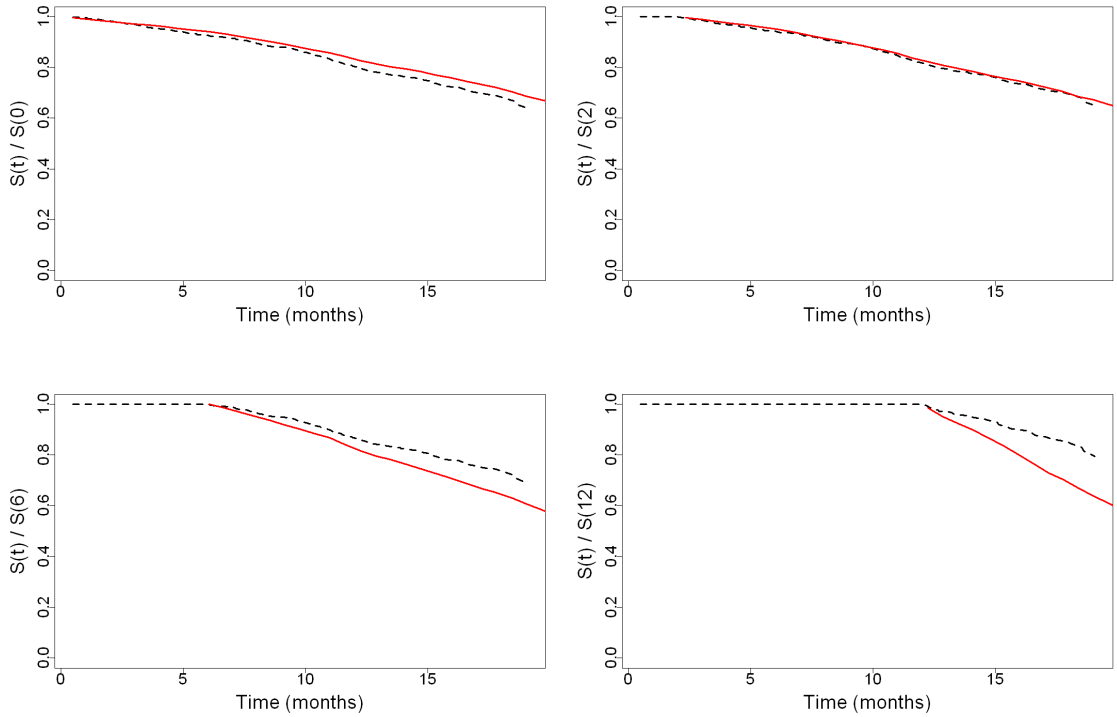
Web Figure 1. Subject-specific longitudinal trajectories for the $\sqrt{\text{CD4}}$ cell count for the ddI and ddC groups, separately.



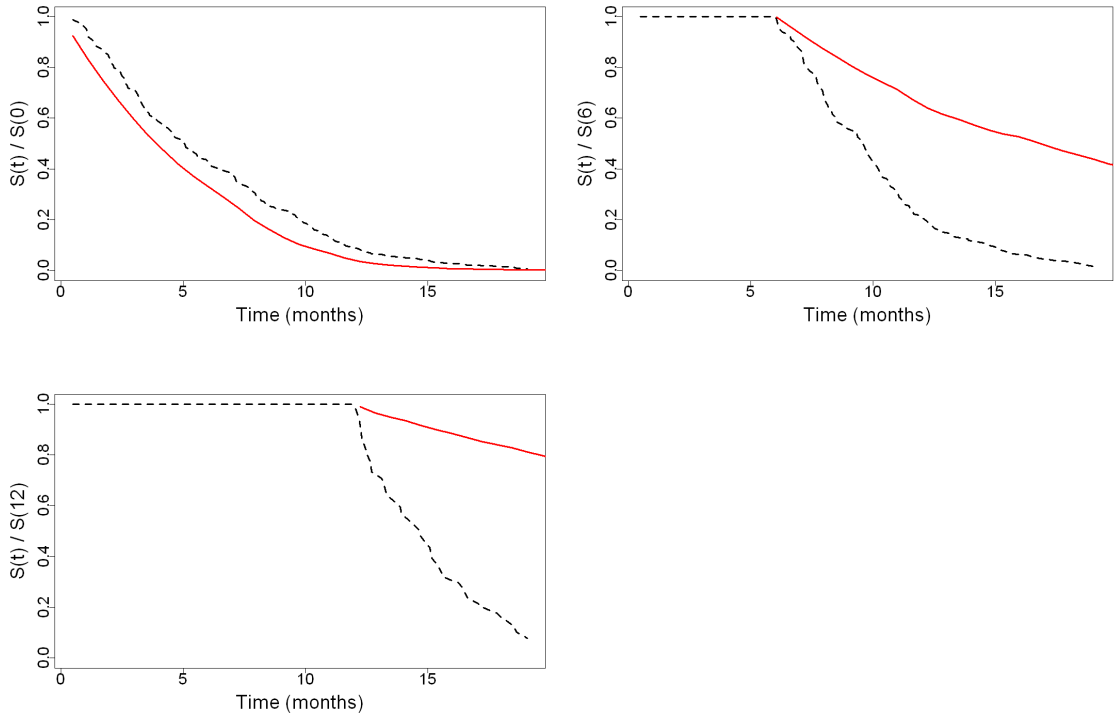
Web Figure 2. Smooth average longitudinal evolutions for the $\sqrt{\text{CD4}}$ cell count for the ddI and ddC groups, separately.



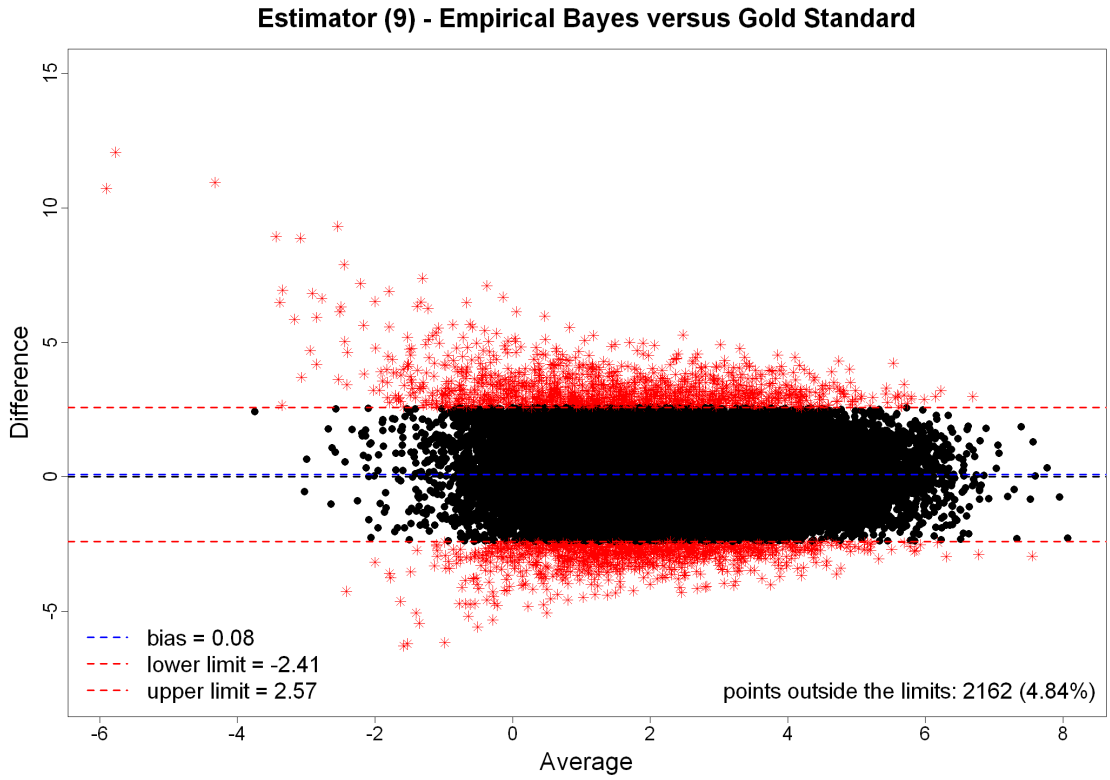
Web Figure 3. Kaplan-Meier estimates of the survival functions of the ddI and ddC treatment groups.



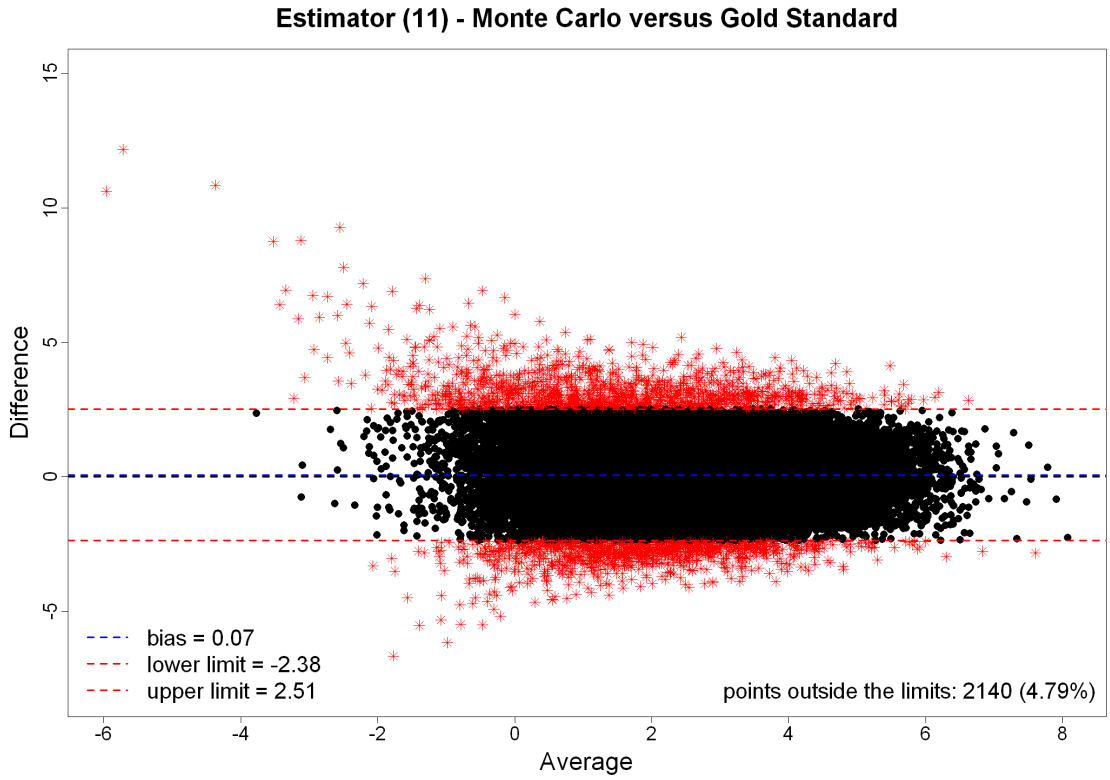
Web Figure 4. Conditional survival probabilities from the Cox (dashed line) and joint model (solid line) for Patient 68.



Web Figure 5. Conditional survival probabilities from the Cox (dashed line) and joint model (solid line) for Patient 185.



Web Figure 6. Bland-Altman plot for the comparison of the first order estimator of $\pi_i(u \mid t)$ with the gold standard estimator over the whole follow-up period. Simulation results on the logit scale based on 200 data sets. The blue dashed line denotes the average bias. The red dashed lines denote the limits of agreement.



Web Figure 7. Bland-Altman plot for the comparison of the Monte Carlo estimator of $\pi_i(u \mid t)$ with the gold standard estimator over the whole follow-up period. Simulation results on the logit scale based on 200 data sets. The blue dashed line denotes the average bias. The red dashed lines denote the limits of agreement.

Web Table 1

Parameter estimates and standard errors for the joint model fitted to the AIDS data set. The parameters D_{11}, D_{22}, and D_{12} correspond to the variance of the random intercepts b_{i0}, the variance of the random slopes b_{i1}, and their covariance, respectively.

Event Process			Longitudinal Process		
	est.	std. err.		est.	std. err.
ddI	0.3450	0.1526	intercept	2.9396	0.1145
failure	0.1083	0.1689	time	−0.0323	0.0122
prevOI	0.6316	0.2409	ddI	0.0277	0.0727
male	−0.3524	0.2555	failure	−0.0309	0.0852
α	−0.9769	0.1263	prevOI	−0.8758	0.0818
$\log(\xi_1)$	−2.1345	0.4182	male	0.1612	0.1215
$\log(\xi_2)$	−1.7771	0.4086	time × failure	0.0009	0.0085
$\log(\xi_3)$	−1.4316	0.4339	time × prevOI	−0.0107	0.0080
$\log(\xi_4)$	−1.9657	0.4940	time × male	−0.0029	0.0121
$\log(\xi_5)$	−1.8453	0.4737	time × ddI	0.0050	0.0064
$\log(\xi_6)$	−1.7844	0.5310	σ	0.3711	0.0384
$\log(\xi_7)$	−1.7766	0.6253			
			Random Effects		
			D_{11}	0.5926	0.0449
			D_{12}	−0.0007	0.0023
			D_{22}	0.0013	0.0003

Web Table 2

Parameter estimates and standard errors for the joint model fitted to the AIDS data set, under the time-dependent value + derivative parameterization. The parameters D_{11}, D_{22}, and D_{12} correspond to the variance of the random intercepts b_{i0}, the variance of the random slopes b_{i1}, and their covariance, respectively.

Event Process			Longitudinal Process		
	est.	std. err.		est.	std. err.
ddI	0.3527	0.1563	intercept	2.9396	0.1136
failure	0.1172	0.1736	time	−0.0327	0.0122
prevOI	0.6188	0.2440	ddI	0.0257	0.0726
male	−0.3724	0.2639	failure	−0.0352	0.0852
α	−0.9769	0.1275	prevOI	−0.8732	0.0822
α_d	−1.7600	4.5497	male	0.1660	0.1201
$\log(\xi_1)$	−2.1939	0.4656	time × failure	0.0008	0.0085
$\log(\xi_2)$	−1.8441	0.4637	time × prevOI	−0.0113	0.0081
$\log(\xi_3)$	−1.4965	0.4833	time × male	−0.0027	0.0121
$\log(\xi_4)$	−2.0288	0.5367	time × ddI	0.0047	0.0064
$\log(\xi_5)$	−1.9049	0.5142	σ	0.3710	0.0384
$\log(\xi_6)$	−1.8357	0.5616			
$\log(\xi_7)$	−1.8177	0.6462	Random Effects		
			D_{11}	0.5926	0.0449
			D_{12}	−0.0005	0.0027
			D_{22}	0.0013	0.0003

Web Table 3

Bias and lower and upper limits of agreement in the Bland-Altman plots for $t = 2$, 6, 12 and 18 months, and $\Delta t = 2$, 4 and 8 months. The top results correspond to the Empirical Bayes estimator, and the bottom ones to the Monte Carlo estimator. Simulation results on the logit scale based on 200 data sets.

	$\Delta t = 2$	$\Delta t = 4$	$\Delta t = 8$
	bias (low; upp)	bias (low; upp)	bias (low; upp)
$t = 2$	0.061 (−2.02; 2.15)	0.053 (−2.07; 2.18)	0.033 (−2.18; 2.25)
$t = 6$	0.001 (−2.15; 2.15)	−0.008 (−2.21; 2.19)	−0.029 (−2.34; 2.28)
$t = 12$	−0.052 (−2.27; 2.17)	−0.062 (−2.35; 2.23)	−0.086 (−2.55; 2.38)
$t = 18$	−0.280 (−2.71; 2.15)	−0.298 (−2.82; 2.23)	−0.341 (−3.12; 2.44)
$t = 2$	0.058 (−2.00; 2.11)	0.058 (−2.03; 2.15)	0.042 (−2.14; 2.22)
$t = 6$	0.001 (−2.11; 2.11)	−0.006 (−2.16; 2.15)	−0.016 (−2.28; 2.25)
$t = 12$	−0.054 (−2.23; 2.13)	−0.064 (−2.31; 2.18)	−0.080 (−2.49; 2.33)
$t = 18$	−0.272 (−2.66; 2.12)	−0.289 (−2.77; 2.19)	−0.329 (−3.06; 2.41)