

Joint Modelling of Longitudinal and Time-to-Event Data: Challenges and Future Directions

Dimitris Rizopoulos

Abstract In longitudinal studies measurements are often collected on different types of outcomes for each subject. These may include several longitudinally measured responses (such as blood values relevant to the medical condition under study) and the time at which an event of particular interest occurs (e.g., death, development of a disease or dropout from the study). These outcomes are often separately analyzed; however, in many instances, a joint modeling approach is either required or may produce a better insight into the mechanisms that underlie the phenomenon under study. In this paper we provide a general overview of the joint modelling framework, discuss its main features and we refer to future directions.

Key words: dropout, longitudinal data analysis, missing data, shared parameter models, survival analysis, time-dependent covariates

1 Introduction

Longitudinal studies often produce two types of outcome, namely a set of longitudinal response measurements and the time to an event of interest, such as death, development of a disease or dropout from the study. Two typical examples of this setting are HIV and cancer studies. In HIV studies patients who have been infected are monitored until they develop AIDS or die, and they are regularly measured for the condition of the immune system using markers such as the CD4 lymphocyte count or the estimated viral load. Similarly in cancer studies the event outcome is the death or metastasis and patients also provide longitudinal measurements of antibody levels or of other markers of carcinogenesis, such as the PSA levels for prostate cancer.

Dimitris Rizopoulos
Department of Biostatistics, Erasmus University Medical Center, PO Box 2040, 3000 CA Rotterdam, the Netherlands, e-mail: d.rizopoulos@erasmusmc.nl

These two outcomes are often separately analyzed using a mixed effects model for the longitudinal outcome and a survival model for the event outcome. However, in mainly two settings a joint modelling approach is required. First, when interest is on the event outcome and we wish to account for the effect of the longitudinal outcome as a time-dependent covariate, traditional approaches for analyzing time-to-event data (such as the partial likelihood for the Cox proportional hazards models) are not applicable. In particular, standard time-to-event models require that time-dependent covariates are external; that is, the value of this covariate at time point t is not affected by the occurrence of an event at time point u , with $t > u$ [11, Sec. 6.3]. However, the type of time-dependent covariates encountered in longitudinal studies do not satisfy this condition, due to the fact that they are the output of a stochastic process generated by the subject, which is directly related to the failure mechanism. Therefore, in order to produce valid inferences a model for the joint distribution of the longitudinal and survival outcomes is required instead. The second setting in which joint modelling is required is when interest is on the longitudinal outcome. In this case the occurrence of events causes dropout since no longitudinal measurements are available at and after the event time. When this dropout is non-random (i.e., when the probability of dropout depends on unobserved longitudinal responses), then bias may arise from an analysis that ignores the dropout process [13, Ch. 15]. To avoid this problem and obtain valid inferences the longitudinal and dropout process must be jointly modelled. One of the modelling frameworks that have been proposed in the missing data literature for handling nonrandom dropout is the shared parameter model [8], which postulates a mixed effects model for the longitudinal outcome and time-to-dropout model for the missingness process. In both settings the joint distribution of the event times and the longitudinal measurements is modelled via a set of random effects that are assumed to account for the associations between these two outcomes. Excellent overviews of this area of biostatistics and statistics research are given by [25] and [30].

In this paper we present the basics of the joint modelling framework, we refer to the key features and assumptions behind these models, and we provide some insight regarding future directions. To illustrate the virtues of joint modelling, we use a real data set concerning patients with primary biliary cirrhosis. The main interest lies in investigating the strength of the association between the longitudinally measured marker serum bilirubin and the time-to-death.

2 Joint Modelling Framework

2.1 Submodels Specification

In the following we will present the joint modelling framework motivated by the time-to-event point of view (i.e., in the setting in which we want to incorporate a time-dependent covariate measured with error in a survival model) – a more direct

connection with the missing data framework is made in Section 3. Let T_i denote the observed failure time for the i th subject ($i = 1, \dots, n$), which is taken as the minimum of the true event time T_i^* and the censoring time C_i , i.e., $T_i = \min(T_i^*, C_i)$. Furthermore, we define the event indicator as $\delta_i = I(T_i^* \leq C_i)$, where $I(\cdot)$ is the indicator function that takes the value 1 if the condition $T_i^* \leq C_i$ is satisfied, and 0 otherwise. Thus, the observed data for the time-to-event outcome consist of the pairs $\{(T_i, \delta_i), i = 1, \dots, n\}$. For the longitudinal responses, let $y_i(t)$ denote the value of the longitudinal outcome at time point t for the i th subject. We should note here that we do not actually observe $y_i(t)$ at all time points but only at the very specific occasions t_{ij} at which measurements were taken. Thus, the observed longitudinal data consist of the measurements $y_{ij} = \{y_i(t_{ij}), j = 1, \dots, n_i\}$.

Our aim is to associate the true and unobserved value of the longitudinal outcome at time t , denoted by $m_i(t)$, with the event outcome T_i^* . Note that $m_i(t)$ is different from $y_i(t)$, with the latter being the contaminated with measurement error value of the longitudinal outcome at time t . To quantify the effect of $m_i(t)$ on the risk for an event, a standard option is to use a relative risk model of the form [22]:

$$\begin{aligned} h_i(t \mid \mathcal{M}_i(t), w_i) &= \lim_{dt \rightarrow 0} dt^{-1} \Pr\{t \leq T_i^* < t + dt \mid T_i^* \geq t, \mathcal{M}_i(t), w_i\} \\ &= h_0(t) \exp\{\gamma^\top w_i + \alpha m_i(t)\}, \end{aligned} \quad (1)$$

where $\mathcal{M}_i(t) = \{m_i(u), 0 \leq u < t\}$ denotes the history of the true unobserved longitudinal process up to time point t , $h_0(\cdot)$ denotes the baseline risk function, and w_i is a vector of baseline covariates (such as a treatment indicator, history of diseases, etc.) with a corresponding vector of regression coefficients γ . Similarly, parameter α quantifies the effect of the underlying longitudinal outcome to the risk for an event; for instance, in the AIDS example mentioned in Section 1, α measures the effect of the number of CD4 cells to the risk for death. To avoid the impact of parametric assumptions, the baseline risk function $h_0(\cdot)$ is typically left unspecified. However, within the joint modelling framework [10] have recently noted that leaving this function completely unspecified leads to an underestimation of the standard errors of the parameter estimates. In particular, problems arise due to the fact that the nonparametric maximum likelihood estimate for this function cannot be obtained explicitly under the full joint modelling approach. To avoid this problem, we could either opt for the hazard function of a standard survival distribution (such as the Weibull or Gamma) or for more flexible models in which $h_0(t)$ is sufficiently approximated using step functions or spline-based approaches.

An alternative modelling framework for event time data, especially when the proportionality assumption in (1) fails, is the accelerated failure time model. In order to incorporate a time-dependent covariate within this framework, we let $\mathcal{S}_0$ denote an absolutely continuous baseline survival function, and we follow the formulation of [4, Sec. 5.2] that postulates

$$\left\{ \int_0^{T^*} \exp\{\gamma^\top w + \alpha m(s)\} ds \right\} \sim \mathcal{S}_0.$$

This can be reexpressed in terms of the risk rate function for subject i as:

$$h_i(t | \mathcal{M}_i(t), w_i) = h_0\{V_i(t)\} \exp\{\gamma^\top w_i + \alpha m_i(t)\}, \quad (2)$$

with

$$V_i(t) = \int_0^t \exp\{\gamma^\top w_i + \alpha m_i(s)\} ds.$$

Similarly, to (1), the baseline risk function $h_0(\cdot)$ can be assumed to be of a specific parametric form or modelled flexibly. An important difference of (2) compared to (1) is that in the former the entire covariate history $\mathcal{M}_i(t)$ is assumed to influence the subject-specific risk, whereas in the latter the subject-specific risk depends only on the current value of the time-dependent covariate $m_i(t)$. Within the joint modelling framework accelerated failure time models have been discussed by [23].

In the definitions of the survival models presented above we used $m_i(t)$ to denote the true value of the underlying longitudinal covariate at time point t . However and as mentioned earlier, longitudinal information is actually collected intermittently and with error at a set of few time points t_{ij} for each subject. Therefore, in order to measure the effect of this covariate to the risk for an event we need to estimate $m_i(t)$ and successfully reconstruct the complete longitudinal history $\mathcal{M}_i(t)$, using the available measurements $y_{ij} = \{y_i(t_{ij}), j = 1, \dots, n_i\}$ of each subject and a set of modelling assumptions. For the remaining of this paper we will focus on normal data and we will postulate a linear mixed effects model to describe the subject-specific longitudinal evolutions. In particular, we have

$$\begin{aligned} y_i(t) &= m_i(t) + \varepsilon_i(t) \\ &= x_i^\top(t)\beta + z_i^\top(t)b_i + \varepsilon_i(t), \quad \varepsilon_i(t) \sim \mathcal{N}(0, \sigma^2), \end{aligned} \quad (3)$$

where β denotes the vector of the unknown fixed effects parameters, b_i denotes a vector of random effects, $x_i(t)$ and $z_i(t)$ denote row vectors of the design matrices for the fixed and random effects, respectively, and $\varepsilon_i(t)$ is the measurement error term, which is assumed independent of b_i , and with variance σ^2 . We should note that care should be taken in the specification of $x_i(t)$ and $z_i(t)$ in order to produce a good estimate of $\mathcal{M}_i(t)$. The main reason for this is that, as we will see later in Section 2.2, in the definition of the likelihood of the joint model the complete longitudinal history is required for the computation of the survival function, and of the risk function under the accelerated failure time formulation. Therefore, in applications in which subjects show highly nonlinear longitudinal trajectories, it is advisable to consider flexible representations for $x_i(t)$ and $z_i(t)$ using a possibly high-dimensional vector of functions of time t , expressed in terms of high-order polynomials or splines [2, 5]. Finally, in order to complete the specification of the longitudinal submodel a suitable distributional assumption for the random effects component is required. A standard choice for this distribution is the multivariate normal distribution; however, within the joint modelling framework and mainly because of the nonrandom dropout (see also Section 3), there is the concern that relying on a standard parametric distribution may influence the derived inferences especially when this distribu-

tion differs considerably from the true random effects distribution. This motivated [21] to propose a more flexible model for the distribution of the random effects that is expressed as a normal density times a polynomial function. However, the findings of these authors suggested that the parameter estimates and standard errors of joint models fitted under the normal assumption for the random effects were rather robust to misspecification. This feature has been further theoretically corroborated by [18], who showed that as the number of repeated measurements per subject n_i increases, a misspecification of the random effects distribution has a minimal effect in parameter estimators and standard errors. Thus, here we will assume that $b_i \sim \mathcal{N}(0, D)$ and we will not further investigate this assumption.

2.2 Maximum Likelihood Estimation

The main estimation methods that have been proposed for joint models are (semi-parametric) maximum likelihood [9, 10, 28] and Bayes using MCMC techniques [1, 3, 27, 29]. Moreover, [24] have proposed a conditional score approach in which the random effects are treated as nuisance parameters, and they developed a set of unbiased estimating equations that yields consistent and asymptotically normal estimators. Here we give the basics of the maximum likelihood method for joint models as the one of the more traditional approaches.

Maximum likelihood estimation for joint models is based on the maximization of the log-likelihood corresponding to the joint distribution of the time-to-event and longitudinal outcomes $\{T_i, \delta_i, y_i\}$. To define this joint distribution we will assume that the vector of time-independent random effects b_i underlies both the longitudinal and survival processes. This means that these random effects account for both the association between the longitudinal and event outcomes, and the correlation between the repeated measurements in the longitudinal process (conditional independence). Formally we have that,

$$p(T_i, \delta_i, y_i | b_i; \theta) = p(T_i, \delta_i | b_i; \theta) p(y_i | b_i; \theta) \quad (4)$$

$$p(y_i | b_i; \theta) = \prod_j p\{y_i(t_{ij}) | b_i; \theta\}, \quad (5)$$

where θ denotes the parameter vector, y_i is the $n_i \times 1$ vector of longitudinal responses of the i th subject, and $p(\cdot)$ denotes an appropriate probability density function. Due to the fact that the survival and longitudinal submodels share the same random effects, joint models of this type are also known as shared parameter models. Under the modelling assumptions presented in the previous section, and these conditional independence assumptions the joint log-likelihood contribution for the i th subject can be formulated as

$$\log p(T_i, \delta_i, y_i; \theta) = \log \int p(T_i, \delta_i | b_i; \theta) \left[\prod_j p\{y_i(t_{ij}) | b_i; \theta\} \right] p(b_i; \theta) db_i, \quad (6)$$

where the likelihood of the survival part is written as

$$p(T_i, \delta_i | b_i; \theta) = \{h_i(T_i | \mathcal{M}_i(T_i); \theta)\}^{\delta_i} \mathcal{S}_i(T_i | \mathcal{M}_i(T_i); \theta), \quad (7)$$

with $h_i(\cdot)$ given by either (1) or (2), and

$$\begin{aligned} \mathcal{S}_i(t | \mathcal{M}_i(t), w_i) &= \Pr(T_i^* > t | \mathcal{M}_i(t), w_i) \\ &= \exp\left\{-\int_0^t h_i(s | \mathcal{M}_i(s); \theta) ds\right\}, \end{aligned} \quad (8)$$

$p\{y_i(t_{ij}) | b_i; \theta\}$ is the univariate normal density for the longitudinal responses, and $p(b_i; \theta)$ is the multivariate normal density for the random effects.

Maximization of the log-likelihood function corresponding to (6) with respect to θ is a computationally challenging task. This is mainly because both the integral with respect to the random effects in (6), and the integral in the definition of the survival function (8) do not have an analytical solution, except in very special cases. Standard numerical integration techniques such as Gaussian quadrature and Monte Carlo have been successfully applied in the joint modelling framework [9, 21, 28]. Furthermore, Rizopoulos, Verbeke and Lesaffre [17] have recently discussed the use of Laplace approximations for joint models, that can be especially useful in high-dimensional random effects settings (e.g., when splines are used in the random effects design matrix). For the maximization of the approximated log-likelihood the EM algorithm has been traditionally used in which the random effects are treated as ‘missing data’. The main motivation for using this algorithm is the closed-form M-step updates for certain parameters of the joint model. However, a serious drawback of the EM algorithm is its linear convergence rate that results in slow convergence especially near the maximum. Nonetheless, [17] have noted that a direct maximization of the observed data log-likelihood, using for instance, a quasi-Newton algorithm [12], requires very similar computations to the EM algorithm. Therefore hybrid optimization approaches that start with EM and then continue with direct maximization can be easily employed.

3 Connection with Missing Data

As mentioned in Section 1, joint models for longitudinal and time-to-event are also used when the main interest is on the longitudinal outcome and we wish to correct for the dropout caused by the occurrence of events. Under the categorization of missing data mechanisms proposed by [20], joint models correspond to missing not at random mechanisms. To show this more clearly, we define for each subject the observed and missing part of the longitudinal response vector. The observed part $y_i^o = \{y_i(t_{ij}) : t_{ij} < T_i, j = 1, \dots, n_i\}$ contains all observed longitudinal measurements of the i th subject before the observed event time, whereas the missing part $y_i^m = \{y_i(t_{ij}) : t_{ij} \geq T_i, j = 1, \dots, n_i'\}$ contains the longitudinal measurements that would

have been taken until the end of the study, had the event not occurred. Under these definitions and the assumptions of the joint modelling framework, we can derive the dropout mechanism, which is the conditional distribution of the time-to-dropout given the complete vector of longitudinal responses (y_i^o, y_i^m) ,

$$p(T_i^* | y_i^o, y_i^m; \theta) = \int p(T_i^* | b_i; \theta) p(b_i | y_i^o, y_i^m; \theta) db_i. \quad (9)$$

We observe that the time-to-dropout depends on y_i^m through the posterior distribution of the random effects $p(b_i | y_i^o, y_i^m; \theta)$, which illustrates the nonrandom nature of the missing data mechanism implied by joint models. In practice, the major implication of this feature is that the observed data upon which the joint model is fitted are not a random sample of the target population, which in turn implies that the characteristics of the observed data distribution cannot be compared with the output of the joint model. Therefore, residuals plots based on the observed data alone can be deceptive regarding the fit of a joint model to the data at hand. To overcome this problem, Rizopoulos, Verbeke and Molenberghs [19] proposed to take dropout into account by multiply imputing the longitudinal responses that would have been observed had the event not occurred. The advantage of this approach is that after the imputation, residuals can be computed for the complete data facilitating thus the use of standard diagnostic plots can for the longitudinal [15, 26] and event time outcomes [22].

4 Mayo Clinic Primary Biliary Cirrhosis Data

As an illustrative example of joint modelling we consider the primary biliary cirrhosis (PBC) data collected by the Mayo Clinic from 1974 to 1984 [14]. 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 154 placebo, and we are interested in testing for a treatment effect on survival after adjusting for the longitudinal bilirubin levels. Due to the right skewness of the observed serum bilirubin level, we will work with the natural logarithm of serum bilirubin for the remainder of this analysis.

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 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 (std. deviation 3.77 visits), resulting in a total of 1945 observations. By the end of the study 140 patients had died that corresponds to 55.1% 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 their interaction for the fixed effects part, and with random intercepts and random slopes for the random-effects part. For the survival submodel, we include as a time-independent covariate the treatment effect, and as time-dependent one the true underlying effect of log serum bilirubin count as estimated from the longitudinal model. The baseline risk function is assumed piecewise constant

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

with $v_0 = 0$, six internal knots placed at equally spaced percentiles of the observed event times, and v_7 taken larger than the largest observed time. The results are presented in Table 1. We observe a strongly significant effect of the log serum bilirubin levels to risk for an event. In particular, the hazard ratio for one unit increase in log serum bilirubin for patients randomized in the same treatment equals 3.52, which indicates the high importance of this marker in quantifying the risk for death. To check the fit of the model we produce multiple imputation residuals for the longitudinal outcome, discussed in [19], and the Cox-Snell residuals for the survival outcome. We observe that the fitted loess curves in the plots of the standardized marginal residuals based on the observed data alone versus the fitted values show a systematic trend (red line). Note, however that high levels of serum bilirubin indicate a worsening of patients' condition resulting in higher death rates (i.e., dropout). Thus, the residuals corresponding to large fitted values are only based on patients with a 'good' health condition, which explains why we observe the systematic trend. On the contrary the loess curves based on both the multiply imputed residuals and

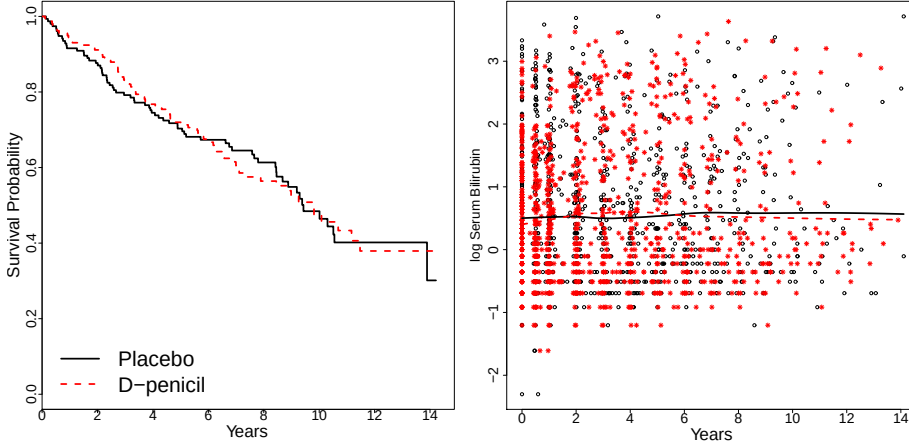


Fig. 1 PBC data: the left panel shows the Kaplan-Meier estimate of the survival function for the two treatment groups, and the right panel shows the observed data and superimposed loess fits for the two treatment groups.

Table 1 Parameter estimates and standard errors for the joint model analysis of the PBC data set. The top part contains the results for the longitudinal submodel, the middle part contains the results for the survival submodel, and the bottom part for the random-effects covariance matrix; $\xi_1, \dots, \xi_7$ denote the parameters of the baseline risk function, D_{11} the variance for the random intercepts, D_{22} the variance for the random slopes, and D_{12} their covariance.

	Value	Std.Err	z-value	p-value
(Intercept)	0.556	0.033	16.77	< 0.0001
D-penicil	0.028	0.047	0.60	0.5461
year	0.185	0.007	26.99	< 0.0001
D-penicil $\times$ year	-0.013	0.009	-1.38	0.1690
$\log(\sigma)$	-0.963	0.018		
D-penicil	0.076	0.180	0.42	0.6743
Assoct	1.258	0.094	13.32	< 0.0001
$\log(\xi_1)$	-4.472	0.261		
$\log(\xi_2)$	-4.329	0.280		
$\log(\xi_3)$	-4.656	0.331		
$\log(\xi_4)$	-4.612	0.383		
$\log(\xi_5)$	-4.229	0.344		
$\log(\xi_6)$	-3.770	0.349		
$\log(\xi_7)$	-4.645	0.490		
D_{11}	1.027	0.088		
D_{12}	0.072	0.015		
D_{22}	0.030	0.004		

the residuals corresponding to y_i^o (blue line) do not exhibit any strange shapes, indicating that after taking dropout into account the fitted joint model does seem to be a plausible model for this data set.

5 Conclusion

Joint modeling of longitudinal and time-to-event data is one of the most rapidly evolving areas of current biostatistics research, with several extensions of the standard joint model presented here already proposed in the literature. These include, among others, handling multiple failure types [6], considering categorical longitudinal outcomes [7], assuming that several longitudinal outcomes affect the time-to-event [1, 3], and associating the two outcomes via latent classes instead of random effects [16].

Furthermore, one very promising subfield which has emerged within the general joint modelling framework, is the use of these models in personalized medicine. In particular, there is lately a great need for tools that can help physicians take better informed decisions regarding their actions for the specific patients that they treat and not for an ‘average’ patient. Two features of joint models that allow them to become such a flexible dynamic tool, is the use of random effects and their time-dependent nature. For instance, as longitudinal information is collected for patients, we can

continuously update the predictions of their survival probabilities, and therefore be able to discern between patients with low and high risk for an event.

Finally, one of the main practical limitations for joint modeling in finding its way into the toolbox of modern statisticians was the lack of free and reliable software. The R package **JM** (it can be downloaded from the CRAN website <http://cran.r-project.org/>). has been developed to fill this gap to some extent. **JM** has a user-friendly interface to fit joint models and also provides several supporting functions that extract or calculate various quantities based on the fitted model (e.g., residuals, fitted values, empirical Bayes estimates, various plots, and others). At <http://rwiki.sciviews.org/doku.php?id=packages:cran:jm>, more information can be found.

References

1. Brown, E., Ibrahim, J.: A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics* **59**, 221–228 (2003)
2. Brown, E., Ibrahim, J., DeGruttola, V.: A flexible b-spline model for multiple longitudinal biomarkers and survival. *Biometrics* **61**, 64–73 (2005)
3. Chi, Y.Y., Ibrahim, J.: Joint models for multivariate longitudinal and multivariate survival data. *Biometrics* **62**, 432–445 (2006)

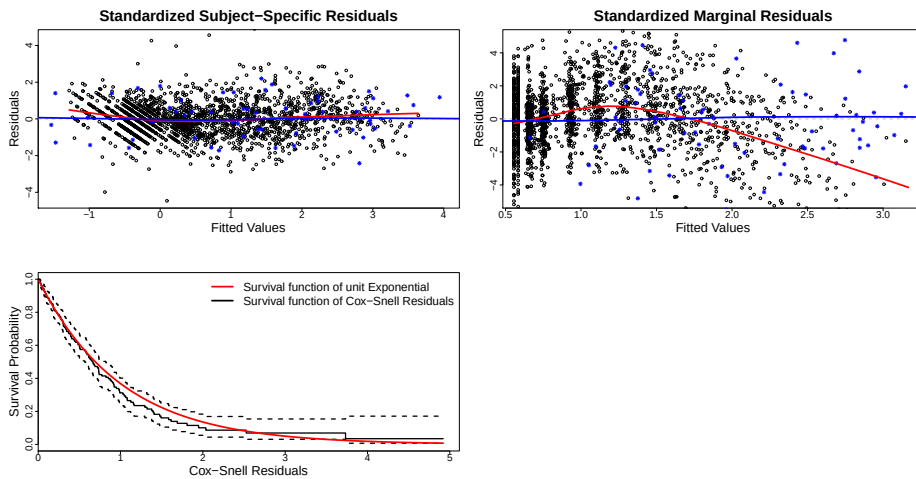


Fig. 2 Residuals for the joint model fitted to the PBC data: the top left panel depicts the standardized subject-specific residuals versus the subject-specific fitted values for the longitudinal outcome; the top right panel depicts the standardized marginal residuals versus the marginal fitted values for the longitudinal outcome; the bottom left panel depicts the Kaplan-Meier estimate of the survival function of the Cox-Snell residuals for the event time outcome. For the top panels the black points correspond to the residuals based on the observed data alone, the blue points to the residuals for the multiply-imputed data, the red line depicts the loess fit for the observed data alone, and the blue line the loess fit based on both the observed and multiply-imputed data.

4. Cox, D., Oakes, D.: Analysis of Survival Data. Chapman & Hall, London (1984)
5. Ding, J., Wang, J.L.: Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. *Biometrics* **64**, 546–556 (2008)
6. Elashoff, R., Li, G., Li, N.: A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics* **64**, 762–771 (2008)
7. Faucett, C., Schenker, N., Elashoff, R.: Analysis of censored survival data with intermittently observed time-dependent binary covariates. *Journal of the American Statistical Association* **93**, 427–437 (1998)
8. Follmann, D., Wu, M.: An approximate generalized linear model with random effects for informative missing data. *Biometrics* **51**, 151–168 (1995)
9. Henderson, R., Diggle, P., Dobson, A.: Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1**, 465–480 (2000)
10. Hsieh, F., Tseng, Y.K., Wang, J.L.: Joint modeling of survival and longitudinal data: Likelihood approach revisited. *Biometrics* **62**, 1037–1043 (2006)
11. Kalbfleisch, J., Prentice, R.: The Statistical Analysis of Failure Time Data, 2nd edn. Wiley, New York (2002)
12. Lange, K.: Optimization. Springer-Verlag, New York (2004)
13. Little, R., Rubin, D.: Statistical Analysis with Missing Data, 2nd edn. Wiley, New York (2002)
14. Murtaugh, P., Dickson, E., Van Dam, G., Malincho, M., Grambsch, P., Langworthy, A., Gips, C.: Primary biliary cirrhosis: prediction of short-term survival based on repeated patient visits. *Hepatology* **20**, 126–134 (1994)
15. Nobre, J., Singer, J.: Residuals analysis for linear mixed models. *Biometrical Journal* **6**, 863–875 (2007)
16. Proust-Lima, C., Joly, P., Dartigues, J.F., Jacqmin-Gadda, H.: Joint modelling of multivariate longitudinal outcomes and a time-to-event: a nonlinear latent class approach. *Computational Statistics and Data Analysis* **53**, 1142–1154 (2009)
17. Rizopoulos, D., Verbeke, G., Lesaffre, E.: Fully exponential laplace approximations for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society, Series B* **71**, 637–654 (2009)
18. Rizopoulos, D., Verbeke, G., Molenberghs, G.: Shared parameter models under random effects misspecification. *Biometrika* **95**, 63–74 (2008)
19. Rizopoulos, D., Verbeke, G., Molenberghs, G.: Multiple-imputation-based residuals and diagnostic plots for joint models of longitudinal and survival outcomes. *Biometrics* **66**, 20–29 (2010)
20. Rubin, D.: Inference and missing data. *Biometrika* **63**, 581–592 (1976)
21. Song, X., Davidian, M., Tsiatis, A.: A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics* **58**, 742–753 (2002)
22. Therneau, T., Grambsch, P.: Modeling Survival Data: Extending the Cox Model. Springer-Verlag, New York (2000)
23. Tseng, Y.K., Hsieh, F., Wang, J.L.: Joint modelling of accelerated failure time and longitudinal data. *Biometrika* **92**, 587–603 (2005)
24. Tsiatis, A., Davidian, M.: A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika* **88**, 447–458 (2001)
25. Tsiatis, A., Davidian, M.: Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica* **14**, 809–834 (2004)
26. Verbeke, G., Molenberghs, G.: Linear Mixed Models for Longitudinal Data. Springer-Verlag, New York (2000)
27. Wang, Y., Taylor, J.: Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association* **96**, 895–905 (2001)
28. Wulfsohn, M., Tsiatis, A.: A joint model for survival and longitudinal data measured with error. *Biometrics* **53**, 330–339 (1997)
29. Xu, J., Zeger, S.: Joint analysis of longitudinal data comprising repeated measures and times to events. *Applied Statistics* **50**, 375–387 (2001)
30. Yu, M., Law, N., Taylor, J., Sandler, H.: Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica* **14**, 835–832 (2004)