THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

A shared frailty model for multivariate longitudinal data on adverse event of radiation therapy

OPEN ACCESS

A time-to-event approach for dealing with multivariate-longitudinal data: with a special application to adverse event of radiation therapy

Sung Won Kim[*,1,*], **Martin Schumacher** [1,*], and **Nicole H. Augustin** [2]

[*] Contributed equally to this work

[1] Institute of Medical Biometry and Statistics, Medical Center-University of Freiburg, Freiburg, 79104, Germany

[2] School of Mathematics, University of Edinburgh, James Clark Maxwell Building, The King's Buildings, Edinburgh EH9 3FD, UK

Oral mucositis is an inflammatory adverse event when treating head and neck cancer patients with radiation therapy (RT). The severity of its occurrence is believed to mainly depend on its site and the distribution of a cumulative radiation dose in the mouth area. The motivating study investigating differences in radiosensitivities (mucositis progression) at distinct sites where the severity of mucositis is assessed regularly at eight distinct sites on an ordinal scale results in *multivariate longitudinal data* and thus poses certain challenges. To deal with the multivariate longitudinal data in this particular setting, we take a time-to-event approach focusing on the first occurrence of severe mucositis at the distinct sites using the fact that the site-specific cumulative radiation dose thought to be the main driver of oral mucositis develops over time. Thereby, we may address multivariate longitudinal processes in a simpler and more compact fashion. In this article, to find out differences in mucositis progression at eight distinct sites we propose *a shared frailty model* for multivariate parallel processes within individuals. The shared frailty model directly incorporating "process indicators" as covariates turns out to adequately explain the differences in the parallel processes (here, mucositis progressions at distinct sites) while taking individual effects into account. The parallel result with the one from the previous analysis based on the same data but conducted with an alternative statistical methodology shows adequacy of the proposed approach.

Key words: Shared fraily model; Multivariate longitudinal process; Multivariate counting process; Parallel processes;

Supporting Information for this article is available from the author or on the WWW under http://dx.doi.org/10.1022/bimj.XXXXXXX

1 Introduction

Oral mucositis (briefly, mucositis) is an inflammatory adverse event when treating head and neck cancer patients with radiation therapy (RT). The motivating study seeks to investigate differences in radiosensitivities at eight distinct sites in the mouth area. Since the radiosensitivities can be represented by the severity of mucositis, in the study mucositis progression is assessed through time at eight different sites in the mouth area, i.e., the data set consists of longitudinal observations for the eight dimensional mucositis progressions along with other covariates, most importantly site-specific cumulative (radiation) dose and site-specific (radiation) dose fraction. The analysis of the *multivariate longitudinal* data taking other covariates into account, however, poses certain challenges.

Longitudinal data is characterized by repeated measurements on individuals through time. While such a data set enables us to directly capture the within-individual variability through time and to relate them

*Corresponding author: e-mail: kim@imbi.uni-freiburg.de, Phone: +49-761-50318-636

to inter-individual differences in covariates of interest, it also poses challenges of correlations within an individual, which are to be taken into account to draw valid inferences (Digg *et al.*, 2002). See also Digger *et al.* (2002), Laird *et al.* (1982), Jennrich and Schluchter (1986), Liang and Zeger (1986), Zeger and Liang (1986), and Long (2011) for an overview. In a longitudinal data set, it is not uncommon that multivariate outcomes are observed for an assessment time point. Such a multivariate longitudinal data set introduces another kind of correlation - between multivariate outcomes for each assessment time point. For example, Verbeke *et al.* (2014) reviewed existing methodologies addressing such multivariate longitudinal data.

Clearly, the multivariate longitudinal analysis becomes very complicated as soon as there are more than two outcomes at an assessment time point because it requires an approach which can address two different kinds of correlations: (i) within individuals over longitudinal assessments and (ii) between multivariate processes.

To this end, we notice two particular features of the motivating data set. Firstly, the site-specific cumulative dose thought to be the main driver of oral mucositis progression during RT develops proportionally to time, i.e. there is a rough relationship of (site-specific cumulative dose) $\approx$ (time) $\times$ (site-specific dose fraction). This particular feature of the mucositis progression allows us to directly incorporate 'time' into the model. Secondly, the main clinical interest lies in the occurrence of severe mucositis defined as mucositis grade ≥ 3. These two features allow us to take a time-to-event approach by incorporating a "site-specific dose fraction" and "time" effects explicitly into the model instead of a "site-specific *cumulative* dose" and by focusing on the first occurrence of severe mucositis at the eight distinct sites. By taking a time-to-event approach we can take advantage of two useful properties: (i) Possible correlations due to repeated measurements of longitudinal data are addressed, and also (ii) the model is able to consider the time aspect explicitly in the model; we can thus estimate how the hazard of experiencing the event (severe mucositis at a specific site) evolves over time.

Thus, by adopting a time-to-event approach we can deal with *multivariate time-to-event data.* There are several possibilities to analyse multivariate time-to-events data: for example, a marginal model, a fixed effects model, a stratified model, a copula model, a shared frailty model, or even a model ignoring dependence could be employed. Among them, the shared frailty model is the most popular choice, owing to its simplicity and easy interpretation without having to specify a complicated dependency structure, the frailty term in the shared frailty model would explain the excessive individual effects which cannot be explained by covariates but still exist among the multivariate outcomes within individuals.

We are primarily motivated by the mucositis progression (an indicator of radiosensitivities which can be seen as a "process") at eight distinct sites. Thus, to find out the significant differences in the radiosensitivities at eight distinct sites we investigate in this article a model which describes *multivariate parallel processes* with a simple correlation structure. Here, the main interest is whether there are significant differences among the multivariate parallel processes. To this end, other possible confounders as well as the possible correlations (within individuals and between the processes) should be taken into account in the model.

This article is structured as follows: Section 2 introduces the motivating study in details. In Section 3 we propose a shared frailty model for multivariate parallel processes. The model incorporates "process effects" as indicator variables to explain the distinctions between the parallel processes while taking account of confounders and individual effects (frailties). In Section 4 the proposed model is applied to the motivating study, and it is shown that the shared frailty model adequately accounts for the multivariate parallel processes. The result is compared with that of an extended proportional odds model based on the same data set, which is presented in Augustin *et al.* (2017). While both models give fairly coherent results, the similarities and the differences of the two models are further investigated. In Section 5, we discuss two different kinds of possible correlations in the multivariate parallel processes; (i) within individuals and (ii) between the processes.

2   Motivating Study

(Oral) mucositis is a common adverse event of cancer treatments, such as radiation therapy (RT). It is a condition characterized by pain and inflammation of the surface of the mucous membrane (mucosa) and commonly occurs especially when treating head and neck cancer patients with RT. When the head and neck area is treated, temporary soreness and ulceration commonly occur in the mouth and throat area. Such damage to the epithelial surfaces like skin or oral mucosa is called (oral) mucositis. Although it is not a fatal adverse event of RT, it accompanies serious troubles with ingestion and is associated with pain, and in the worst case, the interruption of RT has to be considered. Although modern RT aims to minimize occurrence of the mucositis on healthy tissues by using a specialized treatment planning software, its occurrence is still inevitable. The mucositis that occurs during the treatment mainly depends on the amount of the radiation dose but also on the site irradiated, the type of radiation, the fractionation, the concurrency of chemotherapy, the patient and so on.

Interestingly, the healthy tissues in the mouth are not uniform in their structure and the rates of the damage and recovery from mucositis depend on the turnover rate of epithelial cells. So, radiation oncologists are interested in examining how sensitive each part (site) of the mouth area is to RT. Therefore, to promote a better RT planning minimizing mucositis occurrence, radiation oncologists in the radiation oncology department of the Medical Center of the University of Freiburg conducted several studies (only two of them are included in this work) to ascertain the different radiosensitivities of eight distinct sites to RT in the mouth area. While treating the head and neck cancer patients with RT, radiation oncologists have longitudinally assessed mucositis progression on eight distinct sites of the treated area in order to answer the following two questions *(i) Are there significant differences in mucositis development at the eight distinct sites?* and *(ii) if any, how are they different?* Answering these questions is the aim of this work. The study on mucositis progressions at eight distinct sites in the mouth area is unprecedented and we expect that this work can provide the radiation oncologists with more precise empirical knowledge on radiosensitivities of eight sites in developing mucositis, thereby being able to avoid more sensitive sites in the planning stage to minimize the severity of mucositis.

The site-specific cumulative dose is thought to be the main driver of mucositis and increases proportionally to the lapse of time, which can be seen in Fig. 2. Being a main driver of mucositis, the site-specific cumulative dose could be a main confounder in clarifying the effects of distinct ”sites” and should be taken into account in the model formulation. In Fig. 2, boxplots show the distribution of the site-specific cumulative doses and the gray dots indicate the amounts at which point severe mucositis were first observed. Apparently, there is a tendency of more (severe) mucositis occurrences at sites with higher site-specific cumulative doses. However, it can be seen that also other factors play a role why (severe) mucositis occurs. Our hypothesis is that one of these factors could be the effects of the distinct ”sites”.

Observational data were collected on a total of 75 patients receiving RT for head and neck cancer between September 2006 and February 2012 in the Radiation Oncology Department, Medical Center, University of Freiburg. The data set consists of two studies that have different protocols, and one has to take potential differences between the studies into account in the analysis. Three-dimensional radiation planning techniques and standard fractionation ($5 \times 1.8 \sim 2.2$ Gray/Week, mostly 2 Gy per week) were followed, according to individual planning. The total dose to the tumour was from 60 to 80 Gy, thus for the duration of six or seven weeks of RT in general. Any RT interruption was not allowed in principle. For every individual the planning (the direction and the amount of dose) is varying depending on the tumour site as well as its size, so the amount of the dose reaching each site is different for each individual. To achieve the main aim of the study, the oral mucosa was assessed at eight sites in the mouth area: hard palate (`hard`), soft palate (`soft`), tongue, mouth floor (`floor`), upper lip (`up`), lower lip (`low`), left cheek (`left`), and, right cheek (`right`), based on the five level NIH-CTC oral toxicity scale ((0) no mucositis, (1) erythema, (2) ulcer, (3) confluent ulcer, (4) necrosis). The grade is given according to its severity: grade 0 indicates mucositis free, grade 1 and 2 are regarded as minor state, and grade 3 and 4 are categorized as *severe mucositis* (Henke *et al.*, 2011), with severe mucositis being of our main interest. Assessments were conducted twice weekly,
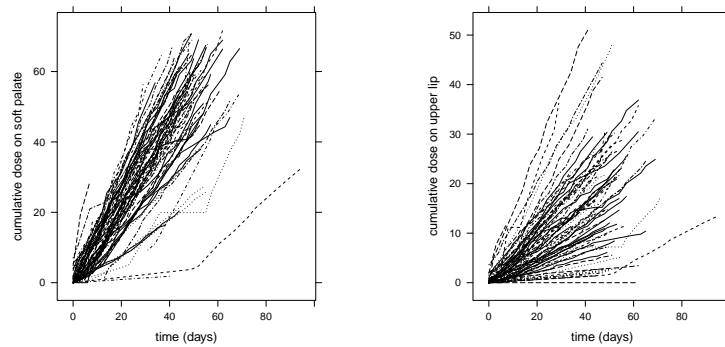
Figure 1   Site-specific cumulative dose increases proportionally to time. (Left) Cumulative dose on soft palate, (Right) Cumulative dose on upper lip.
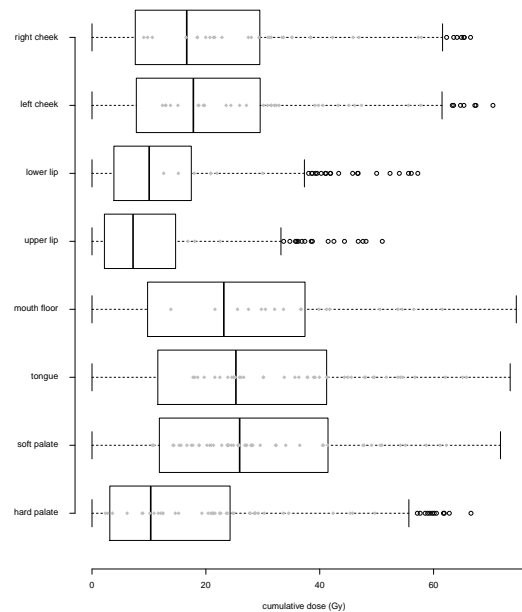
Figure 2   The eight box plots describe the distributions of the cumulative site-specific radiation dose for each site. The gray dots indicate the corresponding values of the site-specific cumulative dose at which severe mucositis occur. (Augustin *et al.*, 2017)

at least at three day intervals, throughout the RT duration. At each assessment the cumulative dose was recorded so that with the median value of the percentage of the site-specific dose *the site-specific cumulative dose* (=`cumulative dose` × `median percentage of site-specific dose`) could be derived. In addition all relevant patient characteristics were recorded, including age, sex, study (the data set consists of the data sets from two different studies). All study procedures were reviewed by the local ethics committee and all patients provided their written consent.

3 Methods

3.1 Notation

Let $\boldsymbol{N}$ be a multivariate counting process with components N_{il} where $N_{il}(t)$ is the number of observed events in $[0, t]$ for individual i, $i = 1, \ldots, n$, and process l, $l = 1, \ldots, k$, i.e.,

$$\boldsymbol{N} = (N_{il}; i = 1, \ldots, n; l = 1, \ldots, k).$$

We assume that the counting process N_{il} jumps from 0 to 1 so it takes value of $\{0, 1\}$ with intensity

$$\alpha_i^{(l)}(t) = Z_i Y_{il}(t) \alpha^{(l)}(t)$$

where $\alpha^{(l)}$ is a common process for process l, and $Y_{il}(t)$ is an observable predictable process for individual i, and process l whose value at time t is known at latest infinitesimally before t. Here, the inclusion of an individual frailty variable Z_i is necessary since we assume the multivariate counting processes share the same frailty variable and model positive statistical dependence among them (Andersen *et al.*, 1993).

Then, for individual i the data consists of

$$\{(N_{i1}(t), Y_{i1}(t), X_{i1}(t)), \ldots, (N_{il}(t), Y_{il}(t), X_{il}(t)), \ldots, (N_{ik}(t), Y_{ik}(t), X_{ik}(t))\}$$

where $X_{il}(t)$ is a vector of covariates for individual i and process l. It can be also time-dependent.

3.2 Strategy

Motivated by the fact that the site-specific cumulative dose thought to be the main driver of oral mucositis evolves proportionally to time, we employ a time-to-event approach focusing on the first occurrence of severe mucositis. This approach supposes that the interest lies in a specific event occurrence, at which time point a longitudinal process exceeds a pre-specified value. From this point of view, the multivariate longitudinal data can be analysed by a multivariate time-to-event approach. By employing this approach, time can be modelled explicitly, and one can see how the instantaneous (or cumulative) hazard to experiencing the event evolves over time.

Before moving to the model proposal, we list several challenges in the data set. Multivariate outcomes within individuals might show (i) correlations within individuals. And considering our primary aim it is necessary to find out the possible (ii) effects leading to differences among the multivariate outcomes, i.e., "process-specific effects". Furthermore (iii) possible confounders such as the amount of the site-specific (radiation) dose are to be taken into account.

Considering point (iii), a hazard-based regression approach is suitable where we start with the best-known one: the proportional hazard model (Cox model) (Cox, 1992). To address challenges (i) and (ii), we propose a shared frailty model for the *multivariate parallel processes*: the *parallel processes* (Cook and Lawless, 2018) are generated by similar mechanisms whereas they can differ by the "process-specific (fixed) effects" that might possibly lead to different outcomes. To sum up, the multivariate parallel processes still share all the covariates other than the process-specific effects. By directly incorporating process-specific effects, it is possible to estimate how differently each process evolves, while retaining the basic structure which all the parallel processes share. Here, the shared frailty term in the model would allow the model to account for possible individual effects which are shared among the parallel processes within individuals. In this way, the multivariate processes are simultaneously incorporated in one model with the process-specific effects while possible individual effects are accounted by the frailty term.

Note that so far the shared frailty model has generally been employed where the frailty term was incorporated to address individual effects penetrating multivariate processes within individuals and those multivariate processes were identical (not only parallel). Since the processes are not distinct from each other, in such a general shared frailty model the differences of the multivariate processes can be adequately

explained by the frailty term without the need of other effects but common (shared) effects. However, the shared frailty model for the multivariate *parallel* processes can be more complex. If the multivariate parallel processes are not adequately accounted by the frailty term, the correlations between the multivariate processes are also to be addressed in the model. There have only been a few applications where the correlated multivariate processes were considered through the shared frailty model. For example Lie *et al.* (2004) proposed a joint modeling of recurrence and mortality processes with a frailty term shared by the intensities of the two processes. However, those works do not address the challenges of our data set.

The proposed shared frailty model approach is appropriate for the above-mentioned motivating study where the mucositis progression is assessed at eight sites and the event of experiencing severe mucositis (of grade larger than or equal to 3) for each site can be regarded as a counting process. Then there are eight-dimensional multivariate parallel counting processes with site-specific (process-specific) covariates. Site-specific processes are modelled by directly incorporating each *site* and the corresponding estimates would explain the effects of the site-specific processes.

3.3　Shared frailty model approach for multivariate counting processes

For univariate time-to-event data, a frailty term explains additional (individual) heterogeneity which cannot be explained by covariates in the model. On the other hand, for the multivariate counterpart the frailty term can also be used to explain an individual effect or correlations of multiple events within an individual. Assume that there is a frailty term Z_i for individual i. It measures the individual specific risk level and the time-to-events within the individual are related through the frailty: assuming that the frailty term sufficiently explains the possible individual effect, conditional on the frailty term the multivariate time-to-events are independent of each other.

We further assume that the hazard rate of an individual is given as the product of an individual frailty and a basic rate $\alpha(t)$ (Aalen *et al.*, 2006). A proportional hazard model with the proportional frailty term shared by all the times-to-events within an individual is called the *shared frailty model*. Unless we are interested in the values of individual frailty terms, the frailty term is usually regarded as a nuisance parameter. In general, interest lies in the degree of heterogeneity between the individuals which is measured by the variance of the frailty terms: a variance of the frailty terms far larger than zero indicates dependencies among the time-to-events within an individual, implying a necessity of the frailty term in the model.

We suppose the situation in which each of the parallel processes is distinguished by its "process-specific (fixed) effect". Such a situation can be observed, for example, in a study on biological progressive processes involving paired or multiple organ systems. Then, the question that naturally arises would be whether such a "process-specific effect" of each process leads to any significant difference in its outcome. While in a general shared frailty model all the covariates are shared by multivariate outcomes within individuals, in our specific setting also the "process-specific covariates" which are not shared by multivariate outcomes are to be incorporated in the model. Such process-specific covariates enable us to estimate the "process-specific" effects which possibly make distinctions between the multivariate outcomes within individuals.

To this end, we suppose the k-dimensional multivariate counting process for the i-th individual

$$\{N_{i1}(t), \ldots, N_{ik}(t)\}$$

where the k processes are assumed to be *parallel*, i.e., the multivariate processes have the same number of states and the identical possible transitions. Then, for individual i, $i = 1, \ldots, n$, the data consists of

$$\{(N_{i1}(t), Y_{i1}(t), X_{i1}(t)), \ldots, (N_{il}(t), Y_{il}(t), X_{il}(t)), \ldots, (N_{ik}(t), Y_{ik}(t), X_{ik}(t))\},$$

where $\mathbf{X}_{il}(t)$ is a vector of $(X_{il1}(t), \ldots, X_{ilm}(t), \ldots, X_{ilp}(t), (0, \ldots, 1, \ldots, 0)^T)$. Here the last component of the data is a process-specific length-$((k-1))$ vector of covariates of indicator variables: the entry of the vector takes a value 1 only for the corresponding process, otherwise a 0. For the reference process

$l = 1$ its vector consists of all zero values. Note that any covariates $X_{ilm}(t)$ might contain process-specific values.

In this setting, the hazard for the k-dimensional multivariate processes can be simultaneously formulated within a shared frailty model framework in a compact manner: for individual i, $i = 1, \ldots, n$, process l, $l = 1, \ldots, k$, and counting process $N_{il}(t)$, the hazard rate of the shared frailty model can be written as

$$\alpha_i^{(l)}(t) = \alpha_0(t) Z_i \exp(\mathbf{X}_{il}(t)^T \boldsymbol{\beta}_l), \tag{1}$$

where Z_i is the frailty term for individual i, $\mathbf{X}_{il}(t)$ is a $(p + k - 1)$-length vector whose values are allowed to be varying by process and $\boldsymbol{\beta}_l$ is a length $(p + k - 1)$ vector that contains site-specific effects as well. We assume that Z_i is independently and identically distributed with gamma or log-Gaussian distribution, with density function, say $p(z; \theta)$, having the variance θ and mean 1. (We may test the null hypothesis of $H_0 : \theta = 0$, meaning no frailty. The larger θ is, the more varying the individuals are, and the more crucial the frailty term is.)

Again note that a general shared frailty model assumes that all the covariates are "shared" within individuals, i.e., the same applies for the multivariate processes within individual. Meanwhile, in our specific setting, other than the "shared" covariates for all the k processes, there are the "process-specific" covariates which are only applicable to the corresponding process.
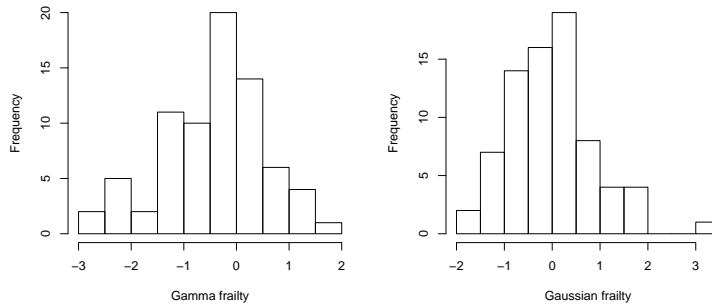
3.4 Estimation

The estimation is done using the `coxph` function in the `Survival` R-package (Therneau, 2020). For a description of the estimation process one refers to Therneau and Grambsch (2000) on which the `coxph` function is based. Therneau *et al.* (2003) shows that an estimation of the frailty model can be based on a penalized Cox model, where, instead of regarding a frailty term as an unobserved covariate, one treats the frailty as an additional coefficient which is constrained by a penalty function added to the log partial likelihood. The maximization of the penalized partial likelihood for fitting the Cox models can be readily done: the score equations for the fixed covariates are identical to those of the ordinary Cox model, and so only the score equations for the frailty terms are to be solved. The score equations for the frailty terms depend on the choice of the frailty distribution which defines the penalty function.

Since the `coxph` function uses the penalized likelihood method on the view of the penalized fitting procedure, the variance of the frailty terms θ plays a role as a "penalty parameter", or a "tuning parameter", of the computation. We might control the degrees of the penalty, either by fixing the value of the variance of the frailty or by assigning the degrees of freedom, but we let the function choose the degrees of the penalty: the function chooses the degrees of the penalty based on the variance of the frailty. It chooses the large penalty (in the case of gamma or Gaussian frailty, $1/\theta$, i.e., small value of θ) for the less varying frailties and the small penalty (large value of θ) for the more varying frailties.

By choosing the penalty function as $g(\boldsymbol{z}'; \theta) = (1/\theta) \Sigma [z_i' - \exp(z_i')]$, where $z = \exp(z')$, the penalized Cox model is equivalent to the gamma frailty model. For the computation of the gamma frailty model `coxph` uses inner and outer loops. In the inner loop the Newton-Raphson method is used to solve the penalized partial log likelihood for a fixed θ, and the outer loop finds $\hat{\theta}$ which maximizes the profile likelihood for θ. Therneau and Grambsch (2000) derived the log profile likelihood for θ, then θ can be estimated by maximizing the log profile likelihood. The Gaussian frailty model is equivalent to a penalized Cox model with the penalty function $g(\boldsymbol{z}'; \theta) = (1/2\theta) \Sigma \left[z_i'^2 \right]$. In the case of the Gaussian frailty model, in the outer loop θ is chosen based on an approximated residual maximum likelihood (REML) and for a given θ the inner loop solves the estimating equations for $\boldsymbol{\beta}$ and $\boldsymbol{z}'$ using the Newton-Rhapson iteration. See Ripatti and Palmgren (2000) and Therneau and Grambsch (2000) for the details.

Figure 3 Histograms of the estimated individual frailty. The frailty is assumed to be distributed as (Left) log gamma, and (Right) Gaussian. Both densities are scaled to have mean zero and variance one.

4 Application

The shared frailty model (1) for multivariate parallel time-to-event processes is applied to the motivating data set from Sec. 2. It is evident from Fig. 4 that the individuals have varying frailties: some individuals are more frail and some others are less frail than others, i.e., experiencing events (severe mucositis) are influenced by individual frailty. Note that, if the `site` effects were not of our main interest, it would be possible to assume the frailty for the `site` effects, i.e., frailties of the sites. In our case, individual effects are nuisance for our purpose so that individual effects are better to be regarded as frailty effects. Furthermore, in the survival analysis context frailty generally indicates frailty of the individual.

The assumption that the frailty term sufficiently explains the possible individual effect, conditional on the frailty term the multivariate time-to-events, is deemed realistic in this application; the frailty term represents the "basic radiosensitivity" which is a specific feature of a particular patient. The frailty structure might be elaborated by, for instance, incorporating history of events in other parallel processes. This so-called proportional hazard (PH) Markov model enables us to characterize the dependence between each pair of the processes and to possibly explain pairwise interrelationship that cannot be taken into account through the frailty term. However, such overly parameterized model comes with the cost of non-valid estimates and practical inefficiency. See discussion in Sec. 5.

4.1 Application to adverse events of radiation therapy on eight distinct sites

The shared frailty model for individual i, $i = 1, \ldots, n$, site l, $l = 1, \ldots, 8$, and the counting process $N_{il}(t)$ takes the form of

$$\alpha_i^{(l)}(t|x) = z_i \alpha_{s0}(t) \exp\left(\boldsymbol{\beta}_{\text{site}.l}^T \mathbf{x}_{site.l} + \beta_{\text{dosl}} x_{\text{dosl}.il}\right), \quad s = 1, 2 \tag{2}$$

where $\alpha_{s0}(t)$ is the baseline hazards stratified by `study` effects. Although the `study` effect is a nuisance variable, it has a significant effect. Thus, it makes sense to include it in the baseline hazard. z_i is the frailty term for individual i, and $\beta_{\text{dosl}.\text{l}}$ is the coefficient of site-specific dose fraction for site l. Other variables of `sex`, `age`, and `cumdosl` (site-specific cumulative dose) turned out to be insignificant; as long as `dosl` is in the model `cumdosl` is not significant anymore. (2) can be equivalently written as

$$\alpha_{ij}^{(l)}(t|x) = \alpha_{s0}(t) \exp(\boldsymbol{\beta}_{\text{site}}^T \mathbf{x}_{\text{site}.l} + \beta_{\text{dosl}} x_{dosl.il} + z_i' w_{ij}), \quad s = 1, 2 \tag{3}$$

mainly for the estimation purpose. Here w_{ij} equals 1 if assessment j belongs to individual i, otherwise 0 and $z_i = \exp(z_i')$. Note that the frailty z_i in (2) follows gamma or log Gaussian distribution, i.e., Gaussian is random effect on the scale of the linear predictor.

4.1.1 Result

Based on (3), the shared frailty model is fitted assuming the gamma frailty distribution, i.e., $\exp(z_i')$ is gamma distributed, and the output is given in Table 4.1.1. We let the `coxph` function find an overall best variance θ which minimizes the Akaike Information Criterion (AIC) : as mentioned in Sec. 3.4 the outer loop finds the maximizer $\hat{\theta}$ of the profile likelihood for θ (Therneau and Grambsch, 2000). The resulting variance of the frailty is estimated to be 1.0798, indicating a fairly large biological effect; this implies an average $\exp(1.0798) = 2.9441$-fold difference in the risks between observations randomly chosen from different individuals (Therneau and Grambsch, 2000). The approximate Wald test statistics for the frailty term of 195.99 with 56.69 degrees of freedom indicates highly significant frailty terms, meaning strong dependencies of the event times among sites within an individual. It implies that the frailty term is not to be disregarded in the model. Notice that the shared frailty model estimates the relative risk within individuals whereas the marginal model (the shared frailty model is a conditional model in the sense that the coefficients of the shared frailty model are estimated, conditional on the frailty term of the corresponding individual) estimates the population averaged risk. Comparing to the marginal model -results not shown-, except for the `site` effects, in the shared frailty model all the estimates of the coefficients as well as the corresponding standard errors are inflated. With the frailty term being included in the model, the estimated values of the coefficients increase with the variance of the frailty term, and this is theoretically expected (Henderson, 1975): omitting the frailty term leads to an underestimation (attenuation towards zero) of the estimates when such individual effects actually exist, i.e., the variance of the frailty is large. And by including the frailty term such bias is corrected so the estimates are inflated. The addition of a frailty term rectifies the bias. However, at the price of some increase in variance over the marginal model which gives the larger variance of the estimates of the coefficients, i.e., larger standard errors. Such standard errors can be even larger if the variance of the frailty terms θ is unknown. The standard error estimated by the `coxph` function through the penalized Cox model is computed assuming that θ is fixed (Therneau and Grambsch, 2000), which is not the case in our application, and this leads to the underestimated standard errors.

Based on (3) the shared frailty model with Gaussian frailty is fit using the `coxph` function whose output is given in Table 4.1.1. There are no significant changes for the Gaussian frailty model from the gamma frailty fit. The estimated variance of the frailty term $\hat{\theta}$ again gives a highly significant value of the approximate Wald test statistic. Here again the variance, or equivalently, the degrees of freedom, was not specified, and by default the `coxph` function estimated the variance of the frailty term using the REML criterion. Since the estimates of both the gamma frailty and the Gaussian frailty models in Table 4.1.1 provide parallel results, we focus on the gamma frailty model for further discussion on this application.

The estimates for the `site` effects in Table 4.1.1 are the values with respect to the reference site, `hard`, and those effects are also graphically illustrated on the right panel of Fig. 4.2.2. The most sensitive sites turn out to be `soft` and `hard` whereas `up` and `low` result in the two smallest values of the `site` effects, after adjusting for `dosl` and `study` effects. This finding is consistent with the previous analysis (Augustin *et al.*, 2017) using a different modelling approach but with the same data set, replacing the time factor with the site-specific cumulative dose (`cumdosl`). See Fig. 4.2.2. We will investigate this point in the following section.

4.2 Comparison with a previous approach

In the above-mentioned study based on the same motivating data set, no time-to-event approach was employed and the original mucositis grades, $r = 0, 1, 2, 3$ (here grade 3 includes both of grades 3 and 4 due to only a few observations of grade 4), were preserved, i.e., longitudinally observed mucositis grades in ordinal scale (Augustin *et al.*, 2017). The approach is termed *a flexible multivariate random effects proportional odds model*, hereafter *"FMRP"*, which is briefly explained in Sec. 4.2.1. The two models are based on different frameworks: the shared frailty model is based on the counting process approach whereas FMRP is an extension of a proportional odds model focusing on ordinal outcomes. In fact, although FMRP deals with longitudinal data, the "time" aspect is considered only indirectly through the covariate of the

Table 1 Estimates for the gamma frailty model and the Gaussian frailty model. Estimated coefficients for fixed coefficients, and corresponding standard errors, Chi-squared values, and p-values are presented. Estimated individual frailties are not shown, but their chi-squared values and p-value are given.

	gamma frailty			Gaussian frailty		
Variables	exp(coef)	se(coef)	p-value	exp(coef)	se(coef)	p-value
site soft	0.97	0.26	0.91	1.06	0.26	0.82
site tongue	0.54	0.28	0.03	0.62	0.27	0.08
site floor	0.15	0.33	<0.001	0.16	0.32	<0.001
site up	0.04	0.61	<0.001	0.04	0.61	<0.001
site low	0.10	0.42	<0.001	0.09	0.43	<0.001
site left	0.45	0.26	0.002	0.48	0.26	0.005
site right	0.42	0.26	<0.001	0.44	0.26	0.002
dosl	2.66	0.22	<0.001	2.24	0.19	<0.001
frailty(id)	Chi-sq=195.99		≈ 0		Chi-sq=243.95	≈ 0

site-specific cumulative dose (`cumdosl`), detailed in Sec. 4.2.2. Thus, it is interesting that the two models still lead to parallel results. In this subsection, a comparison between the two approaches is made.

4.2.1 A flexible multivariate random effects proportional odds model, FMRP

Motivated by the study introduced in Sec. 2, Augustin *et al.* (2017) proposed a novel flexible multivariate random effects proportional odds model (FMRP) that takes the longitudinal course of oral mucositis at different mouth sites into account. As in this article, the clinical aim of the study is to find out whether there are significant differences in radiosensitivities among the eight distinct sites in the mouth area. The FMRP is an extension of the proportional odds model for an ordinal response variable, the original mucositis grade scale. Besides addressing the ordinal multivariate response of the mucositis grades for each site and at each assessment time, the model includes random intercepts for individual effects and a nonlinear function of the site-specific cumulative (radiation) dose (`cumdosl`). Focusing on the difference in radiosensitivities on the sites, `site` effects are also directly incorporated into the model as covariates. Thereby, the model allows to test whether the radiosensitivity differs by site after having adjusted for site-specific cumulative dose. It is also possible to check whether and how the (nonlinear) effect of the site-specific cumulative dose differs by site.

Let R_{ijl} denote the mucositis grade in the range $0, ..., 3$ for individual i, assessment time index j, site l. Then, seeing the proportional odds model as a special case of the cumulative logit model for ordinal outcomes, where the proportional assumption is met, the FMRP model for mucositis grade $r = 0, 1, 2, 3$ is specified as

$$\text{logit}\left(\text{P}(R_{ijl} \leq r)\right) = \log \frac{\text{P}(R_{ijl} \leq r)}{\text{P}(R_{ijl} > r)} = \gamma_r - \eta_{ijl}$$

where

$$\eta_{ijl} = \beta_0 + \beta_{\text{study}} X_{\text{study}i} + X_{\text{site}.l}^T \boldsymbol{\beta}_{\text{site}} + f(\texttt{cumdosl}_{\texttt{ijl}}) + \texttt{b}_{\texttt{i}} \tag{4}$$

is selected to be the best for the mucositis data set (Augustin *et al.*, 2017). Here, γ_r is the cut-point for a mucositis grade r on the unobserved continuous mucositis severity scale, $\boldsymbol{\beta}_{\text{site}}$ is the parameter vector for the eight sites, and b_i is a random intercept for individual i with $b_i \sim N(0, \sigma_b^2)$. To model the effect of the site-specific cumulative dose (`cumdosl`), a flexible function $f(\cdot)$ is used. See Augustin *et al.* (2017). Based on the model above, it is found out that after adjusting for the site-specific cumulative dose and

study effect, upper, lower lips, and mouth floor are related to the lowest mucositis grades and hard and soft palates tend to have the highest mucositis grades.

4.2.2   Comparison

The parallel result of the estimated hazard ratios of the shared frailty model and the estimated odds ratios of the FMRP can be explained by several similar features of the two models. First, the two models contain random effect terms, or, termed frailty term in the shared frailty model, to take individual effects into account. Also, in both models the `site` effects for the eight sites are simultaneously incorporated into one model to address the clinical question as to whether the `site` effects in the RT exist. Furthermore, the odds/hazard ratios for the `site` effects in both models are on the *exponential* scale, which makes the `site` effects in the two models well comparable.

On the other hand, some differences between the two models can also be noted. As mentioned earlier, the original ordinal scale of the mucositis grades, $0, 1, 2, 3$ is preserved in the FMRP whereas in the shared frailty model only the occurrence of the severe mucositis 'event' is considered by takings values of either 0 or 1. Another difference is that the FMRP can also be seen as an additive model in that it allows the variable `cumdosl` to vary in a flexible way. In the FMRP `cumdosl` is incorporated as an additive flexible function of it. Furthermore, the `study` effect (nuisance, but significant) is incorporated as a stratum variable and therefore considered in the baseline hazard in the shared frailty model while it is included parametrically in the FMRP. Finally, the two models account for the radiation dose with different covariates: the site-specific cumulative dose (`cumdosl`) in the FMRP and the site-specific dose fraction (`dosl`) in the shared frailty model. Note that `cumdosl` in the FMRP is a time-dependent covariate while the `dosl` in the shared frailty model is a time-invariant covariate. Indeed, the time-dependent covariate `cumdosl` plays a role as a reflection of the *time* in the FMRP since it is almost exactly proportional to the *time* variable, as Fig. 2 indicates.

Due to the different outcome scales (binary vs. ordinal), a fair comparison between the two approaches can be made when the ordinal scales of the outcomes in the FMRP are collapsed into binary categories, i.e., when the flexible multivariate random effects *logistic* model, hereafter *the reduced FMRP*, is fit with the binary categories of $R \leq 2$ and $R > 2$ instead while all the others in the model are kept the same.

The rationale for the comparison between the proposed model and the reduced FMRP can be drawn from, for instance, Cupples *et al.* (1988), Green and Symons (1983), Elandt-Johnson (1980) and Abbott (1985). Those studies sought to clarify the relationship between the Cox model (on which the shared frailty model is based) and the logistic model (on which the reduced FMRP is based). Green and Symons (1983) found out that under the conditions of the short observation period, a rare occurrence of the event of interest, and a constant baseline hazard, the coefficients of the two models approximate one another. And Elandt-Johnson (1980) showed that under the same conditions as in Green and Symons (1983) the likelihood of the logistic regression approximates the partial likelihood of the Cox regression. The regression coefficients have similar estimated standard errors. In the general setting without the aforementioned conditions, however, the Cox regression tends to outperform the logistic model since it additionally looks at the time-to-event as well as the dropouts during the observation period; Cupples *et al.* (1988) showed that with increasing observation period the coefficients of the logistic regression becomes uncertain and less reliable whereas the coefficients obtained by the Cox regression tend to be invariant to the observation period and the standard errors decrease with increasing observation period.

In our study what makes the (reduced) FMRP more comparable to the shared frailty model is the fact that the model is based on the repeated measurements over the time with the time-dependent covariate `cumdosl` where the assessment time points and the dropouts can be taken into account. Thus, the arguments in the studies comparing the logistic regression and the Cox model allows us to compare the two models specially because the (reduced) FMRP incorporates the repeated measurements (`cumdosl`) over the time.

The two models give the parallel results as Fig. 4.2.2 suggests. The right panel of Fig. 4.2.2 shows the estimated hazard ratios of the shared frailty model (Sec. 4.1.1), and the left panel illustrates the estimated odds ratios of the reduced FMRP where the ordinal grades are collapsed into two categories of $R \leq 2$ and $R > 2$. The estimates of the hazard and odds ratios for the same site are only slightly different, with rather prolonged confidence intervals of the estimates of the hazard ratios: the difference in the confidence intervals originates from the fact that the FMRP allows for recurrence of the events whereas only the first occurrence of the event is considered in the shared frailty model. In addition, although `cumdosl` (used in the FMPR) reflects *time* (used in the shared frailty model) it is not identical to *time*. For a fair comparison it is worth noting that the shared frailty model (2) assumes Gamma frailty and that the Gamma frailty turns out to be comparable to the *log Gaussian* frailty (Gaussian is random effects on the scale of the linear predictor) in this application: Tab. 4.1.1 shows that in this application the two different frailty distributions do not influence much on the estimates and their standard errors. Despite the slight differences the two plots in Fig. 4.2.2 show parallel patterns. The orders of the estimates of the eight `site` effects (representation of radiosensitivity) coincide in the two models: although in the FMRP with the original mucositis scale there are slightly changed orders between `soft` and `hard`, it is not conspicuous. The estimated odds ratios of the FMRP for the original scale can be found in Augustin *et al.* (2017), and here the result from only the reduced FMRP (logistic regression) is shown for the purpose of the comparison.

These two different approaches which led to the same conclusion about the site effects can be seen as some kind of sensitivity analysis to each other.
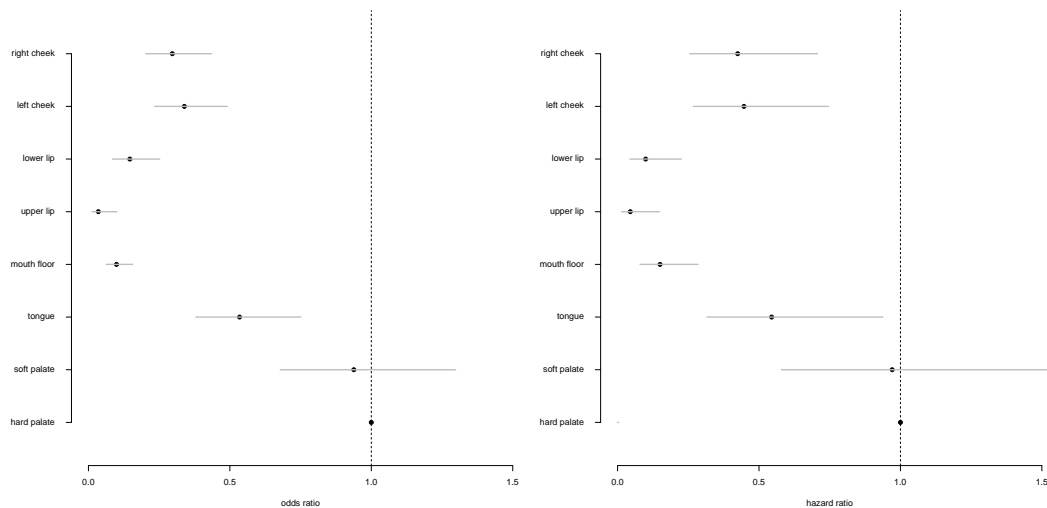
Figure 4   Comparison of (left) the estimated odds ratios of the site effects with respect to the hard palate from the reduced FMRP (edited from Augustin *et al.* 2017) focusing on the severe mucositis, i.e., odds ratios for $R > 2$, (right) the estimated hazard ratios of the site effects with respect to the hard palate from the gamma shared frailty model. The gray lines indicate 95% confidence intervals.

5   Discussion

In this article we have introduced a time-to-event approach for the analysis of complex multivariate and longitudinal data and have chosen a shared frailty model as the most suitable one for the mucositis data structure. An alternative, direct time-to-event approach could be the multi-state modelling where the mucositis grade itself represents a *state* and any change in grade can be described by a *transition* between two

states (extended illness-death model). In fact, it turns out that a *site-specific* Cox model incorporating the site-specific dose fraction (`dosl`) and `study` effect, separately fitted for each site, adequately explains the mucositis progressions for each site. Thereby, the sensitivity of each site to the incremental change in radiation dose fraction can be estimated. However, for the purpose of the comparisons of the `site` effects itself, which is of our main interest, a model incorporating eight `site` effects simultaneously is required, which is impossible in the multi-state model setting especially for a relatively small data set. Therefore, in the case of the mucositis data set, based on the fact that the event of severe mucositis (grade ≥ 3) is of main interest we regard the severe mucositis occurrences in the eight distinct sites as the multivariate time-to-event and take a joint modelling approach by focusing on the severe mucositis occurrences. The shared frailty model proposed in this article can be seen as a joint modelling approach where the eight parallel binary processes are simultaneously incorporated into one model.

The joint modelling approach entails correlation issues, and the mucositis data set has a multivariate longitudinal structure where, in general, complicated correlations between the repeated measurements as well as between multivariate outcomes within an individual are to be addressed. And with the proposed shared frailty model the correlation issues can be easily dealt with by introducing a frailty term to explain individual effects which cannot be explained by covariates.

Notice that in the multivariate parallel processes, there are possibly two different kinds of correlations: (i) between the observations within an individual and (ii) between the parallel processes. The frailty term accounts for the former, but it does not necessarily take into account the latter. Being interested in the latter, a proportional hazard (PH) Markov model introduced by Cook and Lawless (2014) can be employed. This model is motivated by interrelated, multiple progressive disease processes. Under the Markovian and no tied event assumptions, the *extended* state space can be defined, and this structure enables us to model the pairwise interrelationships of the processes where any event is assumed to be explained by the history of another paired process. This means that the model incorporates all the pairwise interrelationships and those terms can tell us which pair of processes is more interrelated to each other. This is what the frailty term cannot possibly explain; such a model added by the frailty term still results in a statistically significant variance of the frailty. There is some overlapping information between the frailty term and the pairwise interrelationship term, but they cannot replace each other. Apparently, however, the PH Markov model is overly parameterized, yielding some non-valid estimates, and is not practically efficient. Still, the PH Markov model brings up an issue of distinction between interrelationship among the parallel processes and individual effect (frailty), and it might be worthy to check the interrelationship between the processes.

Although the frailty term might not reflect all the pairwise interrelationships, unless the interest lies in the interrelationship between the sites, the proposed shared frailty model adequately explains the main effects of interest, and the frailty term sufficiently accounts for the heterogeneity among the observations. The biggest advantage of the proposed approach is its simplicity: less data collection is required while the model gives the coherent result. Furthermore, despite its simple model formulation the shared frailty model sufficiently addresses the challenges listed in Sec. 3.2. Also, a simple formulation would show its strength when it comes to making a prediction on a new data set. A clear inference of the shared frailty model and the simplicity of its model formulation makes its practical application easy. The time-to-event approach has another advantage of enabling us to model *time* explicitly. The shared frailty model is essentially a time-varying model and looks at the process in the whole period of the study until the event happens. Notice that the frailty term is updating with time, and in the case of gamma frailty this "updating formula" can be explicitly given (Nielsen *et al.*, 1992; Hougaard, 2000; Aalen *et al.*, 2006).

The proposed shared frailty model in this article is limited to a rather specific situation where the special interest lies in a time-to-event, i.e., when the longitudinal process exceeds a pre-specified value; the occurrence of severe mucositis (grade ≥ 3) is of main interest rather than the ordinal mucositis grade itself. However, such simplification can be justified by the coherent results between the FMRP and the proposed shared frailty model, and the further parallel results between the reduced FMRP (longistic model) and the FMRP. Thus, in this specific application this multivariate time-to-event approach can be seen as adequate for our purpose.

In this work, we did not encounter any censoring issues. Thus, we did not address any possible complications owing to the censoring or its assumptions, which should be preceded in dealing with survival data in general. Also, a direct comparison between the shared frailty model with the original outcome in ordinal scale and the one with a binary outcome was not possible to be carried out, due to the small size of the data set. It is also worth noting that the model was not adjusted for the absolute volume of the sites in the final model. We assume that the volume does not differ substantially between sites.

It is expected that for a similar kind of data set in the future the proposed shared frailty model can be applied to compare parallel processes with distinct outcomes. More generally, the approach might be considered for a multivariate longitudinal data which is comprised of the risk factors as well, given that each parallel process can be described as a binary process. Although the approach proposed in this article may seem specific for the application to the mucositis adverse event in radiation therapy, as other specifically tailored methodology -e.g. on personalized dose selection in radiation therapy (Schipper *et al.*, 2014) it should also be applicable in other settings when parallel processes are observed in each individual.

Conflict of Interest
The authors have declared no conflict of interest.

References

Aalen, O. and Borgan, Ø. and Gjessing, H. (2006). *Event History Analysis: A Process Point of View*. Springer-Verlag Inc., New York.

Abbott, R. (1985). Logistic regression in survival analysis. *American journal of epidemiology* **121**, 465–471.

Andersen P. and Borgan Ø. and Gill R. and Keiding N. (1998). *Statistical models based on counting processes*. Springer-Verlag Inc., New York.

Augustin, N. and Kim, S. and Uhlig, A. and Hanser, C. and Henke, M. and Schumacher, M.(2017).A flexible multivariate random effects proportional odds model with application to adverse effects during radiation therapy. *Biometrical Journal* **59**, 1339–1351.

Cox, D. (1992). Regression models and life-tables. *Breakthroughs in statistics* Springer, 527–541.

Cook, R. and Lawless, J. (2014). Statistical issues in modeling chronic disease in cohort studies. *Statistics in Biosciences* **6**, 127–161.

Cook, R. and Lawless, J. (2018)]bib7Cook, R. and Lawless, J. (2018). Multivariate models for the analysis of life history data *Chapman and Hall/CRC*

Cupples, L. and D'Agostino, R. and Anderson, K. and Kannel, W. (1988). Comparison of baseline and repeated measure covariate techniques in the Framingham Heart Study. *Statistics in medicine* **7**, 205–218.

Diggle, P. and Heagerty, P. and Liang, K. and Zeger, S. (2002). Analysis of longitudinal data *Oxford Statistical Science Series*

Elandt-Johnson, R. (1980). *Time Dependent Logistic Models in Follow-up Studies and Clinical Trials*. Department of Biostatistics [University of North Carolina at Chapel Hill].

Green, M. and Symons, M. (1983). A comparison of the logistic risk function and the proportional hazards model in prospective epidemiologic studies. *Journal of chronic diseases* **36**, 715–723.

Henderson C.(1975).Best linear unbiased estimation and prediction under a selection model. *Biometrics* 423–447.

Henke, M. and Alfonsi, M. and Foa, P. and Giralt, J. and Bardet, E. and Cerezo, L. *et al.*. Palifermin decreases severe oral mucositis of patients undergoing postoperative radiochemotherapy for head and neck cancer: a randomized, placebo-controlled trial. *Journal of Clinical Oncology* **29**, 2815–2820.

Hougaard, P. (2000). *Analysis of multivariate survival data*. Speinger-Verlag.

Jennrich, R. and Schluchter, M (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*, 805–820.

Laird N. and Ware J. and James H. (1982). Random-effects models for longitudinal data, *Biometric*s, JSTOR, 963–974.

Liang, K. and Zeger, S.(1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

Liu, L. and Wolfe, R. A. and Huang, X. (2004). *Shared frailty models for recurrent events and a terminal events*. *Biometrics* **60**, 747–756.

Long, J. (2011). *Longitudinal data analysis for the behavioral sciences using R*. Sage.

Nielsen G and Gill R. and Andersen P. and Sørensen T. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian journal of Statistics* 25–43.

Ripatti, S. and Palmgren, J.(2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics* **56**, 1016–1022.

Schipper, M. and Taylor, J. and TenHaken, R. and Matuzak, M., Kong, F. and Lawrence, T.). Personalized dose selection in radiation therapy using statistical models for toxicity and efficacy with dose and biomarkers as covariates. *Statistics in medicine* **33**, 5330–5339.

Therneau, T.M.. *A Package for Survival Analysis in R*. https://CRAN.R-project.org/package=survival, R package version 3.1-12, New York.

Therneau, T. and Grambsch, P. (2000). *Event History Analysis: A Process Point of View*. Springer-Verlag Inc., New York.

Therneau, T. and Grambsch, P. and Pankratz, V. (2003). Penalized survival models and frailty. *Journal of computational and graphical statistics* **12**, 156–175.

Verbeke, G. and Fieuws, S. and Molenberghs, G. and Davidian, M. (2014). The analysis of multivariate longitudinal data: A review. *Statistical methods in medical research* **23**, 42–59.

Zeger, S. and Liang, K. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 121–130.