Weighted NPMLE for the Subdistribution of a Competing Risk

Anna Bellach[1], **Michael R Kosorok**[2], **Ludger Rüschendorf**[3], and **Jason P Fine**[2]

[1]Department of Biostatistics at University of Copenhagen

[2]Department of Biostatistics and Department of Statistics and Operations Research at University of North Carolina at Chapel Hill. These authors shared seniorauthorship

[3]Department of Mathematics at Albert Ludwigs University of Freiburg im Breisgau

Abstract

Direct regression modeling of the subdistribution has become popular for analyzing data with multiple, competing event types. All general approaches so far are based on non-likelihood based procedures and target covariate effects on the subdistribution. We introduce a novel weighted likelihood function that allows for a direct extension of the Fine-Gray model to a broad class of semiparametric regression models. The model accommodates time-dependent covariate effects on the subdistribution hazard. To motivate the proposed likelihood method, we derive standard nonparametric estimators and discuss a new interpretation based on pseudo risk sets. We establish consistency and asymptotic normality of the estimators and propose a sandwich estimator of the variance. In comprehensive simulation studies we demonstrate the solid performance of the weighted NPMLE in the presence of independent right censoring. We provide an application to a very large bone marrow transplant dataset, thereby illustrating its practical utility.

Keywords

cumulative incidence function; Fine-Gray model; nonparametric maximum likelihood estimation; semiparametric transformation models; time-varying covariates

1 Introduction

Competing risks occur when subjects are exposed to multiple, mutually exclusive event types. The model may formally be defined as a jump process, with a single transient state and several absorbing states that we refer to as the competing events. In such settings one only observes the time of the first event, denoted by T, and the type of that event, denoted by $\varepsilon \in \{1, \ldots, k\}$. As an example, in the bone marrow transplant dataset analyzed in Section 5, recurrence of leukemia and death in remission are the competing events of interest.

Let C denote a possible censoring time and Z a vector of covariates. It is assumed that C is independent of (T, ε) given Z. A common approach is to model the cause specific hazards, defined as $\alpha_j^*(t \mid Z) := \lim_{\Delta t \to 0} P\{t < T \le t + \Delta t, \varepsilon = j \mid T > t, Z\}$ for $j = 1, \ldots, k$, which can be interpreted as the instantaneous risk of dying from the event of type j at time T conditional on not having experienced any of the competing events prior to T. Modeling the cause specific hazard for the event of interest is straightforward from the point of view that it

relies on a traditional risk set of individuals not having experienced any of the *k* events. In many applications however, the cumulative incidence function (CIF), defined as $F_j(t|Z) = P(T \leq t, \varepsilon = j|Z)$ and also referred to as the absolute risk of an event over time, is of interest. An efficient prediction of the CIF may be achieved by modeling the cause specific hazards of all event types, see e.g. Kalbfleisch and Prentice (1980, 2002). The relation between the cause specific hazards and the CIF, however, is complex and an accurate prediction of $F_j(t|Z)$ requires correctly specified models with appropriate sample sizes for all cause specific hazards $\alpha_1^*(t \mid Z), \ldots, \alpha_k^*(t \mid Z)$.

A common pitfall in the analysis of competing risks is the use of a product limit estimator based on the cause specific hazard to estimate the corresponding CIF, as illustrated e.g. by Tai at al. (2001). This generally leads to a systematic overestimation of the CIF in the presence of competing events. Similar issues arise in regression modeling on the cause specific hazard, since the effect of a covariate on the cause specific hazard for a specific event type is not interpretable with regard to the corresponding CIF, see Latouche et al. (2013). In extreme cases a covariate may have a strong effect on $\alpha_j^*(t \mid Z)$ but no effect on $F_j(t|Z)$.

For modeling the cause specific hazard, competing risk events are removed from the risk set in the same way as censored individuals, despite a conceptual difference between censorings and competing events: per definition, a competing risk event precludes the possibility of a subsequent event of interest, while for independently censored individuals it is assumed that an event of interest will occur with the same probability as for the individuals with no prior event.

An alternative approach to competing risks regression modeling is based on the idea that individuals are retained in the risk set instead of being removed. Gray (1988) has first defined the subdistribution hazard

$$\alpha(t \mid Z): = \lim_{\Delta t \to 0} P\{t < T \leq t + \Delta t, \varepsilon = 1 \mid (T > t) \cup (T \leq t, \varepsilon \neq 1), Z\},$$

which may be interpreted as the hazard of the improper random variable $T^* = T \cdot \mathbb{1}\{\varepsilon = 1\} + \infty \cdot \mathbb{1}\{\varepsilon \neq 1\}$, as in Fine and Gray (1999). The risk set for the subdistribution hazard is unconventional but interpretable with a cure model intuition, where individuals are no longer exposed to the event of interest after a competing risk event and therefore regarded as cured with regard to the event of interest. For statistical inference on $F_1(t|Z)$, cured individuals are included in the risk set, as is typically done in cure modeling without competing risks. The resulting risk set consists of those individuals who either did not experience an event of interest in the past or will never experience an event of interest in the future. Fine and Gray (1999) propose a proportional hazards regression model for the subdistribution hazard based on a weighted partial likelihood function. For the Fine-Gray model, estimated regression parameters are directly interpretable with regard to the corresponding CIF, as with

$F_1(t \mid Z) = 1 - \exp\left(-\int_0^t \alpha(u \mid Z)du\right)$, the usual one to one relationship between the hazard and the failure probability is obtained.

It should be emphasized that neither for the cause specific hazard, nor for the subdistribution hazard approach, the dependency structure between the different event types is of relevance.

The extension of the Fine-Gray model to a larger class of semiparametric regression models is of particular interest as the proportional hazards assumption is not valid in general, as exhibited in Section 5, where alternative models are shown to yield improved model fit for leukemic relapse. Theoretical concerns have been expressed regarding the ability to simultaneously model multiple CIFs using the Fine-Gray approach. For example, with two event types, one requires $F_1(t/Z) + F_2(t/Z) \le 1$ for all t and all Z, with equality at $t = \infty$. The condition may hold for finite t but not at $t = \infty$. By considering a general class of semiparametric regression models this constraint can be considerably relaxed. In Section 3, we propose such a general class to address potential lack of fit in practice and theoretical issues of model compatibility. This class includes the models of Fine (2001), Klein and Andersen (2005), Scheike, Zhang and Gerds (2008), and others. Unlike these earlier approaches which model $F_1(t/Z)$, our model directly targets the subdistribution hazard, accommodating time-dependent covariates in the hazard model. This modeling strategy permits likelihood based inferences, which have not been possible with previous approaches to direct competing risks regression.

In Section 2, a weighted nonparametric likelihood for competing risks data is developed from the cure model point of view that the contribution of individuals with a competing event is the probability of not having an event of interest until the endpoint of the study. The proposed weighted likelihood function yields a Nelson-Aalen type estimator of the subdistribution hazard in the absence of covariates and enables parameter estimation in regression models targeting the subdistribution hazard in the presence of Z, as further discussed in Section 3. The method is widely applicable to models for the subdistribution hazard, in particular to the classes of Box-Cox transformation models and logarithmic transformation models that include the Fine-Gray model and the proportional odds model. The estimators are shown to be uniformly consistent and weakly convergent to a Gaussian distribution. We present a sandwich variance estimator that is theoretically justified by a new lemma in connection to the proof of weak convergence, as provided in Section A.3.

Comprehensive simulation studies are conducted in Section 4 to assess the performance of the weighted NPMLEs in finite sample sizes. The bone marrow transplant data is reanalyzed in Section 5.

2 Competing risks analysis based on pseudo risk sets

The subdistribution hazard is related to a cure model interpretation of the competing risks setting, where competing risk events are assigned to the risk set in a similar way as cured individuals. For statistical inference about the CIF via the subdistribution hazard, an adjustment for cured individuals is required to account for unobserved censoring. Fine-Gray (1999) propose a weight function that induces the appropriate pseudo risk set. In the

following paragraphs we derive a Nelson-Aalen type estimator for the subdistribution hazard from the weighted Doob decomposition of the underlying counting process for the events of interest. The same Nelson-Aalen type estimator is then obtained from a pseudo likelihood function that reduces to the common counting process likelihood for settings without competing risks.

2.1 Complete data

Let T_i denote the event time and ε_i the failure type of the i–th individual, $i = 1, \ldots, n$. All data is observed on a time interval $[0, \tau]$ with τ denoting the duration of the study. With $N_i(t) = \mathbb{1}\{T_i \le t, \varepsilon_i = 1\}$, let $N(t) = \sum_{i=1}^n N_i(t)$ denote the process counting the events of interest and define $Y_i(t) = 1 - N_i(t-)$. Then $Y(t) = \sum_{i=1}^n Y_i(t)$ is the cardinality of the risk set $\mathcal{R}(t) = \{i : (T_i \ge t) \cup (T_i \le t, \varepsilon_i \ne 1)\}$. Notice that $\mathcal{R}(t)$ contains those individuals who have not experienced any event until t or who will never experience an event of interest. It is easy to check that $\Lambda(t) := \int_0^t \alpha(u)Y(u)du$ is the compensator of $N(t)$ with respect to the filtration $\mathcal{F}(t) := \sigma\{N_i(s), s \le t, i = 1, \ldots, n\}$. From the Doob decomposition $dN_i(t) = Y_i(t)\alpha(t)dt + dM_i(t)$ we obtain the Nelson-Aalen estimator of the cumulative subdistribution hazard, $\hat{A}(t) := \int_0^t \{Y(u)\}^{-1}dN(u)$.

This estimator can also be derived from a novel likelihood function. Defining the increments of the cumulative subdistribution hazard by $A\{T_i\} := A(T_i) - A(T_{i-1})$, the contribution for an event of interest at T_i is $P(T^* \in (T_{i-1}, T_i], \varepsilon = 1) = A\{T_i\}S_1(T_{i-1})$, while the contribution for a competing risk event is $P(T^* \ge \tau) = S_1(\tau)$. The likelihood function therefore takes the form

$$
\begin{aligned}
L_n &= \prod_{i=1}^n [A\{T_i\}S_1(T_{i-1})]^{\mathbb{1}\{\varepsilon_i = 1\}} \cdot \left[S_1(\tau)\right]^{\mathbb{1}\{\varepsilon_i \ne 1\}} \\
&= \prod_{i : T_i \le \tau} \left[[Y_i(T_i)A\{T_i\}]^{dN_i(T_i)} \cdot \exp\left(-\int_0^\tau Y_i(u)dA(u)\right) \right].
\end{aligned}
$$

2.2 Model with administrative censoring

We consider a setting where the censoring time C is observed even for individuals with a previous competing risk event. Such data arises for example in administrative databases where all individuals are censored at a fixed calendar time with complete follow-up to that time point.

Let C_i denote the censoring time of the i-th individual, $X_i = T_i \wedge C_i$ and $\Delta_i = \mathbb{1}\{T_i \le C_i \wedge \tau\}$. The process counting the events of interest is denoted by $N^a(t) = \sum_{i=1}^n N_i^a(t)$, $N_i^a(t) = \int_0^t \mathbb{1}\{C_i \ge u\}dN_i(u)$. The process $Y^a(t) = \sum_{i=1}^n Y_i^a(t)$ with $Y_i^a(t) = 1\{C_i \ge t\}Y_i(t)$ is the cardinality of the risk set $\mathcal{R}^a(t) = \{i : (X_i \ge t)\} \cup \{i : (X_i < t, \Delta_i = 1, \varepsilon_i \ne 1, C_i \ge t)\}$. In this case $\Lambda^a(t) = \int_0^t \alpha(u)Y^a(u)du$ is the compensator of $N^a(t)$ with respect

to $\mathcal{A}^a(t) = \sigma\{N_i^a(s), \ \mathbb{1}\ (C_i \geq s), s \leq t, i = 1, ..., n\}$ and the Nelson-Aalen estimator

$\hat{A}(t) = \int_0^t \{Y^a(u)\}^{-1} dN^a(u)$ can be derived from the likelihood function

$$L_n^a = \prod_{i:X_i \leq \tau} \left[[Y_i^a(X_i) A_n\{X_i\}]^{dN_i^a(X_i)} \cdot \exp\left(-\int_0^\tau Y_i^a(u) dA_n(u)\right)\right].$$

2.3 Model with independent right censoring

For the Fine-Gray approach, individuals with a competing risk event are regarded as a cure fraction of the risk set, as they are no longer exposed to the event of interest. This slightly differs from the ordinary cure model, where individuals in the cure fraction would be subject to independent right censoring which, per definition, cannot be observed after a competing risk event. To maintain the common cure model structure, Fine and Gray (1999) use inverse probability of censoring weighting (IPCW) technique, thereby applying the estimated weight $w_i(t) := \mathbb{1}\{C_i \ \ T_i \wedge t\} \cdot \hat{G}_C(t)/\hat{G}_C(T_i \wedge t)$ in place of $\tilde{w}_i(t) := \mathbb{1}\{C_i \ \ T_i \wedge t\} \cdot G_C(t)/G_C(T_i \wedge t)$, where $\hat{G}_C$ is the product limit estimator of $G_C(t) = P(C > t)$.

A Nelson-Aalen type estimator for the subdistribution hazard can be derived from the weighted Doob decomposition $w_i(t) dN_i(t) = w_i(t) Y_i(t) a(t) dt + w_i(t) dM_i(t)$, with

$$\sum_{i=1}^n w_i(t) Y_i(t) = \sum_{i=1}^n \mathbb{1}\{X_i \geq t\} + \sum_{i=1}^n \mathbb{1}\{X_i < t, \Delta_i = 1, \varepsilon_i \neq 1\} \cdot \hat{G}_c(t)/\hat{G}_c(T_i)$$

representing the expected number of individuals in the pseudo risk set. Alternatively $\hat{A}_n$ can be derived from the pseudo likelihood function

$$\begin{aligned} L_n &= \prod_{i=1}^n \left\{ [A\{X_i\}]^{\Delta_i \mathbb{1}\{\varepsilon_i = 1\}} \cdot \prod_{\ell:X_\ell < X_i} (1 - A\{X_\ell\}) \right. \\ &\times \left. \prod_{\ell:X_\ell \geq X_i} (1 - A\{X_\ell\})^{\Delta_i \mathbb{1}\{\varepsilon_i \neq 1\} \hat{G}(X_\ell)/\hat{G}(X_i \wedge X_\ell)} \right\} \\ &= \prod_{i=1}^n \left[[w_i(X_i) Y_i(X_i) dA_n(X_i)]^{\Delta_i dN_i(X_i)} \exp\left(-\int_0^\tau w_i(u) Y_i(u) dA(u)\right)\right]. \end{aligned}$$

Our proposed likelihood function is consistent with the general likelihood function in Andersen et al. (1993) and it can be ascertained that the product limit estimator based on the estimated cumulative subdistribution hazard is equivalent to the fully efficient Aalen-Johansen estimator, see e.g. Antolini, Biganzoli and Boracchi (2007), Zhang, Zhang and Fine (2009), Geskus (2011). This indicates that the weighted likelihood is a promising candidate for inference in semiparametric regression models, an idea which is explored in subsequent sections.

3 General subdistribution hazard regression model

Fine-Gray (1999) propose a proportional hazards model for the subdistribution with an estimation procedure based on a weighted partial likelihood function. The method accommodates time-varying covariate effects on the subdistribution hazard and yields the usual nonparametric estimators in the absence of Z.

Various other approaches have been developed based on modeling $F_1(t|Z)$ instead of $\alpha(t|Z)$. Fine (2001) proposed a general regression model for $F_1(t|Z)$ in which the subdistribution is assumed to satisfy a generalized linear model after transformation, with a time-varying intercept. Pairwise rank estimators were utilized, again using inverse probability of censoring weighting. A simple method for estimation of the model based on $F_1(t|Z)$ was proposed by Andersen and Klein (2007). Using jackknife ideas, pseudo values may be employed, which enable the construction of estimating equations. While computationally appealing owing to its ease of implementation in standard software, the approach is ad hoc, requiring choices of time points at which to evaluate the pseudo-values, and its general theoretical properties with right censoring, including efficiency, are unclear. Scheike et al (2008) adapted binomial regression techniques to regression modeling of $F_1(t|Z)$, also using inverse probability of censoring weighting. None of the above methods adapt easily to time-varying covariates, which are most naturally accommodated in models for the hazard function, as with survival data without competing risks. Moreover, these methods do not reduce to the usual nonparametric estimators without covariates.

3.1 Model formulation

As a general model for the cumulative subdistribution hazard we propose

$$A(t) = \mathscr{G}\left(\int_0^t e^{\beta^T Z(s)} dA_0(s)\right),$$

where $\beta \in \mathbb{R}^d$ is a vector of unknown regression parameters, A_0 is an unspecified increasing function, $Z(s)$ is a vector of possibly time dependent covariates of bounded variation and $\mathscr{G}$ is a thrice continuously differentiable and strictly increasing function with $\mathscr{G}(0) = 0$, $\mathscr{G}'(0) > 0$ and $\mathscr{G}(\infty) = \infty$. Additional regularity conditions for the existence of the weighted NPMLE, which are considerably weaker than those in Zeng and Lin (2006), are specified in Section A.1. In particular, our general model entails not only the class of Box-Cox transformation models with link function $\mathscr{G}(x) = \{(1 + x)^\rho - 1\} /\rho$ for ρ 0 but also the class of logarithmic transformation models with link function $\mathscr{G}(x) = \log(1+rx)/r$ for r 0 (Chen, Jin, Ying, 2002). Important special cases of these classes are firstly the Fine-Gray model, which is the Box-Cox transformation model with $\rho = 1$ and the limiting logarithmic transformation model for $r \rightarrow 0$ and secondly the proportional odds model as the logarithmic transformation model with $r = 1$ and the limiting Box-Cox transformation model for $\rho \rightarrow 0$.

With independent right censoring our general model is related to the multiplicative intensity process

$$\alpha(t, Z)w(t)Y(t) = \alpha_0(t)w(t)Y(t)e^{\beta^T Z(t)}\mathcal{G}'\left(\int_0^t e^{\beta^T Z(s)}dA_0(s)\right), \quad (1)$$

while in the case of administrative right censoring the multiplicative intensity process

$$\alpha(t, Z)\,\mathbb{1}\,(C > t)Y(t) = \alpha_0(t)\,\mathbb{1}\,(C > t)Y(t)e^{\beta^T Z(t)}\mathcal{G}'\left(\int_0^t e^{\beta^T Z(s)}dA_0(s)\right)$$

is obtained, which yields the cumulative intensity

$$\Lambda(t) = \mathcal{G}\left(\int_0^t \mathbb{1}\,(C > t)Y(t)e^{\beta^T Z(s)}dA_0(s)\right),$$

akin to the model proposed by Zeng and Lin (2006) for settings without competing risks.

The model incorporates time-dependent covariates into a general subdistribution hazard regression model, extending Fine and Gray's (1999) formulation to the non-proportional hazard setting. Challenges arise for individuals with a competing event, where the covariate process $Z(t)$ should be well defined for $t \in [X, \tau]$. This is always the case for baseline covariates and for external time dependent covariates, where $Z(t)$ can be observed after a competing risks event. For internal time dependent covariates, the issues are more complicated, as discussed in Section 6.

3.2 Weighted nonparametric maximum likelihood estimation

With independent right censoring, the observed data needed for estimation consists of iid samples of $(X, \ , \varepsilon, Z(t), t \ (X \wedge \tau)\,\mathbb{1}(\ \varepsilon \in \{0, 1\}) + \tau\,\mathbb{1}(\ \varepsilon \notin \{0, 1\}))$ denoted by $\{(X_i, \ _i, \ _i\varepsilon_i, Z_i(t), t \ (X_i \wedge \tau) \cdot \mathbb{1}(\ _i\varepsilon_i \in \{0, 1\}) + \tau \cdot \mathbb{1}(\ _i\varepsilon_i \notin \{0, 1\})), i = 1, ..., n\}$. Notice that the covariate process $Z_i(t)$ is observed on $[0, X_i]$ for all individuals and observed on $[X_i, \tau]$ for those individuals with $\ _i\varepsilon_i \notin \{0, 1\}$. We thus obtain the weighted log-likelihood function under the general semiparametric regression model:

$$\ell(\beta, A_0) = \sum_{i=1}^n \left[\int_0^\tau \log\left(e^{\beta^T Z_i(t)}\alpha_0(t)\mathcal{G}'\left(\int_0^t e^{\beta^T Z_i(u)}dA_0(u)\right)\right)\mathbb{1}\,(C_i \geq t)Y_i(t)dN_i(t) \quad (2)\right.$$
$$\left. - \int_0^\tau w_i(t)Y_i(t)e^{\beta^T Z_i(t)}\mathcal{G}'\left(\int_0^t e^{\beta^T Z_i(u)}dA_0(u)\right)dA_0(t)\right].$$

Decomposition of the weighted risk set $w_i(t)Y_i(t) = \mathbb{1}\,(X_i \geq t) + w_i^*(t)\,\mathbb{1}\,(X_i \leq t, \Delta_i\varepsilon_i \notin \{0, 1\})$ with a simplified weight function $w_i^*(t) = \mathbb{1}\,(C_i \geq T_i)\hat{G}_c(t)/\hat{G}_c(T_i)$ yields the equivalent representation

$$\ell(\beta, A_0) = \sum_{i=1}^{n} \left[\int_0^\tau \log \left(e^{\beta^T Z_i(t)} \alpha_0(t) \mathcal{G}' \left(\int_0^t e^{\beta^T Z_i(u)} dA_0(u) \right) \right) \mathbb{1} (C_i \geq t) Y_i(t) dN_i(t) \right] \quad (3)$$

$$- \sum_{i=1}^{n} \mathcal{G} \left(\int_0^{X_i \wedge \tau} e^{\beta^T Z_i(t)} dA_0(t) \right)$$

$$- \sum_{i: \Delta_i \varepsilon_i \neq 1, X_i \leq \tau} \left[\int_{X_i}^\tau w_i^*(t) e^{\beta^T Z_i(t)} \mathcal{G}' \left(\int_0^t e^{\beta^T Z_i(u)} dA_0(u) \right) dA_0(t) \right].$$

The cumulative baseline hazard A_0 is approximated by a sequence of step functions A_n^0, with jumps at the observed events of interest. Those increments are interpretable as a finite dimensional parameter. Let $\{0 < \tilde{T}_1 < \tilde{T}_2 < \ldots < \tilde{T}_{k(n)} < \tau\}$ denote the ordered times and $k(n)$ the number of the observed events of interest. Replacing A_0 by A_n^0 we obtain a modified likelihood function $\ell_n(\beta, \hat{A}_n^0)$ and maximization yields an estimator $\hat{\theta}_n^* = (\hat{\beta}_n, \hat{A}_n^0\{\tilde{T}_1\}, \ldots, \hat{A}_n^0\{\tilde{T}_{k(n)}\})$, with $A_n^0\{\tilde{T}_i\} = A_n^0(\tilde{T}_i) - A_n^0(\tilde{T}_{i-1})$ and $\hat{\beta}_n \in \mathbb{R}^d$. An estimator for the parameter of interest $\hat{\theta}_n = (\hat{\beta}_n, \hat{A}_n^0(\tilde{T}_1), \ldots, \hat{A}_n^0(\tilde{T}_{k(n)}))$ is obtained by the linear transformation $\hat{\theta}_n^T = C \hat{\theta}_n^{*T}$, with C being specified in our technical report.

For $\mathcal{G}(x) = x$, corresponding to the proportional subdistribution hazard model, the weighted log-likelihood function takes the simple form

$$\ell(\beta, A_0) = \sum_{i=1}^{n} \left[\int_0^\tau \log \left(e^{\beta^T Z_i(t)} \alpha_0(t) \right) \mathbb{1} (C_i \geq t) Y_i(t) dN_i(t) - \int_0^\tau w_i(t) Y_i(t) e^{\beta^T Z_i(t)} dA_0(t) \right]$$

and in the same way as for the proportional hazards model without competing risks, it can be factorized into the Fine-Gray partial likelihood function and a second term,

$$L_n = \prod_{i: \Delta_i \varepsilon_i = 1} \left[\frac{w_i(X_i) Y_i(X_i) e^{\beta^T Z_i(X_i)}}{\sum_{j=1}^n w_j(X_i) Y_j(X_i) e^{\beta^T Z_j(X_i)}} \right] \cdot \left(\sum_{j=1}^{n} w_j(X_i) Y_j(X_i) e^{\beta^T Z_j(X_i)} A_0\{X_i\} \right) \times \exp$$

$$\left(- \int_0^\tau w_i(u) Y_i(u) e^{\beta^T Z_i(u)} dA_0(u) \right).$$

Accordingly parameter estimates derived from the weighted log-likelihood function with $\mathcal{G}(x) = x$ are identical to those derived from the Fine-Gray model.

For administratively censored data we obtain the log-likelihood function

$$\ell^a(\beta, A_0) \tag{4}$$

$$= \sum_{i=1}^{n} \left[\int_0^\tau \log \left(e^{\beta^T Z_i(t)} \alpha_0(t) \mathscr{G}' \left(\int_0^t e^{\beta^T Z_i(u)} dA_0(u) \right) \right) \mathbb{1}\,(C_i \ge t) Y_i(t) dN_i(t) - \mathscr{G} \right.$$
$$\left. \left(\int_0^\tau \mathbb{1}\,(C_i \ge t) Y_i(t) e^{\beta^T Z_i(t)} dA_0(t) \right) \right],$$

with the same structure as in Zeng and Lin (2006) for the model without competing risks. Because there are no estimated weights, the estimators inherit those same asymptotic properties. In particular efficiency is obtained, and the variance estimator simplifies to the empirical inverse Fisher information.

3.3 Asymptotic inference and variance estimator

Theorem 1—*The estimator* $\hat{\theta}_n = (\hat{\beta}_n, \hat{A}_n^0)$ *derived from maximizing the weighted likelihood function is uniformly consistent.*

To prove consistency we adapt arguments established by Murphy (1994), (1995) and developed e.g. by Parner (1998), Kosorok, Lee and Fine (2004) and by Zeng and Lin (2006), (2010). An additional challenge for our model are the estimated weights that require an application of IPCW technique. To our knowledge this issue has not previously been addressed in connection with nonparametric maximum likelihood estimation of parameters in semiparametric regression models. As in Fine and Gray (1999), we first prove consistency for settings with administrative censoring. It is then argued that the weighted NPMLEs for settings with independent right censoring are asymptotically equivalent.

Theorem 2— $\sqrt{n}(\hat{\beta}_n - \beta_0, \hat{A}_n - A_0)$ converges weakly to a Gaussian process.

To establish weak convergence we apply our new Lemma 3, as stated in Section A.3, that is based on Theorem 3.3.1. of van der Vaart and Wellner (1996).

We consider linear functionals of the form
$\sqrt{n}(\hat{\theta}_n - \theta_0)(h) = \sqrt{n} h_1^T(\hat{\beta}_n - \beta_0) + \sqrt{n} \int_0^\tau h_2(t) d(\hat{A}_n(t) - A_0(t))$, with $h = (h_1, h_2)$, h_2 being an arbitrary element in the Skorohod space $\mathscr{D}[0, \tau]$ and $h_1 \in \mathbb{R}^d$. Defining a corresponding vector $\tilde{h}_2 \in \mathbb{R}^{k(n)}$ with $\tilde{h}_2^j = h_2(\tilde{T}_i)$, $\tilde{A}_0\{\tilde{T}_i\} = A_0(\tilde{T}_i) - A_0(\tilde{T}_{i-1})$ for $i = 1, \dots, k(n)$ and $\tilde{h} \equiv (h_1, \tilde{h}_2^1, \dots, \tilde{h}_2^{k(n)})$, we propose the sandwich estimator

$$\widehat{\mathrm{Var}}\sqrt{n}(\hat{\theta}_n - \theta_n)(\tilde{h}) = \tilde{h}^T \mathscr{I}_n^{-1T} \widehat{\Sigma}_n \mathscr{I}_n^{-1} \tilde{h}$$

for the asymptotic variance of $\sqrt{n}h_1^T(\hat{\beta}_n - \beta_0) + \sqrt{n}\sum_{j=1}^{k(n)}\tilde{h}_2^j(\hat{A}_n\{\tilde{T}_j\} - \tilde{A}_0\{\tilde{T}_j\})$, where $\mathscr{I}_n$ is the observed Fisher information with respect to β and jump sizes, and $\widehat{\Sigma}_n = n^{-1}\sum_{i=1}^n(\hat{\eta}_i + \hat{\psi}_i)^{\otimes 2}$ is the estimator for the variance of the score with respect to β and jump sizes, with $\hat{\eta}_i$ and $\hat{\psi}_i$ denoting the components of the iid decomposition as given in Section A.4. Analogously we obtain that a consistent estimator for the covariances is

$$\widehat{\text{Cov}}\left(\sqrt{n}(\hat{\theta}_n - \theta_0)(\tilde{h}), \sqrt{n}(\hat{\theta}_n - \theta_0)(\tilde{g})\right) = \tilde{h}^T\mathscr{I}_n^{-1^T}\widehat{\Sigma}_n\mathscr{I}_n^{-1}\tilde{g}.$$

The sandwich estimator contains a term that reflects the additional variability resulting from estimating the weights. Similarly to Fine and Gray (1999), this additional variability cannot be ignored, even though in practice the adjusted sandwich estimator may be quite close to the empirical inverse Fisher information.

4 Simulation studies

Simulation studies were conducted for the Fine-Gray model and for the proportional odds model with sample sizes $n = 50$, 200 and 500. Event times were generated using the model formulations in Fine and Gray (1999) and in Fine (2001), with two distinct failure types. For each setting, 1000 samples were generated. We defined censoring times uniformly distributed on subsets of $[0, \tau]$. Covariates Z_{i1} and Z_{i2} were generated independently from a standard normal distribution for $i = 1, \ldots, n$. The subdistribution for the events of interest was defined by $F_1(t|Z_i) = P(T_i \leq t, \varepsilon_i = 1|Z_i) = 1 - [1 - p\{1 - \exp(-t)\}]^{\exp(Z_{i1}\beta_{11} + Z_{i2}\beta_{12})}$ for the Fine-Gray model and by $F_1(t|Z_i) = \exp[p + \log\{1 - \exp(-1)\} + Z_{i1}\beta_{11} + Z_{i2}\beta_{12}] (1 + \exp[p + \log\{1 - \exp(-t)\}] + Z_{i1}\beta_{11} + Z_{i2}\beta_{12})^{-1}$ for the proportional odds model. The subdistribution for the competing risk events was in both scenarios obtained from the conditional distribution $P(T_i \leq t|\varepsilon_i = 2, Z_i)$ being exponential distributed with rate $\exp(Z_{i1}\beta_{21} + Z_{i2}\beta_{22})$ and by taking $P(\varepsilon_i = 2|Z_i) = 1 - P(\varepsilon_i = 1|Z_i)$. The parameter vector was defined as $(p, \beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}) = (0.3, 0.5, -0.5, 0.5, 0.5)$.

In Tables 2 and 3 we report the average deviation from the true parameter values (bias), empirical standard errors, the mean of the Fine-Gray sandwich estimator and of the inverse Fisher information, and the coverage probabilities for both variance estimators. The bias is generally small, the standard errors decrease at rate $\sqrt{n}$ with the empirical and model based variances in agreement, and the coverage probabilities are close to the nominal level 0.95, particularly with sample sizes $n = 200$ and 500. The solid performance of the weighted NPMLE with realistic sample sizes indicates that the proposed method is reliable for use in real applications. For the Fine-Gray model, the estimated parameters and variances were identical with those derived from the crr function (cmpsrk package). Interestingly, the performances of the proposed sandwich estimator and of the inverse Fisher information, which lacks theoretical justification, are rather similar for all sample sizes, with both approaches yielding diminished coverage with $n = 50$. Additional simulation studies, histograms and qq-plots of the parameter estimates are presented in our technical report.

For the proportional odds model, we compare in table 4 weighted NPMLEs for β_1 with the corresponding parameter estimates derived from the regression model of the competing

crude failure probability, as reported in Fine (2001). Bias and variances of the weighted NPMLEs are smaller, with comparable coverage probabilities.

This is somewhat expected for a likelihood based procedure, which is efficient with complete data and with censoring complete data.

5 Bone marrow transplant dataset

A common side effect of a bone marrow transplant (bmt) as a treatment of leukemia is the occurrence of graft-versus-host disease (GVHD), where white blood cells from the donor attack cells in the patient's body. This complication is closely related to the matching of the human leukocyte antigen (HLA) type between recipient and donor.

We reanalyzed bmt data of 1715 patients who received stem cells from either an HLA-identical sibling, an HLA matched or an HLA mismatched unrelated donor. During 74 months of follow-up, 311 patients had a relapse and 557 patients died in remission. As in Gerds et al. (2011), the main objective was to determine how the transplant of stem cells from unrelated donors would influence the risk for relapse and for death in remission, in comparison to having an HLA matched sibling as a donor.

We defined binary covariates "match" and "mism" indicating if the donor was HLA matched and unrelated or HLA mismatched and unrelated. Seven additional covariates were included in the analysis as displayed in Tables 5 and 6.

Weighted NPMLEs were calculated for transformation models from the logarithmic-and the Box-Cox class. As a simple tool for model selection we applied the Akaike criterion, by ranking models based on the value of the weighted log-likelihood.

As illustrated in Figure 1, the logarithmic transformation model with parameter $r \approx 5$ is the best fitting model for relapse, while the optimal choice for death in remission is the Box-Cox transformation model with parameter $\rho = 1.5$, which is only slightly different from the Fine-Gray model. In Tables 5 and 6 we report the estimated regression parameters for different link functions. Pp-plots, as displayed in our technical report, indicate that none of the models selected by the Akaike criterion exhibits a significant lack of fit. For the Fine-Gray model the estimated values were identical with the results from the crr function.

For relapse and $r = 5$ we can see in Table 5 that "match" is not significant while "mism" is significant with a negative estimated parameter value, indicating that an HLA mismatched and unrelated donor will decrease the risk of relapse in comparison to an HLA matched sibling. This was somewhat expected since GVHD might decrease the risk of relapse, if white blood cells from the donor attack the patients cancer cells.

For death in remission and $r = 0$ the covariates "match" and "mism" are both significant with positive estimated parameter values, as displayed in Table 6. This indicates that an HLA matched unrelated donor will increase the risk of dying in remission in comparison to an HLA matched sibling. For patients receiving stem cells from an HLA mismatched and unrelated donor, the increase in the risk of dying is greater.

The magnitude of the parameter and standard error estimates in Tables 5 and 6 vary across models with different choices of parameter r, owing to the fact that the parameters are defined on different scales. The ratios of the parameter estimates and the estimated standard errors are, however, relatively constant across these models. As established by Kosorok, Lee and Fine (2004) such constancy holds under regularity conditions for misspecified models within classes of semiparametric regression models.

The logarithmic transformation model may be regarded as a proportional hazards frailty model for the CIF, in which the frailty has gamma distribution with parameter r. The estimated covariate effects are thus interpretable as subdistribution hazard ratios in a frailty regression model, where the frailty accounts for important covariates that are omitted from the regression model.

To illustrate the importance of model selection in prediction, we plotted in figure 2 the aggregated CIFs for subgroups of the 727 female patients from different regression models in comparison to the Aalen-Johansen estimator. The baseline hazards for the events relapse and death in remission were obtained by maximizing the weighted log-likelihood functions over the entire population.

Aggregated CIFs for both events were then obtained for the three subgroups by averaging the predicted probabilities across individuals in the subgroup.

For relapse it seems that the prognosis of female patients with a sibling as a donor and female patients with an HLA matched donor is similar. Having an HLA mismatched unrelated donor seems to decrease the probability of a relapse in comparison to the other two subgroups. For the largest subgroup of 530 female patients with an HLA matched sibling as a donor and for the subgroup of 37 female patients with an HLA mismatched donor, the CIFs from the logarithmic transformation model with $r = 5$ are located closest to the nonparametric Aalen Johansen estimator. This means that the prediction of the best model under AIC, $r = 5$, clearly improves on the Fine-Gray model. For the second group of 160 female patients with an HLA matched donor, the performance of the prediction for the logarithmic transformation model is better for the first 30 months. For this subgroup no relapse was observed at a time later than $t = 15$ months.

For death in remission the plots for the three subgroups illustrate how the choice of an unrelated donor leads to a dramatic increase in the risk of dying in remission. For the largest subgroup of 530 female patients with a sibling as a donor and 129 events of interest, all regression models perform equally well, evidenced by good agreement with the Aalen-Johansen estimator. For the subgroups of female patients with an HLA matched unrelated donor or an HLA mismatched unrelated donor, the Fine-Gray model and the Box-Cox transformation model with $\rho = 1.5$ perform similarly well and are clearly superior to the other preselected models. The prediction curves for both subgroups were closest to the Aalen-Johansen estimator.

6 Discussion

We propose a novel weighted likelihood function and establish a general semiparametric regression model for the subdistribution hazard. We thereby provide a flexible likelihood framework for the Fine-Gray model that is extendable in several directions.

The new approach is applicable to a general class of semiparametric transformation models, as previously considered by Zeng and Lin (2006) for simple survival settings without competing events. A major point is that in cases where the proportional hazards assumption cannot be validated, nonproportionality may be captured via the choice of the link function. Alternatively, the use of interaction terms in the proportional hazards model can lead to an improvement with regard to the fit of the model. We present results from such a regression analysis with interaction terms for the bmt dataset in Section B.5.2 of our technical report. Difficulties in interpretation arise from the fact that covariates and the corresponding interaction terms have in many cases opposing directions. With our proposed general regression model nonproportionality can be addressed more parsimoniously and estimated covariate effects are immediately interpretable.

A useful feature of the weighted likelihood approach is that the Akaike criterion can easily be adapted as a simple tool for model selection, as we demonstrate for the bmt dataset. To address the overall question of fit we propose a method based on pp-plots in Section B.5.1 of our technical report. Another method to check the fit of the model might be to compare the prediction plots for different transformation models to the Aalen-Johansen estimator within particular subgroups of the dataset, as conducted in Section 5. Investigating other model selection techniques for semiparametric transformation models would be an important topic for further research.

Furthermore, the weighted likelihood approach enables modeling time-dependent covariate effects on the subdistribution hazard, as opposed to earlier work on direct regression modeling for competing risks, where covariate effects target the CIF.

The one-two-one correspondence between the subdistribution hazard and the CIF holds with either time-independent, eg, baseline covariates, or with external time-dependent covariates in the same way as for the survival model without competing risk (Kalbfleisch and Prentice, 2002). External time-dependent covariates might include interactions of time-independent covariates and time used to capture nonproportionality or exogenous covariates. As an example of a nontrivial external covariate, in evaluating hospital performance, where death serves as a competing risk for discharge, operational characteristics of the hospital, including staffing factors and level of hospital utilization, change over time. The cumulative incidences of discharge and death may be derived from the subdistribution hazard conditionally on these covariates, which are observed after a patient's event.

For internal time-dependent covariates, it is not possible to establish a direct relation between the subdistribution hazard and the CIF. The problem is well known even for survival settings without competing risks, where regression parameters for the hazard rate in the presence of internal time-dependent covariates are not interpretable with regard to the underlying distribution, see Kalbfleisch and Prentice (2002). For procedures which directly

model the absolute risk conditionally on time-dependent covariates instead of the subdistribution hazard, similar problems occur. Regression modeling of the subdistribution hazard with internal time-dependent covariates is a more delicate issue as covariates are not generally available after a competing risk event. Similar challenges arise when directly modeling the absolute risk of an event by a particular time point conditionally on the value of internal time-dependent covariates, see Klein and Andersen (2005) and Scheike, Zhang and Gerds (2008). The approach proposed by Beyersmann and Schumacher (2008) or alternative methods of extrapolation may be applied in our weighted NPMLE procedure. In contrast to that, modeling the cause specific hazard with internal time dependent covariates is straightforward. The approach, however, does not alleviate the problem of interpretability of the estimated regression parameters with regard to the CIF.

A heuristic approach to the subdistribution hazard modeling is elaborated in Section 2. The idea of modeling the subdistribution hazard from a cure model perspective leads to a hazard ratio interpretation which has been accepted by many practitioners, with widespread application in substantive biomedical papers.

To evaluate the options of either modeling the subdistribution hazard or the cause specific hazard, the main consideration should be the scientific objectives of a clinical trial. Modeling the cause-specific hazards is considered as intuitive with regard to a traditional risk set of individuals with no prior event. The sum of the cause specific hazards is the total hazard, which is directly related to the overall survival function.

An advantage of a regression models targeting either the subdistribution hazard or directly the CIF is that it quantifies the absolute risk of a particular event type. Lau et al (2009) suggest that cause specific hazard models are "better suited for studying the etiology of diseases, while the subdistribution hazard model has use in predicting an individual's absolute risk". Similarly, Wolbers et al (2009) suggested that subdistribution hazard methods are preferable when the focus is on actual risks and prognosis, which are not captured by a single cause specific hazard function.

With regard to the theoretical foundation of the weighted NPMLE, several new results are presented in Section A.2 and in our technical report. A proof for the existence of the weighted NPMLE and the uniform boundedness of the baseline hazard is provided that requires only one of the weak model assumptions M4a) or M4b), whereas Zeng and Lin's model assumptions are rather restrictive. A challenge thereby is that the fundamental theorem of calculus cannot be routinely applied, which has previously been neglected in theoretical work on NPMLE.

The proposed sandwich estimator is theoretically justified by a new lemma for weighted $\mathscr{Z}$–estimators that may be regarded as a Wald type argument in the light of Theorem 3.3.1 of van der Vaart and Wellner being based on a Taylor series expansion in abstract spaces. The arguments for weighted NPMLEs are similar but considerably more complex than those for weighted parametric maximum likelihood estimators.

Finally, an extension of the weighted NPMLE to the clinically relevant setting of recurrent events with competing terminal events appears to be an obvious next step. Let D denote the

time of the competing terminal event and let $N(t)$ denote the process counting the number of recurrent events before D. The marginal mean intensity, defined as $\lambda(t) = E[dN(t)|N(s), s < t]$, has the desirable property that the usual relationship with the marginal mean is retained, that is $E[N(t)] = \int_0^t \lambda(u)du$. With a weight function defined as in Ghosh and Lin (2002) the general model (1) in Section 3.1 may be proposed for the marginal mean intensity. Even for the particular choice of the link function $\mathscr{G}(x) = x$, this approach does not represent a Markov model and, as for the Andersen Gill model, it would be suitable in cases where the dependency structure between the recurrent events for a particular individual can be mediated by time dependent covariates. A general regression model for the marginal mean intensity is then obtained from the modified weighted log-likelihood function. It will be of interest for further research to investigate a general regression model for the marginal mean intensity in greater depth.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Andersen, PK, Borgan, ∅, Gill, RD, Keiding, N. Statistical Models Based on Counting Processes. New York: Springer-Verlag; 1993.

2. Antolini L, Biganzoli E, Boracchi P. 2007; Crude cumulative incidence as a Horvitz-Thompson like and a Kaplan-Meier like estimator. Am J Epidemiol. 3:1–8.

3. Andersen PK, Klein JP. 2007; Regression analysis for multistate models based on a pseudo-value approach, with applications to bone marrow transplantation studies. Scand J Stat. 34(1):3–16.

4. Bellach, A. master's thesis. University of Freiburg; Breisgau: 2011. Models for the hazard rate of the subdistribution in competing risks settings. née Barz

5. Beyersmann J, Schumacher M. 2008; Time-dependent covariates in the proportional subdistribution hazards model for competing risks. Biostatistics. 9(4):765–776. [PubMed: 18434297]

6. Chen K, Jin Z, Ying Z. 2002; Analysis of transformation models with censored data. Biometrika. 89:659–668.

7. Cheng SC, Wei LJ, Ying Z. 1997; Predicting survival probabilities with semiparametric transformation models. JASA. 92(437):227–235.

8. Fine JP. 2001; Regression models of competing crude failure probabilites. Biostatistics. 2(1):85–97. [PubMed: 12933558]

9. Fine JP, Gray RJ. 1999; A proportional hazards model for the subdistribution of a competing risk. JASA. 94(446):496–509.

10. Gerds TA, Scheike TH, Andersen PK. 2011; Absolute risk regression for competing risks: interpretation, linc functions, and prediction. Stat Med. 31:3921–3930.

11. Geskus RB. 2010; Cause-specific cumulative incidence estimation and the Fine and Gray model under both left truncation and right censoring. Biometrics. 67(1):39–49.

12. Ghosh D, Lin DY. 2002; Marginal regression models for recurrent and terminal events. Statistica Sinica. 12:663–688.

13. Gray RJ. 1988; A class of *k*-sample tests for comparing the cumulative incidence of a competing risk. Ann Statist. 16:1141–1154.

14. Kalbfleisch, JD, Prentice, RL. Wiley Series in Probability and Statistics. 1980. The Statistical Analysis of Failure Time Data. 2002

15. Klein JP, Andersen PK. 2005; Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. Biometrics. 61(1):223–229. [PubMed: 15737097]

16. Kosorok, MR. Introduction to Empirical Processes and Semiparametric inference. New York: Springer-Verlag; 2008.

17. Kosorok MR, Lee BL, Fine JP. 2004; Robust inference for univariate proportional hazards frailty regression models. Ann Statist. 32(4):1448–1491.

18. Latouche A, Allignol A, Beyersmann J, Labopin M, Fine JP. 2013; A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. J Clin Epidemiol. 66(6):648–653. [PubMed: 23415868]

19. Lau B, Cole SR, Gange SJ. 2009; Competing risk regression models for epidemiologic data. Am J Epidemiol. 170(2):244–56. [PubMed: 19494242]

20. Li, C, Gray, RJ, Fine, JP. Technical Report Series, Working Paper 29. The University of North Carolina at Chapel Hill, Department of Biostatistics; 2012. Reader reaction: on variance estimation for the Fine-Gray model.

21. Murphy SA. 1994; Consistency in a proportional hazards model incorporating a random effect. Ann Statist. 22(2):712–731.

22. Murphy SA. 1995; Asymptotic theory for the frailty model. Ann Statist. 23(1):182–198.

23. Parner E. 1998; Asymptotic theory for the correlated gamma-frailty model. Ann Statist. 26(1):183–214.

24. Prentice R, Kalbfleisch J, Peterson A, Flournoy N, Farewell V, Breslow N. 1978; The analysis of failure times in the presence of competing risks. Biometrics. 34:541–554. [PubMed: 373811]

25. Scheike TH, Zhang MJ, Gerds TA. 2008; Predicting cumulative incidence probability by direct binomial regression. Biometrika. 95:205–220.

26. Sydlo R, Goldman JM, Klein JP, Gale RP, Ash RC, Bach FH, et al. 1997; Result of allogeneic bone marrow transplants for leukemia using donors other than HLA-identical siblings. J Clin Oncol. 15(5):1767–77. [PubMed: 9164184]

27. Tai BC, Machin D, White I, Gebski V. 2001; Competing risks analysis by patients with osteosarcoma: a comparison of four different approaches. Statist Med. 20:661–684.

28. van der Vaart, AW. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press; 1998. Asymptotic Statistics.

29. van der Vaart, AW, Weller, JA. Weak Convergence and Empirical Processes. New York: Springer-Verlag; 1996.

30. Wolbers M, Koller MT, Witteman JC, Steyerberg EW. 2009; Prognostic models with competing risks: methods and application to coronary risk prediction. Epidemiology. 20(4):555–561. DOI: 10.1097/EDE.0b013e3181a39056 [PubMed: 19367167]

31. Zeng D, Lin DY. 2006; Efficient estimation of semiparametric transformation models for counting processes. Biometrika. 93(3):627–640.

A Appendix

A.1 Model conditions

M1) The cumulative baseline $A_0(t)$ is a strictly increasing and continuously differentiable function and β_0 lies in the interior of a compact set $\mathscr{C}$.

M2) The vector of covariates $Z(t)$ is P-almost surely of bounded variation on the observed interval $[0, \tau]$.

M3) The endpoint of the study τ is chosen in a way that P-almost surely there exists a constant $\delta > 0$ such that $P(C \quad \tau|Z) > \delta$ and $P(X \quad \tau|Z) > \delta$.

M4) $\mathscr{G}$ is a thrice continuously differentiable and strictly increasing function with $\mathscr{G}(0) = 0$, $\mathscr{G}'(0) > 0$ and $\mathscr{G}(\infty) = \infty$. In addition to that one of the following conditions is required:

  a. $) \mathscr{G}''(x) \quad 0$ for $x > 0$ or

  b. $\mathscr{G}''(x) \quad 0$ for $x > 0$. In addition to that for any $a \in (0,\infty)$ and for any sequence $(x_n) \subset \mathbb{R}$ with $x_n \to \infty$ as $n \to \infty$,

$$\lim_{n \to \infty} \mathbb{P}(X \le \tau, \Delta\varepsilon = 1) \left[\frac{\log (ax_n)}{\mathscr{G}(a^{-1}x_n)} + \frac{\log \mathscr{G}'(ax_n)}{\mathscr{G}(a^{-1}x_n)} \right] < P(X \ge \tau). \quad (*)$$

M5) *Identifiability condition*: If $h_1 \in \mathbb{R}^d$ and $h_2 \in \mathscr{D}[0, \tau]$ exist such that $h_1^T Z(t) + h_2(t) = 0$ P-almost surely, then $h_1 = 0$ and $h_2(t) = 0 \ \forall t \in [0, \tau]$.

M6) For any $h_1 \in \mathbb{R}^d$ and for any $h_2 \in \mathscr{D}[0, \tau]$ exists a subset $\mathscr{S} \subset [0, \tau]$ of nonzero Lebesgue measure such that $\forall t \in \mathscr{S}$

$$h_1^T Z(t) + h_2(t) \ne -\frac{\mathscr{G}''\left(\int_0^t e^{\beta_0^T Z(u)} dA_0(u) \right)}{\mathscr{G}'\left(\int_0^t e^{\beta_0^T Z(u)} dA_0(u) \right)} \int_0^t e^{\beta^T Z(u)}(h_1^T Z(u) + h_2(u)) dA_0(u).$$

For the consistency of the weighted NPMLE, model conditions M1)-M5) and twice continuous differentiability of $\mathscr{G}$ are sufficient. Model conditions M1)-M2),M5)-M6) and thrice continuous differentiability of $\mathscr{G}$ are sufficient to obtain weak convergence. For the proportional hazards (Fine-Gray) model conditions M5) and M6) are equivalent. Condition M4a) is satisfied for example for the class of logarithmic transformation models and for the Box-Cox transformation models with $\rho \in [0, 1]$. M4b) holds for the Box-Cox transformation models with $\rho > 1$. It is sufficient for (*) to prove that $\log(ax_n)/\mathscr{G}(a^{-1}x_n) \to 0$ and $\log \mathscr{G}'(ax_n)/\mathscr{G}(a^{-1}x_n) \to 0$ for $x_n \to \infty$.

A.2 Consistency

We show that $\hat{A}_n^0(\tau)$ is bounded P-almost surely, thereby ascertaining the existence of the weighted NPMLE. Then $\hat{A}_n^0(t)$ is bounded on $[0, \tau]$ uniformly P-almost surely. From Helly's selection theorem it is then obtained that $\exists n_k \subset \mathbb{N} : \hat{A}_{n_k}^0 \overset{*}{\to} A^*$, for a limit A^*, with * denoting outer almost sure convergence. With a Kullback-Leibler argument it is ascertained that every

subsequence $(\hat{\beta}_{nk}, \hat{A}^0_{n_k})$ converges to the true parameter (β_0, A_0). $\hat{A}_n(t)$ is a sequence of monotone increasing functions, and the limit $A_0(t)$ is continuous. From this $\|\hat{A}_n - A_0\|\ell^\infty_{[0,\tau]} \rightarrow 0 \, [P]$ and $|\hat{\beta}_n - \beta_0| \rightarrow 0 \, [P]$ where $\ell^\infty[0, \tau]$ denotes the space of bounded functions on $[0, \tau]$.

Existence of $(\hat{\beta}_n, \hat{A}^0_n)$ and boundedness of $\hat{A}^0_n(\tau)$ under model condition M4a)

We define $\widetilde{A}^0_n(t) := n^{-1}\sum_{i=1}^n N_i(t)$. By Jensen's inequality it is ascertained that

$$n^{-1}[\ell(\hat{A}^0_n, \hat{\beta}_n) - \ell(\widetilde{A}^0_n, \beta_0)] \leq \left[\mathbb{P}_n \mathbb{1}\{X \leq \tau, \Delta\varepsilon = 1\}\frac{\log \mathscr{G}\left(e^{-M\hat{A}^0_n(\tau)}\right)}{\mathscr{G}\left(e^{-M\hat{A}^0_n(\tau)}\right)} - \mathbb{P}_n \mathbb{1}\{X \geq \tau\}\right]\mathscr{G}(e^{-M\hat{A}^0_n(\tau)}) + O_p(1)$$

.

Existence of $(\hat{\beta}_n, \hat{A}^0_n)$ and boundedness of $\hat{A}^0_n(\tau)$ under model condition M4b)

$$n^{-1}[\ell(\hat{A}^0_n, \hat{\beta}_n) - \ell(\widetilde{A}^0_n, \beta_0)] \leq \left\{\mathbb{P}_n \mathbb{1}\{X \leq \tau, \Delta\varepsilon = 1\}\left[\frac{\log \mathscr{G}'(e^{M\hat{A}^0_n(\tau)})}{\mathscr{G}(e^{-M\hat{A}^0_n(\tau)})} + \frac{\log(e^{M\hat{A}^0_n(\tau)})}{\mathscr{G}(e^{-M\hat{A}^0_n(\tau)})}\right] - \mathbb{P}_n \mathbb{1}\{X \geq \tau\}\right\} \times \mathscr{G}$$

$(e^{-M\hat{A}^0_n(\tau)}) + O_p(1)$.

The conclusion in both cases is that if $\hat{A}^0_n(\tau)$ became infinitely large, the right hand side would go to $-\infty$, which contradicts the definition of $(\hat{A}_n, \hat{\beta}_n)$ as a maximum likelihood estimator.

Competing risks setting with administrative censoring

From differentiating the discretized log-likelihood with respect to jumps sizes we obtain

$$\hat{A}^0_n(t) = \sum_{i: X_i \leq t, \Delta_i \varepsilon_i = 1}\left[n \cdot \Phi^a_n(X_i, \hat{A}^0_n, \hat{\beta}_n)\right]^{-1} = \int_0^t \frac{\frac{1}{n}\sum_{i=1}^n \mathbb{1}\,(C_i \geq s)dN_i(s)}{|\Phi^a_n(s, \hat{A}^0_n, \hat{\beta}_n)|},$$

with $\Phi^a_n(s, \hat{A}^0_n, \hat{\beta}_n)$ as provided in our technical report. Substituting estimated parameters by the true model parameters we obtain

$$\widetilde{A}_n^a(t) = \int_0^t \frac{\frac{1}{n}\Sigma_{i=1}^n \mathbb{1}(C_i \geq s)dN_i(s)}{|\widetilde{\Phi}_n^a(s, A_0, \beta_0)|},$$

and by Doob decomposition of the counting process $N_i(t)$

$$\widetilde{\Phi}_n^a(s, \beta_0, A_0) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}(C_i \geq s)Y_i(s)e^{\beta_0^T Z_i(s)}\mathscr{G}'\left(\int_0^s e^{\beta_0^T Z_i(u)}dA_0(u)\right) + o_p(1).$$

By the Glivenko-Cantelli theorem $\widetilde{\Phi}_n^a(s, \beta_0, A_0)$ converges uniformly to $E[\eta(s, \beta_0, A_0)]$ with

$$\eta(s, \beta_0, A_0) \equiv \mathbb{1}(C \geq s)Y(s)e^{\beta_0^T Z(s)}\mathscr{G}'\left(\int_0^s Y(u)e^{\beta_0^T Z(u)}dA_0(u)\right)$$

and another application of Doob decomposition implies

$$\widetilde{A}_n^a(t) = A_0(t) + n^{-1}\sum_{i=1}^n \int_0^t \frac{\mathbb{1}(C_i \geq s)dM_i(s)}{E[\eta(s, \beta_0, A_0)] + o_p(1)} + o_p(1).$$

By the Glivenko-Cantelli theorem $\widetilde{A}_n^a(t) \rightarrow A_0(t)$ for $n \rightarrow \infty$ uniformly almost surely in t. Further the Glivenko-Cantelli theorem implies that $\Phi_n^a(s, \hat{A}_n^0, \hat{\beta}_n)$ converges uniformly to a continuously differentiable function $\Phi_*(s, A_*, \beta_*)$ and the limit is bounded away from zero. $\hat{A}_n^0(t)$ is absolutely continuous with respect $\widetilde{A}_n^a(t)$ and $\hat{A}_n^0(t)/\widetilde{A}_n^a(t)$ converges to $a_*(t) = E[\eta(s, \beta_0, A_0)]/\Phi_*(s, A_*, \beta_*)$. Therefore,

$$\hat{A}_n^0(t) \rightarrow A_*(t) = \int_0^t \frac{E[\eta(s, \beta_0, A_0)]}{\Phi_*(s, A_*, \beta_*)}dA_0(s).$$

With a Kullback-Leibler argument it is then obtained that the limit $A_*(t)$ is P-almost surely the true baseline hazard $A_0(t)$. For the setting with administrative censoring we define the Kullback-Leibler distance

$$\mathscr{K}_a(\theta_0, \theta_*) = E_{F_{\theta_0}, G}\left[\log\left[\left(\frac{f_{\theta_0}(X)}{f_{\theta_*}(X)}\right)^{\Delta\mathbb{1}(\varepsilon=1)}\left(\frac{S_{\theta_0}(C \wedge \tau)}{S_{\theta_*}(C \wedge \tau)}\right)^{1-\Delta}\left(\frac{S_{\theta_0}(C \wedge \tau)}{S_{\theta_*}(C \wedge \tau)}\right)^{\Delta\mathbb{1}(\varepsilon \neq 1)}\right]\right],$$

thereby denoting $\theta_0 = (\beta_0, A_0)$ and $\theta_* = (\beta_*, A_*)$ and with $f_\theta(t) = a_\theta(t) S_\theta(t)$ being the subdensity for the event of interest. As derived in Section B.3 this Kullback-Leibler distance is nonnegative. On the other hand, as $\hat{\theta}_n = (\hat{\beta}_n, \hat{A}_n^0(t))$ maximizes the log-likelihood function it is obtained that $\mathscr{K}_a(\theta_0, \theta_*) \leq 0$ and thus $\mathscr{K}_a(\theta_0, \theta_*) = 0$. It is then ascertained Section B.3 of our technical report that the Kullback-Leibler distance takes the value zero if and only if $\beta_* = \beta_0$ and $A_* = A_0$. From this we conclude that $\hat{\beta}_n \to \beta_0$ P-almost surely and $\hat{A}_{n_k}^0 \to A_0$ uniformly P-almost surely.

Competing risks setting with independent right censoring

From maximizing the discretized likelihood function we obtain

$$\hat{A}_n(t) = \sum_{i: X_i \leq t, \Delta_i \varepsilon_i = 1} \left[n \cdot \Phi_n(X_i, \hat{A}_n, \hat{\beta}_n) \right]^{-1} = \int_0^t \frac{\frac{1}{n}\sum_{i=1}^n \mathbb{1}\,(C_i \geq s) dN_i(s)}{|\,\Phi_n(s, \hat{A}_n, \hat{\beta}_n)\,|},$$

with $\Phi_n(s, \hat{A}_n, \hat{\beta}_n)$ as provided in our technical report. Substituting estimted parameters by the true model parameters we define

$$\tilde{A}_n(t) = \int_0^t \frac{\frac{1}{n}\sum_{i=1}^n \mathbb{1}\,(C_i \geq s) dN_i(s)}{|\,\tilde{\Phi}_n(s, A_0, \beta_0)\,|}.$$

With an application of empirical process theory it is ascertained that $\tilde{\Phi}_n(s, \beta_0, A_0) = \tilde{\Phi}_n^a(s, \beta_0, A_0) + o_p(1)$ and thus $\tilde{A}_n(t) \to A_0(t)$. By the Glivenko-Cantelli theorem $\Phi_n(s, \hat{\beta}_n, \hat{A}_n)$ converges uniformly to a continuously differentiable function $\Phi^*(s, A^*, \beta^*)$ with $\Phi^*(s, A^*, \beta^*) > 0$ for $s \in [0, \tau]$ and

$$\hat{A}_n(t) \to A^*(t) = \int_0^t \frac{E[\eta(s, \beta_0, A_0)]}{\Phi^*(s, \beta^*, A^*)} dA_0(s).$$

$A^*(t)$ is absolutely continuous with respect to the Lebesgue measure and the Radon–Nikodym derivative takes the form $a^*(t) = E[\eta(t, \beta_0, A_0)]/\Phi^*(t, \beta^*, A^*) a_0(t)$.

We define the Kullback-Leibler distance corresponding to the weighted log-likelihood function as

$$\mathscr{K}_{w^*}(\theta_0, \theta_*)$$

$$= E_{F_{\theta_0}, G}\left[\log \left\{ \left(\frac{f_{\theta_0}(X)}{f_{\theta_*}(X)}\right)^{\Delta \mathbb{1}(\varepsilon = 1)} \left(\frac{S_{\theta_0}(C \wedge \tau)}{S_{\theta_*}(C \wedge \tau)}\right)^{1 - \Delta} \times \left(\frac{\exp\left(-\int_0^\tau \widetilde{w}(t) dA_{\theta_0}(t)\right)}{\exp\left(-\int_0^\tau \widetilde{w}(t) dA_{\theta_*}(t)\right)}\right)^{\Delta \mathbb{1}(\varepsilon \neq 1)} \right\} \right].$$

From the asymptotic equivalence of $\mathscr{K}_a(\theta_0, \theta_*)$ and $\mathscr{K}_{w^*}(\theta_0, \theta_*)$ it can be concluded that $\beta_* = \beta_0$ and $A_* = A_0$.

A.3 Weak convergence

Let $\mathscr{H}$ denote the space of elements $h = (h_1, h_2)$ with $h_1 \in \mathbb{R}^d$ and $h_2 \in \mathscr{D}[0, \tau]$. A norm on $\mathscr{H}$ is then defined by $\|h\|_{\mathscr{H}} = \|h_1\| + \|h_2\|_v$, with $\|\cdot\|$ denoting the Euclidean norm and $\|\cdot\|_v$ denoting the total variation norm. For $p < \infty$ we define $\mathscr{H}_p = \{h \in \mathscr{H} : \|h\|_{\mathscr{H}} \leq p\}$. The parameter space is denoted by $\Theta = \{\theta = (\beta, A_0)$, with $\beta \in \mathbb{R}^d$ and A_0 being a monotone increasing element of $\mathscr{D}[0, \tau]\}$. For $h \in \mathscr{H}_p$ we define $\theta(h) = h_1^T \beta + \int_0^\tau h_2(u) dA_0(u)$, so that $\Theta \subset \ell^\infty(\mathscr{H}_p)$. One-dimensional submodels of the form $t \to \theta_t = \theta + t(h_1, \int_0^{(\cdot)} h_2(u) dA_0(u))$ are considered with $h \in \mathscr{H}_p$ to define the empirical score operator

$$\Psi_n(\theta, h, w) = \frac{\partial}{\partial t} \ell_n(\theta_t, h, w)\Big|_{t = 0} = \mathbb{P}_n \psi(\theta, h, w) = \mathbb{P}_n(\psi_1(\theta, h, w) + \psi_2(\theta, h, w)),$$

for a measurable function $\psi(\theta, h, w)$, where $\Psi_{n1}^w(\theta, h, w) = \mathbb{P}_n \psi_1(\theta, h, w)$ is the component related to the derivative with regard to β and $\Psi_{n2}^w(\theta, h, w) = \mathbb{P}_n \psi_2(\theta, h, w)$ is the component related to the derivative along the submodel for A_0. The limiting version Ψ is defined by replacing the empirical measure $\mathbb{P}_n$ by the probability measure $\mathscr{P}$.

Weak convergence is ascertained by a new lemma for weighted $\mathscr{Z}$–estimators, that is based on Theorem 3.3.1. of van der Vaart and Wellner (1996):

Lemma 3

Let the parameter set Θ be a subset of a Banach space. Let $\widetilde{w}(t)$ be a bounded deterministic weight function and let $\hat{w}_n(t)$ be a sequence of bounded random weight functions with values in $\mathbb{R}_+$. Let Ψ_n and Ψ be a linear random map and a linear deterministic map, respectively from $\Theta \times \mathbb{R}_+$ into a Banach space such that

    a) $\sqrt{n}(\Psi_n - \Psi)(\hat{\theta}_n, \hat{w}_n) - \sqrt{n}(\Psi_n - \Psi)(\theta_0, \hat{w}_n) = o_p^*\left(1 + \sqrt{n}\|\hat{\theta}_n - \theta_0\|\right),$

such that $\sqrt{n}(\Psi_n - \Psi)(\theta_0, \widetilde{w})$ converges to a tight limit $\mathcal{Z}_1$ and

$\sqrt{n}\big((\Psi_n - \Psi)(\theta_0, \widehat{w}_n) - (\Psi_n - \Psi)(\theta_0, \widetilde{w})\big)$ converges to a tight limit $\mathcal{Z}_2$, and the sequences jointly converge to $(\mathcal{Z}_1, \mathcal{Z}_2)$.

Let $(\theta, w) \to \Psi(\theta, w)$ be Fréchet-differentiable at $(\theta_0, \widetilde{w})$ with a continuously invertible derivative $\dot{\Psi}_{\theta_0}^{\widetilde{w}}$. *If $\Psi(\theta_0, \widetilde{w})$ and $\hat{\theta}_n$ satisfies $\Psi_n(\hat{\theta}_n, \widehat{w}_n) = o_p^*(n^{-1/2})$, if $\Psi_n(\hat{\theta}_n, \widetilde{w}) = \Psi_n(\hat{\theta}_n, \widehat{w}_n) + o_p(1)$ and if $\Psi_n(\hat{\theta}_n, \widetilde{w})$ converges in outer probability to $\Psi(\theta_0, \widetilde{w})$, then*

$$\sqrt{n}\dot{\Psi}_{\theta_0}^{\widetilde{w}}(\hat{\theta}_n - \theta_0) = -\sqrt{n}\big((\Psi_n - \Psi)(\theta_0, \widetilde{w}) + (\Psi_n - \Psi)(\theta_0, \widehat{w}) - (\Psi_n - \Psi)(\theta_0, \widetilde{w})\big) + o_p^*(1) \quad (5)$$

and $\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow -\Big[\dot{\Psi}_{\theta_0}^{\widetilde{w}}\Big]^{-1}(\mathcal{Z}_1 + \mathcal{Z}_2)$, with $\rightsquigarrow$ denoting weak convergence.

As our model is based on iid observations, for condition *a)* in the above Lemma 3 it is sufficient to verify the two conditions of Lemma 5 in our technical report.

By the Donsker theorem $\sqrt{n}(\Psi_n - \Psi)(\theta_0, \widetilde{w})$ converges in distribution to the tight random element $\mathcal{Z}_1$. Also by Donsker theorem $\sqrt{n}\big((\Psi_n - \Psi)(\theta_0, \widehat{w}_n) - (\Psi_n - \Psi)(\theta_0, \widetilde{w})\big)$ converges in distribution to a tight random element $\mathcal{Z}_2$. Joint convergence follows from the asymptotic linearity of the two components marginally, combined with the fact that the composition of two Donsker classes is also Donsker. Per definition $\Psi_n(\hat{\theta}_n, \widehat{w}_n) = 0$. As argued in Parner (1998) from the Kullback-Leibler information being positive and by interchanging expectation and differentiation we obtain $\Psi(\theta_0, \widetilde{w}) = 0$. Arguments for the continuous invertibility of $\dot{\Psi}_0$ and $\dot{\Psi}_0^a$ under model conditions *M6),M7)* are provided in our technical report.

From Lemma 3 we obtain weak convergence of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ and the covariances for the limiting process $-\dot{\Psi}_0^{-1}(\mathcal{Z}_1 + \mathcal{Z}_2)$, with $\mathcal{Z}_1(h) = \lim_{n \to \infty} \mathbb{P}_n \, \psi(\hat{\beta}_n, \hat{A}_n, h, \widetilde{w})$ and $\mathcal{Z}_2(h) = \lim_{n \to \infty} \mathbb{P}_n[(\psi(\hat{\beta}_n, \hat{A}_n, h, \widehat{w}_n) - \psi(\hat{\beta}_n, \hat{A}_n, h, \widetilde{w})]$, are given by

$$\mathscr{P}\left\{\Big[(\mathcal{Z}_1 + \mathcal{Z}_2)\big(\dot{\Psi}_0^{-1}(g)\big)\Big]\Big[(\mathcal{Z}_1 + \mathcal{Z}_2)\big(\dot{\Psi}_0^{-1}(h)\big)\Big]\right\}$$

for $g, h \in \mathcal{H}$ (Kosorok, 2008, page 302).

A.4 Variance estimation

The middle part of the sandwich variance estimator, as proposed in section 3.3, is obtained from the iid decomposition of the score with $\eta_{i,n} = \big(\eta_i^1, \ldots, \eta_i^{d+k(n)}\big)$ and $\psi_{i,n} = \big(\psi_i^1, \ldots, \psi_i^{d+k(n)}\big)$ defined as

$$\eta_{i,n}^\ell = \left[Z_{i\ell}(X_i) + \sum_{\substack{j:X_j \le X_i \\ \Delta_j \varepsilon_j = 1}} Z_{i\ell}(X_j) e^{\beta^T Z_i(X_j)} A_n\{X_j\} \frac{\mathcal{G}''\left(\int_0^{X_i} e^{\beta_0^T Z_i(u)} dA_n(u)\right)}{\mathcal{G}'\left(\int_0^{X_i} e^{\beta^T Z_i(u)} dA_n(u)\right)} \right] \mathbb{1}(\Delta_i \varepsilon_i = 1)$$

$$- \sum_{\substack{j:X_j \le X_i \wedge \tau \\ \Delta_j \varepsilon_j = 1}} Z_{i\ell}(X_j) e^{\beta^T Z_i(X_j)} A_n\{X_j\} \mathcal{G}'\left(\int_0^{X_i \wedge \tau} e^{\beta^T Z_i(u)} dA_n(u)\right)$$

$$- \left[\int_{X_i}^\tau \widetilde{w}_i^*(t) e^{\beta^T Z_i(t)} Z_{i\ell}(t) \mathcal{G}'\left(\int_0^t e^{\beta^T Z_i(u)} dA_n(u)\right) dA_n(t) \right] \mathbb{1}(\Delta_i \varepsilon_i = 2)$$

$$- \left[\int_{X_i}^\tau \widetilde{w}_i^*(t) e^{\beta^T Z_i(t)} \left(\int_0^t Z_{i\ell}(u) e^{\beta^T Z_i(u)} dA_n(u)\right) \mathcal{G}''\left(\int_0^t e^{\beta^T Z_i(u)} dA_n(u)\right) dA_n(t) \right] \times \mathbb{1}(\Delta_i \varepsilon_i = 2)$$

for $\ell \in \{1, \ldots, d\}$ and

$$\eta_{i,n}^\ell = \left[\mathbb{1}(X_\ell = X_i)[A_n\{X_\ell\}]^{-1} + \mathbb{1}(X_\ell \le (X_i \wedge \tau)) e^{\beta^T Z_i(X_\ell)} \times \mathcal{G}''\left(\int_0^{X_i} e^{\beta^T Z_i(u)} dA_n(u)\right) \right]$$

$$\left| \mathcal{G}'\left(\int_0^{X_i} e^{\beta^T Z_i(u)} dA_n(u)\right) \right| \mathbb{1}(\Delta_i \varepsilon_i = 1)$$

$$- \mathbb{1}(X_\ell \le (X_i \wedge \tau)) e^{\beta^T Z_i(X_\ell)} \mathcal{G}'\left(\int_0^{X_\ell} e^{\beta^T Z_i(u)} dA_n(u)\right)$$

$$- \mathbb{1}(X_i \le X_\ell) \left[\widetilde{w}_i^*(X_\ell) e^{\beta^T Z_i(X_\ell)} \mathcal{G}'\left(\int_0^{X_\ell} e^{\beta^T Z_i(u)} dA_n(u)\right) \right] \mathbb{1}(\Delta_i \varepsilon_i = 2)$$

$$- \left[e^{\beta^T Z_i(X_\ell)} \int_{X_i \wedge X_\ell}^\tau \widetilde{w}_i^*(t) e^{\beta^T Z_i(t)} \mathcal{G}''\left(\int_0^t e^{\beta^T Z_i(u)} dA_n(u)\right) dA_n(t) \right] \mathbb{1}(\Delta_i \varepsilon_i = 2)$$

for $\ell \in \{d+1, \ldots, d+k(n)\}$. To calculate $\psi_{i,n}^\ell$ we apply

$$\frac{\widehat{G}_c(t)}{\widehat{G}_c(X_j)} - \frac{G_c(t)}{G_c(X_j)} = -\frac{G_c(t)}{G_c(X_j)} \sum_{i=1}^n \int_{X_j}^t \frac{1}{\sum_{k=1}^n \mathbb{1}(X_k \ge u)} \times dM_i^c(u) + o_p(n^{-1/2}),$$

where $M_i^c(t) = \mathbb{1}(X_i \leq t, \Delta_i = 0) - \int_0^t \mathbb{1}(X_i \geq u)dA^c(u)$ is the martingale associated with the censoring process and $A^c(t)$ is the cumulative hazard of the censoring distribution. From this we obtain the representation

$$\psi_{i,n}^\ell = \int_0^\infty q_n^\ell(u)\{\pi_n(u)\}^{-1}dM_i^c(u), \text{ with}$$

$$q_n^\ell(u) = \frac{1}{n}\sum_{j:\Delta_j\varepsilon_j = 2}\left[\int_0^\tau \widetilde{w}_j^*(t)\,\mathbb{1}(X_j \leq u \leq t)e^{\beta^T Z_j(t)}Z_{j\ell}(t)\mathscr{G}'\left(\int_0^t e^{\beta^T Z_j(s)}dA_n(s)\right)dA_n(t)\right.$$

$$+ \frac{1}{n}\sum_{j:\Delta_j\varepsilon_j = 2}\left[\int_0^\tau \widetilde{w}_j^*(t)\,\mathbb{1}(X_j \leq u \leq t)e^{\beta^T Z_j(t)}\left(\int_0^t Z_{j\ell}(s)e^{\beta^T Z_j(s)}dA_n(s)\right) \times \mathscr{G}''\left(\int_0^t e^{\beta^T Z_j(s)}dA_n(s)\right)dA_n(t)\right.$$

$$\left.\vphantom{\int_0^\tau} \right]$$
for $\ell \in \{1, ..., d\}$,

$$q_n^\ell(u) = \frac{1}{n}\sum_{j:\Delta_j\varepsilon_j = 2}\left[\widetilde{w}_j^*(X_\ell)\,\mathbb{1}(X_j \leq u \leq X_\ell)e^{\beta^T Z_j(X_\ell)}\mathscr{G}'(\int_0^{X_\ell} e^{\beta^T Z_j(s)}dA_n(s))\right]$$

$$+ \frac{1}{n}\sum_{j:\Delta_j\varepsilon_j = 2}\left[e^{\beta^T Z_j(X_\ell)}\int_0^\tau \mathbb{1}(X_\ell \leq t)\,\mathbb{1}(X_j \leq u \leq t)\widetilde{w}_j^*(t)e^{\beta^T Z_j(t)}\mathscr{G}''(\int_0^t e^{\beta^T Z_j(s)}dA_n(s))dA_n(t)\right]$$

for $\ell \in \{d+1, ..., d+k(n)\}$ and $\pi_n(u) = n^{-1}\sum_{j=1}^n \mathbb{1}(X_j \geq u)$.
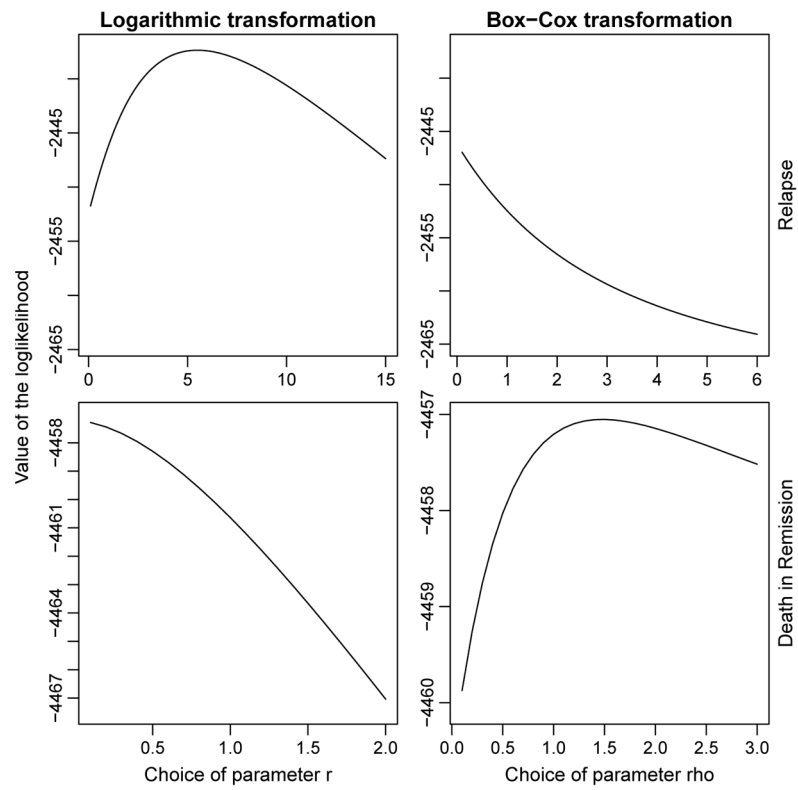
Figure 1.
Model selection with Akaike criterion for the bmt dataset. Value of the loglik in relation to the choice of parameter r for the logarithmic transformation models (left hand side) and of parameter ρ for the Box-Cox class of transformation models (right hand side). First row: relapse, second row: death in remission.
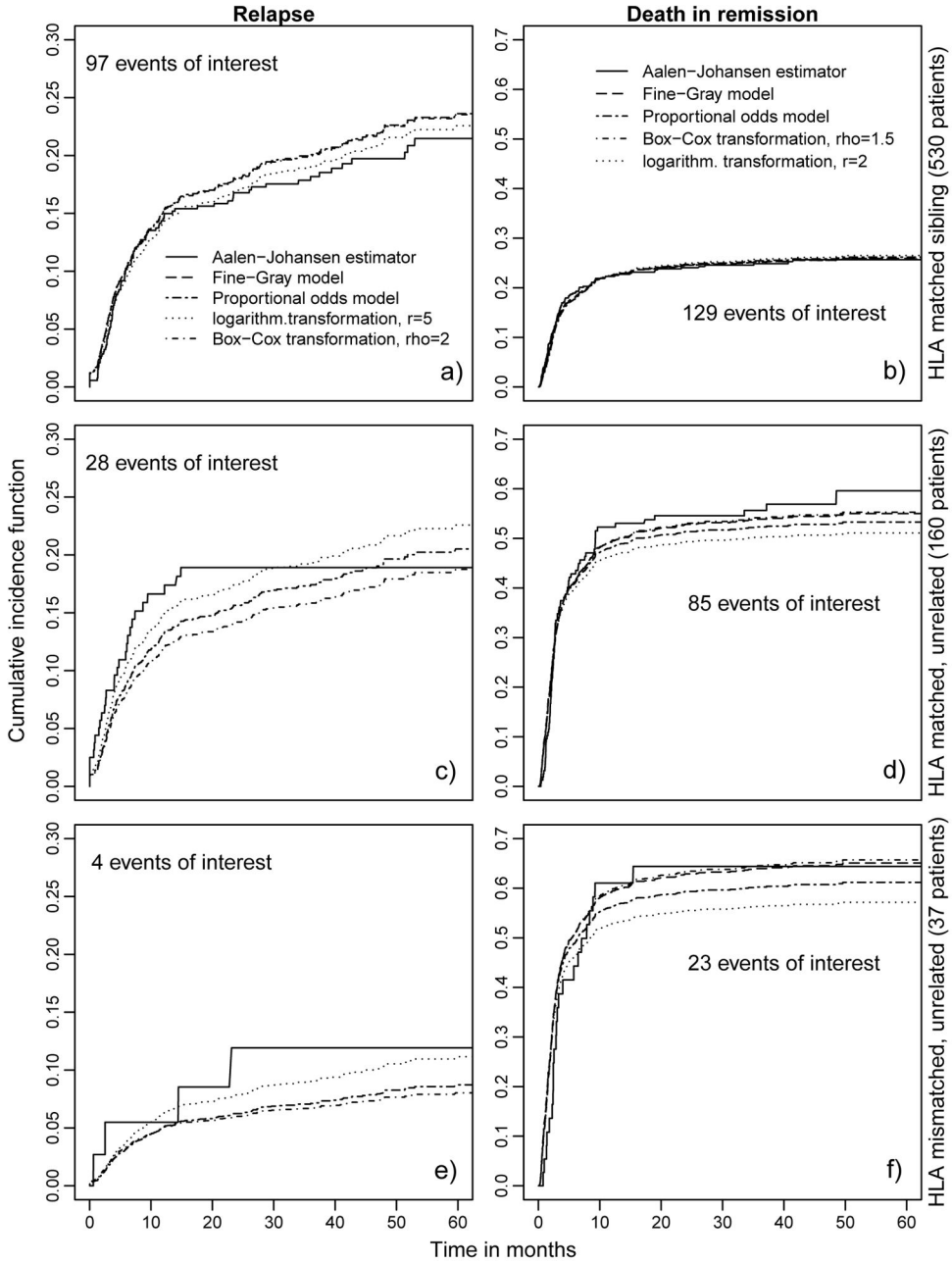
Figure 2.
Prediction plots for the bmt dataset for different subsets of the 727 female patients. For 530 female patients the donor was an HLA matched sibling, for 160 female patients the donor was HLA matched and unrelated, for 37 female patients the donor was HLA mismatched and unrelated.

Table 1

Simulations studies for the weighted NPMLE with $(\beta_{11}, \beta_{12}) = (0.5, -0.5)$. Censorings were uniformly distributed on $[a, b]$, with $b = [\tau]$ indicating that there was only administrative censoring at τ. Columns 5–7: events of interest, competing risk events and censorings in %.

Model	$\mathscr{G}(x)$	a	b	τ	type 1	type 2	cens.
Fine-Gray	x	-	$[\tau]$	4	33	62	5
Fine-Gray	x	1	2	1.9	26	49	25
Fine-Gray	x	0.5	1	0.95	18	36	46
prop.odds	$\log(1 + x)$	-	$[\tau]$	3.5	56	39	5
prop.odds	$\log(1 + x)$	0.75	1.5	1.4	47	29	24
prop.odds	$\log(1 + x)$	0.5	3.5	2.5	51	33	16

Table 2

Simulation studies for the Fine-Gray model, parameter of interest (β_{11}, β_{12}) = (0.5,−0.5). Censorings were uniformly distributed on [a, b], with [τ] denoting that there was only administrative censoring at τ. Bias),%) deviation from true parameter (absolute values and in %), SE) empirical standard error, SEE_1) sandwich estimator, SEE_2) inverse Fisher information, Cov_1, Cov_2) coverage probability of 0.95 confidence intervals for the variance estimators.

size	[a, b]	parameter	Bias and Standard errors					Coverage	
			Bias	%	SE	SEE_1	SEE_2	Cov_1	Cov_2
50	[τ]	β_{11}	0.023	4.7	0.295	0.269	0.280	0.936	0.956
		β_{12}	−0.029	5.8	0.295	0.267	0.279	0.936	0.952
		$A_{r}(\tau/4)$	−0.006	3.0	0.074	0.069	0.070	0.888	0.894
		$A_{r}(\tau/2)$	−0.006	1.9	0.093	0.088	0.090	0.908	0.916
		$A_{r}(\tau)$	−0.008	2.3	0.103	0.097	0.099	0.918	0.923
200	[τ]	β_{11}	0.004	0.7	0.131	0.129	0.130	0.950	0.960
		β_{12}	−0.009	1.8	0.127	0.128	0.130	0.954	0.959
		$A_{r}(\tau/4)$	−0.002	1.1	0.035	0.035	0.035	0.935	0.938
		$A_{r}(\tau/2)$	−0.003	1.1	0.044	0.044	0.045	0.948	0.952
		$A_{r}(\tau)$	−0.005	1.3	0.048	0.049	0.049	0.950	0.946
500	[τ]	β_{11}	−0.002	0.4	0.080	0.081	0.081	0.954	0.953
		β_{12}	−0.001	0.2	0.086	0.081	0.081	0.932	0.938
		$A_{r}(\tau/4)$	−0.001	0.3	0.023	0.022	0.022	0.938	0.939
		$A_{r}(\tau/2)$	0.000	0.2	0.028	0.028	0.028	0.941	0.940
		$A_{r}(\tau)$	−0.001	0.4	0.031	0.031	0.031	0.951	0.952
50	[1, 2]	β_{11}	0.047	9.5	0.360	0.298	0.317	0.920	0.935
		β_{12}	−0.048	9.7	0.362	0.303	0.319	0.913	0.940
		$A_{r}(\tau/4)$	−0.008	6.9	0.050	0.048	0.048	0.866	0.869
		$A_{r}(\tau/2)$	−0.013	6.3	0.072	0.068	0.069	0.890	0.897
		$A_{r}(\tau)$	−0.013	4.5	0.110	0.101	0.100	0.880	0.888

size	[a, b]	parameter	Bias and Standard errors					Coverage	
			Bias	%	SE	SEE_1	SEE_2	Cov_1	Cov_2
200	[1, 2]	β_{11}	0.004	0.7	0.153	0.143	0.145	0.928	0.936
		β_{12}	−0.012	2.4	0.147	0.143	0.145	0.943	0.952
		$A_r(\tau/4)$	−0.002	2.0	0.026	0.025	0.025	0.928	0.927
		$A_r(\tau/2)$	−0.004	2.1	0.036	0.035	0.035	0.934	0.934
		$A_r(\tau)$	−0.012	4.0	0.048	0.048	0.048	0.920	0.924
500	[1, 2]	β_{11}	0.000	0.1	0.091	0.089	0.090	0.940	0.047
		β_{12}	0.000	0.1	0.092	0.089	0.090	0.944	0.951
		$A_r(\tau/4)$	0.001	0.5	0.016	0.016	0.016	0.956	0.954
		$A_r(\tau/2)$	0.000	0.1	0.022	0.022	0.022	0.952	0.955
		$A_r(\tau)$	−0.007	2.3	0.030	0.031	0.031	0.938	0.940
50	[0.5, 1]	β_{11}	0.041	8.2	0.438	0.361	0.385	0.916	0.944
		β_{12}	−0.044	8.7	0.417	0.354	0.383	0.902	0.941
		$A_r(\tau/4)$	−0.017	26.3	0.124	0.040	0.042	0.832	0.847
		$A_r(\tau/2)$	−0.010	8.3	0.051	0.049	0.050	0.855	0.864
		$A_r(\tau)$	−0.016	7.9	0.085	0.079	0.081	0.865	0.860
200	[0.5, 1]	β_{11}	0.003	0.6	0.173	0.170	0.172	0.932	0.949
		β_{12}	−0.007	1.3	0.172	0.170	0.172	0.948	0.954
		$A_r(\tau/4)$	−0.001	1.6	0.018	0.018	0.018	0.935	0.930
		$A_r(\tau/2)$	−0.002	2.0	0.026	0.026	0.026	0.931	0.934
		$A_r(\tau)$	−0.003	1.5	0.045	0.043	0.043	0.912	0.914
500	[0.5, 1]	β_{11}	0.004	0.7	0.110	0.106	0.107	0.936	0.939
		β_{12}	−0.004	0.8	0.105	0.107	0.107	0.954	0.957
		$A_r(\tau/4)$	−0.001	1.5	0.011	0.011	0.011	0.935	0.935
		$A_r(\tau/2)$	−0.001	1.0	0.016	0.017	0.017	0.941	0.941
		$A_r(\tau)$	−0.001	0.4	0.028	0.028	0.028	0.941	0.943

Table 3

Simulation studies for the proportional odds model, parameter of interest (β_{11}, β_{12}) = (0.5,−0.5). Censorings uniformly distributed on [a, b], with [τ] denoting that there was only administrative censoring at τ. Bias),%) deviation from true parameter (absolute values and in %), SE) empirical standard error, SEE$_1$) sandwich estimator, SEE$_2$) inverse Fisher information. Cov$_1$, Cov$_2$) coverage probability of 0.95 confidence intervals for the variance estimators.

size	[a, b]	parameter	Bias	%	SE	SEE$_1$	SEE$_2$	Cov$_1$	Cov$_2$
50	[τ]	β_{11}	0.010	2.0	0.285	0.276	0.287	0.942	0.965
		β_{12}	−0.020	3.9	0.296	0.278	0.288	0.937	0.955
		$A_n(\tau/4)$	−0.005	0.6	0.239	0.237	0.241	0.920	0.923
		$A_n(\tau/2)$	0.007	0.7	0.344	0.339	0.346	0.917	0.920
		$A_n(\tau)$	0.001	0.0	0.390	0.399	0.408	0.923	0.925
200	[τ]	β_{11}	0.007	1.4	0.137	0.136	0.138	0.956	0.959
		β_{12}	0.000	0.0	0.139	0.137	0.138	0.941	0.943
		$A_n(\tau/4)$	−0.004	0.5	0.113	0.117	0.117	0.951	0.951
		$A_n(\tau/2)$	0.005	0.5	0.169	0.167	0.167	0.948	0.950
		$A_n(\tau)$	−0.001	0.0	0.198	0.196	0.197	0.943	0.944
500	[τ]	β_{11}	0.002	0.5	0.085	0.084	0.087	0.934	0.949
		β_{12}	−0.002	0.4	0.085	0.083	0.087	0.950	0.961
		$A_n(\tau/4)$	0.002	0.3	0.068	0.066	0.068	0.947	0.951
		$A_n(\tau/2)$	0.003	0.3	0.099	0.096	0.099	0.944	0.949
		$A_n(\tau)$	0.004	0.3	0.122	0.117	0.122	0.940	0.947
50	[0.75, 1.5]	β_{11}	0.009	1.8	0.310	0.289	0.301	0.938	0.956
		β_{12}	−0.023	4.7	0.315	0.291	0.303	0.934	0.957
		$A_n(\tau/4)$	−0.013	3.3	0.130	0.127	0.129	0.908	0.914
		$A_n(\tau/2)$	−0.012	1.8	0.211	0.205	0.209	0.917	0.918
		$A_n(\tau)$	0.012	1.1	0.467	0.383	0.386	0.898	0.904
200	[0.75, 1.5]	β_{11}	0.004	0.8	0.147	0.143	0.144	0.947	0.951

size	[a, b]	parameter	Bias	%	SE	SEE_1	SEE_2	Cov_1	Cov_2
		β_{12}	−0.011	2.3	0.145	0.142	0.144	0.943	0.950
		$A_{ir}(\tau 4)$	−0.005	1.4	0.063	0.064	0.065	0.940	0.940
		$A_{ir}(\tau 2)$	0.000	0.1	0.103	0.103	0.104	0.946	0.945
		$A_{ir}(\tau)$	0.005	0.5	0.183	0.178	0.178	0.943	0.944
500	[0.75, 1.5]	β_{11}	−0.002	0.3	0.091	0.089	0.090	0.944	0.946
		β_{12}	0.000	0.1	0.087	0.089	0.090	0.961	0.963
		$A_{ir}(\tau 4)$	−0.003	0.6	0.041	0.041	0.041	0.937	0.940
		$A_{ir}(\tau 2)$	−0.003	0.4	0.066	0.065	0.065	0.940	0.939
		$A_{ir}(\tau)$	−0.005	0.5	0.109	0.111	0.110	0.935	0.937
50	[0.5, 3.5]	β_{11}	0.005	0.9	0.294	0.285	0.296	0.945	0.959
		β_{12}	−0.019	3.9	0.298	0.289	0.297	0.944	0.953
		$A_{ir}(\tau 4)$	−0.011	1.8	0.194	0.190	0.193	0.909	0.918
		$A_{ir}(\tau 2)$	0.007	0.7	0.322	0.302	0.309	0.926	0.925
		$A_{ir}(\tau)$	0.039	3.1	0.479	0.450	0.454	0.917	0.920
200	[0.5, 3.5]	β_{11}	0.004	0.9	0.137	0.140	0.141	0.953	0.960
		β_{12}	−0.004	0.8	0.143	0.140	0.141	0.954	0.956
		$A_{ir}(\tau 4)$	−0.006	0.9	0.094	0.095	0.095	0.935	0.938
		$A_{ir}(\tau 2)$	−0.009	0.9	0.145	0.147	0.147	0.942	0.944
		$A_{ir}(\tau)$	−0.009	0.7	0.197	0.210	0.206	0.948	0.944
500	[0.5, 3.5]	β_{11}	0.001	0.3	0.089	0.088	0.088	0.948	0.948
		β_{12}	−0.005	1.0	0.091	0.088	0.088	0.946	0.950
		$A_{ir}(\tau 4)$	−0.001	0.1	0.059	0.060	0.060	0.952	0.951
		$A_{ir}(\tau 2)$	0.000	0.0	0.092	0.093	0.093	0.951	0.951
		$A_{ir}(\tau)$	0.002	0.2	0.129	0.133	0.130	0.955	0.954

Table 4

Weighted NPMLEs for the proportional odds model in comparison to Fine (2001). Censorings uniformly distributed on [*a*, *b*], with [τ] denoting administrative censoring at τ. Bias),%) deviation from true parameter (absolute values and in %), SE) empirical standard error, SEE) sandwich estimator, Cov) coverage probability of 0.95 confidence intervals for the sandwich estimator.

| size | [*a*, *b*] | weighted NPMLE | | | | | Fine (2001) | | | | |
		Bias	%	SE	SEE	Cov	Bias	%	SE	SEE	Cov
50	[τ]	0.010	2.0	0.285	0.276	0.942	0.029	5.8	0.409	0.377	0.955
150	[τ]	0.003	0.7	0.163	0.159	0.940	0.017	3.4	0.205	0.197	0.938
50	[0.75, 1.5]	0.009	1.8	0.310	0.289	0.938	0.093	18.6	0.567	0.482	0.956
150	[0.75, 1.5]	0.003	0.7	0.165	0.164	0.948	0.036	7.2	0.239	0.230	0.954

Table 5

Bmt dataset, weighted NPMLEs for Relapse.

model	Box-Cox transformation		Logarithmic transformation		
		Fine-Gray	prop. odds		
parameter	$\rho = 0.5$	$\rho = 1$	$r = 1$	$r = 2$	$r = 5$
loglik	2449.68	−2452.46	−2446.21	−2442.00	−2437.43
match	−0.233(0.174)	−0.236(0.162)	−0.222(0.191)	−0.181(0.215)	0.001(0.272)
mism	−1.392(0.404)	−1.326(0.381)	−1.473(0.437)	−1.573(0.499)	**−1.678(0.628)**
AML	−0.273(0.163)	−0.249(0.150)	−0.305(0.179)	−0.342(0.200)	−0.380(0.248)
CML	−0.803(0.166)	−0.752(0.158)	−0.869(0.176)	−0.976(0.193)	−1.240(0.239)
wtime	−0.010(0.004)	−0.010(0.004)	−0.011(0.004)	−0.013(0.005)	−0.015(0.006)
sex	−0.068(0.126)	−0.068(0.118)	−0.066(0.135)	−0.065(0.148)	−0.082(0.182)
karn	0.156(0.167)	0.179(0.153)	0.108(0.183)	0.019(0.204)	−0.223(0.254)
sint	0.733(0.178)	0.689(0.169)	0.785(0.189)	0.873(0.208)	1.129(0.267)
sadv	1.700(0.166)	1.557(0.155)	1.885(0.182)	2.173(0.204)	2.835(0.264)
$\hat{A}_r(\tau 4)$	0.184(0.042)	0.172(0.037)	0.244(0.067)	0.244(0.074)	0.395(0.134)
$\hat{A}_r(\tau 2)$	0.231(0.053)	0.212(0.046)	0.324(0.090)	0.324(0.099)	0.580(0.200)
$\hat{A}_r(\tau)$	0.286(0.066)	0.258(0.056)	0.330(0.084)	0.424(0.119)	0.841(0.297)

Table 6

Bmt dataset, weighted NPMLEs for Death in remission.

model parameter	Box-Cox transformation		Fine-Gray	prop. odds	Logarithmic transformation
	$\rho=2$	$\rho=1.5$	$\rho=1$	$r=1$	$r=2$
loglik	−4457.15	**−4457.05**	−4457.21	−4460.64	−4467.04
match	0.598(0.085)	**0.638(0.092)**	0.685(0.099)	0.811(0.125)	0.905(0.170)
mism	0.940(0.113)	**1.005(0.122)**	1.086(0.134)	1.274(0.166)	1.384(0.232)
AML	−0.036(0.129)	−0.043(0.137)	−0.054(0.147)	−0.100(0.178)	−0.147(0.228)
CML	0.329(0.109)	0.342(0.116)	0.355(0.124)	0.363(0.152)	0.350(0.198)
wtime	0.006(0.001)	0.007(0.001)	0.008(0.002)	0.010(0.002)	0.012(0.003)
sex	0.003(0.076)	0.003(0.081)	0.002(0.087)	−0.006(0.106)	−0.019(0.138)
karn	−0.349(0.091)	−0.387(0.098)	−0.437(0.106)	−0.595(0.136)	−0.733(0.188)
sint	0.250(0.094)	0.267(0.100)	0.287(0.108)	0.337(0.131)	0.382(0.175)
sadv	0.367(0.114)	0.398(0.121)	0.438(0.132)	0.567(0.164)	0.678(0.220)
$\hat{A}_f(\tau 4)$	0.224(0.034)	0.239(0.038)	0.258(0.044)	0.326(0.067)	0.425(0.112)
$\hat{A}_f(\tau 2)$	0.238(0.036)	0.255(0.041)	0.276(0.047)	0.356(0.072)	0.472(0.124)
$\hat{A}_f(\tau)$	0.244(0.037)	0.262(0.042)	0.286(0.049)	0.370(0.075)	0.496(0.131)

Tables 5 and 6: Parameter estimates and estimated standard errors (in brackets) for different link functions from the classes of logarithmic- and Box-Cox transformation models. Covariates included were "match" and "mism" indicating if the type of donor was HLA matched and unrelated or HLA mismatched and unrelated in comparison to a having an HLA matched sibling as a donor, "AML" and "CML" indicating the type of leukemia in comparison to ALL, waiting time until transplant ("wtime"), "sex" indicating if the patient was female, "karn" indication if the Karnovsky index was 90, "sint" and "sadv" indicating if the stage of the disease was intermediate or advanced in comparison to early.