Multiple-Imputation-Based Residuals and Diagnostic Plots for Joint Models of Longitudinal and Survival Outcomes

DOI: 10.1111/j.1541-0420.2009.01273.x
Handle: http://hdl.handle.net/1942/10813

Biometrics 00, 1–25 DOI: 00.0000/0.0000-0000.0000.00000.0
00 0000

Multiple-Imputation-Based Residuals and Diagnostic Plots for Joint Models of Longitudinal and Survival Outcomes

Dimitris Rizopoulos[1,*], Geert Verbeke[2], and Geert Molenberghs[2]

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

[2]Interuniversity Institute for Biostatistics and statistical Bioinformatics,
Katholieke Universiteit Leuven, Kapucijnenvoer 35, Blok D, bus 7001, B3000 Leuven, Belgium, and
Universiteit Hasselt, Agoralaan 1, B3590 Diepenbeek, Belgium

**email:* d.rizopoulos@erasmusmc.nl

Summary: The majority of the statistical literature for the joint modelling of longitudinal and time-to-event data has focused on the development of models that aim at capturing specific aspects of the motivating case studies. However, little attention has been given to the development of diagnostic and model-assessment tools. The main difficulty in using standard model diagnostics in joint models is the non-random dropout in the longitudinal outcome caused by the occurrence of events. In particular, the reference distribution of statistics, such as the residuals, in missing data settings is not directly available and complex calculations are required to derive it. In this paper we propose a multiple-imputation-based approach for creating multiple versions of the completed data set under the assumed joint model. Residuals and diagnostic plots for the complete data model can then be calculated based on these imputed data sets. Our proposals are exemplified using two real data sets.

Key words: Dropout; Joint Modelling; Longitudinal Data; Model Diagnostics; Residuals; Survival Data.

1. Introduction

In many longitudinal studies the outcomes recorded on the subjects under study include both a set of repeated measurements and the time at which an event of particular interest occurs: for instance, death, development of a disease or dropout from the study. These two outcomes are often separately analyzed. However, when interest is in measuring the effects of a longitudinal covariate measured with error in the survival outcome, a joint modelling approach is required. This has lead to a new and very active area of biostatistical research that deals with the joint modelling of longitudinal and time-to-event data (Tsiatis and Davidian, 2004).

The majority of the joint modelling literature has focused on the development of models that capture specific aspects of the motivating case studies. Examples are found in Ding and Wang (2008), Elashoff et al. (2008), Liu et al. (2008), Larsen (2004), Lin et al. (2002), Wang and Taylor (2001), and references therein. Little attention has been given to the development of diagnostics and model-assessment tools for such joint models, the only exception being the conditional residuals proposed by Dobson and Henderson (2003), hereafter abbreviated as DH. A reason for the lack of such tools in this area is attrition in the longitudinal profiles. In particular, when patients experience the event, they dropout from the study and no longitudinal measurements are available after this time point. Thus, a direct connection can be drawn between the missing data area and the joint modelling of longitudinal and survival data. Specifically, in the joint modelling framework, it is assumed that the occurrence of events is related with the underlying evolution of the subject-specific longitudinal profiles, which corresponds to a non-random dropout mechanism (Little, 1995). Under this non-randomness setting, the use of standard goodness-of-fit measures and residuals becomes problematic due to the fact that their reference distribution is not directly available. Thus, residual plots based on the observed data alone can be misleading, because these residuals

should not be expected to exhibit standard properties, such as zero mean and independence. For a more thorough discussion on the difficulties of model assessment in the missing data context we refer to Verbeke et al. (2008).

In this paper, we propose a new method for calculating residuals and producing diagnostic plots in joint models, based on the idea of multiply imputing the missing longitudinal responses under the fitted joint model, thus creating random versions of the completed data set. These completed data sets can then be used to extract conclusions regarding the modelling assumptions, and how these assumptions are affected by the non-random dropout. Contrary to the conditional residuals of DH, we feel that our approach is simple to use in practice, since it aims at validating assumptions about the complete data model, and not about the observed data model as in DH. Our method shares similarities with the approach of Gelman et al. (2005) who used multiple imputation for posterior predictive checks in missing data and latent variable contexts. Our proposals are exemplified using two real data sets. The first concerns the joint modelling of CD4 cell counts and time to death in HIV infected patients, and the second one deals with the joint modelling of serum bilirubin levels and time to death in patients with primary biliary cirrhosis. In addition, motivated by these case studies, we propose different multiple imputation schemes depending on the nature of the visiting process (i.e., the stochastic mechanism that generates the time points at which the longitudinal measurements are collected). In particular, in the first data set patients are asked to provide CD4 cell counts measurements at fixed time points, whereas in the second one serum bilirubin is recorded at random visit times. Finally, the practicality of the proposed methods can be explored within the R environment (R Development Core Team, 2008), with the publicly available package `JM` (developed by the first author) that can be downloaded from `http://cran.r-project.org`.

The rest of the paper is organised as follows. Section 2 presents the submodels specification

for the longitudinal and survival processes. Section 3 describes the two real data examples that are used throughout the paper and the joint models fitted to them. Section 4 starts by illustrating the problems in interpreting residual plots based on the observed data alone, and presents the multiple imputation schemes that are used to recreate random versions of the completed data set. Finally, Section 5 utilizes the multiple imputation residuals for checking the joint models fit in the two data sets, and Section 6 refers to the results of a simulation study.

2. Joint Modelling Framework

Let $y_{ij} = \{y_i(t_{ij}), j = 1, \ldots, n_i\}$ denote the longitudinal response measurements for the ith subject ($i = 1, \ldots, n$) taken at time points t_{ij}. We focus on continuous responses and we postulate a linear mixed effects model to capture the subject-specific evolutions

$$y_i(t_{ij}) = W_i(t_{ij}) + \varepsilon_{yi}(t_{ij}), \quad \varepsilon_{yi}(t_{ij}) \sim \mathcal{N}(0, \sigma_y^2), \tag{1}$$

where $W_i(t_{ij}) = x_i^\top(t_{ij})\beta + z_i^\top(t_{ij})b_i$, and X_i and Z_i are the design matrices (with corresponding row vectors $x_i^\top(t_{ij})$ and $z_i^\top(t_{ij})$) for the fixed and random effects, β and b_i, respectively. The random effects b_i are assumed to follow a multivariate normal distribution with mean zero and variance-covariance matrix D.

For the event process, we denote by T_i^* the true failure time for the ith subject and by $T_i = \min(T_i^*, C_i)$ the observed failure time, where C_i is the corresponding censoring time. Further, we define the event indicator as $\delta_i = I(T_i^* \leqslant C_i)$, where $I(\cdot)$ is the indicator function. Relative risks models of the form:

$$h(t \mid b_i) = h_0(t) \exp\{\gamma^\top x_{ti} + \alpha W_i(t)\},$$

have been traditionally used for the event outcome within the joint modelling framework (Wulfsohn and Tsiatis, 1997), where the baseline risk function $h_0(t)$ is typically left unspecified, x_{ti} denotes the vector of baseline covariates, $W_i(t)$ denotes the hypothetical true value of

the longitudinal time-dependent covariate at time t, and γ and α are regression coefficients. However, there are two issues with joint models, in which $h_0(t)$ is left unspecified. First, Hsieh et al. (2006) have recently noted that the nonparametric maximum likelihood estimate for the baseline hazard cannot be obtained explicitly under the random effects structure, and thus, when the profile score vector (that remains a function of the baseline hazard) is used for the estimation of standard errors, it leads to their underestimation. Second, the multiple-imputation approach for model diagnostics, which is going to be introduced in more detail in Section 4, requires a complete likelihood specification, and thus it cannot be applied in joint models with infinite dimensional baseline risk functions. Mainly for the latter reason, we are going to concentrate on parametric models for the survival outcome. Available options are either common survival distributions, such as the Weibull or Gamma, or more flexible models in which $h_0(t)$ is approximated using a piecewise-constant function by appropriately discretizing the time scale into a number of intervals or using splines-based approaches. In this paper and for illustrative purposes that will become apparent in Section 4.1, we postulate a Weibull model for the survival times under the accelerated failure time formulation; the possibility of using the developments of Section 4 with a more flexible survival model is presented in Section 7. In particular, let

$$\log T_i^* = \gamma^\top x_{ti} + \alpha W_i(T_i^*) + \sigma_t \varepsilon_{ti}, \tag{2}$$

where σ_t is a scale parameter, and ε_{ti} follows a standard extreme value distribution of a minimum. Note that the regression parameters γ and α have a different interpretation than in the relative risk model, and they measure the effect of the baseline covariates and of the time-dependent longitudinal outcome to the expected log survival time.

Combining (1) and (2), the joint likelihood contribution of the two outcomes for the ith subject is given by:

$$p(T_i, \delta_i, y_i; \theta) = \int \{p(T_i \mid b_i; \theta)^{\delta_i} S(T_i \mid b_i; \theta)^{1-\delta_i}\} p(y_i \mid b_i; \theta) p(b_i; \theta)\, db_i \tag{3}$$

$$\propto \int [\exp\{\zeta_i(b_i) - \log\sigma_t\}]^{\delta_i} \exp[-\exp\{\zeta_i(b_i)\}] \times$$
$$(\sigma_y^2)^{-n_i/2} \exp\Big(- \parallel y_i - X_i\beta - Z_i b_i \parallel^2 /2\sigma_y^2\Big) \times \det(D)^{-1/2} \exp(-b_i^\top D^{-1} b_i/2)\, db_i,$$

where $p(\cdot)$ denotes appropriate probability density functions, for the event outcome $S(\cdot)$ denotes the survival function conditional on the random effects, y_i is the $n_i \times 1$ vector of longitudinal responses of the ith subject, $\theta^\top = (\gamma^\top, \alpha, \sigma_t, \beta^\top, \sigma_y^2, \text{vech}(D))$ is the parameter vector, with vech(D) denoting the unique elements of D, $\zeta_i(b_i) = \{\log T_i - \gamma^\top x_{ti} - \alpha W_i(T_i)\}/\sigma_t$, with $W_i(T_i)$ being a function of b_i, and $\parallel \cdot \parallel$ denotes the Euclidean vector norm. Furthermore, we make the local independence assumption that is, the random effects are assumed to account for all the correlations between the longitudinal repeated measurements, i.e., $p(y_i \mid b_i; \theta) = \prod_j p\{y_i(t_{ij}) \mid b_i; \theta\}$. Extensions to more complex error structures for the longitudinal outcome, by e.g., including serial correlation terms, are straightforward but are not considered here. Finally, the censoring and visiting processes are assumed noninformative given the observed history of longitudinal responses (Tsiatis and Davidian, 2004).

Maximization of the log-likelihood function corresponding to (3) with respect to θ requires a combination of numerical integration and optimization algorithms due to the fact that the (multidimensional) integral with respect to the random effects does not have a closed-form solution. Standard numerical integration techniques such as Gaussian quadrature and Monte Carlo have been successfully applied in the joint modelling framework (Henderson et al., 2000; Wulfsohn and Tsiatis, 1997). Furthermore, Rizopoulos et al. (2009) have recently discussed the use of Laplace approximations for joint models, that can be especially useful in high-dimensional random effects settings. These integration techniques have been traditionally combined with an EM algorithm (treating the random effects as missing data), mainly because of the appealing feature of closed-form M-step updates for certain parameters.

3. Description of Case Studies and Fitted Joint Models

3.1 *AIDS Data Set*

Longitudinal CD4 cell counts measurements and the time to death have been recorded for 467 HIV infected patients who had failed or were intolerant of zidovudine therapy. The aim of this study was to compare the efficacy and safety of two alternative antiretroviral drugs, namely didanosine (ddI) and zalcitabine (ddC). Patients were randomly assigned to receive either ddI or ddC, and CD4 cell counts were recorded at study entry, where randomization took place, as well as 2, 6, 12, and 18 months thereafter. More details about this data set can be found in Guo and Carlin (2004). The Kaplan-Meier estimate for the time to death as well as sample smooth average profiles for the two treatment groups are displayed in Web Figure 1.

Out of the 2335 planned measurements, 1408 were actually recorded, leading to 39.7% intermittent missingness; moreover, until the end of the study 188 patients died, resulting in 59.7% censoring. Taking advantage of the randomization set-up of the study, we fit a joint model in which we only correct for treatment. In particular, for the longitudinal process, we assume a linear mixed model, with main effects for treatment and time and with their interaction for the fixed effects part, and with random intercepts and random slopes for the random effects part. For the survival process, we assume the Weibull model (2), where x_{ti} contains an intercept and the treatment dummy variable:

$$\begin{aligned} y_i(t_{ij}) &= \beta_0 + \beta_1 t_{ij} + \beta_2 \texttt{Treat}_i + \beta_3 \texttt{Treat}_i \times t_{ij} + b_{i0} + b_{i1} t_{ij} + \varepsilon_{yi}(t_{ij}), \\ \log T_i^* &= \gamma_0 + \gamma_1 \texttt{Treat}_i + \alpha W_i(T_i^*) + \sigma_t \varepsilon_{ti}, \end{aligned}$$

where $\texttt{Treat}_i$ denotes the treatment indicator taking the value '1' if the ith subject received ddI, and '0' otherwise. Furthermore, for the random effects covariance matrix D, we set $d_{11} = \text{var}(b_{i0})$, $d_{22} = \text{var}(b_{i1})$, and $d_{12} = \text{cov}(b_{i0}, b_{i1})$. The parameter estimates and standard errors for this joint model can be found in Table 1.

[Table 1 about here.]

3.2 *PBC Data Set*

As a second example, we also consider the primary biliary cirrhosis (PBC) data collected by the Mayo Clinic from 1974 to 1984 (Murtaugh et al., 1994). PBC is a chronic, fatal, but rare liver disease characterized by inflammatory destruction of the small bile ducts within the liver, which eventually leads to cirrhosis of the liver. Patients with PBC have abnormalities in several blood tests, such as elevated levels of serum bilirubin. In this study 312 patients are considered of whom 158 were randomly assigned to receive D-penicillamine and 158 placebo, and we are interested in testing for a treatment effect on survival after adjusting for the longitudinal bilirubin levels. The Kaplan-Meier estimate for the time to death as well as the sample smooth average profiles for the two treatment groups are depicted in Web Figure 1.

Patients did not return to the study centers at prespecified time points to provide serum bilirubin measurements, and thus we observe great variability between their visiting patterns. In particular, patients made on average 6.23 visits ($s.d. = 3.77$), resulting in a total of 1945 observations. The joint model fitted to the PBC data is of the same form as for the AIDS data set presented in the previous section. Now $\texttt{Treat}_i$ takes the value '1' for D-penicillamine, and '0' otherwise. The parameter estimates and standard errors are also shown in Table 1.

4. Calculating Residuals for Joint Models

4.1 *Residuals based on the Observed Data*

A traditional approach to check model assumptions is the inspection of residual plots. Properties and features of residuals, when longitudinal and survival outcomes are separately modelled, have been extensively studied in the literature. For instance, different types of residuals for linear mixed models are discussed in Nobre and Singer (2007) and Verbeke and Molenberghs (2000), whereas residuals for parametric and semiparametric survival models

are presented in Harrell (2001) and Therneau and Grambsch (2000). For our purposes, we consider two types of residuals for each of the two processes. In particular, for the longitudinal process we use the standardized marginal and subject-specific residuals defined as

$$\begin{aligned} r_i^{(ym)} &= \hat{V}_i^{-1/2}(y_i - X_i\hat{\beta}), \text{ and} \\ r_i^{(ys)}(t_{ij}) &= \{y_i(t_{ij}) - x_i^\top(t_{ij})\hat{\beta} - z_i^\top(t_{ij})\hat{b}_i\}/\hat{\sigma}_y, \end{aligned} \tag{4}$$

where $\hat{\beta}$, $\hat{\sigma}_y$, and $\hat{D}$ denote the maximum likelihood estimates under model (1), $\hat{b}_i$ are the empirical Bayes estimates for the random effects, and $\hat{V}_i = Z_i\hat{D}Z_i^\top + \hat{\sigma}_y^2\mathrm{I}$, with I denoting the identity matrix of appropriate dimensions. The marginal residuals $r_i^{(ym)}$ predict the marginal errors $y_i - X_i\beta = Z_ib_i + \varepsilon_{yi}$, and can be used to investigate misspecification of the mean structure $X_i\beta$ as well as to validate the assumptions for the within-subjects covariance structure V_i. The subject-specific residuals $r_i^{(ys)}(t_{ij})$ predict the conditional errors $\varepsilon_{yi}(t_{ij})$, and can be used for checking the homoscedasticity and normality assumptions. For survival models the martingale residuals are commonly used for a direct assessment of excess events (i.e., to reveal subjects that are poorly fit by the model), and for evaluating whether the appropriate functional form for a covariate is used in the model. The accelerated failure time formulation of the Weibull model (2) also allows the calculation of standardized residuals of the form

$$r_i^{(t)} = \{\log T_i - \hat{\gamma}^\top x_{ti} - \hat{\alpha}\hat{W}_i(T_i)\}/\hat{\sigma}_t, \tag{5}$$

where $\hat{\gamma}$, $\hat{\alpha}$, and $\hat{\sigma}_t$ denote the maximum likelihood estimates under model (2). Note that when T_i is right-censored, $r_i^{(t)}$ is also right-censored. Thus, in order to use $r_i^{(t)}$ for practical purposes, censoring must be taken into account by displaying for instance, Kaplan-Meier estimates based on groups of residuals rather than showing individual residuals (Harrell, 2001, Sec. 17.3.5). These residuals can be used to investigate the appropriateness of the assumed parametric survival model, by comparing their Kaplan-Meier estimate with the survival function of the assumed distribution for the error terms ε_{ti}.

The problem in using the above defined residuals for inspecting the fit of joint models is that their reference distribution is not directly evident. Complications arise due to the non-random dropout in the longitudinal process caused by the occurrence of events. That is, the observed data, upon which the residuals are calculated, are not a random sample of the target population. To clarify this, we define for each subject the observed and missing part of the longitudinal response vector. The observed part $y_i^o = \{y_i(t_{ij}) : t_{ij} < T_i, j = 1, \ldots, n_i\}$ contains all observed longitudinal measurements of the ith subject before the observed event time, whereas the missing part $y_i^m = \{y_i(t_{ij}) : t_{ij} \geqslant T_i, j = 1, \ldots, n_i'\}$ contains the longitudinal measurements that would have been taken until the end of the study, had the event not occurred. Under these definitions, the dropout mechanism corresponding to (3) has the form

$$p(T_i^* \mid y_i^o, y_i^m; \theta) = \int p(T_i^* \mid b_i; \theta) p(b_i \mid y_i^o, y_i^m; \theta)\, db_i, \tag{6}$$

which still depends on y_i^m through the posterior distribution $p(b_i \mid y_i^o, y_i^m; \theta)$. It is this feature of joint models that complicates inspection of residual plots, because a potential systematic behaviour is not necessarily indicative of a model misfit. Thus, conclusions from common residual plots in the joint model framework should be drawn with extreme caution.

To depict how the non-random dropout affects the use of residuals based on the observed data alone, we show in Figure 1 plots of the standardized marginal and standardized subject-specific residuals (4) versus the fitted values, for the AIDS and PBC data sets.

[Figure 1 about here.]

We observe that the fitted loess curves in the plots of the standardized marginal residuals versus the fitted values show a systematic trend. Note, however that, small numbers of CD4 cell counts on the one hand and high levels of serum bilirubin on the other, indicate a worsening of patients' condition resulting in higher death rates (i.e., dropout). This is also reflected in the different signs for the association parameter α in the two data sets, presented

in Table 1. Thus, the residuals corresponding to small and large fitted values, respectively, for the two data sets are only based on patients with a 'good' health condition.

Finally, we should mention that, in non-random missing data contexts, the form of the dropout mechanism cannot be verified from the observed data (Molenberghs and Kenward, 2007). That is, we cannot definitively test whether the time to dropout depends on the underlying evolution of the disease (i.e., joint model assumption) or, for instance, it depends on the actual value of the longitudinal outcome at the event time (i.e., selection model assumption). Thus, in order to proceed, we make the assumption that joint model (3) is the correct modelling framework and that the non-random dropout mechanism is of the form (6). Our interest is then focused on investigating whether the formulation of this model (e.g., functional form of covariates, assumptions for the error terms, etc.) is adequate for a specific data set at hand. The effects of misspecifying the missing data mechanism is empirically investigated in Section 6.

4.2 *Multiple Imputation Based Residuals with Fixed Visit Times*

To produce residuals that can be readily used in diagnostic plots for joint models, we propose to augment the observed data with randomly imputed longitudinal responses under the complete data model, corresponding to the longitudinal outcomes that would have been observed had the patients not dropped out. The multiple imputation approach properly accounts for the uncertainty in the imputed values due to missingness. Furthermore, note that in some clinical studies in which the terminating event is death, it may not be conceptually reasonable to consider the values of the longitudinal outcome after the event time; for instance see Kurland and Heagerty (2005). However, in our setting, we merely use multiple imputation as a mechanism to help us investigate the fit of the model, and we are not actually interested in inferences after the event time.

In the following, we assume that the joint model has been fitted to the data set at hand,

and that we have obtained the maximum likelihood estimates $\hat{\theta}$ and an estimate of their asymptotic covariance matrix, say $\hat{\mathcal{H}}$. Moreover, we assume that longitudinal measurements are planned to be taken at prespecified time points $t_0, t_1, \ldots, t_{max}$, and that for the ith subject measurements are available up to the last prespecified visit time earlier than T_i. Since multiple imputation has Bayesian grounds (Little and Rubin, 2002, Ch. 10), we adopt a Bayesian point of view for the joint model, even though we have fitted it using maximum likelihood. The multiple imputation method is based on repeated sampling from the posterior distribution of y_i^m given the observed data, averaged over the parameter values. Under joint model (3) and dropout mechanism (6), the density for this distribution can easily be found to be

$$p(y_i^m \mid y_i^o, T_i, \delta_i) = \int p(y_i^m \mid y_i^o, T_i, \delta_i; \theta) p(\theta \mid y_i^o, T_i, \delta_i)\, d\theta. \tag{7}$$

The first part on the right hand side of (7) can be derived from (6) taking also into account the local independence assumption, i.e.,

$$\begin{aligned} p(y_i^m \mid y_i^o, T_i, \delta_i; \theta) &= \int p(y_i^m \mid b_i, y_i^o, T_i, \delta_i; \theta) p(b_i \mid y_i^o, T_i, \delta_i; \theta)\, db_i \\ &= \int p(y_i^m \mid b_i; \theta) p(b_i \mid y_i^o, T_i, \delta_i; \theta)\, db_i. \end{aligned} \tag{8}$$

For the second part, which is the posterior distribution of the parameters given the observed data, we use arguments of standard asymptotic Bayesian theory (Cox and Hinkley, 1974, pp. 299–300), and assume that the sample size n is sufficiently large such that $\{\theta \mid y_i^o, T_i, \delta_i\}$ can be well approximated by $\mathcal{N}(\hat{\theta}, \hat{\mathcal{H}})$. This assumption, combined with (7) and (8), suggests the following simulation scheme:

Step 1. draw $\theta^{(\ell)} \sim \mathcal{N}(\hat{\theta}, \hat{\mathcal{H}})$.

Step 2. draw $b_i^{(\ell)} \sim \{b_i \mid y_i^o, T_i, \delta_i, \theta^{(\ell)}\}$.

Step 3. draw $y_i^{m(\ell)}(t_{ij}) \sim \mathcal{N}\left\{\mu_i^{(l)}(t_{ij}),\, \hat{\sigma}_y^{2,(\ell)}\right\}$, for the prespecified visit times $t_{ij} \geqslant T_i$, $j =$

$1, \ldots, n_i'$ that were not observed for the ith subject, where $\mu_i^{(l)}(t_{ij}) = x_i^\top(t_{ij})\hat{\beta}^{(\ell)} + z_i^\top(t_{ij})\hat{b}_i^{(\ell)}$.

Step 4. repeat Steps 1–3 for each subject, $\ell = 1, \ldots, L$ times, where L denotes the number of imputations.

Steps 1 and 2 account for uncertainties in the parameter and empirical Bayes estimates, respectively, whereas Step 3 imputes the missing longitudinal responses. Steps 1 and 3 are straightforward to perform since they require sampling from a multivariate normal distribution; on the contrary, the posterior distribution for the random effects in Step 2 is of a non-standard form, and thus a more sophisticated approach is required to sample from it. We propose the use of a Metropolis-Hastings algorithm with independent proposals from a multivariate t distribution centered at the empirical Bayes estimates $\hat{b}_i$, with scale matrix $\text{var}(\hat{b}_i)$, and four degrees of freedom. A similar approach has been used by Booth and Hobert (1999) to simulate from the posterior distribution of the random effects in the generalized linear mixed models context. In the joint modelling framework, our justification for a multivariate t proposal is two fold. First, Rizopoulos et al. (2008) have recently shown that, as n_i increases, the leading term of the log posterior distribution of the random effects is the linear mixed model $\log p(y_i \mid b_i; \theta^{(\ell)}) = \sum_j \log p\{y_i(t_{ij}) \mid b_i; \theta^{(\ell)}\}$, which is quadratic in b_i and will resemble the shape of a multivariate normal distribution, and second, for small n_i, the heavier tails of the t distribution will ensure sufficient coverage.

The simulated $y_i^{m(\ell)}(t_{ij})$ values together with y_i^o can now be used to calculate residuals according to (4). A key advantage of the multiply imputed residuals is that they inherit the properties of the complete data model. This facilitates common graphical model checks, without requiring formal derivation of their reference distribution. In contrast, if we used only the observed residuals, as in the approach of DH, then it is required to explicitly compute characteristics of their distribution. For instance, DH computed the first two moments of the

marginal residuals $r_i^{(ym)}$ conditional on the dropout time and event status. Finally, regarding the choice of L, we expect that a moderate number of multiple imputations, say between 10 and 100, will be sufficient for the propagation of uncertainty.

4.3 *Multiple Imputation Based Residuals with Random Visit Times*

In observational studies and in some randomized trials (such as the PBC study), the time points at which the longitudinal measurements are taken are not fixed by design but rather determined by the physician or even the patients themselves. For instance, for the PBC data set, the patients' visit patterns are illustrated in the bottom-left panel of Web Figure 1. We observe that for the first two years of follow-up, measurements of serum bilirubin are taken at baseline, 0.5, 1, and 2 years, with little variability, whereas, in the latter years the variability in the visit times increases considerably. Under the noninformativeness assumption mentioned in Section 2, and provided that the joint model is correctly specified, the visiting process can be ignored in the modeling process without influencing the asymptotic properties of the maximum likelihood estimates.

However, the possibility of random visit times complicates the methodology presented in Section 4.2. In particular, the time points at which the ith subject was supposed to provide measurements after the observed event time T_i are not available, and thus the corresponding rows $x_i^\top(t_{ij})$ and $z_i^\top(t_{ij})$, for $t_{ij} \geqslant T_i$, of the design matrices X_i and Z_i, respectively, required in Step 3, cannot be computed. In addition, a 'simplistic' approach of imputing y_i^m at arbitrary specified fixed time points may contaminate the residuals plots by producing either too many or too few, say positive, residuals in areas with few observed data. An example can be found in Web Section 2.

To overcome this problem and use the multiple imputation idea introduced in Section 4.2, we propose postulating a suitable model for the visiting process, and use it to simulate future visit times for each individual. Formally, let u_{ik} $(k = 2, \ldots, n_i)$ denote the time elapsed

between visit $k-1$ and visit k for the ith subject, and without loss of generality assume that all subjects have at least one measurement. Let also $\{y_i^*(t)\}$ denote the complete version of the longitudinal response vector. Using these definitions, the noninformativeness assumption for the visiting process can be formulated as

$$p(u_{ik} \mid u_{i2}, \ldots, u_{i,k-1}, \{y_i^*(t)\}; \theta_v) = p\{u_{ik} \mid u_{i2}, \ldots, u_{i,k-1}, y_i(t_1), \ldots, y_i(t_{k-1}); \theta_v\}, \qquad (9)$$

where θ_v is the vector parameterizing the visiting process density, and $\{\theta, \theta_v\}$ have disjoint parameter spaces. For the multivariate elapsed visit times $u_i^\top = (u_{i2}, \ldots, u_{in_i})$, we propose a Weibull model with a multiplicative Gamma frailty

$$\lambda(u_{ik} \mid x_{vi}, \omega_i) = \lambda_0(u_{ik})\omega_i \exp(x_{vi}^\top \beta_v), \;\; \omega_i \sim \text{Gamma}(\eta, \eta), \qquad (10)$$

where $\lambda(\cdot)$ is the risk function conditional on the frailty term ω_i, x_{vi} denotes the covariate vector that may contain a functional form of the observed longitudinal responses $y_i(t_{i1}), \ldots, y_i(t_{i,k-1})$, β_v is the vector of regression coefficients, and η^{-1} is the unknown variance of ω_i's. The Weibull baseline risk function is given by $\lambda_0(u_{ik}) = \phi\psi u_{ik}^{\psi-1}$, with $\psi, \phi > 0$. Our choice for this model is motivated, not only by its flexibility and simplicity, but also by the fact that the posterior distribution of the frailty term, given the observed data, is of standard form (Sahu et al., 1997), which as will be shown below facilitates simulation.

Similarly to Section 4.2, we assume that both models (3) and (10) have been fitted to the data at hand, and that the maximum likelihood estimates $\hat{\theta}$ and $\hat{\theta}_v$, and their corresponding asymptotic covariance matrices, $\hat{\mathcal{H}}$ and $\hat{\mathcal{H}}_v$, respectively, have been obtained. Let also t_{max} denote the end of the study, and $\delta_{v,ik}$ the event indicator corresponding to u_{ik}. Furthermore, taking into consideration the noninformativeness assumption (9), the future elapsed visit time u_{i,n_i+1} can be simulated independently from $y_i^m(t_{i,n_i+1})$. Thus, the simulation scheme under the random visit times setting takes the following form:

Step 1. Parameter Values

a. draw $\theta_v^{(\ell)} \sim \mathcal{N}(\hat{\theta}_v, \hat{\mathcal{H}}_v)$.

b. draw $\theta^{(\ell)} \sim \mathcal{N}(\hat{\theta}, \hat{\mathcal{H}})$.

Step 2. Frailties and Random Effects

a. draw $\omega_i^{(\ell)} \sim \text{Gamma}\Big\{\eta^{(\ell)} + \sum_{k=2}^{n_i} \delta_{v,ik},\ \eta^{(\ell)} + \phi^{(\ell)} \sum_{k=2}^{n_i} u_{ik}^{\psi^{(\ell)}} \exp(x_{vi}^\top \beta_v^{(\ell)})\Big\}$ for subjects with two or more visits, and $\omega_i^{(\ell)} \sim \text{Gamma}(\eta^{(\ell)}, \eta^{(\ell)})$ for subjects with one visit.

b. draw $b_i^{(\ell)} \sim \{b_i \mid y_i^o, T_i, \delta_i, \theta^{(\ell)}\}$.

Step 3. Outcomes

a. draw $u_i^{(\ell)} \sim \text{Weibull}\Big\{\psi^{(\ell)},\ \phi^{(\ell)} \omega_i^{(\ell)} \exp(x_{vi}^\top \beta_v^{(\ell)})\Big\}$.

b. set $\tilde{t}_i = u_i^{(\ell)} + t_{in_i}$, where t_{in_i} denotes the last observed visit time for the ith subject. If $\tilde{t}_i > t_{max}$, no y_i^m need to be imputed for this subject; otherwise draw $y_i^{m(\ell)}(\tilde{t}_i) \sim \mathcal{N}\Big\{\mu_i^{(l)}(\tilde{t}_i),\ \hat{\sigma}_y^{2,(\ell)}\Big\}$, where $\mu_i^{(l)}(\tilde{t}_i) = x_i^\top(\tilde{t}_i)\hat{\beta}^{(\ell)} + z_i^\top(\tilde{t}_i)\hat{b}_i^{(\ell)}$.

c. set $t_{in_i} = \tilde{t}_i$, and repeat a–b until $t_{in_i} > t_{max}$ for all i.

Step 4. Repeat Steps 1–3 for $\ell = 1, \ldots, L$ times.

As in Section 4.2, Steps 1–3 simultaneously account for uncertainties in both the joint and visiting process models. Furthermore, note that subjects who have only one longitudinal measurement provide no information to the visiting process model. For these cases, in Step 3a, we can only simulate future elapsed visit times using a simulated frailty value from the Gamma prior distribution (Step 2a).

The form of the linear predictor of the visiting model can have an effect on the simulated future visit times for each subject. Therefore, we would like to note that assumption (9) is the weakest assumption under which the joint model provides valid inferences even if the visiting process is ignored; however, the visiting model corresponding to (9) may be unstable because it involves many parameters. A set of stronger but maybe more plausible assumptions is

$$p(u_{ik} \mid u_{i2}, \ldots, u_{i,k-1}, \{y_i^*(t)\}; \theta_v) = p\{u_{ik} \mid u_{i2}, \ldots, u_{i,k-1}, y_i(t_{k-1}); \theta_v\}, \tag{11}$$

or

$$p(u_{ik} \mid u_{i2}, \ldots, u_{i,k-1}, \{y_i^*(t)\}; \theta_v) = p\{u_{ik} \mid y_i(t_{k-1}); \theta_v\}. \quad (12)$$

Equation (11) posits that the time elapsed between visit $k-1$ and visit k depends on the previous elapsed times and the last observed longitudinal measurement, whereas under (12) it depends only on the last observed longitudinal measurement. These assumptions describe the situation in which physicians base their decision for a future visit for a patient on the last observed outcome and possibly the past visiting pattern.

5. MI Based Residuals for the AIDS and PBC Data

Based on the fitted joint models presented in Section 3, we simulated $y_i^m(t_{ij})$ values for each of the AIDS and PBC data sets. The implementation of the multiple imputation scheme for the PBC data set is based on the visiting process model (10) with linear predictor $x_{vi}^\top \beta_v = \beta_{v1} \texttt{Treat}_i + \beta_{v1} y_i(t_{k-1})$, which corresponds to assumption (11). The parameter estimates and standard errors for this model can be found in Web Table 1. For the AIDS data we use $L = 50$, and for the PBC data $L = 10$. The reason for using less imputations for the PBC data set is that, as we observed, in each imputation many future visit times are simulated for subjects that dropped out quite early. This inevitably results in busy residual plots, and should we have used $L = 50$ for this case as well, then their usefulness would be compromised.

As we argued in Section 4.2, the multiply imputed residuals combined with the residuals corresponding to y_i^o can be used to investigate the validity of the underlying assumptions of the complete data model. For instance, in Web Figure 2 we present the plots of standardized marginal and standardized subject-specific residuals versus the fitted values for the first imputation, for the two data sets. From these plots, we observe that the systematic trends that were present in the residual plots based on the observed data alone (i.e., Figure 1) are

alleviated. However, it might prove difficult in some cases to extract conclusions by examining the residual plots for each imputation separately. This can also be seen in Web Figure 3, which shows that the variability due to missingness in the loess smoother can be considerable in the areas with few observed data. To overcome this problem, we propose to include in one plot the residuals from all imputations. Then we can check for systematic trends using weighted loess fits, with weight one for the observed residuals, and $1/L$ for the imputed ones. This approach is illustrated in Figures 2 and 3 for the AIDS and PBC data, respectively.

[Figure 2 about here.]

[Figure 3 about here.]

These plots corroborate the conclusions made using the first imputation alone. However, Figure 3 also reveals that the variability in the standardized subject-specific residuals increases with the fitted values. This feature is more clearly illustrated in Web Figure 4, which depicts the square root of the absolute residuals versus the fitted values. Note that this issue is only revealed by an inspection of the multiply imputed residuals, since the residuals for the observed data show constant variance. Based on this finding, we could suggest that a possible extension of the joint model for the PBC data set is to consider heteroscedastic error terms $\varepsilon_{yi}(t_{ij})$.

Figures 2 and 3 also include residual plots for the survival submodel. In particular, we present Kaplan-Meier estimates per treatment group for the standardized accelerated failure time residuals (5), and the scatterplots of the martingale residuals versus the fitted values of the longitudinal outcomes evaluated at the observed event times. We observe that, for both data sets, the Kaplan-Meier estimates of the accelerated failure time residuals are in close agreement with the survival function of the standard extreme value distribution. This suggests that (2) is a suitable model for the time to death for both studies. Furthermore, the martingale residuals also show that the assumed relation between the longitudinal time-

dependent covariate and the hazard function is adequate, since the loess smoother does not show severe discrepancies from zero.

Finally, we would like to note that the methodology presented in Sections 4.2 and 4.3 can be directly used to perform posterior predictive checks as have been formalized by Gelman et al. (2005). We applied this procedure to the AIDS data set. In particular, 60 data sets are simulated from the fitted joint model, and time-specific samples averages are compared to the averages obtained from augmenting the observed data using the multiple imputations $y_i^{m(\ell)}$. The results are shown in Web Figure 5 for the two treatment groups separately. Both plots suggest that the posited joint model is in agreement with the AIDS data set. Moreover, the comparison of the simulated data from the fitted joint model (dashed grey lines) with the sample averages based on the observed data alone (dashed black line) elucidates why plots based on only the observed data can lead to misleading conclusions.

A similar to the posterior predictive checks approach, suggested by a referee, is to simulate the reference distribution of the observed data residuals. In particular, with this approach we have two options. First, based on $\hat{\theta}$, complete data sets can be simulated based on which we can compute the *error terms* of the longitudinal outcomes that are measured before the simulated event times. Second, for the simulated data sets of the first option we can fit the joint model in order to form *residuals* for the longitudinal outcomes that are measured before the simulated event times. These simulated error terms or residuals can then be directly compared with the observed residuals corresponding to the original data set. To explore this idea, we simulated 200 data sets based on the maximum likelihood estimates and the structure of the AIDS data set, and then calculated the standardized subject-specific and standardized marginal residuals for the observed part of the simulated longitudinal responses. Web Figure 6 illustrates Q-Q plots of the simulated residuals versus the observed ones from the AIDS data set as well as Q-Q plots of the simulated error terms versus the observed

residuals. From these plots we observe that standardized marginal residuals for the AIDS data set lie within the envelope of their empirical distribution, whereas the distribution of the observed standardized subject-specific residuals seems to have longer tails compared to the empirical reference distribution. Regarding the use of the simulated distribution of the error terms versus the one of the residuals, we should note that, from a statistical point of view, the latter is more appropriate because in the calculation of the residual terms we also take into account the variability in $\hat{\theta}$ (see also Section 7). This difference is also apparent from the comparison between the simulated distributions of the subject-specific error terms and the subject-specific residuals, presented in the top-left and bottom-left panels of Web Figure 6, respectively. Therefore, the appealing feature of this approach is that it directly checks assumptions about the observed residuals that are generally much easier to compute. However, as mentioned above, to simulate the empirical distribution for the observed residuals (that are more appropriate than the error terms), it is required that we fit the joint model for each simulated data set to form residuals that hinders the practicality of this method, especially for large data sets.

6. Simulation Study

Since in a non-random dropout context the joint modelling assumption (6) cannot be verified from the observed data, we have performed a number of simulations in order to empirically evaluate the performance of the proposed multiple-imputation-based residuals, especially in the case of misspecification. In particular, 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 (1) and (2). Second, we considered misspecification of the missing data mechanism (6) by positioning joint models as a special case of the general selection modelling framework formulated as

$$p(T_i^*, y_i^o, y_i^m, b_i; \theta) = p(T_i^* \mid y_i^o, y_i^m, b_i; \theta) p(y_i^o, y_i^m \mid b_i; \theta) p(b_i; \theta),$$

i.e., the event time T_i^* could depend on y_i^m and/or b_i. In each case, residuals were calculated based on a misspecified joint model and plots were produced to check for systematic behavior. A detailed description of the set up of the simulation study as well as a discussion of the results can be found in Web Section 3. 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.

7. Discussion

We have proposed a new approach for calculating residuals for joint models of longitudinal and survival data based on multiple imputation. A key advantage of this method is that it requires simple simulation steps that can be easily performed, using the components of the fitted joint model. Moreover, the practical use of our proposals can be directly explored using the publicly available R package `JM` that was used to fit the joint models considered in Section 3, and to compute the residuals presented in Section 4.

Even though we have focused on linear mixed models for the longitudinal responses and a Weibull survival model for the dropout process, the proposed method can easily be extended to other types of joint models. For instance, the Weibull assumption in (2) can be relaxed by

assuming more flexible models. In particular, the methodology developed in Section 4 has also been implemented in `JM` for a relative risk model, that is related to the time-varying Cox model, and in which the covariates affect linearly the log cumulative hazard ratio. In order to allow for flexibility, the log cumulative baseline risk function $\log H_0(t) = \log \int_0^t h_0(s)\, ds$ is expanded into B-spline basis functions, i.e.,

$$\begin{aligned} \log H(t \mid b_i) &= \log H_0(t) + \gamma^\top x_{ti} + \alpha W_i(t), \\ \log H_0(t) &= \kappa_0 + \sum_{d=1}^{m} \kappa_d B_d(\log t, q), \end{aligned}$$

where $\kappa^\top = (\kappa_0, \kappa_1, \ldots, \kappa_m)$ are the spline coefficients, q denotes the degree of the B-splines basis functions $B(\cdot)$, and $m = \ddot{m} + q - 1$, with $\ddot{m}$ denoting the number of interior knots. More information regarding this model can been found in Rizopoulos et al. (2009) as well as in the manual of `JM`.

Another issue that we have not considered in this paper is the studentization of residuals (4) and (5). That is, we have assumed that the variance of $r_i^{(ym)}$, $r_i^{(ys)}(t_{ij})$, and $r_i^{(t)}$ is the same as the variance of the corresponding error terms of submodels (1) and (2). This assumption, although commonly made for the residuals of a variety of statistical models, is unfortunately not correct. Under our approach, the residuals corresponding to the observed data, augmented with the multiply imputed residuals, inherit the properties of the complete data model, which implies that we may use known results for studentization of residuals, ignoring the dropout process. For example, Nobre and Singer (2007) present formulas for calculating studentized subject-specific residuals for linear mixed models. However, under the joint model (3), it is difficult to apply these formulas, because we do not have closed-form solutions for $\hat{\theta}$, and thus $\text{cov}(y_i, \hat{\theta})$ and $\text{cov}(\{T_i, \delta_i\}, \hat{\theta})$ cannot be easily derived. A possible solution to this issue could be to adapt the developments of Cox and Snell (1968) to the joint modelling framework.

Acknowledgements

The authors gratefully acknowledge support from the IAP research network grant nr. P6/03 of the Belgian government (Belgian Science Policy).

Supplementary Materials

Web Figure 1 referenced in Sections 3 and 4.3, Web Table 1 and Web Figures 2 to 6, referenced in Section 5, Web Section 2 referenced in Section 4.3, and Web Section 3 referenced in Section 6, as well as the R programs used for the analyses presented in this paper are available as Supplementary Material.

References

Booth, J. and Hobert, J. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society, Series B* **61,** 265–285.

Cox, D. and Hinkley, D. (1974). *Theoretical Statistics*. Chapman & Hall, London.

Cox, D. and Snell, E. (1968). A general definition of residuals. *Journal of the Royal Statistical Society, Series B* **30,** 248–275.

Ding, J. and Wang, J.-L. (2008). Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. *Biometrics* **64,** 546–556.

Dobson, A. and Henderson, R. (2003). Diagnostics for joint longitudinal and dropout time modeling. *Biometrics* **59,** 741–751.

Elashoff, R., Li, G., and Li, N. (2008). A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics* **64,** 762–771.

Gelman, A., Van Mechelen, I., Verbeke, G., Heitjan, D., and Meulders, M. (2005). Multiple imputation for model checking: completed-data plots with missing and latent data. *Biometrics* **61,** 74–85.

Guo, X. and Carlin, B. (2004). Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician* **58,** 16–24.

Harrell, F. (2001). *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer-Verlag, New York.

Henderson, R., Diggle, P., and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1,** 465–480.

Hsieh, F., Tseng, Y.-K., and Wang, J.-L. (2006). Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics* **62,** 1037–1043.

Kurland, B. and Heagerty, P. (2005). Directly parameterized regression conditioning on being alive: analysis of longitudinal data truncated by deaths. *Biostatistics* **6,** 241–258.

Larsen, K. (2004). Joint analysis of time-to-event and multiple binary indicators of latent classes. *Biometrics* **60,** 85–92.

Lin, H., Turnbull, B., McCulloch, C., and Slate, E. (2002). Laten class models for joint analysis of longitudinal biomarker and event process data: application to longitudinal prostate-specific antigen readings and prostate cancer. *Journal of the American Statistical Association* **97,** 53–65.

Little, R. (1995). Modeling the dropout mechanism in repeated measures studies. *Journal of the American Statistical Association* **90,** 1112–1121.

Little, R. and Rubin, D. (2002). *Statistical Analysis with Missing Data*. Wiley, New York, 2nd edition.

Liu, L., Huang, X., and O'Quigley, J. (2008). Analysis of longitudinal data in the presence of informative observational times and a dependent terminal event, with application to medical cost data. *Biometrics* **64,** 950–958.

Molenberghs, G. and Kenward, M. (2007). *Missing Data in Clinical Studies*. Wiley, New York.

Murtaugh, P., Dickson, E., Van Dam, G., Malincho, M., Grambsch, P., Langworthy, A., and Gips, C. (1994). Primary biliary cirrhosis: prediction of short-term survival based on repeated patient visits. *Hepatology* **20,** 126–134.

Nobre, J. and Singer, J. (2007). Residuals analysis for linear mixed models. *Biometrical Journal* **6,** 863–875.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Rizopoulos, D., Verbeke, G., and Lesaffre, E. (2009). Fully exponential laplace approximations for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society, Series B* , to appear.

Rizopoulos, D., Verbeke, G., and Molenberghs, G. (2008). Shared parameter models under random effects misspecification. *Biometrika* **95,** 63–74.

Sahu, S., Dey, D., Aslanidou, H., and Sinha, D. (1997). A Weibull regression model with gamma frailties for multivariate survival data. *Lifetime Data Analysis* **3,** 123–137.

Therneau, T. and Grambsch, P. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag, New York.

Tsiatis, A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica* **14,** 809–834.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer-Verlag, New York.

Verbeke, G., Molenberghs, G., and Beunckens, C. (2008). Formal and informal model selection with incomplete data. *Statistical Science* **23,** 201–218.

Wang, Y. and Taylor, J. (2001). Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association* **96,** 895–905.

Wulfsohn, M. and Tsiatis, A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53,** 330–339.

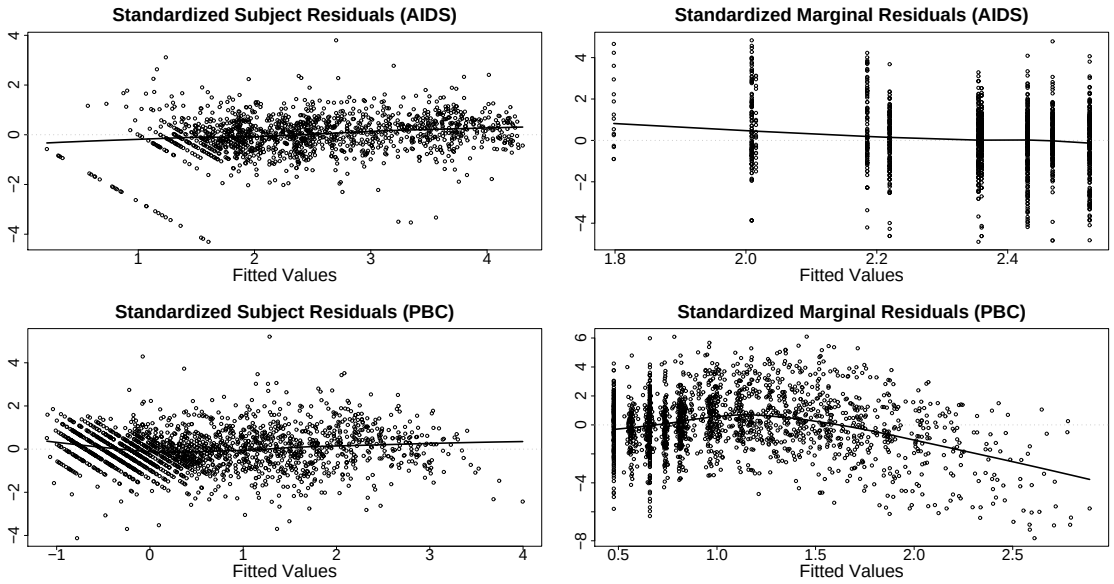

Figure 1. Standardized marginal and subject-specific residuals for the AIDS and PBC data set based on the observed data only. The superimposed lines represent the fit of the loess smoother.

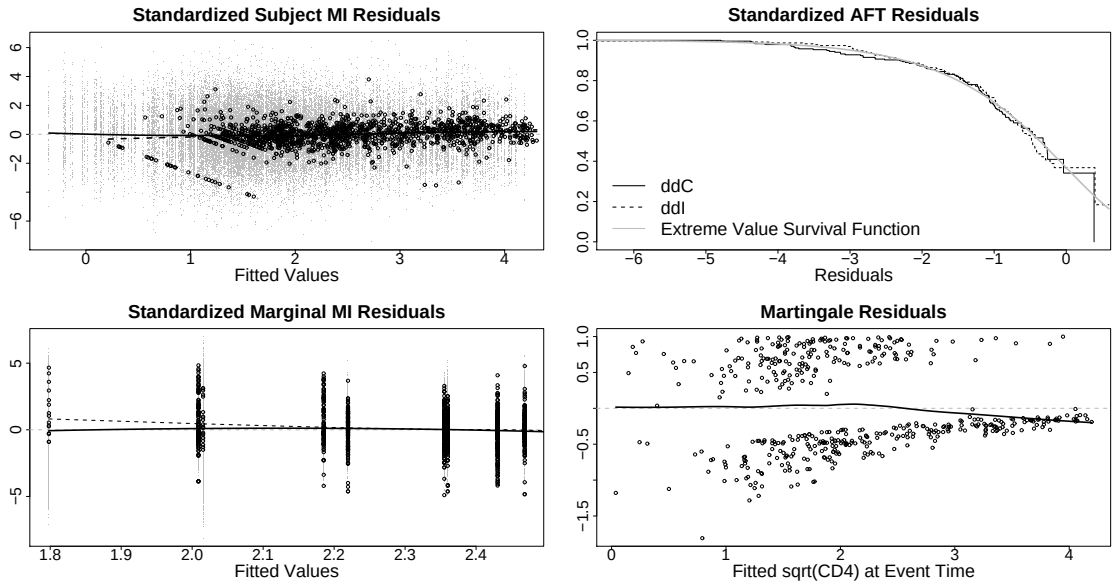

Figure 2. Residual plots for the AIDS data set. The top and bottom left panels show the observed standardized subject-specific and standardized marginal residuals (black circles), augmented with all the multiply imputed residuals produced by the $L = 50$ imputations (grey points). The superimposed dashed lines represent a loess fit based only on the observed residuals; the superimposed solid lines represent a weighted loess fit based on all residuals. The top right panel shows Kaplan-Meier estimates of the accelerated failure time residuals (5) for each treatment group; the superimposed solid grey line is the survival function of the standard extreme value distribution. The bottom right panel shows martingale residuals versus the fitted values of $\sqrt{\text{CD4}}$ evaluated at the observed event times.

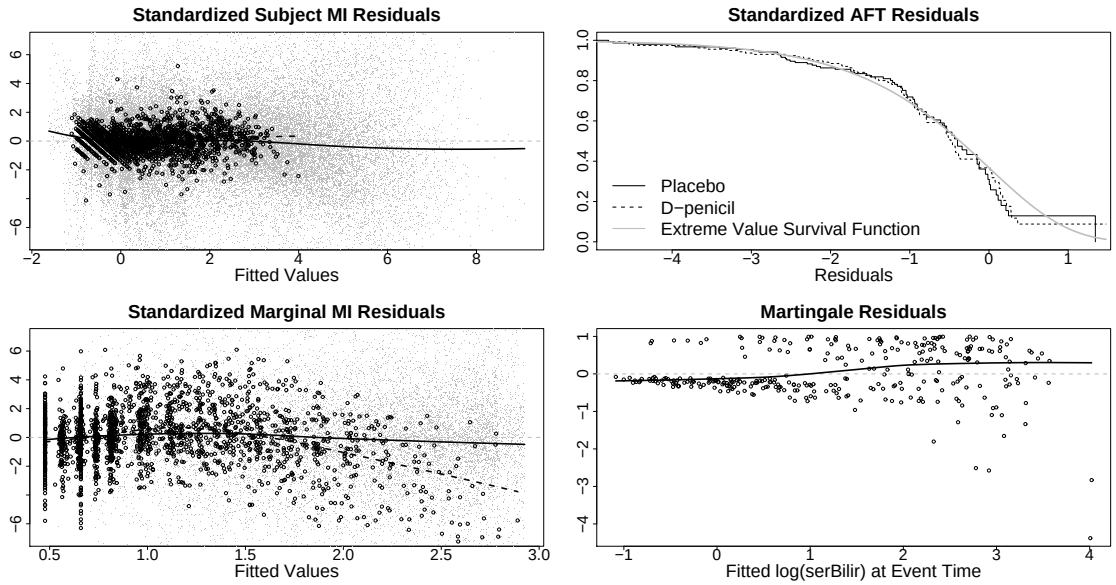

Figure 3. Residual plots for the PBC data set. The top and bottom left panels show the observed standardized subject-specific and standardized marginal residuals (black circles), augmented with all the multiply imputed residuals produced by the $L = 10$ imputations (grey points). The superimposed dashed lines represent a loess fit based only on the observed residuals; the superimposed solid lines represent a weighted loess fit based on all residuals. The top right panel shows Kaplan-Meier estimates of the accelerated failure time residuals (5) for each treatment group; the superimposed solid grey line is the survival function of the standard extreme value distribution. The bottom right panel shows martingale residuals versus the fitted values of log serum bilirubin levels evaluated at the observed event times.

Table 1
Parameter estimates and standard errors (in parenthesis) for the joint models fitted to the AIDS and PBC data sets, respectively.

	Longitudinal Process			Survival Process	
	AIDS	PBC		AIDS	PBC
β_0	2.430 (0.052)	0.660 (0.037)	γ_0	2.548 (0.145)	3.296 (0.177)
β_1	−0.035 (0.005)	0.151 (0.006)	γ_1	−0.233 (0.110)	0.054 (0.146)
β_2	0.095 (0.080)	−0.183 (0.065)	α	0.407 (0.075)	−0.514 (0.067)
β_3	0.007 (0.006)	0.023 (0.016)	$\log \sigma_t$	−0.332 (0.067)	−0.178 (0.073)
$\log \sigma_y$	−0.988 (0.029)	−0.998 (0.018)			
d_{11}	0.077 (0.037)	1.025 (0.043)			
d_{12}	0.000 (0.003)	0.066 (0.013)			
d_{22}	0.001 (0.124)	0.027 (0.076)			