Literature review

September 18, 2018

Contents

1 Notation

Same notation across thesis? Or can be different in literature review nad method section

2 Introduction

There are many studies collect information about a survival outcome, together with longitudinal measurements of a disease marker. Longitudinal data are data in which a response variable is measured at different time points. For example, HIV patients may be followed over time and CD4 counts are coolected monthly. Survival data describe the length of time from a time origin to an endpoint of interest. For example, individuals might be followed from the diagnosis of lung cancer to death. The endpoint in survival data does not have to be death. It could also be particular events such as onset of a disease, discharge from hospital after surgery, etc. There are many studies collect information about a survival outcome, together with longitudinal measurements of a disease marker. The goal of these study might quite variable. The study may be accessing how the longitudinal measurement changes with time and how it is affect by other covariates; it may be designed to improve efficiency and reduce bias in the survival component; it may be characterizing the relationship between the survival and longitudinal processes.

3 Longitudinal submodel

3.1 Linear mixed model

The idea of linear mixed model was started by (Harville, 1977) and (Laird & Ware, 1982) introduced a unified approach and popularized the idea. In linear mixed models, each individual is assumed to have a same form of distribution but with different parameters. Explanatory covariates are specified to be either fixed or random. Fixed effects are considered to be the population-average effects while random effects refer to subject-specific effects. This model allows explicit identification of the between- and within-subject variation. The covariance matrix of within subject measurements can have flexible structures. The timing and number of measurements of each individual are not required to be the same.

The following model was proposed by (Laird & Ware, 1982):

$$Y_i = X_i\beta + Z_ib_i + e_i \tag{3.1}$$

where:

- Y_i is the $n_i \times 1$ outcome vector for subject i.

- $\boldsymbol{\beta}$ is the $p \times 1$ vector of unknown population parameters or fixed effect coffeicients.

- $\boldsymbol{X}_i$ is the $n_i \times p$ known design matrix for fixed effects.

- $\boldsymbol{b}_i$ is the $k \times 1$ unknown vector of individual effects. The $\boldsymbol{b}_i$ are distributed as $N(0, D)$, where D denotes an $k \times k$ positive-definite covariance matrix. They are assumed to be independent of each other and of $\boldsymbol{e_i}$.

- $\boldsymbol{Z}_i$ is the known $n_i \times k$ design matrix for random effects.

- $\boldsymbol{e_i}$ is the $n_i \times 1$ vector of random error terms. It is distributed as $N(0, R_i)$, where R_i denotes an $n_i \times n_i$ positive-definite covariance matrix.

$\boldsymbol{Y}_i$ can then be expressed as independent normals with mean $\boldsymbol{X}_i\boldsymbol{\beta}$ and covariance matrix $\boldsymbol{R_i} + \boldsymbol{Z_i} D \boldsymbol{Z}_i^T$.

If the covariance matrix are completely known, the expressions for the fixed effects and random effects are

$$\hat{\boldsymbol{\beta}} = (\sum_1^m \boldsymbol{X}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{X}_i)^{-1} (\sum_1^m \boldsymbol{X}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{y_i}) \tag{3.2}$$

and

$$\hat{\boldsymbol{b_i}} = D \boldsymbol{Z_i}^T \boldsymbol{V_i}^{-1} (\boldsymbol{y_i} - \boldsymbol{X_i}\hat{\boldsymbol{\beta}}) \tag{3.3}$$

where $\boldsymbol{V_i} = \mathrm{var}(\boldsymbol{y_i}) = \boldsymbol{R_i} + \boldsymbol{Z_i} D \boldsymbol{Z_i}^T$.

Their variances can be expressed as

$$\mathrm{var}(\hat{\boldsymbol{\beta}}) = \left(\sum_1^m \boldsymbol{X}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{X}_i \right)^{-1} \tag{3.4}$$

and

$$\mathrm{var}(\hat{\boldsymbol{b_i}}) = D \boldsymbol{Z_i}^T \left\{ \boldsymbol{V_i}^{-1} - \boldsymbol{V_i}^{-1} \boldsymbol{X_i} \left(\sum_1^m \boldsymbol{X}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{X}_i \right) \boldsymbol{X}_i^T \boldsymbol{V_i}^{-1} \right\} \boldsymbol{Z_i} D. \tag{3.5}$$

If the covariance matrix are unknown, $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{b_i}}$ can be obtained by replacing $\boldsymbol{V_i}$ by its estimate $\hat{\boldsymbol{V_i}}$ and using the same equations as equation(3.2) and equation(3.3), where $\hat{\boldsymbol{V_i}} = \hat{\boldsymbol{R_i}} + \boldsymbol{Z_i} \hat{D} \boldsymbol{Z_i}^T$. (Laird & Ware, 1982) showed that this intuitive approach is actually the maximum likelihood for $\boldsymbol{\beta}$ and empirical Bayes estimates for $\boldsymbol{b}$.

Let $\boldsymbol{\theta}$ denote the vector for variance and covariance parameters in $\boldsymbol{R_i}$ and $\boldsymbol{D}$. (Laird & Ware, 1982) provided two approaches to estimate $\boldsymbol{\theta}$. The maximum likelihood estimate $\hat{\boldsymbol{\theta}}_{\boldsymbol{ML}}$ was based on maximum likelihood estimation of $\boldsymbol{\theta}$ from the marginal distribution of $\boldsymbol{y^T}$. The restricted maximum likelihood estimate $\hat{\boldsymbol{\theta}}_{\boldsymbol{REML}}$ was derived from a Bayesian formulation. A flat prior for $\boldsymbol{\beta}$ was introduced and $\boldsymbol{\theta}$ was estimated from the marginal likelihood of $\boldsymbol{y}$ after integrating out $\boldsymbol{\beta}$ and $\boldsymbol{bi}$. The ML approach have downwards biased estimates of variance componets, while REML

approach can produce unbiased estimates (Laird & Ware, 1982). The EM algorithm (Dempster, Laird, & Rubin, 1977) is a suitable approach for parameter estimation in both formulations. It is used to estimate unobservable parameters, no data was considered missing.

3.2   Shape invariant model with random effects

(Lawton, Sylvestre, & Maggio, 1972) first introduced the concept of shape invariant model and applied the model on spirometry, spectrophotometric and optical density data. (Beath, 2007) developed a shape invariant model with random effects to analyze infant growth data. In this approach, the poplation is assumed to have a common characteristic function and individual curves can be obtained by shifting and scaling the common curve. The mathmatical form of the model can be expressed as

$$f(\alpha_{0i}, \alpha_{1i}, \beta_{0i}, \beta_{1i}; t) = \alpha_{0i} + e^{\alpha_{1i}} g\left(\frac{t - \beta_{0i}}{e^{\beta_{1i}}}\right). \tag{3.6}$$

Function $g(\cdot)$ represents the common characteristic curve. $\alpha_{0i}, \alpha_{1i}, \beta_{0i}, \beta_{1i}$ are the parameters for SIM. They have direct graphical interpretations. α_{0i} and β_{0i} are location parameters for y-axis and x-axis, respectively. α_{1i} and β_{1i} are scale parameters. Exponentiation is used so that to scale parameters are constrained to be greater than zero. (?, ?) also applied a very similar model to height in puberty data except α_1 was set to zero. They provided practical interpretation for the location and scale parameters. α_0 was termed size, which represented a random height intercept that adjusts for differences in mean height. β_0 was termed tempo, which represented a random age intercept to adjust for different timing of the pubertal growth spurt. (?, ?) used $\gamma = -\beta_1$ to represent the random age scaling which adjusts for the duration of the growth spurt. It was termed velocity.

The common characteristic function may be parametric or non-parametric function. (Lawton et al., 1972) used a linear spline and (Beath, 2007) used a natural cubic spline. A spline is a function defined by piecewise polynomial joined by knots. It is usually fitted over the range of the data. However, in SIM the data may be transformed by shifting and scaling outside the boundary of original data. A natural spline is constrained to be linear beyond the boundary knots, which allows fitting of transformed data beyond the boundaries (Beath, 2007). The internal knots can be set at the time points when visits occur. Individual knots can then be removed in order to produce a parsimonious model.

Growth curves can be fitted for each subject separately and population charactistics are obtained by summarizing individual parameter estimates. However, when each subject has only a small number of measurements, this apporach may produce large variability of estimates. A mixed-effects model is more advantages. (Beath, 2007) proposed to have the coefficients in the natural spline as fixed effects, which were constant across the population;

and the scale-location parameters as random effects, which described the variability among subjects. The non-linear mixed-effcts SIM can be expressed as

$$y_{ij} = \alpha_{0i} + e^{\alpha_{1i}} g\left(\frac{t_j - \beta_{0i}}{e^{\beta_{1i}}}\right) + e_{ij}. \tag{3.7}$$

where

$$\begin{pmatrix} \alpha_{0i} \\ \alpha_{1i} \\ \beta_{0i} \\ \beta_{1i} \end{pmatrix} \sim N(0, \Psi)$$

and

$$e_{ij} \sim N(0, \sigma^2)$$

Individual curves can be predicted using best linear unbiased prediction (BLUP) of random effects. Population curves can be obtained using fixed effects and setting random effects to zero.

Time independent covariates can be included in the model by modifying parameters, such as allowing the location parameters to be the sum of random effects and a linear function of the covariates and modifying the time directly for scale parameter. For example, including a single covariate z_i which affects size and growth rate, the model can be expressed as

$$y_{ij} = \alpha_{0i} + \alpha z_i + e^{\alpha_{1i}} g\left(\frac{e^{\beta z_i} t_j - \beta_{0i}}{e^{\beta_{1i}}}\right) + e_{ij}. \tag{3.8}$$

This model is very interpretable. A unit increase in z_i affects subjects size and growth rate by e^α and e^β.

This kind of models can be fitted using nlme package (?, ?) in R. It was recommended to fit nonlinear least-squares model for initial parameter estimates for fixed effects (Beath, 2007).

4   Survival submodel

Survival analysis is a set of methods for analyzing data where the outcome variable is the time to the occurrence of an event of interest. The outcome does not need to be death. "Survival" here means free of a particular event over time. The event may be heart attack, cancer remission, machine part failure, etc. The time-to-event can be measured in days, months, years, etc. The questions in survival analysis includes: do some variables (e.g., age, sex, treatment, etc) affect participants' chances of survival, what is the probability for an individual survives past a certain time, etc.

One of the major differences between survival analysis and ordinary linear regression is that the latter is able to handle censoring of observations. Censoring occurs when some information is available for survival time but the information is not complete.There are three types of censoring: right, left, and interval. Right censoring occurs if an individual experience the event after the observation time. Left censoring occurs an individual experience the event before the observation time. Interval censoring occurs if we don't know the exact event time but we know the event occurred within some given time period. Right censoring is the most common type. Censoring is often assumed to be non-informative in order to make use of the censored data. The underlying mechanisms to censoring should be unrelated to the event occurance so that the observations which are censored would have the same survival time distribution as if they were observed. Many sort of survival analysis, including Kaplan-Meier estimation and Cox model, can be invalidated if this assumption is not satisfied.

In survival analysis, two functions are used to describe the event times: survival function and hazard function. Let T be a non-negative random variable representing the time to the occurrence of the event with probability density function $f(t)$ and cumulative distribution function $F(t)$. The survival function gives the probability that a subject will 'survive' up to time t:

$$\begin{aligned} S(t) &= Pr\{T \geq t\} \\ &= 1 - F(t) \\ &= \int_t^\infty f(x)dx. \end{aligned} \tag{4.1}$$

The survival function is non-increasing. At time $t = 0$, $S(t) = 1$, which means the probability of surviving past $t = 0$ is 1. At time $t = \inf$, $S(t) = 0$, as time goes to infinity, the survival probability goes to 0. The hazard function gives the instantaneous rate of occurrence of the event at time t given that the event has not occurred before:

$$\begin{aligned} h(t) &= \lim_{dt \to 0} \frac{Pr\{t \leq T \leq t + dt | T \geq t\}}{dt} \\ &= \frac{f(t)}{S(t)} \end{aligned} \tag{4.2}$$

The survival and hazard functions provide equivalent characterizations of the distribution of survival time. If we know any one of the two functions, we can derive the other function:

$$\begin{aligned} S(t) &= exp\{-\int_0^t h(x)dx\} \\ h(t) &= -\frac{d}{dt}logS(t) \end{aligned} \tag{4.3}$$

Many quantities of interest can then be estimated from either the hazard or survival function.

4.1   Cox proportional hazards model

One of the most popular regression methods for survival analysis is Cox proportional hazards model (COX, 1972). It allows estimation of the relationship between survival and several variables through the hazard function without any consideration of the shape of the baseline hazard function.

In a Cox proportional hazards model, the relationship between predictors and the time-to-event is investigated through the hazard function. A Cox model assumes that

$$h(t) = h_0(t)exp(\beta_1 z_1 + \beta_2 z_2 + ... + \beta_p z_p) \tag{4.4}$$

where

- $h(t)$ is the expected hazard at time t for a subject.

- $h_0(t)$ is the hazard when all independent variables are euqal to zero. It is called the baseline hazard.

- $z_1, ..., z_p$ are independent variables and $\beta_1, ..., \beta_p$ are their coefficients. We assume the effect of the covariatesis the same at all times.

The above model assumes that effect of predicors on the hazard is multiplicative and it is constant over time. The basline hazard function $h_0(t)$ can have any shape as a functin of t so Cox proportional hazards model is called a semiparametric model.

A Cox model is usually interpreted using hazard ratio which is the ratio of the hazard rates corresponding to the conditions determined by two sets of predictors. The estimated coefficients in the Cox model represent the change in the expected log of the hazard ratio relative to a one unit change in respective covariates. For example, if there is a one unit increase in z_k while other covariates being fixed, then

$$
\begin{aligned}
log\left[\frac{h(t|z_k + 1)}{h(t|z_k)}\right] &= log(h(t|z_k + 1)) - log(h(t|z_k)) \\
&= \beta_k(z_k + 1) - \beta_k z_k \\
&= \beta_k
\end{aligned}
\tag{4.5}
$$

Therefore,$exp(\beta_k)$ is the hazard ratio associated with one unit incease in z_k. If the hazard ratio is less than 1,

then the predictor is associated with improved survival. If the hazard ratio is greater than 1, then the predictor is associated with decreased survival. A hazard ratio close to 1 means the predictor has no effect on survival.

One of the crucial assumptions in Cox model is that the hazards are proportional over time (so the name of proportional hazards model), which means the effect of covariates is constent over time. For example, with two different sets of predictors $\boldsymbol{z}$ and $\boldsymbol{z'}$,

$$\frac{h(t|\boldsymbol{z})}{h(t|\boldsymbol{z'})} = \frac{h_0(t)exp(\boldsymbol{z\beta})}{h_0(t)exp(\boldsymbol{z'\beta})} = e^{(\boldsymbol{z-z'})\boldsymbol{\beta}}, \tag{4.6}$$

which is independent of time for all t. Violation of this assumtpion may result questionable result so this assumption must be checked. Several approaches are available to assess the proportionality assumption, some methods involve graphical tools and others are built on statistical tests. For categorical variables, "log-log" transformation of the Kaplan-Meier survival curves can be plotted for different categories. If the proportionality assumption holds, the curves should not intersect with each other and roughly be parallel over time. A plot of Schoenfeld residuals versus time can be used as a graphical diagnostic for continuous predictors (Grambsch & Therneau, 1994). Under proportional hazards, the Schoenfeld residuals should scatter around zero randomly. A statistical test which correlates scaled Schoenfeld residuals with time is also available to test whether the slope of residuals on time is zero or not (Grambsch & Therneau, 1994). A non-zero slope means that the proportional hazard assumption has been violated. Violation of the proportional hazard assumption may be resolved by fitting a stratified Cox model in which the baseline hazard function can be different in stratums or including time-varying covariates in the model (George, Seals, & Aban, 2014).

Suppose we have n subjects with survival time described by a survivor function $S(t)$ with a hazard function $h(t)$ and density $f(t)$. Let t_i denote the observed time for subject i and d_i denote the indicator which equals to 1 if the subject is not censored and d_i equals to 0 otherwise. For non-censored subjects, t_i is the event time so the likelihood function can be expressed as

$$L_i = f(t_i) = S(t_i)h(t_i). \tag{4.7}$$

For censored events, we only know that the subject i is still free of event at time t_i so the likelihood is just $S(t_i)$. Both types shares the same survival function under non-informative censoring. Therefore the likelihood can be written as

$$L = \prod_i L_i = \prod_i S(t_i)h(t_i)^{d_i}. \tag{4.8}$$

Ordinary likelihood approaches cannot be applied to estimated parameters in Cox model since the baseline hazard

function $h_0(t)$ is unspecified. (COX, 1972) developed a 'conditional likelihood' to estimate the parameters. The approach was later formalized and renamed to 'partial likelihood' in (Cox, 1975). Cox's analysis goes conditionally under the set of $\{(t_i)\}$ in which failure occurs. Assuming there is no ties in event time, let $t_{(1)} < t_{(2)} < ... < t(n)$ be the ordered observed failure times and $R_{t_{(i)}}$ be the risk set of subjects at time $t_{(i)}$. For a specific failure time $t_{(i)}$, conditionally on the fact that a subject failed and risk set $R_{t_{(i)}}$, the probability of the observed event can be expressed as

$$\frac{exp(\boldsymbol{z}'_{(i)}\boldsymbol{\beta})}{\sum\limits_{l \in R_{t_{(i)}}} exp(\boldsymbol{z}'_{(l)}\boldsymbol{\beta})} \tag{4.9}$$

Multiplying factors of all failure times results the partial likelihood.

$$\sum_{i=1}^{n} \left\{ exp(\boldsymbol{z}'_{(i)}\boldsymbol{\beta}) - log \sum_{l \in R_{t_{(i)}}} exp(\boldsymbol{z}'_{(l)}\boldsymbol{\beta}) \right\} \tag{4.10}$$

(COX, 1972) obtained maximum-likelihood estimates of the above using the usual iterative way. In this approach, the baseline hazard function $h_0(t)$ is considered to be a nuisance function and removed completely from the inference process. (Efron, 1977) showed that the Cox's partial likelihood approach was asympotically equivalent to the likelihood approach based on all the data under very realistic conditions.

4.2 Extension: time-dependent covariates

4.3 aft?

5 Classical approaches to joint modeling

A joint model is comprised of two linked submodels: a model for event occurrence and a model for longitudinal measurements. There are three major ways to link these two parts together: (1) a simple approach using the observed longitudinal measurements as time-dependent covariates in a survival model; (2) a two-stage approach where the longitudinal process is first fitted separately, then the longitudinal fitted values are used as covariates in the survival model; (3) a shared random effects approach where in the models for the longitudinal and time-to-event likelihoods.

Suppose longitudinal measurements and time-to-event data are available for m subjects. Let $\boldsymbol{y}_i = \{y_i(t_{ij}), j = 1, ..., n_i\}$ denote the longitudinal responses for the ith subject at times $\{t_{ij}, j = 1, ..., n_i\}$. Denote T_i^* to be a true failure time, C_i to be a censoring time, $T_i = min(T_i^*, C_i)$ to be an observed failure time and $\delta_i = I(T_i^* \leq C_i)$ to be an event indicator.

5.1 A survival model with time-dependent covariates

One of the simplest way to incorporate association between the longitudinal and time-to-event processes is to include the observed longitudinal measurements as time-dependent covariates in a survival model. The hazard of failure at time t for patient i is modelled as (Sweeting & Thompson, 2011):

$$h_i(t) = h_0(t)exp(\alpha y_i(t) + \boldsymbol{x}_i^T(t)\boldsymbol{\beta}), \tag{5.1}$$

where $h_0(t)$ is a baseline function, α is a coefficient for the observed longitudinal measurement and $\boldsymbol{\beta}$ is a set of regression parameters for other explanatory variables $\boldsymbol{x}_i$.

There are several drawbacks of this approach. As discussed by (Sweeting & Thompson, 2011), (A. A. Tsiatis & Davidian, 2004), (Yu, Law, Taylor, & Sandler, 2004) and many other papers, in order to plug-in the longitudinal measurements into the survival model and use partial likelihood to estimate the hazard rate, we must know the values of the time varying measurements for all subjects at risk at any failure time point. However, longitudinal measurements typically are only collected at certain examination time points t_{ij}, and thus no measurements exist for subjects at risk when a failure occurs between scheduled visits. One popular approach to address this difficulty is to replace the unobserved, true longitudinal measurement by the observed, most recent value. Hence the hazard function become

$$h_i(t) = h_0(t)exp(\alpha y_i(t_{ij}) + \boldsymbol{x}_i^T(t_{ij})\boldsymbol{\beta}) \tag{5.2}$$

$t_{ij} \leq t < t_{i,j+1}$. Sweeting and Thompson found that naively using the observed data as covariates led to severe underestimation (Sweeting & Thompson, 2011). (Gail, 1981) adapted this approach to use the most recent value only if it was measured within certain time range prior to the failure time, otherwise the subject was excluded for that event. This substitution approach is problematic since it does not account for measurement error of the longitudinal value (A. Tsiatis, Degruttola, & Wulfsohn, 1995). The estimated relative risk in the Cox model will be underestimated due to measurement error (Prentice, 1982). This problem is exacerbated if visit times are sparse.

5.2 A two-stage model

An alternative approach is a two-stage model. In the first stage, a model, usually a random effect model (Laird & Ware, 1982), is fit to the longitudinal data and the estimators and imputed longitudinal values are obtained for each time point. Then in the second stage, these imputed values are inserted into the survival model as covariates. This approach estimates the regression coefficient in survival model using approximations of the true longitudinal values. (A. Tsiatis et al., 1995) adapted this model in a HIV trial in order to investigate the association between CD4 counts and survival. A linear mixed model was fit to the longitudinal trajectory and then the fitted values

were treated as true values and inserted in the Cox model as a time varying covariate. Two-stage models were also investigated by (Dafni & Tsiatis, 1998) and (Bycott & Taylor, 1998).

This approach is better than naively using the observed longitudinal data in the Cox model. This approach allows interpolating longditudinal values between visits and it also adjusts measurement error by using the entire covariate history (Faucett & Thomas, 1996). Simulation shows that this approach yields improved efficiency of statistical inferences and reduced bias of the parameter estimates in the Cox model (Dafni & Tsiatis, 1998),(Bycott & Taylor, 1998). The two-stage approach is also quick to implement using standard software (Sweeting & Thompson, 2011). However, the two-stage approach also have some drawbacks. No survival information is used in the longitudinal submodel, which could result in efficiency loss. Also, this approach does not account for event-dependent drop-out. NOT FINISHED

(Sweeting & Thompson, 2011) conducted extensive simulations in order to assess the perfomrances of several approaches. The models were fitted using a linear mixed-effects model for the longitudinal process, and a parametric model with constant baseline hazard or a semi-parametric Cox model for the survival data. They found that two-stage approach generally underestiamted the asssociation parameter and had poor coverage properties.

5.3   Shared random effects approach

model detail

reduce bias

improve efficiency

correct for informative drop out

6   Association Structures

6.1   Time-dependent associations

current;

lag;

rate/derivate

6.2   Time-independent associations

random intercept/ slope

random change point

7 Estimation?

8 Application examples

better inference on longitudinal process;

better inference on survival process;

association between the two, surrogate marker.

References

Beath, K. J. (2007). Infant growth modelling using a shape invariant model with random effects. *Statistics in medicine*, *26*(12), 2547–2564.

Bycott, P., & Taylor, J. (1998). A comparison of smoothing techniques for cd4 data measured with error in a time-dependent cox proportional hazards model. *Statistics in medicine*, *17*(18), 2061–2077.

COX, D. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, *34*(2), 87–22.

Cox, D. R. (1975). Partial likelihood. *Biometrika*, *62*(2), 269–276.

Dafni, U. G., & Tsiatis, A. A. (1998). Evaluating surrogate markers of clinical outcome when measured with error. *Biometrics*, 1445–1462.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, 1–38.

Efron, B. (1977). The efficiency of cox's likelihood function for censored data. *Journal of the American statistical Association*, *72*(359), 557–565.

Faucett, C. L., & Thomas, D. C. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: a gibbs sampling approach. *Statistics in medicine*, *15*(15), 1663–1685.

Gail, M. H. (1981). Evaluating serial cancer marker studies in patients at risk of recurrent disease. *Biometrics*, 67–78.

George, B., Seals, S., & Aban, I. (2014). Survival analysis and regression models. *Journal of Nuclear Cardiology*, *21*(4), 686–694.

Grambsch, P. M., & Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, *81*(3), 515–526.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, *72*(358), 320–338.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 963–974.

Lawton, W., Sylvestre, E., & Maggio, M. (1972). Self modeling nonlinear regression. *Technometrics*, *14*(3), 513–532.

Prentice, R. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, *69*(2), 331–342.

Sweeting, M. J., & Thompson, S. G. (2011). Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biometrical Journal*, *53*(5), 750–763.

Tsiatis, A., Degruttola, V., & Wulfsohn, M. (1995). Modeling the relationship of survival to longitudinal data measured with error. applications to survival and cd4 counts in patients with aids. *Journal of the American Statistical Association*, *90*(429), 27–37.

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

Yu, M., Law, N. J., Taylor, J. M., & Sandler, H. M. (2004). Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica*, 835–862.