Supplementary material for
"Joint longitudinal and time-to-event cure models to improve the assessment of being cured"

Antoine Barbieri & Catherine Legrand

Institute of Statistics, Biostatistics and Actuarial sciences,
Université catholique de Louvain,
Belgium.

Correspondence authors: Antoine.Barbieri@gmail.com

1 Conditional probability of being cured for the FJCMs

In this section, we detail some equations associated with the conditional probability of being cured using joint modeling. As in the article, note that all computations are done conditionally on the explanatory variables whatever the submodels. Given the property associated with the conditioning of the normal distribution, the multivariate normal distributions associated with the LMM given the membership to the class $d\,(d = 0, 1)$ are:

$$\begin{aligned}
\boldsymbol{Y_i}|D_i = d &\sim \mathcal{N}\left(\boldsymbol{X_i\beta_d}, \boldsymbol{\Gamma_d}\right), \quad \boldsymbol{\Gamma_d} = \boldsymbol{U_i\Sigma_d U_i^\top} + \sigma_d^2\boldsymbol{I_{n_i}} \\
\boldsymbol{Y_i}|\boldsymbol{b_i}, D_i = d &\sim \mathcal{N}\left(\boldsymbol{X_i\beta_d} + \boldsymbol{U_i b_i}, \sigma_d\boldsymbol{I_{n_i}}\right) \\
\boldsymbol{b_i}|\boldsymbol{Y_i} = \boldsymbol{y_i}, D_i = d &\sim \mathcal{N}\left(\boldsymbol{\Sigma_d U_i^\top \Gamma_d}^{-1}\left(\boldsymbol{y_i} - \boldsymbol{X_i\beta_d}\right),\ \boldsymbol{\Sigma_d} - \boldsymbol{\Sigma_d U_i^\top \Gamma_1^{-1} U_i \Sigma_d}\right) \\
\begin{pmatrix}\boldsymbol{Y_i}\\ \boldsymbol{b_i}\end{pmatrix}|D_i = d &\sim \mathcal{N}\left(\begin{bmatrix}\boldsymbol{X_i\beta_d}\\ \boldsymbol{0}\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Gamma_d} & \boldsymbol{\Sigma_d U_i^\top}\\ \boldsymbol{U_i\Sigma_d} & \boldsymbol{\Sigma_d}\end{bmatrix}\right).
\end{aligned}$$

The conditional probability of being cured given the subject i is event-free at time t is:

$$\begin{aligned}
\pi_{i|t}\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{a}}\right) &= \Pr\left(D_i = 0|T_i > t, \boldsymbol{Y_i} = \boldsymbol{y_{i|t}}; \widehat{\boldsymbol{\theta}}_{\boldsymbol{a}}\right) \\
&= \frac{\Pr\left(D_i = 0, T_i > t, \boldsymbol{Y_i} = \boldsymbol{y_{i|t}}; \widehat{\boldsymbol{\theta}}_{\boldsymbol{a}}\right)}{\Pr\left(T_i > t, \boldsymbol{Y_i} = \boldsymbol{y_{i|t}}; \widehat{\boldsymbol{\theta}}_{\boldsymbol{a}}\right)} \\
&= \frac{\Pr\left(T_i > t, \boldsymbol{Y_i} = \boldsymbol{y_{i|t}}|D_i = 0; \widehat{\boldsymbol{\theta}}_{\boldsymbol{a}}\right)\Pr\left(D_i = 0; \widehat{\boldsymbol{\theta}}_{\boldsymbol{a}}\right)}{\sum_{d=0}^{1}\Pr\left(T_i > t, \boldsymbol{Y_i} = \boldsymbol{y_{i|t}}|D_i = d; \widehat{\boldsymbol{\theta}}_{\boldsymbol{a}}\right)\Pr\left(D_i = d; \widehat{\boldsymbol{\theta}}_{\boldsymbol{a}}\right)} \\
&= \frac{\Omega_0}{\Omega_0 + \Omega_1}
\end{aligned}$$

where,

$$\begin{aligned}
\Omega_0 &= \Pr\left(T_i > t, \boldsymbol{Y_i} = \boldsymbol{y_{i|t}}|D_i = 0; \widehat{\boldsymbol{\theta}}_{\boldsymbol{a}}\right)\Pr\left(D_i = 0; \widehat{\boldsymbol{\theta}}_{\boldsymbol{a}}\right) \\
&= \Pr\left(T_i > t|\boldsymbol{Y_i} = \boldsymbol{y_{i|t}}, D_i = 0; \widehat{\boldsymbol{\theta}}_{\boldsymbol{a}}\right) f\left(\boldsymbol{y_{i|t}}|D_i = 0; \widehat{\boldsymbol{\theta}}_{\boldsymbol{a}}\right)\Pr\left(D_i = 0; \widehat{\boldsymbol{\theta}}_{\boldsymbol{a}}\right) \\
&= f\left(\boldsymbol{y_{i|t}}|D_i = 0; \widehat{\boldsymbol{\theta}}_{\boldsymbol{a}}\right)\Pr\left(D_i = 0; \widehat{\boldsymbol{\theta}}_{\boldsymbol{a}}\right) \qquad (1)
\end{aligned}$$

and,

$$\begin{aligned}
\Omega_1 &= \Pr\left(T_i > t, \boldsymbol{Y_i} = \boldsymbol{y_{i|t}}|D_i = 1; \widehat{\boldsymbol{\theta}}_{\boldsymbol{a}}\right)\Pr\left(D_i = 1; \widehat{\boldsymbol{\theta}}_{\boldsymbol{a}}\right) \\
&= \Pr\left(T_i > t|\boldsymbol{Y_i} = \boldsymbol{y_{i|t}}, D_i = 1; \widehat{\boldsymbol{\theta}}_{\boldsymbol{a}}\right) f\left(\boldsymbol{y_{i|t}}|D_i = 1; \widehat{\boldsymbol{\theta}}_{\boldsymbol{a}}\right)\Pr\left(D_i = 1; \widehat{\boldsymbol{\theta}}_{\boldsymbol{a}}\right) \\
&= \int \Pr\left(T_i > t|\boldsymbol{Y_i} = \boldsymbol{y_{i|t}}, \boldsymbol{b_i}, D_i = 1; \widehat{\boldsymbol{\theta}}_{\boldsymbol{a}}\right) f\left(\boldsymbol{b_i}|\boldsymbol{Y_i} = \boldsymbol{y_{i|t}}, D_i = 1; \widehat{\boldsymbol{\theta}}_{\boldsymbol{a}}\right) d\boldsymbol{b_i} \\
&\quad \times f\left(\boldsymbol{y_{i|t}}|D_i = 1; \widehat{\boldsymbol{\theta}}_{\boldsymbol{a}}\right)\Pr\left(D_i = 1; \widehat{\boldsymbol{\theta}}_{\boldsymbol{a}}\right). \qquad (2)
\end{aligned}$$

We have

$$\boldsymbol{b_i}|\boldsymbol{Y_i} = \boldsymbol{y_{i|t}}, D_i = 1 \quad \sim \quad \mathcal{N}\left(\boldsymbol{\Sigma_1 U_{i|t}^\top \Gamma_1^{-1}}\left(\boldsymbol{y_{i|t}} - \boldsymbol{X_{i|t}\beta_1}\right),\ \boldsymbol{\Sigma_1} - \boldsymbol{\Sigma_1 U_{i|t}^\top \Gamma_1^{-1} U_{i|t}\Sigma_1}\right), \tag{3}$$

with $\boldsymbol{\Gamma_1} = \boldsymbol{U_{i|t}\Sigma_1 U_{i|t}^\top} + \sigma_\varepsilon^2 \boldsymbol{I_{n_{i|t}}}$. Ω_1 defined in equation (2) can be approximated using MCMC,

$$\Omega_1 \quad \approx \quad \frac{1}{M}\sum_{m=1}^{M} \Pr\left(T_i > t|\boldsymbol{b_i}^{(m)}, D_i = 1; \widehat{\boldsymbol{\theta}}_{\boldsymbol{a}}\right) f\left(\boldsymbol{y_{i|t}}|D_i = 1; \widehat{\boldsymbol{\theta}}_{\boldsymbol{a}}\right) \Pr\left(D_i = 1; \widehat{\boldsymbol{\theta}}_{\boldsymbol{a}}\right), \tag{4}$$

with $\boldsymbol{b_i}^{(m)}$ a draw of the M-sample of $\boldsymbol{b_i}$ from the distribution defined in equation (3). Another way for this approximation is to use directly the expectation of $\widetilde{\boldsymbol{b_i}}$ of subject-specific random effects defined by $\widetilde{\boldsymbol{b_i}}_{|\boldsymbol{y_{i|t}}, D_i=1} = \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{1}}\boldsymbol{U_{i|t}^\top}\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{1}}^{\boldsymbol{-1}}\left(\boldsymbol{y_{i|t}} - \boldsymbol{X_{i|t}}\widehat{\boldsymbol{\beta}}_{\boldsymbol{1}}\right)$, at least to make a sum in equation (4).

2 Simulation complements

Dataset generation

We consider an fictive design because the application one is based on linear longitudinal trends while we want to study joint cure models on polynomial form for the longitudinal trajectories across classes. Figure 1 shows the longitudinal profiles of both classes for the reference scenario (scenario 1, on left) and scenario 8 (on right). Note that the longitudinal behavior on left is the same for all scenarios excepted the 8-th and the 9-th. For the scenario 8, we use $\boldsymbol{\beta_1} = (2.2, -0.72, 0.14, 0.006)^\top$, while we use $\boldsymbol{\beta_0} = \boldsymbol{\beta_1} = (2.2, -0.72, 0.14, 0.006)^\top$ for scenario 9. Figure 2 presents the survival profiles for the global and the susceptible populations given scenario 1 (on left) and scenario 2 (on right) using the Kaplan–Meier estimates.

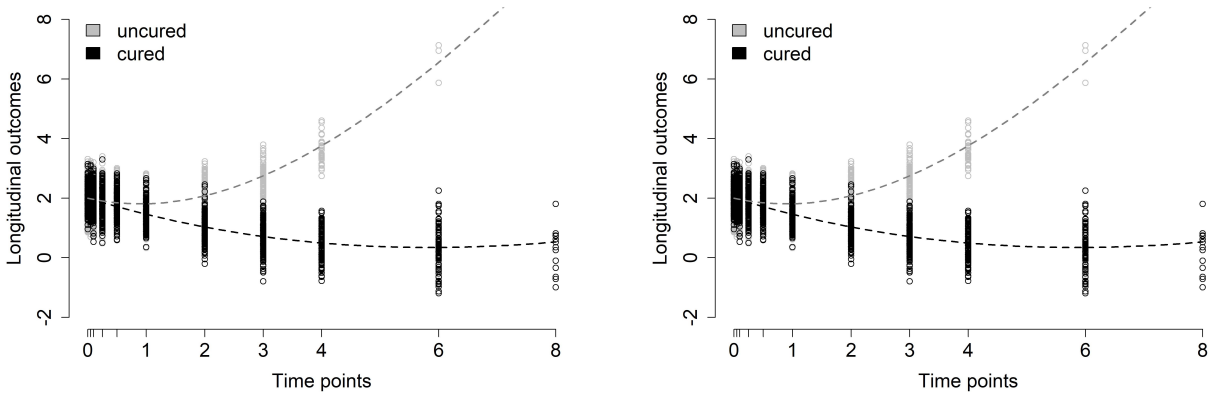

Figure 1 – Example of average longitudinal profiles given the reference scenario (scenario 1 on the left) and given the scenario 8 (on the right) using datesets generated by the FJCM-$\mathcal{A}_2$.

Complements of simulation given datasets generated by FJCM-$\mathcal{A}_2$

Although Figure 3 presents the distributions of $\widehat{\xi}_{b_0}$ and $\widehat{\xi}_{b_1}$ associated with the random intercept and the random slope, respectively, Figure 4 shows the EEs associated with the incidence and the latency models and Figure 5 shows the EEs obtained for the longitudinal parameters for each scenario. To complement the predictive performance study, the results associated with the well-classified rate for all scenarios are given in Figure 6. Moreover, the distributions of the area under the ROC curves (AUC) based on the estimated conditional probability $\widehat{\pi}_{i|t}$ are shown in Figure 7. The ROC curve plots of the sensitivity against the (1-specificity) for the range of all potential cut-offs c. In our case, the prediction is performed by

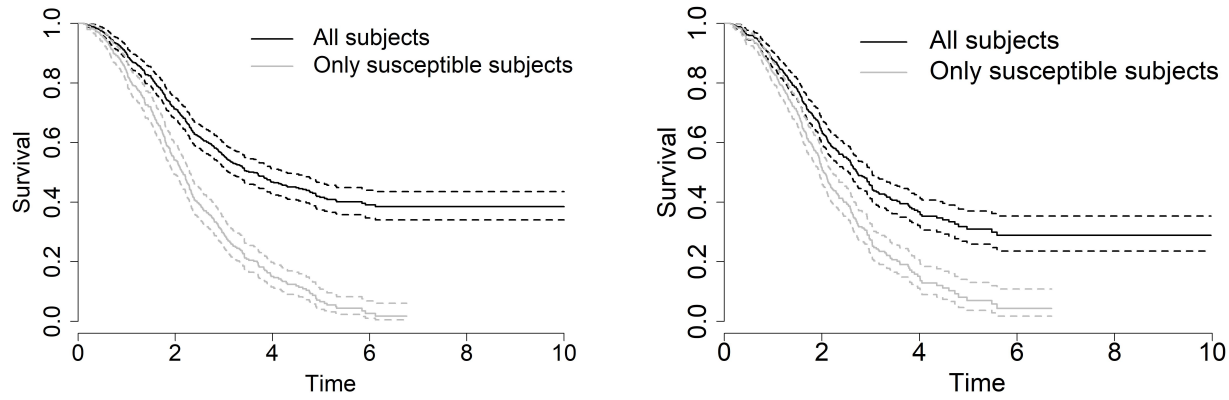

Figure 2 – Example of average survival profiles given the scenario 1 on the left and the scenario 2 on the right using datesets generated by the FJCM-$\mathcal{A}_2$.

taking into account both the estimation of $\pi_{i|t}$ and the performance evaluated for $c \in [0, 1]$. The sensitivity (i.e. true positive rate) is defined by $\Pr\left(\pi_{i|t} > c | D_i = 0\right)$, and the specificity (i.e. true negative rate) by $\Pr\left(\pi_{i|t} \leq c | D_i = 1\right)$. For each setting, the models are compared for both the well-classified rates and the AUCs. The results based on AUC show similar trends in comparison to the ones associated with the well-classified rates showed in the article. For both estimation quality and predictive performance, the three joint cure models give similar or better results than the MCM, even when the longitudinal outcomes do not add information on the class membership (see results of scenario 9).

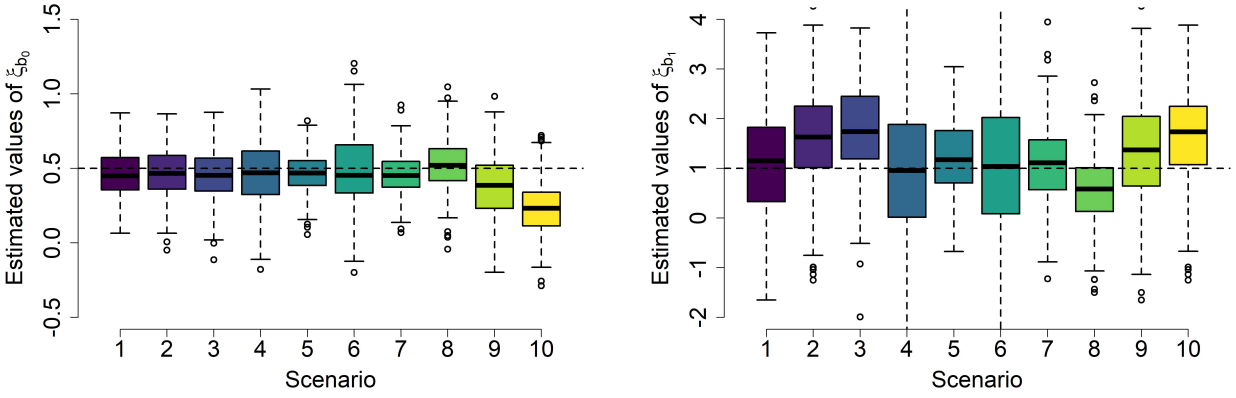

Figure 3 – Estimated values of parameters relative to share associations according to all 10 scenarios. Datasets are generated by FJCM-$\mathcal{A}_2$ given $\xi_{b_0} = 0.5$ and $\xi_{b_1} = 1$ (horizontal dotted lines).

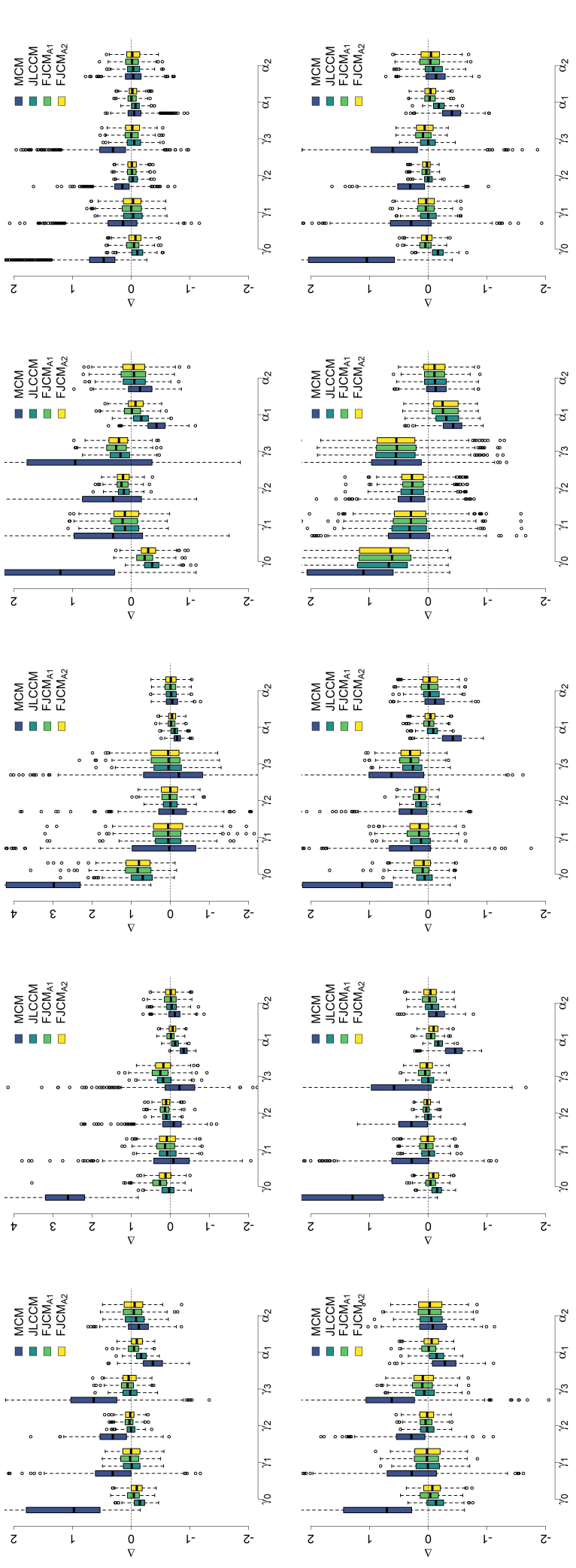

Figure 4 – Estimations errors (EEs) for incidence and latency parameters. From left panel to the right one on the top (resp. bottom), the distribution of EEs are given for each cure models according to the scnenarios 1 to 5 (resp. 6 to 10), respectively.

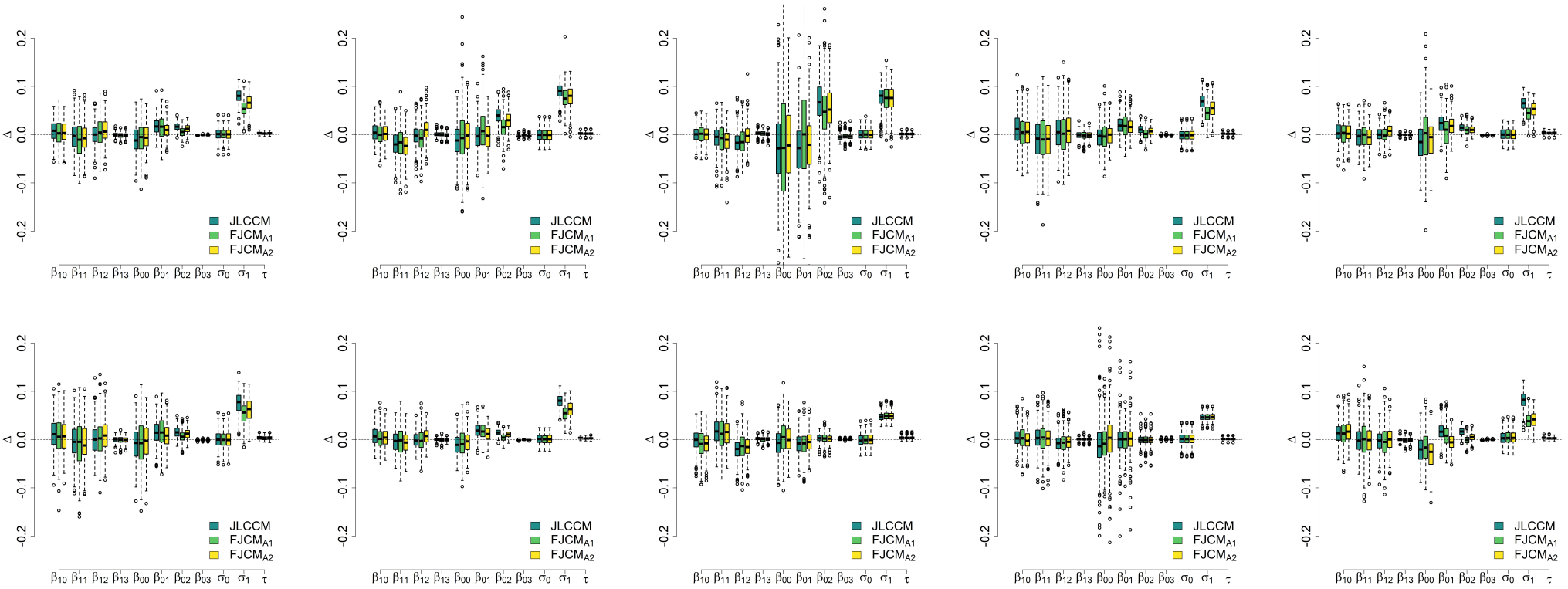

Figure 5 – Estimations errors (EEs) for longitudinal parameters. From left panel to the right one on the top (resp. bottom), the distribution of EEs are given for each cure models according to the scnenarios 1 to 5 (resp. 6 to 10), respectively.

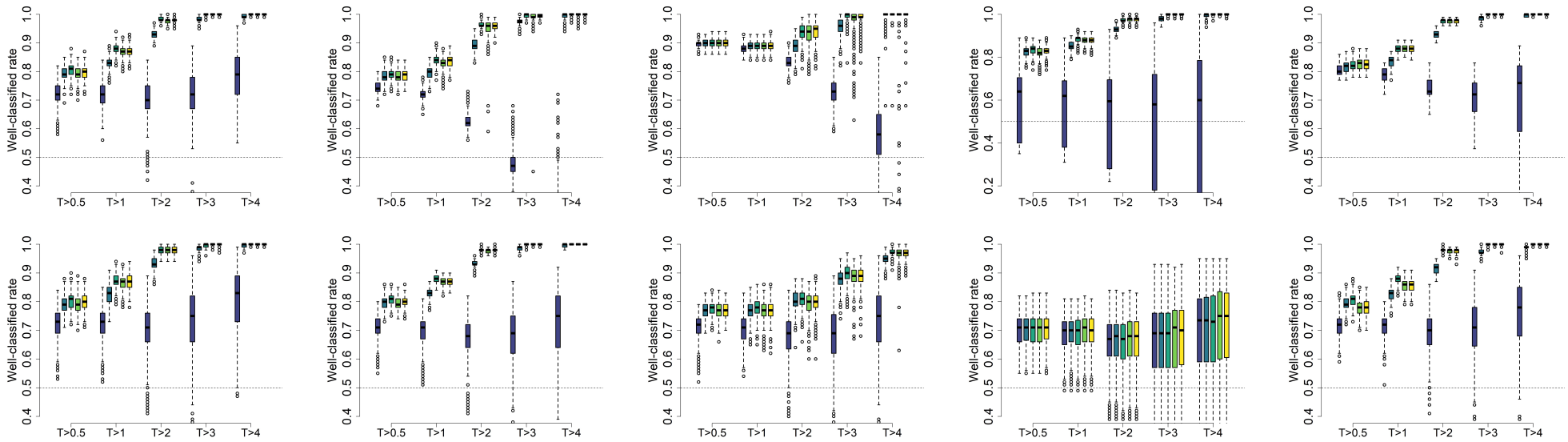

Figure 6 – Distributions of the well-classified rates given $T > t$ such as $d_i = \mathbb{1}\left(\pi_{i|t} < \frac{1}{2}\right)$. From left panel to the right one on the top (resp. bottom), the distribution of EEs are given for each cure models according to the scnenarios 1 to 5 (resp. 6 to 10), respectively. Boxplots from left to right for one time point are associated with MCM, JLCCM$_p$, JLCCM$_s$, FJCM-$\mathcal{A}_1$ and FJCM-$\mathcal{A}_1$, respectively. Note that JLCCM$_p$ is associated with the prediction given only one longitudinal outcome (y_{it}) at t, and JLCCM$_s$ associated with the prediction given $y_{i|t}$.

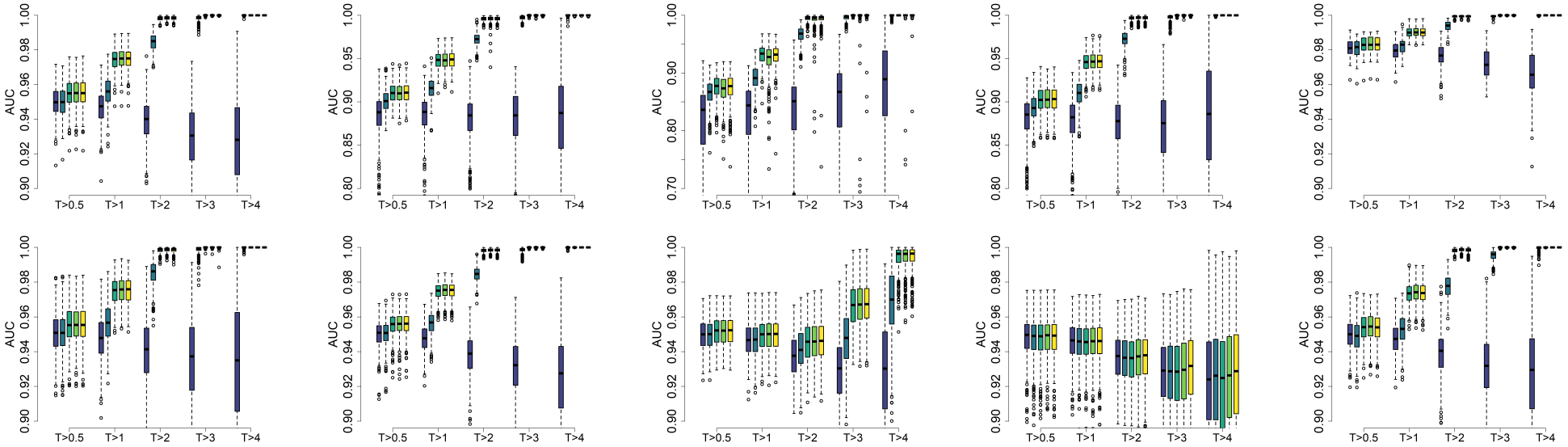

Figure 7 – Distributions of the AUC computed from $\widehat{\pi}_{i|t}$ when datasets were generated from FJCM-$\mathcal{A}_2$. From left panel to the right one on the top (bottom), the distributions are given for each cure models according to the scnenarios 1 to 5 (6 to 10), respectively. Boxplots from left to right for one time point are associated with MCM, JLCCM$_p$, JLCCM$_s$, FJCM-$\mathcal{A}_1$ and FJCM-$\mathcal{A}_2$, respectively.

Simulation results when datasets are from JLCCM or FJCM-$\mathcal{A}_1$

Simulation results when datasets are from JLCCM or FJCM-$\mathcal{A}_1$ are similar to the results obtained using FJCM-$\mathcal{A}_2$ in the generation procedure. According to scenarios 1 to 5, Figures 9 and 10 present the results about the estimation quality, and Figures 11 and 12 present the results about the predictive performance. Moreover, Figure 8 gives the distribution of the estimations of parameters ξ when datasets are generated from JLCCM or FJCM-$\mathcal{A}_1$ and according to scenarios 1 to 5.

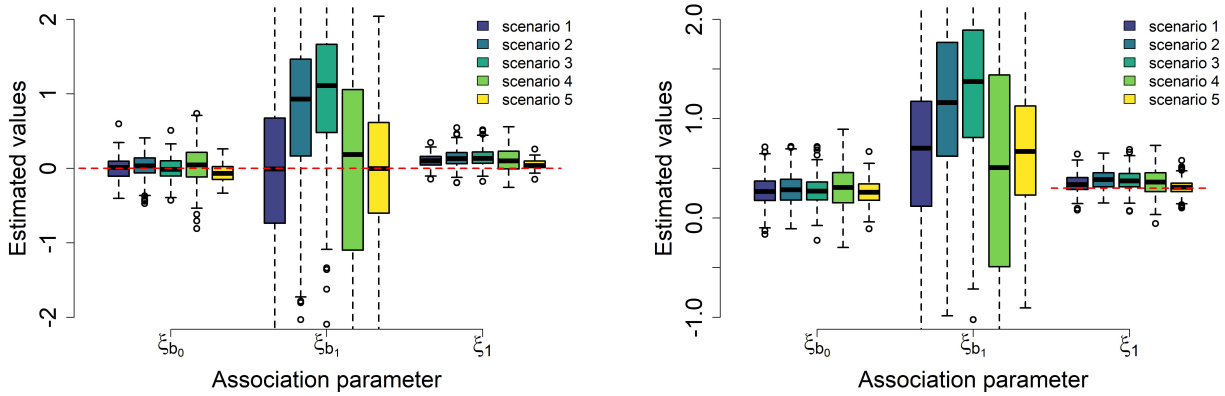

Figure 8 – Estimated values of parameters relative to share associations according to scenarios 1 to 5. Datasets are generated from JLCCM (on left) and from FJCM-$\mathcal{A}_1$ given $\xi_1 = 0.3$ (on rigth). Excepted value is precised by the horizontal dotted lines.

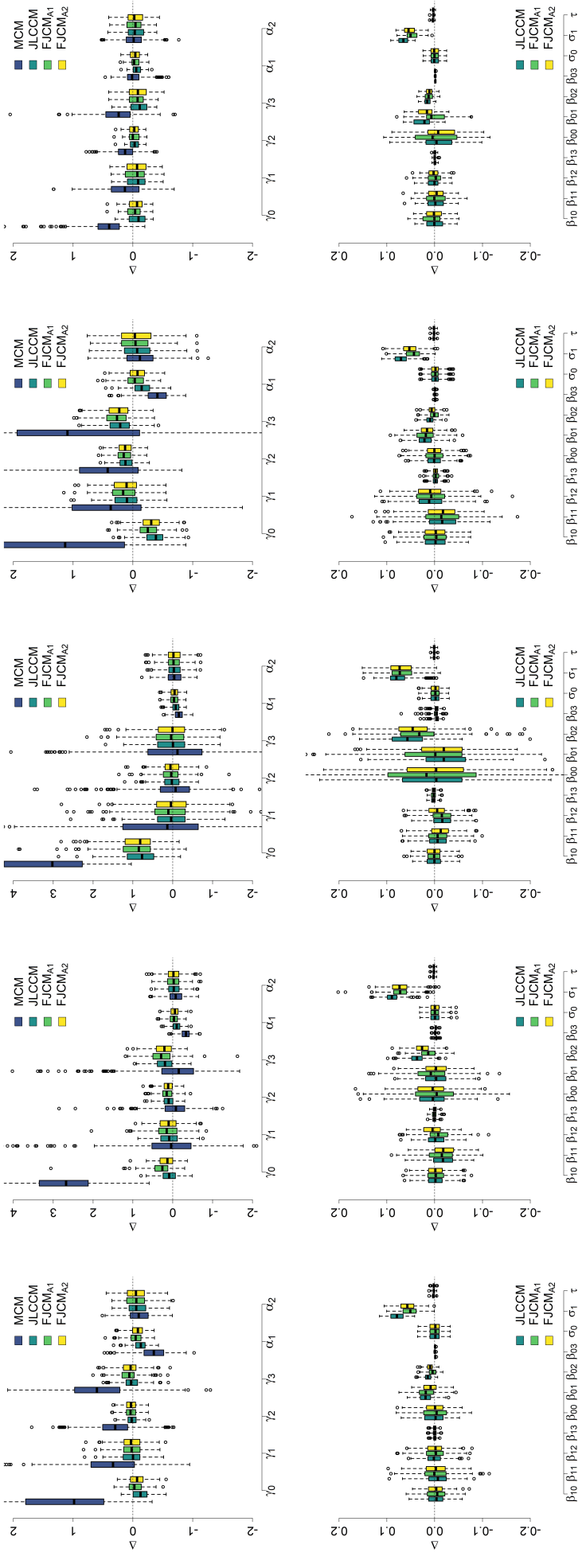

Figure 9 – Distributions of the estimations errors (EEs) for incidence and latency parameters on the top, and longitudinal parameters on the bottom. The plots are associated to the scenarios 1 to 5 from the left to the right, respectively. Datasets are generated from the JLCCM.

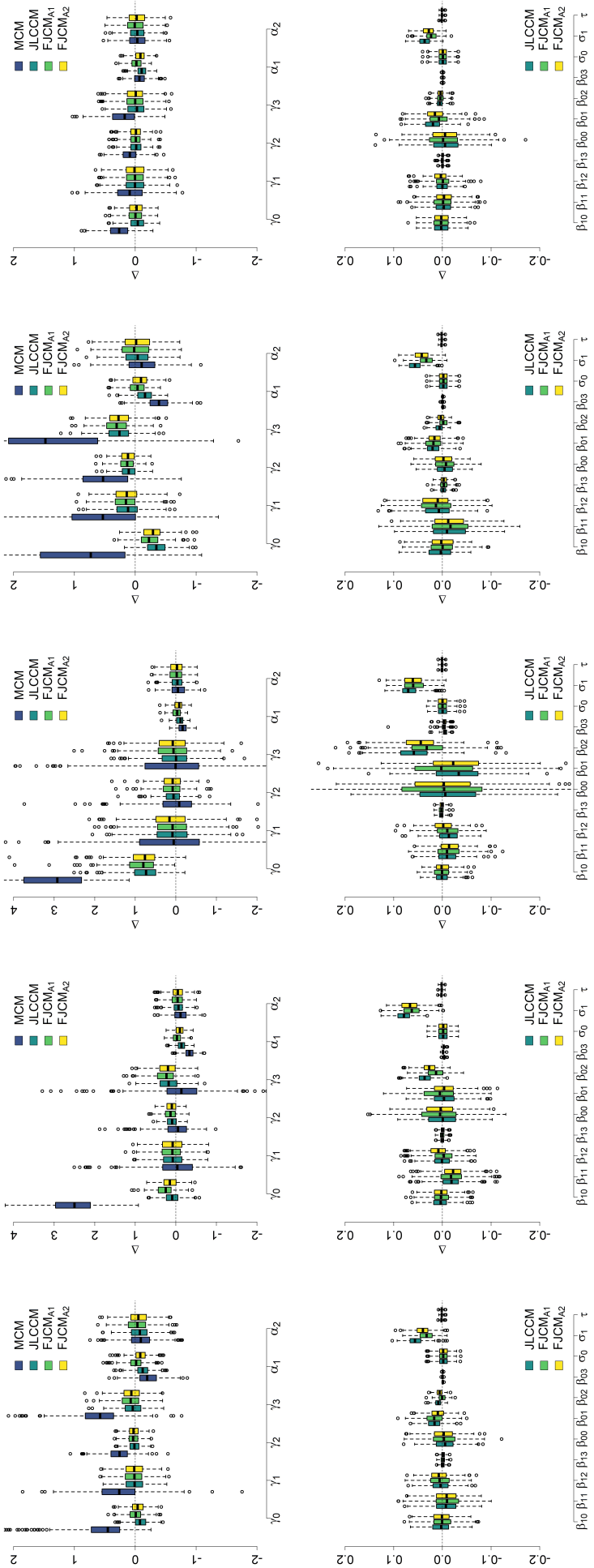

Figure 10 – Distributions of the estimations errors (EEs) for incidence and latency parameters on the top, and longitudinal parameters on the bottom. The plots are associated to the scenarios 1 to 5 from the left to the right, respectively. Datasets are generated from the FJCM-$\mathcal{A}_1$.

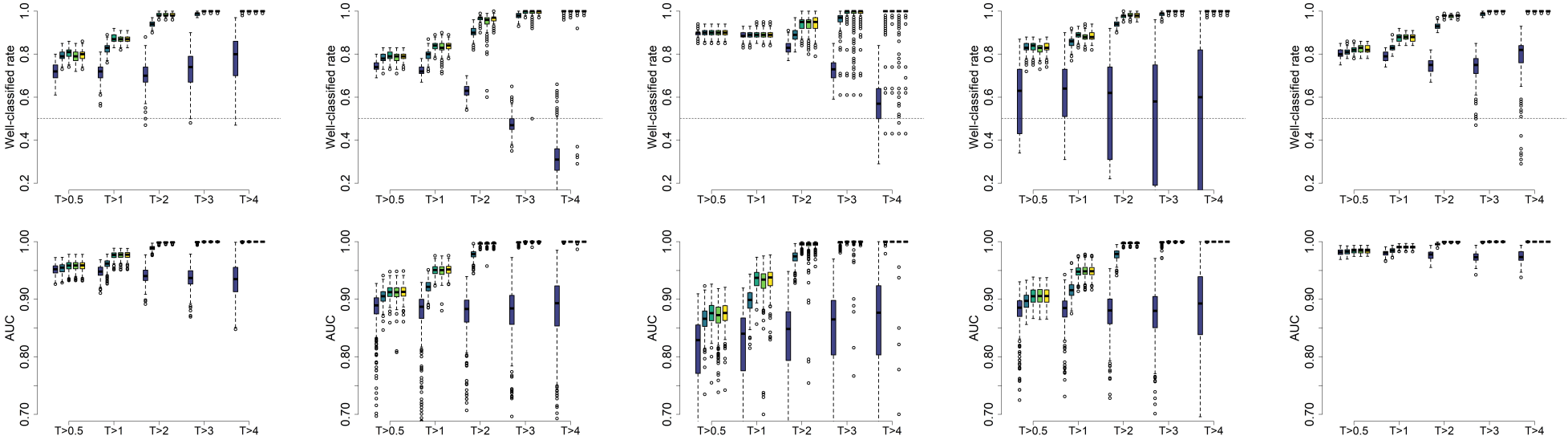

Figure 11 – Distributions of the well-classified rates (on the top), the AUC of ROC curves (on the middle) and the optimal threshold (on the bottom). The plots are associated to the scenarios 1 to 5 from the left to the right, respectively. Datasets are generated from the JLCCM.

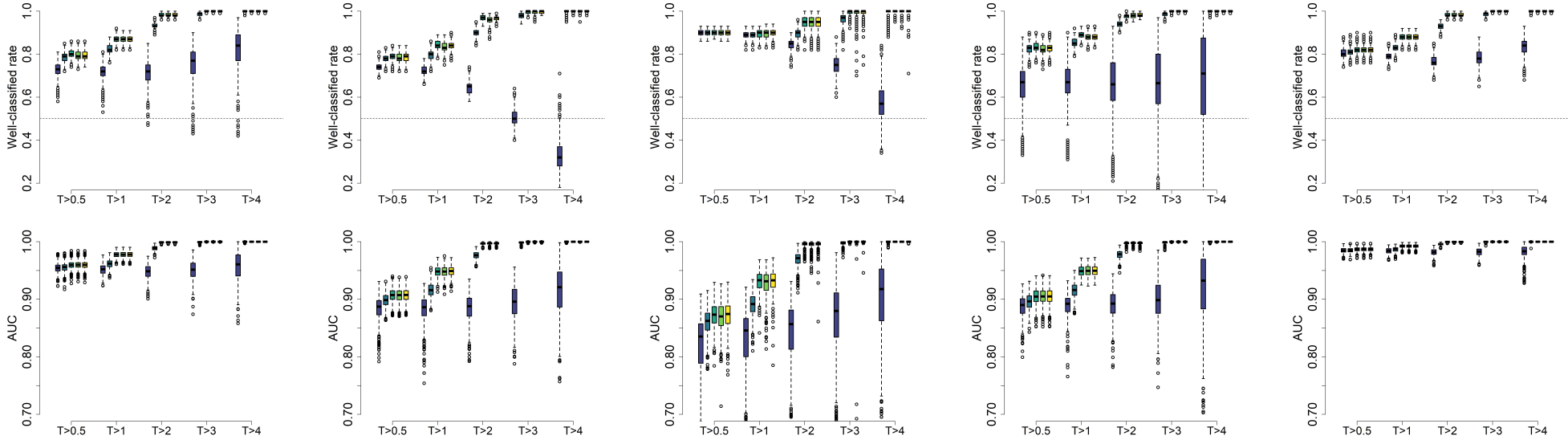

Figure 12 – Distributions of the well-classified rates (on the top), the AUC of ROC curves (on the middle) and the optimal threshold (on the bottom). The plots are associated to the scenarios 1 to 5 from the left to the right, respectively. Datasets are generated from the FJCM-$\mathcal{A}_1$.

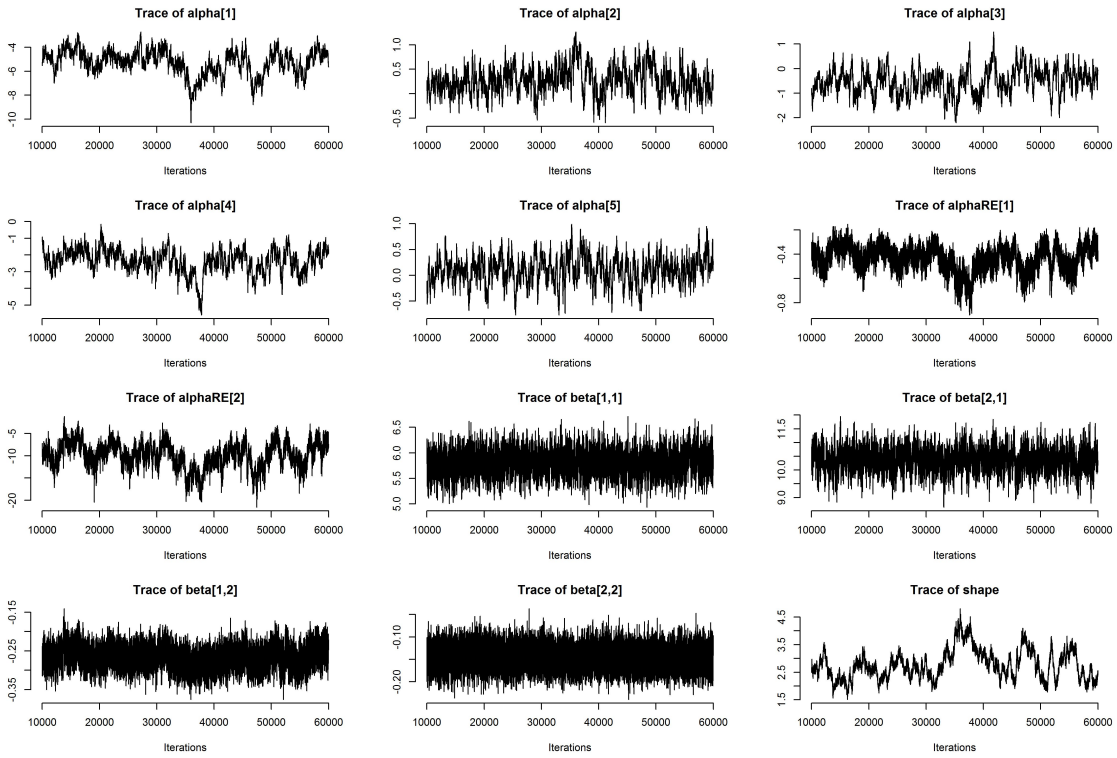

Figure 13 – Trace of the latency parameters and fixed effects of the longitudinal submodel for the FJCM-$\mathcal{A}_2$ with two random effects: intercept (alpha_RE1) and slope (alpha_RE2).

3 Application complements

Estimation

The estimation of FJCM-$\mathcal{A}_2$ given two random effects presents bad convergences. This is caused by the the presence of correlated parameters. Indeed, the traces of associated parameters (alpha_RE1 alpha_RE2) have similar behavior given different scale (Figure 13) and opposite behavior in comparison to the shape parameter. This influences the convergence of the latency parameters, and especially the shape parameter. When only the random intercept is considered, the convergence of Markov chains is improved (Figure 14).

Predictions

Regarding the prediction of the class membership for subjects still event-free at their censoring time, differences between the four considered cure models for 11 subjects are observed. These subjects have a baseline profile of short-term survivors causing high estimations of their incidence probability (Table 1). However, the observed longitudinal trajectories of these subjects show characteristics close to those of long-term survivors. The MCM predicts these 11 subjects as short-term survivors while the JLCCM predicts them as long-term survivors (Figure 15, top-left panel). The longitudinal information gives optimistic expectations for those subjects given their high CD4 counts over time.

Concerning the predictions from FJCMs, both the FJCM-$\mathcal{A}_1$ and the FJCM-$\mathcal{A}_2^\star$ predict 6 out of the 11 subjects as short-term survivors while the JLCCM predicts all of them as long-term survivors (Figure 15). This difference of classification is due to the information from longitudinal outcomes included in the latency submodel, increasing the estimation of later one. Finally, among the five other subjects, only the two patients have a different class membership given the two considered shared associations ($\mathcal{A}_1$ versus $\mathcal{A}_2^\star$). These two subjects (Figure 15) are classified as long-term survivors with the FJCM-$\mathcal{A}_2^\star$ while they are classified as short-term survivors with the FJCM-$\mathcal{A}_1$ due to the difference of weigh of the survival function in the conditional probability.

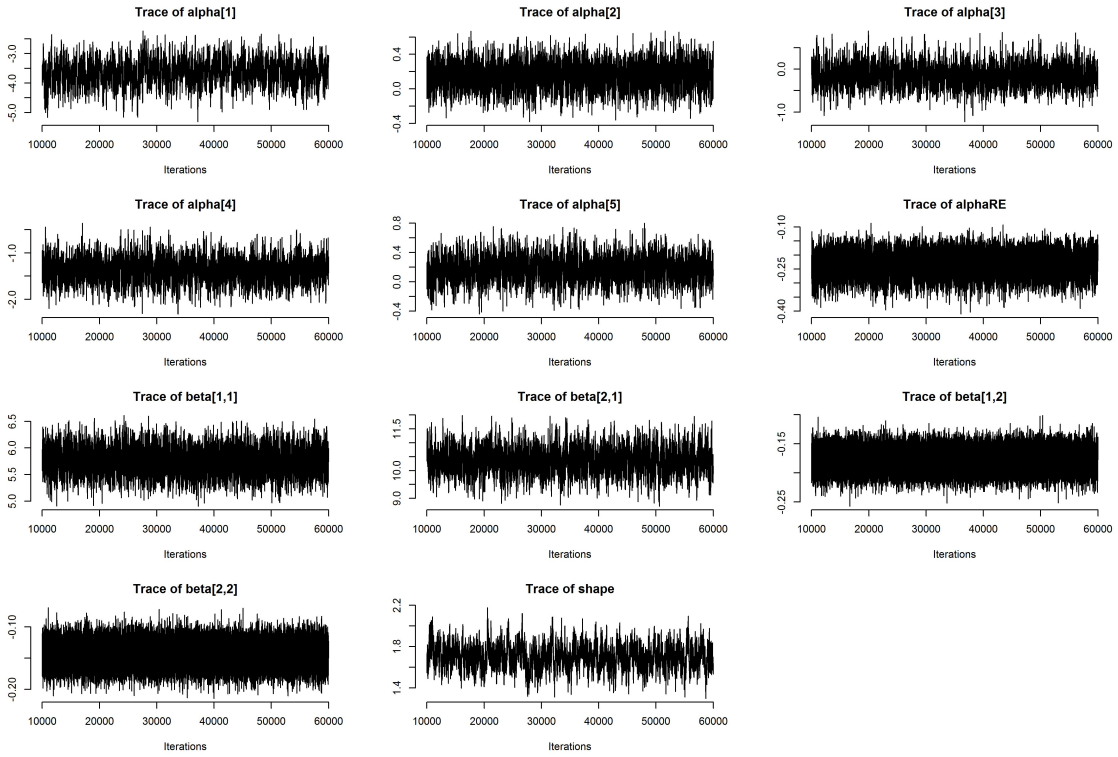

Figure 14 – Trace of the latency parameters and fixed effects of the longitudinal submodel for the FJCM-$\mathcal{A}_2$ with only one random intercept (alpha_RE1).

Table 1 – Individual information for the subjects classified differently according to the cure models. Estimations of the conditional probability of being long-term (LT) survivor from each cure model, estimations of their survival given short-term (ST) survivors, estimations of their incidence probability and the subject-specific covariates are given.

	ID of subject										
	20	28	63	86	110	189	219	346	348	441	460
Conditional probability of being LT survival models											
MCM	0.12	0.12	0.10	0.05	0.06	0.02	0.09	0.08	0.06	0.14	0.06
JLCCM	0.94	0.76	0.57	0.78	0.67	0.62	0.52	0.56	0.66	0.77	0.69
FJCM-$\mathcal{A}_1$	0.84	0.41	0.34	0.57	0.44	0.52	0.23	0.25	0.44	0.49	0.49
FJCM-$\mathcal{A}_2^\star$	0.89	0.45	0.33	0.73	0.49	0.56	0.13	0.22	0.50	0.51	0.60
Incidence probability											
MCM	0.94	0.96	0.94	0.98	0.96	0.98	0.96	0.96	0.96	0.94	0.96
JLCCM	0.93	0.96	0.93	0.98	0.96	0.98	0.96	0.96	0.96	0.93	0.96
FJCM-$\mathcal{A}_1$	0.94	0.96	0.94	0.98	0.96	0.98	0.96	0.96	0.96	0.94	0.96
FJCM-$\mathcal{A}_2^\star$	0.94	0.96	0.94	0.98	0.96	0.98	0.96	0.96	0.96	0.94	0.96
Survival given ST survivor ($D = 1$)											
MCM	0.48	0.27	0.55	0.46	0.53	0.97	0.42	0.40	0.53	0.38	0.52
JLCCM	0.48	0.26	0.55	0.46	0.52	0.97	0.42	0.40	0.52	0.38	0.52
FJCM-$\mathcal{A}_1$	0.98	0.88	0.94	0.98	0.96	1.00	0.86	0.89	0.96	0.92	0.97
FJCM-$\mathcal{A}_2^\star$	0.96	0.84	0.92	0.95	0.94	1.00	0.86	0.88	0.94	0.90	0.95
Covariates:											
Drug	ddC	ddI	ddC	ddI	ddI	ddI	ddC	ddI	ddI	ddC	ddI
Gender	male	male	male	male	male	male	male	male	male	male	male
prevOI	AIDS	AIDS	AIDS	AIDS	AIDS	AIDS	AIDS	AIDS	AIDS	AIDS	AIDS
AZT	failure	failure	failure	intolerance	failure	intolerance	intolerance	failure	failure	failure	failure

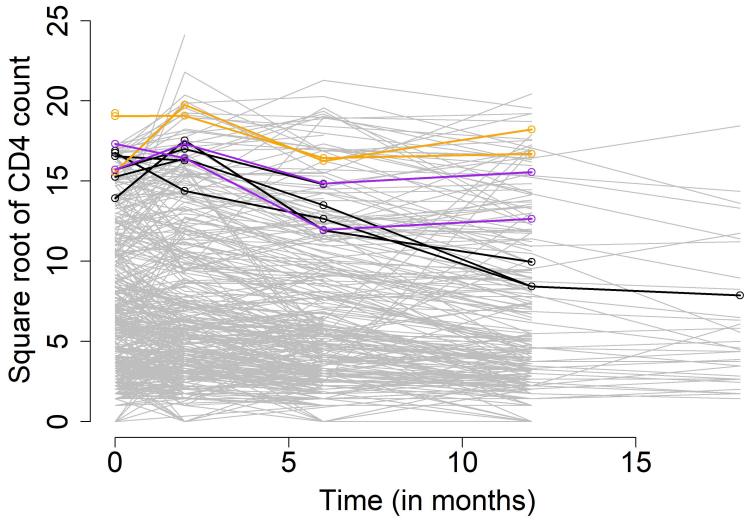

Figure 15 – Classification differences observed between the four cure models concern 11 subjects (all predicted as short-term survivors with MCM): all joint cure models classfied the 3 orange ones as long-term survivors, only the JLCCM and the FJCM-$\mathcal{A}_2^\star$ classified the 2 purple ones as LT survivors, only the JLCCM classifed the 6 black ones as LT survivors.