Joint Modeling of Longitudinal and Survival Data

Kenny Chen

December 12, 2023

Abstract

A common interest in many clinical trials or medical studies involve examining the relationship between longitudinal data and survival data. In this paper, we explore various ways to jointly model longitudinal and survival data in order to account for missingness, measurement, errors and potential dependencies between the two types of data. Three approaches for inference in joint modeling are discussed: a two-stage approach, likelihood-based methods, and Bayesian methods. We then apply these approaches to an HIV/AIDS study, comparing results and emphasizing the importance of careful consideration in joint modeling methodologies due to its complexity and various concerns.

1 Introduction

We often learn about longitudinal and survival analyses in separated settings. The former interested in the data being measured over time and the latter focusing on a time to an event outcome. However, in many clinical studies, we might have data that combines time-to-event data and longitudinal data and instances where longitudinal data such as immune response to a treatment, tumor cell count, or blood pressure could be important predictors of a time-to-event like survival or relapse. Consider the following Figure 1 that shows 100 sampled patients from an HIV/AIDS study done by [1] (we will use this dataset in the Application section as well) that shows CD4 cell counts repeatedly measured over time. The time to event here could be time to dropout from study to time to death to name a few.

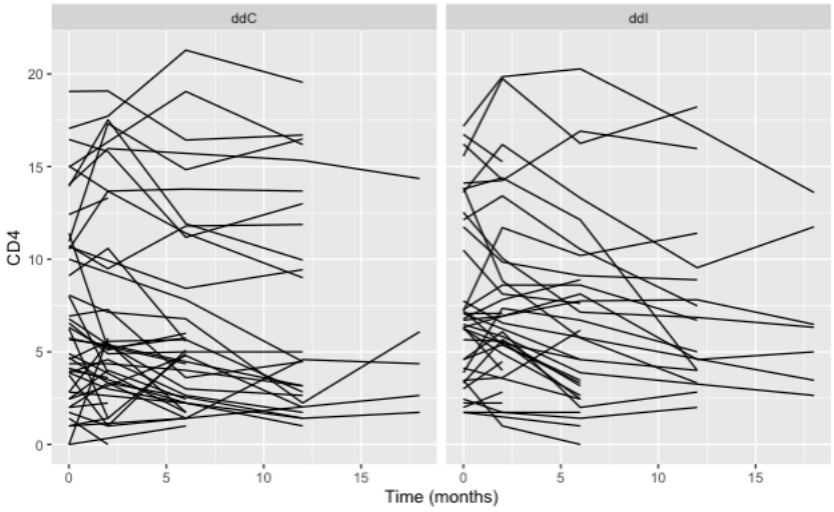

Figure 1: CD4 cell counts over time for two different antiretroviral drug treatments (ddC and ddI)

Popular models such as the linear mixed effects model for longitudinal data and the Cox proportional hazards for survival data do not consider the potential associations between these two types of data. When we perform separate analyses, it could potentially lead to inefficient or biased results thus urging a consideration of jointly modeling longitudinal and survival data with the goal of obtaining valid and efficient inferences. A simulation example is described in [8] which demonstrated that when the true hazard ratio is 0.67 and the longitudinal effect is "moderately correlated" with the time to event, the treatment effect estimate was 0.76 with a standard error of 0.063 in a 400 subject study. Using a joint model however, the estimate was 0.67 with a standard error of 0.051. Having less bias and greater efficiency has massive implications on design study. It can potentially lead to higher power as well as smaller sample sizes when designing and conducting clinical trials [8].

Joint model of longitudinal and survival data are then a type of model that aims to coalesce these two ideologies in order to better assess some treatment effect. It has been shown that joint models provide more efficient estimates of treatment effects on the time to event, more efficient estimates of treatment effects of the longitudinal covariates, and reduce bias in the estimates of the overall treatment effect on survival and the longitudinal marker [8]. Ibrahim et al. [8] further suggest they might also have the potential to allow for lower sample sizes with higher power compared to standard designs based on survival data only [1].

[1] They cite a manuscript submitted for publication but didn't include the title of the manuscript and my own

In this paper, we will describe three approaches for inference in joint modeling and discuss their advantages and disadvantages. The first being a two-stage approach from the likes of [17], [13] and its extensions. Then, we will examine likelihood methods which are based on the likelihood for both the longitudinal and survival data and then briefly touch upon how these joint models can be viewed through a Bayesian framework [20]. We will then list out some other remarks and considerations related to this topic to acknowledge that this field is quite vast with its methodologies, implementations, and various concerns. Finally, we apply the three approaches to the HIV/AIDS study from [1] and compare results.

2 Setting

In this paper, we adopt the notation used in [20] and consider a standard formulation for the joint model where the setup consists of a survival model with measurement errors in time-dependent covariates, in which a linear mixed-effects model is often used to model time-dependent covariates to address covariate measurement errors and a Cox proportional hazards model is used to model the survival data.

Suppose we have a longitudinal study with n individuals. The question of interest pertains to modeling the time to an event or survival time and we are considering incorporating a time-varying covariate in the survival model. Let s_i be the survival time for individual $i = 1, 2, \cdots, n$. Some individuals might not experience the event by the end of the study so their event times are right censored. For simplicity, we assume that the censoring is random and non-informative. For individual $i = 1, 2, \cdots, n$, let c_i be the censoring time and let $\delta_i = I(s_i \leq c_i)$ be the censoring indicator. So $\delta_i = 0$ if it is right censored and $\delta_i = 1$ otherwise. The observed data are then $\{(t_i, \delta_i), i = 1, 2, \cdots, n\}$ with $t_i = \min(s_i, c_i)$.

We will also only consider one time-dependent covariate for simplicity. Let z_{ij} be the observed covariate value for an individual i at time u_{ij}, subject to some measurement error, $i = 1, 2, \cdots, n$ and $j = 1, 2, \cdots, m_i$. We will also let the unobserved true covariates be z^*_{ij} and we denote $\boldsymbol{z}_i = \{z_{i1}, \cdots, z_{im_i}\}^T$ and $\boldsymbol{z^*}_i = \{z^*_{i1}, \cdots, z^*_{im_i}\}^T$. And we let $\boldsymbol{x}_i$ be the baseline covariates with the assumption that $\boldsymbol{x_i}$ has no measurement errors.

We will also work with the Cox model for survival data:

$$\lambda_i(t) = \lambda_0(t) \exp(z^*_i(t)\beta_1 + \mathbf{x}_i^T \boldsymbol{\beta}_2), i = 1, 2, \cdots, n \tag{1}$$

where $\lambda_i(t)$ is the hazard function, $\lambda_0(t)$ is the baseline hazard function and $\boldsymbol{\beta} = (\beta_1, \boldsymbol{\beta}_2^T)^T$ are the unknown true parameters. In addition, in a Cox model, $\exp(\beta)$ is the relative change in the hazard at time t for a one unit change in the corresponding covariate, much like the coefficients in a logistic regression but instead of odds ratios, we are dealing with hazard ratios. If the estimate for a parameter is less than 0 on the log hazards scale, we can intuitively imagine the risk of an event is lower and thus the covariate is beneficial and vice versa. One can perform estimation and inference on the β's via a partial likelihood. I won't get too much into detail here but see [4] for the seminal paper.

In the Cox model, ideally, we have the value of $z_i(t)$ at each event time t but in practice, covariates are observed intermittently at various times, $\{u_{ij}, j = 1, 2, \cdots, m_i\}$ for an individual i with observation times possibly varying from individual to individual. This could lead to "missing" covariates and measurement errors could occur as well which needs to be address [20].

search online failed to show any results so I'm not too confident on this last bit, but it is definitely something worthy of pursuing

To address either issues, a standard approach is to model the time-dependent covariates. One choice is a linear mixed effects model:

$$\boldsymbol{z}_i = U_i\boldsymbol{\alpha} + V_i\boldsymbol{a_i} + \boldsymbol{\epsilon}_i \equiv \boldsymbol{z}_i^* + \boldsymbol{\epsilon}_i, i = 1, \cdots, n \tag{2}$$

where U_i, V_i are design matrices, $\boldsymbol{\alpha}$ is the vector of fixed-effects parameters, $\boldsymbol{a_i}$ is a vector of random effects, $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \cdots, \epsilon_{im_i})^T$ is a vector of measurement errors, and the true unobserved covariates are $\boldsymbol{z}_i^* = U_i\boldsymbol{\alpha} + V_i\boldsymbol{a_i}$.

We also will assume that

$$\boldsymbol{a_i} \sim N(0, \Sigma), \quad \boldsymbol{\epsilon}_i \sim N(0, \sigma^2 I)$$

with $\boldsymbol{a_i}$ and $\boldsymbol{\epsilon}_i$ independent, Σ is a covariance matrix, σ^2 is a parameter and I is the identity matrix. We will also assume that covariate observations are truncated by event time so no covariates are observed after the event occurs or after they drop out.

See that the survival mode in (1) and the longitudinal model in (2) are linked through the shared random effects $\boldsymbol{a_i}$. This shared random effects induces the dependence between the longitudinal and survival processes, thus pointing to the need of joint modeling.

3 Methods

3.1 Two-step Methods

There have been several two-stage methods proposed in the joint modeling literature. We first introduce a basic two-step method and then provide some variations at the end.

1. Fit a linear mixed effects (LME) model to the longitudinal covariate data and estimate the missing or mismeasured covariates based on the fitted model.

2. Fit the survival model separately, with the missing/ unobserved true covariate values substituted by their estimates from the first stage as if they were observed and proceed with typical survival analysis methods.

Some of the earliest implementations of this idea are in from Self and Pawitan [14] and Tsiatis et al. [17]. Self and Pawitan [14] proposed a two-stage method which uses the least-squares method to fit individual longitudinal covariate trajectories and the following estimates were used to impute the covariate values in the survival model and then inference was based on the partial likelihood. Tsiatis et al. [17] approximated the hazard function and used that approximation in the partial likelihood by replacing the true covariate $z_i^*(t)$ by the empirical Bayes estimate of the conditional expectation $E(z_i^*(t)|z_i^H(t), t \leq s_i)$, where $z_i^H(t) = \{z_i(u), u \leq t\}$ is the covariate history up to time t. Other approaches have been proposed in [17], [5], [3].

This approach has the advantage that it is simple and can be implemented using standard statistical software. However, this approach could lead to biased inferences [20] [21]. This is due to a bevy of reasons. When we are estimating the longitudinal covariate model parameters, the truncations from the events are not taken into consideration. In other words, the longitudinal covariate trajectories of subjects who experience the event of interest may be different from those who do not experience the event, which could potentially bias results based only on observed covariate data. In addition, the uncertainty of the estimation in the first stage is not incorporated in the second stage of the survival model estimation. This could lead to underestimation in standard errors of the parameter estimates of the survival model.

Wu et al. [20] points out that the bias in the estimation of the longitudinal model parameters due to not accounting for the informative truncations from the events may depend on the strength of the association between the longitudinal process and the survival process. And the bias stemming from ignoring the estimation uncertainty in Stage 1 may depend on the magnitude of measurement errors in covariates. To address these issues, various modified two-stage methods have been proposed, leading to a better less biased two-stage methods.

Since a lot of the bias from the naive two-stage method is caused by the fact that the covariate trajectory is related to the length of follow up, the bias may be removed if we can recapture these missing covariate measurements due to truncation by incorporating the event time information. Albert and Shih [2] proposed the following method to recapture the missing measurements by generating data from the conditional distribution of the covariate given the event time:

$$f(\boldsymbol{z}_i|s_i,\boldsymbol{\theta}) = \int f(\boldsymbol{z}_i|\boldsymbol{a}_i,\boldsymbol{\theta}) f(\boldsymbol{a}_i|s_i,\boldsymbol{\theta}) d\boldsymbol{a}_i \tag{3}$$

where the covariate $\boldsymbol{z}_i$ and event time s_i are conditionally independent given the random effects $\boldsymbol{a}_i$ and $\boldsymbol{\theta}$ is the vector of unknown parameters. The conditional density is then approximated using a linear mixed effects model and they simulated missing data from $f(\boldsymbol{z}_i|s_i,\boldsymbol{\theta})$. After simulating the missing data, the covariate model is then fitted to the "complete data".

3.2 Likelihood-based Methods

The likelihood approach is more widely used and produces valid and efficient inferences if the assumed models are correct. This method is based on the both the likelihood for the longitudinal data and survival data. Denote the observed data as $\{(t_i,\delta_i,\boldsymbol{z}_i,\boldsymbol{x}_i), i = 1,2,\cdots,n\}$. Let $\boldsymbol{\theta} = (\boldsymbol{\beta},\boldsymbol{\alpha},\sigma,\Sigma,\boldsymbol{\lambda}_0)$ denote the set of unknown parameters in the models where $\boldsymbol{\lambda}_0 = \{\lambda_0(t_i), i = 1,2,\cdots,n\}$. Once again, we act under the assumption that censoring of the survival data and assessment process of longitudinal measurements are non-informative. The overall likelihood for all the observed data is given by:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \int f(t_i,\delta_i|\boldsymbol{z}_i^*,\boldsymbol{\lambda}_0,\boldsymbol{\beta}) f(\boldsymbol{z}_i|\boldsymbol{a}_i,\boldsymbol{\alpha},\sigma^2) f(\boldsymbol{a}_i|\Sigma) d\boldsymbol{a}_i \tag{4}$$

where

$$f(t_i,\delta_i|\boldsymbol{z}_i^*,\boldsymbol{\lambda}_0,\boldsymbol{\beta}) = \Big(\lambda_0(t)\exp(z_i^*(t)\beta_1 + \mathbf{x}_i^T\boldsymbol{\beta}_2)\Big)^{\delta_i} \times \exp\bigg(-\int_0^{t_i} \lambda_0(x)\exp(z_i^*(t)\beta_1 + \mathbf{x}_i^T\boldsymbol{\beta}_2)dx\bigg),$$

$$f(\boldsymbol{z}_i|\boldsymbol{a}_i,\boldsymbol{\alpha},\sigma^2) = (2\pi\sigma^2)^{-m_i/2}\exp\bigg(-\frac{(\boldsymbol{z}_i-\boldsymbol{z}_i^*)^T(\boldsymbol{z}_i-\boldsymbol{z}_i^*)}{2\sigma^2}\bigg),$$

$$f(\boldsymbol{a}_i|\Sigma) = (2\pi|\Sigma|)^{-1/2}\exp\bigg(-\frac{\boldsymbol{a}_i^T\Sigma^{-1}\boldsymbol{a}_i}{2}\bigg)$$

We can either directly maximize the likelihood or use an Expectation-Maximization (EM) algorithm. Since the integral is intractable, maximization is done using numerical integration techniques such as the Gaussian Hermite quadrature methods that we've covered in class or Monte Carlo methods.

3.2.1 Computation

The appeal with two-stage modeling was computation efficiency, especially when considering the time period when people started to publish on joint modeling approaches back in the 90s. Now with blazing fast computers and numerical integration techniques, we can simply ignore everything and just approximate the likelihood right? Well as we saw in the midterm and in class, computational complexity still is a challenge and issue that we must reconcile with. These methods can be computationally intensive if the dimension of the unobservable random effects $\boldsymbol{a}_i$ is not low so a few approaches have been suggested.

Tsiatis and Davidian [18] proposed an approach called the conditional score method that makes no distributional assumptions for the random effects $\boldsymbol{a}_i$. They treat $\boldsymbol{a}_i$ as a nuisance parameter and conditions on an appropriate sufficient statistic for $\boldsymbol{a}_i$. By conditioning on this sufficient statistic, it removes the dependence of the conditional distribution on the random effects $\boldsymbol{a}_i$. They show it is under certain regularity conditions consistent and asymptotically normal. In another approach, Song et al. [15] relaxes the normality assumption of the vector of random effects by assuming they merely follow some distribution in a class of smooth density family.

There have also been other approaches to improving computation speed such as using Laplace approximations rather than Gaussian Hermite Quadratures, see [12]. They demonstrate that their method requires a much smaller number of repeated measurements per individual to produce reliable results compared to standard Laplace approximation.

3.3 Bayesian Methods

One could also consider a Bayesian joint model where the model parameters are assumed to follow some prior distributions and inference is based on the posterior distribution given the observed data. An advantage of Bayesian ideology is that one could incorporate informative priors based on similar studies or from experts to gain additional information.

Let $\boldsymbol{\theta}$ denote the collection of unknown parameters in the joint model and let $f(\boldsymbol{\theta})$ denote the prior distribution. Let $\mathcal{D} = \{(t_i, \delta_i, \boldsymbol{z}_i, \boldsymbol{x}_i), i = 1, 2, \cdots, n\}$ denote the observed data. The joint posterior distribution for all unknown parameters $\boldsymbol{\theta}$ and random effects $\boldsymbol{a} = \{\boldsymbol{a}_i, i = 1, 2, \cdots, n\}$ is given by:

$$f(\boldsymbol{\theta}, \boldsymbol{a}|\mathcal{D}) \propto \prod_{i=1}^{n} f(t_i, \delta_i|\boldsymbol{z}_i^*, \boldsymbol{x}_i, \boldsymbol{\theta}) f(\boldsymbol{z}_i|\boldsymbol{a}_i, \boldsymbol{\theta}) f(\boldsymbol{a}_i|A) f(\boldsymbol{\theta}|\boldsymbol{\theta}_0) \tag{5}$$

where $\boldsymbol{\theta}_0$ are the known hyperparameters. We can then perform Bayesian inference using Monte Carlo samples drawn from the posterior distribution such as the Gibbs sampler. We can estimate posterior means and variances of the parameters based on Monte Carlo samples and Bayesian inference can be conducted based on these estimated posterior means and variances.

4 Other considerations and remarks

In the previous section, we made some stringent assumptions regarding missingness being non-informative or only considering right censoring with truncation and these assumptions usually do not hold. Since the topic of joint modeling is quite extensive, I wanted to also quickly mention other things that researchers have considered in this field.

4.1 Other Survival Models

In our literature review, we only looked at the Cox proportional hazards model but a joint model can be based on other models used for survival data, such as the Accelerated Failure Time (AFT) model where the AFT model can be written in a similar form as the Cox model:

$$h_i(t) = h_0(t)\left(\int_0^t \exp(-z_i^*(u))\beta du \right) \exp(-z_i^*(t)\beta) \tag{6}$$

where $h_i(t)$ is the hazard function of individual i at time t, $h_0(t)$ is the baseline hazard function and $z_i^*(t)$ is the unobserved true covariate value at time t. One might be interested in the accelerated failure time model as an alternative to the Cox model when the proportionality assumption does not hold; however, Tseng et al. [16] shows that the AFT structure in the joint modeling scheme is not the same as the Cox model.

4.2 Nonlinear Models

We also only considered a linear mixed effects model for the longitudinal data, but nonlinear mixed effects model (NLMEs) or generalized linear mixed models (GLMMs) can also be used. For example, Wu et al. [19] considered a nonlinear mixed-effects model for the longitudinal process and the Cox proportional hazards model for the time-to-event process and applied it to an AIDS study by jointly modeling HIV viral dynamics and time to viral rebound.

4.3 Informative Missingness

Since we acted under the assumption that the missingness was non-informative, the missing data mechanism can be ignored in likelihood inference. But when there are informative dropouts (or non-ignorable missing longitudinal responses), the missing data mechanism must be taken into account for valid likelihood inference. Wu et al. [19] accounts for the missingness by assuming a missing response model that allows the missing probability of missing to possibly depend on the random effects $\boldsymbol{a}_i$ and considered Monte Carlo EM algorithms and Laplace approximations.

4.4 A few more remarks

I went down various rabbit holes so I just wanted to briefly highlight some work that has been done in this field that might be interesting to take a look at.

There has been work in considering other types of censoring such as interval censoring [20] where we know the event occurs within some interval but we do not know when or left censoring where we know the event occurred before a time point.

In class, we learned about LMEs vs GEEs so I was also curious to see if GEEs have been considered and so I stumbled upon Zheng [23] who proposed in their dissertation "a joint generalized estimating equation framework using an inverse intensity weighting approach for parameter estimation from joint models".

One might naturally consider avoiding the topic of joint modeling by simply incorporating the longitudinal measures directly into the Cox model as time-varying covariates and then proceed with the Cox model analysis. However, because the longitudinal measures typically have a great deal of random error from subject to subject, this approach will lead to highly biased (typically attenuated) and inefficient estimates of the treatment effect [8].

Zhang et al. [22] proposed a new measure in the model fitting procedure. They made a new decomposition of AIC and BIC (i.e., AIC = AICLong + AICSurv|Long and BIC = BICLong +

BICSurv|Long) to assess the fit of each component in the joint model as well as assess the fit of each model separately.

In our case, we only considered one longitudinal covariate but people have also considered multivariate longitudinal data [6].

5 Data Application

We will consider an HIV study from [1] that was a randomized clinical trial in which both longitudinal and survival data were collected to compare the efficacy and safety of two antiretroviral drugs (ddC and ddI) in treating patients who had failed or were intolerant of zidovudine (AZT) therapy. CD4 cells are white blood cells that fight infection with higher CD4 counts indicating a stronger immune system that is more prepared to resist infection while lower CD4 counts indicate a higher risk of an infection. People have been interested and have looked at the association between the CD4 cell count at time t and risk of death at the same time.

This example was originally considered in [9] but in their application of joint modeling, they only considered the joint modeling approach and compared it to the time-dependent Cox model. We extend their application by also comparing it to two-stage model proposed by [17] as well as the Bayes model using the JMbayes package [11]. They also fit the joint model using a piecewise-baseline hazard but we will also change so that the baseline hazard is left unspecified to align with the method proposed by [21].

We first consider the time-dependent Cox model where we also control for the treatment:

$$\lambda_i(t) = \lambda_0(t) \exp(\gamma \text{ddI}_i + \alpha y_i(t))$$

where ddI_i is the dummy variable for the ddI group and $y_i(t)$ is the observed value of the CD4 cell count. We also fit a two-stage model where we consider the following linear mixed effects model:

$$y_i(t) = m_i(t) + \epsilon_i(t) = \beta_0 + \beta_1 t + \beta_2 \{t \times \text{ddI}_i\} + b_{i0} + b_{i1} t + \epsilon_i(t)$$

where β_0, β_1 are the fixed effects of time and the interaction between time and treatment and b_{i0}, b_{i1} are the random intercept and time effects. Then we plug in estimated values $m_i(t)$ into the Cox model above for $y_i(t)$. Lastly, we also compute using the joint likelihood approach using the JM package [10] and the JMbayes package [11].

Method	CD4	Treatment (ddI)
Time-dependent	-0.19 (0.024)	0.31 (0.15)
Two-Stage	-0.22 (0.027)	0.31 (0.15)
Joint	-0.27 (0.025)	0.32 (0.13)
Bayes	-0.2972 (-0.39, -0.22)	0.33 (-0.06, 0.7)

Table 1: Parameter Estimates and Standard Errors (95% credible intervals shown for Bayes model)

Table 1 displays the estimates of the main parameters and their standard errors (credible intervals for Bayes method). From Table 1, we see that the estimates are quite different, ranging from -0.19 to -0.29 on the log hazards scale. We can interpret the parameter estimate of CD4, say for the joint model as such: For a one unit increase in CD4 cell count, it corresponds to a exp(-0.27) = 0.76-fold

decrease in the risk for death. The p-value also indicates it is significant so the joint model suggests a strong association between CD4 cell count and the risk for death.

The joint model giving a larger estimate is consistent with [21] who reported slightly larger parameter estimates in the joint model compared to the two-stage model. Wulfsohn et al. [21] also reported larger standard errors in their application but here, we do not see the same pattern. Rizopoulos [9] points out this may occur due to the fact that the standard errors calculated using the JM package with the baseline hazard having no specification are underestimated. Rerunning the joint model using a different baseline hazard gave a slightly different parameter estimate but larger standard errors than the two-stage approach. Hsieh et al. [7] recommends a bootstrapped based approach to deal with these underestimated standard errors. Overall, it seems that the time-dependent and two-stage approaches underestimate not only the effect of CD4 but also the standard errors themselves.

6 Conclusion

The joint modeling of longitudinal and survival data is quite an extensive field with various adaptations and methods having been proposed for inference. We were motivated to perform joint modeling under the suspicion that a longitudinal and an event process may be associated and discussed a few ways to perform inference. A naive way to perform inference is to use the longitudinal data as a time-dependent covariate in say a Cox model however this can lead to biased (typically attenuated) estimates. We then moved on to consider two-staged models where we first fit the linear mixed effects model followed by imputing those values into the survival model. This has the advantage of being computationally less intensive and easily conducted using standard statistical software. However, there are various biases that may arise as mentioned previously. Then we covered joint modeling, which is the most widely used one today and one could also appeal to a Bayesian approach using standard Bayesian estimation techniques as well.

References

[1] Donald I. Abrams et al. "A Comparative Trial of Didanosine or Zalcitabine after Treatment with Zidovudine in Patients with Human Immunodeficiency Virus Infection". In: *New England Journal of Medicine* 330.10 (Mar. 1994), 657â662. ISSN: 1533-4406. DOI: 10.1056/nejm199403103301001. URL: http://dx.doi.org/10.1056/NEJM199403103301001.

[2] Paul S Albert and Joanna H Shih. "On estimating the relationship between longitudinal measurements and time-to-event data using a simple two-stage procedure". In: *Biometrics* 66.3 (2010), pp. 983–987.

[3] Paul Bycott and Jeremy Taylor. "A comparison of smoothing techniques for CD4 data measured with error in a time-dependent Cox proportional hazards model". In: *Statistics in medicine* 17.18 (1998), pp. 2061–2077.

[4] D. R. Cox. "Regression Models and Life-Tables". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 34.2 (1972), pp. 187–220. ISSN: 00359246. URL: http://www.jstor.org/stable/2985181 (visited on 12/11/2023).

[5] Urania G Dafni and Anastasios A Tsiatis. "Evaluating surrogate markers of clinical outcome when measured with error". In: *Biometrics* (1998), pp. 1445–1462.

[6] Graeme L. Hickey et al. "Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues". In: *BMC Medical Research Methodology* 16.1 (Sept. 2016). ISSN: 1471-2288. DOI: 10.1186/s12874-016-0212-5. URL: http://dx.doi.org/10.1186/s12874-016-0212-5.

[7] Fushing Hsieh, YiâKuan Tseng, and JaneâLing Wang. "Joint Modeling of Survival and Longitudinal Data: Likelihood Approach Revisited". In: *Biometrics* 62.4 (Apr. 2006), 1037â1043. ISSN: 1541-0420. DOI: 10.1111/j.1541-0420.2006.00570.x. URL: http://dx.doi.org/10.1111/j.1541-0420.2006.00570.x.

[8] Joseph G. Ibrahim, Haitao Chu, and Liddy M. Chen. "Basic Concepts and Methods for Joint Models of Longitudinal and Survival Data". In: *Journal of Clinical Oncology* 28.16 (June 2010), 2796â2801. ISSN: 1527-7755. DOI: 10.1200/jco.2009.25.0654. URL: http://dx.doi.org/10.1200/JCO.2009.25.0654.

[9] DIMITRIS RIZOPOULOS. *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. CRC PRESS, 2013.

[10] Dimitris Rizopoulos. "JM: An R Package for the Joint Modelling of Longitudinal and Time-to-Event Data". In: *Journal of Statistical Software* 35.9 (2010), pp. 1–33. URL: https://doi.org/10.18637/jss.v035.i09.

[11] Dimitris Rizopoulos. "The R Package JMbayes for Fitting Joint Models for Longitudinal and Time-to-Event Data Using MCMC". In: *Journal of Statistical Software* 72.7 (2016), pp. 1–45. DOI: 10.18637/jss.v072.i07.

[12] Dimitris Rizopoulos, Geert Verbeke, and Emmanuel Lesaffre. "Fully Exponential Laplace Approximations for the Joint Modelling of Survival and Longitudinal Data". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 71.3 (June 2009), 637â654. ISSN: 1467-9868. DOI: 10.1111/j.1467-9868.2008.00704.x. URL: http://dx.doi.org/10.1111/j.1467-9868.2008.00704.x.

[13] Steve Self and Yudi Pawitan. "Modeling a Marker of Disease Progression and Onset of Disease". In: *AIDS Epidemiology*. Birkhäuser Boston, 1992, 231â255. ISBN: 9781475712292. DOI: 10.1007/978-1-4757-1229-2_11. URL: http://dx.doi.org/10.1007/978-1-4757-1229-2_11.

[14] Steve Self and Yudi Pawitan. "Modeling a marker of disease progression and onset of disease". In: *AIDS epidemiology: methodological issues*. Springer, 1992, pp. 231–255.

[15] Xiao Song, Marie Davidian, and Anastasios A. Tsiatis. "A Semiparametric Likelihood Approach to Joint Modeling of Longitudinal and Time-to-Event Data". In: *Biometrics* 58.4 (Dec. 2002), 742â753. ISSN: 0006-341X. DOI: 10.1111/j.0006-341x.2002.00742.x. URL: http://dx.doi.org/10.1111/j.0006-341X.2002.00742.x.

[16] Yi-Kuan Tseng, Fushing Hsieh, and Jane-Ling Wang. "Joint modelling of accelerated failure time and longitudinal data". In: *Biometrika* 92.3 (Sept. 2005), pp. 587–603. URL: https://ideas.repec.org/a/oup/biomet/v92y2005i3p587-603.html.

[17] A. A. Tsiatis, Victor Degruttola, and M. S. Wulfsohn. "Modeling the Relationship of Survival to Longitudinal Data Measured with Error. Applications to Survival and CD4 Counts in Patients with AIDS". In: *Journal of the American Statistical Association* 90.429 (Mar. 1995), 27â37. ISSN: 1537-274X. DOI: 10.1080/01621459.1995.10476485. URL: http://dx.doi.org/10.1080/01621459.1995.10476485.

[18] Anastasios A. Tsiatis and Marie Davidian. "JOINT MODELING OF LONGITUDINAL AND TIME-TO-EVENT DATA: AN OVERVIEW". In: *Statistica Sinica* 14.3 (2004), pp. 809–834. ISSN: 10170405, 19968507. URL: http://www.jstor.org/stable/24307417 (visited on 12/06/2023).

[19] L. Wu, X. J. Hu, and H. Wu. "Joint inference for nonlinear mixed-effects models and time to event at the presence of missing data". In: *Biostatistics* 9.2 (July 2007), 308â320. ISSN: 1468-4357. DOI: 10.1093/biostatistics/kxm029. URL: http://dx.doi.org/10.1093/biostatistics/kxm029.

[20] Lang Wu et al. "Analysis of Longitudinal and Survival Data: Joint Modeling, Inference Methods, and Issues". In: *Journal of Probability and Statistics* 2012 (2012), 1â17. ISSN: 1687-9538. DOI: 10.1155/2012/640153. URL: http://dx.doi.org/10.1155/2012/640153.

[21] Michael S. Wulfsohn and Anastasios A. Tsiatis. "A Joint Model for Survival and Longitudinal Data Measured with Error". In: *Biometrics* 53.1 (Mar. 1997), p. 330. ISSN: 0006-341X. DOI: 10.2307/2533118. URL: http://dx.doi.org/10.2307/2533118.

[22] Danjie Zhang et al. "Assessing model fit in joint models of longitudinal and survival data with applications to cancer clinical trials". In: *Statistics in Medicine* 33.27 (July 2014), 4715â4733. ISSN: 0277-6715. DOI: 10.1002/sim.6269. URL: http://dx.doi.org/10.1002/sim.6269.

[23] Mengjie Zheng. "JOINT MODELING OF LONGITUDINAL AND SURVIVAL OUTCOMES USING GENERALIZED ESTIMATING EQUATIONS". In: (2018).