1. Web Case Studies Results

This section presents Web Figures 1 to 6 and Web Table 1 that illustrate several aspects of the fitted joint models to the AIDS and PBC data sets as well as the use of the multiply imputed residuals.

2. Web Example Visiting Process

In some applications, the future visit times may depend on previous visit times and/or the observed history of longitudinal responses. In such cases, a careful modelling approach of the visiting process is advantageous for the calculation of residuals because it accounts for the fact that the visiting behaviour may not be the same for all patients.

To make this clearer, consider the following hypothetical example: say that patients who start at high $y$-values and show a steep increase in their longitudinal profiles are more likely to dropout. Moreover, assume that for all the patients the future visit times depend on the history of their elapsed visit times, i.e., an MAR-type visiting process. Now for high fitted values and because of dropout, we would observe more negative residuals, such as in the PBC data set (i.e., Web Figure 1, bottom-right panel). To randomly impute residuals for the patients who dropout, we need to specify their future visit times. If we would impute $y_m^n$ at a few fixed time points, then there is the danger

*email: d.rizopoulos@erasmusmc.nl
that very few high $y_i^{m}$ values would exist in comparison to many low observed $y$-values, possibly resulting in a greater number of negative residuals. If, on the other hand, we would impute $y_i^{m}$ at many fixed time points, then there is the danger that too many high $y_i^{m}$ values would exist in comparison to fewer low observed $y$-values, possibly resulting in a greater number of positive residuals. This illustrates that there should be a balance between the observed visit times, and the ones that we would have observed had the event not occurred. Intuitively, under the MAR-type mechanism for the visiting process, we should match each patient who dropped out at time point $t^*$ with a patient who did not drop out, while they both share a very similar visit times pattern prior to $t^*$; then, we could simply impute $y_i^{m}$ for the patient who dropped out at the same visit times after $t^*$ of the patient who did not drop out. However, this matching is not always possible, and it is, in fact, in such settings that the modelling approach of the visiting process is advantageous in comparison with the simpler approach of imputing $y_i^{m}$’s at some fixed time points.
3. Web Simulation Study

3.1 Web Simulation Study Setup

In order to empirically evaluate the performance of the proposed multiple-imputation-based residuals we have performed a series of simulation studies. In particular, since in a non-random dropout context the joint modelling assumption cannot be verified from the observed data, we have focused on model misspecification. The effects of misspecification were studied in two directions. First, within the joint modelling framework, where we considered misspecification of the linear predictors and of the error distributions for the two submodels. Second, we considered misspecification of the missing data mechanism by positioning joint models as a special case of the general selection modelling framework formulated as

\[ p(T^*_i, y^o_i, y^m_i, b_i; \theta) = p(T^*_i \mid y^o_i, y^m_i, b_i; \theta)p(y^o_i, y^m_i \mid b_i; \theta)p(b_i; \theta), \]

i.e., the event time \( T^*_i \) could depend on \( y^m_i \) and/or \( b_i \).

In each case, we used the same setup as the one of the AIDS data set, using the MLEs presented in Table 1, and assuming the specification of the joint model presented in Section 3. For the censoring mechanism we assumed an exponential distribution with mean 21. In the following, we present four simulation scenarios, where under each scenario some components on the joint model specification have been altered.

**Scenario I:** this scenario is used as a reference case and, in fact, we simulate data under the correct joint model presented in Section 3 of the manuscript.

**Scenario II:** this scenario considers misspecification of the linear mixed model assumptions for the longitudinal outcomes. In particular, data are simulated under the following model

\[ y_i(t_{ij}) = \beta_0 + \beta_1 t_{ij} + \beta_2 \text{Treat}_i + \beta_3 \text{Treat}_i \times t_{ij} + \epsilon_{yi}(t_{ij}), \]
where $\varepsilon_{yi} \sim N(0, \Sigma)$ with

$$
\Sigma =
\begin{pmatrix}
1.000 & 0.485 & 0.221 & 0.089 & 0.027 \\
0.485 & 0.640 & 0.291 & 0.118 & 0.036 \\
0.221 & 0.291 & 0.360 & 0.146 & 0.044 \\
0.089 & 0.118 & 0.146 & 0.160 & 0.049 \\
0.027 & 0.036 & 0.044 & 0.049 & 0.040 \\
\end{pmatrix}.
$$

This model specification violates the linear mixed model assumptions in two manners. First, we do not include random effects to induce correlation between the repeated measurements (in fact we have assumed a serial correlation component of the exponential type). Second, the variability of the marginal error terms decreases as time increases; this is in contrast with the marginal models induced by a mixed model, i.e., mixed models assume that the marginal error variance increases with time. Thus, the assumed marginal model does not support the existence of random effects.

**Scenario III:** this scenario considers misspecification of the survival model assumptions for the time-to-event outcome. In particular, data are simulated under the following model

$$
\log T_i^* = \gamma_0 + \gamma_1 \text{Treat}_i + \alpha_0 b_{i0} + \alpha_1 b_{i1} + \sigma_t \varepsilon_{ti},
$$

where $\alpha_0 = 0.407$, $\alpha_1 = 0.203$, and $\varepsilon_{ti} \sim N(0, 1)$. This model specification violates the assumptions of the survival submodel presented in Section 3 of the manuscript in two manners. First, we postulate that the time-to-event depends on the underlying random effects $b_i$ and not on the value of the longitudinal outcome $W_i(t)$ at time point $t$. Second, the error terms $\varepsilon_{ti}$ do not follow the standard extreme value distribution.

**Scenario IV:** this scenario considers misspecification of the missing data mechanism. In
particular, data are simulated under a selection model in which the probability of dropping out in each time point depends on the actual value of $y_i$ at this time point, i.e.,

$$\log \frac{\Pr\{D_i(t_{ij}) = 1 \mid D_i(t_{ij}) = 0\}}{1 - \Pr\{D_i(t_{ij}) = 1 \mid D_i(t_{ij}) = 0\}} = \gamma_0 + \gamma_1\text{Treat}_i + \alpha y_i(t_{ij}),$$

where $\{D_i(t_{ij}) = 1 \mid D_i(t_{ij}) = 0\}$ denotes the event of dropping out at time point $t_{ij}$ given that patient $i$ has not dropped prior to $t_{ij}$. This scenario violates the form of the dropout mechanism, presented in Section 4.1, Equation (6), that is induced by the joint model.

3.2 Web Simulation Study Results

The results are presented in Web Figures 7 to 10 for the longitudinal process, and Web Figures 11 to 14 for the survival process. From Scenario I we can extract similar conclusions as we did for the AIDS and PBC data sets in the manuscript. Namely, the observed residuals are clearly affected by the non-random missing data mechanism posited by the joint model, whereas the multiply imputed residuals attempt to compensate for the missing information by exploiting the joint model assumptions. In the remaining scenarios however, we observe evidence of misspecification. For the longitudinal process this is more evident in Scenario II where we observe that the shapes of the multiply imputed residuals clearly indicate that the postulated longitudinal submodel is not appropriate. For the survival outcome the signs of model misspecifications are more clear in Scenario III where we observe that both the AFT and martingale residuals clearly indicate problems in the survival submodel part of the postulated joint model. Furthermore, the residual plots under the misspecified missing data mechanism (Scenario IV), presented in Web Figures 10 and 14, show more extreme behaviour for the survival submodel. This is to be expected because in this scenario, it is the conditional distribution of the survival outcome given the longitudinal one that is misspecified.

The general conclusion that can be extracted from these simulations is that the plots of the observed residuals, in all cases, suggest that the assumptions of the joint model seem to be violated, even when we simulated from the true model. On the other
hand, the multiply imputed residuals show the expected systematic trends mainly in the scenarios where we have misspecified some of the components of the joint model. Therefore, by looking at the observed residuals alone, one cannot be certain if something is indeed wrong with the postulated joint model or if the systematic trends in the observed residuals plots are mainly attributed to the non-random dropout setting. In such cases, the multiply imputed residuals are more insightful regarding the model assumptions, because they explicitly take dropout into account.
Web Figure 1. The top-left and center-left panels show the Kaplan-Meier estimate for the two treatment groups, for the AIDS and PBC data sets, respectively. The top-right and center-right panels show the observed data and superimposed loess fits for the two treatment groups, for the AIDS and PBC data sets, respectively. The bottom-left panel shows the visit times for each one of the patients during the follow-up period in the PBC data set.
**Web Figure 2.** Observed standardized marginal and subject-specific residuals (black circles), augmented with one multiple imputation of the residuals corresponding to the missing longitudinal responses (blue points), for the AIDS and PBC data sets respectively. The black and blue superimposed solid lines represent loess fits, based on the observed residuals and on the observed residuals augmented with multiply imputed residuals, respectively.
Web Figure 3. Variability in the loess smoother due to missingness, for the standardized marginal and subject-specific residuals versus the fitted values, for the AIDS and PBC data sets, respectively. The black dashed lines represent the loess fit for each one of the multiple imputations (i.e., it is based on the observed and multiply imputed residuals together). The red lines represent the loess fit based on the observed data only.
Web Figure 4. Square root of the absolute standardized subject-specific residuals versus fitted values for the PBC data. The black circle depict the residuals corresponding to the observed data, and the grey points the multiply imputed residuals (based on 10 imputations). The dashed line is the loess fit for the observed residuals, and the solid line the weighted loess fit for all the residuals.
Web Figure 5. Posterior predictive checks for the AIDS data set. The grey dashed lines represent the sample average evolutions based on 60 simulated completed data sets. The black solid line denotes the sample average evolution based on the observed data, augmented with 50 multiple imputations for the missing cases. The black dashed line depicts the sample average evolution based on the observed data only (i.e., without augmentation). The red solid line is the fitted line based on the joint model.
Web Figure 6. Simulated reference distribution for observed residuals. In particular, 200 data sets were simulated based on the MLEs and the structure of the AIDS data set. For each data set standardized subject-specific and standardized marginal residuals were calculated for the observed part of the simulated longitudinal responses based on the joint model fit. The top-left and top-right panels show Q-Q plots for the observed residuals from the AIDS data set versus their simulated theoretical distribution. The dashed lines represent an envelope calculated from the simulated residuals. The bottom-left and bottom-right panels show Q-Q for the observed residuals from the AIDS data set versus the simulated theoretical distribution of the error terms. The dashed lines represent an envelope calculated from the simulated errors.
Web Figure 7. Simulation results based on 100 data sets under Scenario I. The left hand side panels show loess smooth curves for the standardized subject-specific and standardized marginal residuals using the observed residuals alone. The right hand side panels show loess smooth curves for the standardized subject-specific and standardized marginal residuals using the observed and the multiply imputed residuals.
Web Figure 8. Simulation results based on 100 data sets under Scenario II. The left hand side panels show loess smooth curves for the standardized subject-specific and standardized marginal residuals using the observed residuals alone. The right hand side panels show loess smooth curves for the standardized subject-specific and standardized marginal residuals using the observed and the multiply imputed residuals.
Web Figure 9. Simulation results based on 100 data sets under Scenario III. The left hand side panels show loess smooth curves for the standardized subject-specific and standardized marginal residuals using the observed residuals alone. The right hand side panels show loess smooth curves for the standardized subject-specific and standardized marginal residuals using the observed and the multiply imputed residuals.
Web Figure 10. Simulation results based on 100 data sets under Scenario IV. The left hand side panels show loess smooth curves for the standardized subject-specific and standardized marginal residuals using the observed residuals alone. The right hand side panels show loess smooth curves for the standardized subject-specific and standardized marginal residuals using the observed and the multiply imputed residuals.
Web Figure 11. Simulation results based on 100 data sets under Scenario I. The left hand side panel shows Kaplan-Meier estimates of the accelerated failure time residuals for each treatment group (solid and dashed lines, respectively). The superimposed red solid line is the survival function of the standard extreme value distribution. The right hand side panel shows loess smooth curves of the martingale residuals versus the fitted values of $y$ evaluated at the observed event times.
Web Figure 12. Simulation results based on 100 data sets under Scenario II. The left hand side panel shows Kaplan-Meier estimates of the accelerated failure time residuals for each treatment group (solid and dashed lines, respectively). The superimposed red solid line is the survival function of the standard extreme value distribution. The right hand side panel shows loess smooth curves of the martingale residuals versus the fitted values of $y$ evaluated at the observed event times.
Web Figure 13. Simulation results based on 100 data sets under Scenario III. The left hand side panel shows Kaplan-Meier estimates of the accelerated failure time residuals for each treatment group (solid and dashed lines, respectively). The superimposed red solid line is the survival function of the standard extreme value distribution. The right hand side panel shows loess smooth curves of the martingale residuals versus the fitted values of $y$ evaluated at the observed event times.
Web Figure 14. Simulation results based on 100 data sets under Scenario IV. The left hand side panel shows Kaplan-Meier estimates of the accelerated failure time residuals for each treatment group (solid and dashed lines, respectively). The superimposed red solid line is the survival function of the standard extreme value distribution. The right hand side panel shows loess smooth curves of the martingale residuals versus the fitted values of $y$ evaluated at the observed event times.
Web Table 1
Parameter estimates and standard errors for the visiting process model fitted to the PBC data set.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{v1}$</td>
<td>$-0.161$</td>
<td>$0.104$</td>
</tr>
<tr>
<td>$\beta_{v2}$</td>
<td>$-0.047$</td>
<td>$0.040$</td>
</tr>
<tr>
<td>$\log \phi$</td>
<td>$0.391$</td>
<td>$0.086$</td>
</tr>
<tr>
<td>$\log \psi$</td>
<td>$1.166$</td>
<td>$0.022$</td>
</tr>
<tr>
<td>$\log \eta^{-1}$</td>
<td>$-0.954$</td>
<td>$0.132$</td>
</tr>
</tbody>
</table>