B Supplemental technical report

Weighted NPMLE for the subdistribution of a competing risk

by Anna Bellach, Michael R Kosorok, Ludger Rüschendorf and Jason P Fine

B.1 Competing risks model

Setting with independent right censoring. By considering the jump sizes of the baseline as a parameter, maximization of the discretized log-likelihood

$$
\begin{aligned}
\ell \;=\; & \sum_{i:\Delta_i\varepsilon_i=1} \left[\log A_n\{X_i\} + \beta^T Z_i(X_i) + \log \mathcal{G}'\Big(\sum_{\substack{j:X_j\le X_i,\\ \Delta_j\varepsilon_j=1}} e^{\beta^T Z_i(X_j)} A_n\{X_j\} \Big) \right] \\
& - \sum_{i=1}^{n} \left[\mathcal{G}\Big(\sum_{k:\Delta_k\varepsilon_k=1} \mathbb{1}\big((X_i\wedge\tau)\ge X_k\big) e^{\beta^T Z_i(X_k)} A_n\{X_k\} \Big) \right] \\
& - \sum_{i:\Delta_i\varepsilon_i=2} \left[\sum_{k:\Delta_k\varepsilon_k=1} w_i^*(X_k)\mathbb{1}(X_i\le X_k) e^{\beta^T Z_i(X_k)} A_n\{X_k\}\, \mathcal{G}'\Big(\sum_{\substack{j:X_j\le X_k,\\ \Delta_j\varepsilon_j=1}} e^{\beta^T Z_i(X_j)} A_n\{X_j\} \Big) \right]
\end{aligned}
$$

yields the estimator $\widehat{\theta}_n^* = \big(\widehat{\beta}_n, \widehat{A}_n\{\widetilde{T}_1\}, \ldots, \widehat{A}_n\{\widetilde{T}_{k(n)}\}\big)$.

Transformation of the baseline hazard An estimator for the parameter of interest $\widehat{\theta}_n = \big(\widehat{\beta}_n, \widehat{A}_n(\widetilde{T}_1), \ldots, \widehat{A}_n(\widetilde{T}_{k(n)})\big)$ is obtained by the linear transformation $\widehat{\theta}_n = C\cdot\widehat{\theta}_n^*$ with

$$
C = \left(\begin{array}{c|c} A_1 & A_2 \\ \hline A_3 & A_4 \end{array}\right), \qquad C^{-1} = \left(\begin{array}{c|c} A_1 & A_2 \\ \hline A_3 & A_4^{-1} \end{array}\right),
$$

where $A_1 = I_d, \quad A_2 = 0, \quad A_3 = 0,$

$$A_4 = \begin{pmatrix} 1 & 0 & 0 & \dots \\ 1 & 1 & 0 & \dots \\ 1 & 1 & 1 & \dots \\ \vdots & \vdots & \vdots & \end{pmatrix}, \quad \text{and} \quad A_4^{-1} = \begin{pmatrix} 1 & 0 & 0 & \dots \\ -1 & 1 & 0 & \dots \\ 0 & -1 & 1 & \dots \\ \vdots & \vdots & \vdots & \end{pmatrix}.$$

The covariances matrix of $\widehat{\theta}_n$ is then obtained by $C^T \mathcal{I}_n^{-1^T} \widehat{\Sigma}_n \mathcal{I}_n^{-1} C$.

B.2 Proof of existence and of $\widehat{A}_n(\tau)$ being uniformly bounded

B.2.1 Fundamental theorem of calculus: version for Stieltjes integrals

The fundamental theorem of calculus is not applicable to Stieltjes integrals in the common way. For our proof of existence we apply the following lemma:

Lemma 1. *Let $B(x) : [0, \infty) \mapsto [0, \infty)$ be a monotone non-decreasing, càdlàg function and let $H(x) : [0, \infty) \mapsto [0, \infty)$ be a continuously differentiable function. The following integration rules apply for $a, b \in \mathbb{R}_+$ with $a < b < \infty$:*

a) If $H'(x) : [0, \infty) \mapsto (0, \infty)$ is monotone non-decreasing, then

$$\int_a^b H'\big(B(x-)\big) dB(x) \leq H(B(b)) - H(B(a)) \leq \int_a^b H'(B(x)) dB(x).$$

b) If $H'(x) : [0, \infty) \mapsto (0, \infty)$ is monotone non-increasing, then

$$\int_a^b H'\big(B(x)\big) dB(x) \leq H(B(b)) - H(B(a)) \leq \int_a^b H'(B(x-)) dB(x).$$

Proof.
a) $dH\big(B(x)\big) = H\big(B(x-)+\Delta B(x)\big) - H\big(B(x-)\big) = \int_0^1 H'\big(B(x-)+s\Delta B(x)\big) dB(x) ds$

$$\Rightarrow H'\big(B(x-)\big)dB(x) \leq dH\big(B(x)\big) \leq H'\big(B(x)\big)dB(x)$$

$b)$ analogously with $H'\big(B(x)\big)dB(x) \leq dH\big(B(x)\big) \leq H'\big(B(x-)\big)dB(x)$.

B.2.2 Existence of the NPMLE under model condition M4a)

According to model conditions $M1, M2)$ $\sup_{\beta \in \mathcal{C}, t \in [0,\tau]} |\beta^T Z(t)| \leq M[P]$ for a constant M. Defining $\widetilde{A}_n(t) := n^{-1} \sum_{i=1}^n N_i(t)$ we consider

$$\begin{aligned}
&n^{-1}\big[\ell(\widehat{A}_n, \widehat{\beta}_n) - \ell(\widetilde{A}_n, \beta_0)\big] \\
&\quad = \frac{1}{n}\sum_{i=1}^n \left[\int_0^\tau \log\left(e^{\hat{\beta}_n^T Z_i(t)} n\Delta\widehat{A}_n(t)\mathcal{G}'\left(\int_0^t e^{\hat{\beta}_n^T Z_i(u)} d\widehat{A}_n(u)\right)\right)\mathbb{1}(C_i \geq t)dN_i(t)\right] \\
&\qquad\qquad -\frac{1}{n}\sum_{i=1}^n \mathcal{G}\left(\int_0^{\tau\wedge X_i} e^{\hat{\beta}_n^T Z_i(t)} d\widehat{A}_n(t)\right) + O_p(1) \\
&\quad \overset{i)}{\leq} \frac{1}{n}\sum_{i=1}^n \left[\int_0^\tau \log\left(e^{-M} n\Delta\widehat{A}_n(t)\mathcal{G}'\left(\int_0^t e^{-M} d\widehat{A}_n(u)\right)\right)\mathbb{1}(C_i \geq t)dN_i(t)\right] \\
&\qquad\qquad -\frac{1}{n}\sum_{i=1}^n \mathcal{G}\left(\int_0^{\tau\wedge X_i} e^{-M} d\widehat{A}_n(t)\right) + O_p(1)
\end{aligned}$$

$$\begin{aligned}
&\overset{ii)}{\leq} \left(\int_0^\tau \mathbb{P}_n dN(t)\right)\log\left(\int_0^\tau e^{-M}\mathcal{G}'\big(e^{-M}\widehat{A}_n(t)\big)\,d\widehat{A}_n(t)\right) - \frac{1}{n}\sum_{i:X_i\geq\tau}\mathcal{G}\big(e^{-M}\widehat{A}_n(\tau)\big) + O_p(1) \\
&\overset{iii)}{\leq} \mathbb{P}_n\mathbb{1}\{x \leq \tau, \Delta\varepsilon = 1\}\log\mathcal{G}\big(e^{-M}\widehat{A}_n(\tau)\big) - \mathbb{P}_n\mathbb{1}\{x \geq \tau\}\,\mathcal{G}\big(e^{-M}\widehat{A}_n(\tau)\big) + O_p(1) \\
&= \left[\mathbb{P}_n\mathbb{1}\{x \leq \tau, \Delta\varepsilon = 1\}\,\frac{\log\mathcal{G}\big(e^{-M}\widehat{A}_n(\tau)\big)}{\mathcal{G}\big(e^{-M}\widehat{A}_n(\tau)\big)} - \mathbb{P}_n\mathbb{1}\{x \geq \tau\}\right] \times \mathcal{G}\big(e^{-M}\widehat{A}_n(\tau)\big) + O_p(1)
\end{aligned}$$

From $\log x/x \to 0$ for $x \to \infty$ we have that

$$\mathbb{P}_n\mathbb{1}\{x \leq \tau, \Delta\varepsilon = 1\} \cdot \log\mathcal{G}\big(e^{-M}\widehat{A}_n(\tau)\big)/\mathcal{G}\big(e^{-M}\widehat{A}_n(\tau)\big)$$

will be arbitrary small for n sufficiently large, while, at the same time, $\mathbb{P}_n \mathbb{1}\{x \geq \tau\}$ will be arbitrary close to $P(X \geq \tau) > 0$. Therefore, if $\widehat{A}_n(\tau)$ gets infinitely large, the right hand side will go to $-\infty$, which contradicts the definition of $(\widehat{A}_n, \widehat{\beta})$ as a maximum likelihood estimator.

In this calculation we have applied in $i)$ the boundedness of $\widehat{\beta}_n^T Z$ according to model conditions M1) and M2). In $ii)$ we apply Jensen's inequality to obtain

$$\frac{1}{n}\sum_{i=1}^{n}\left[\int_0^\tau \log\left(e^{-M} n\Delta\widehat{A}_n(t)\mathcal{G}'\left(\int_0^t e^{-M} d\widehat{A}_n(t)\right)\right)\mathbb{1}(C_i \geq t) dN_i(t)\right]$$

$$\begin{aligned} &< \left(\int_0^\tau \mathbb{P}_n dN(t)\right)\log\left[\frac{\int_0^\tau e^{-M}\mathcal{G}'\big(e^{-M}\widehat{A}_n(t)\big)\, d\widehat{A}_n(t)}{\int_0^\tau \mathbb{P}_n dN(t)}\right] \\ &= \left(\int_0^\tau \mathbb{P}_n dN(t)\right)\log\left(\int_0^\tau e^{-M}\mathcal{G}'\big(e^{-M}\widehat{A}_n(t)\big)\, d\widehat{A}_n(t)\right) + O_p(1). \end{aligned}$$

$iii)$ is a consequence of the Stieltjes integration rules as stated in our supplemental technical report.

B.2.3 Existence of the NPMLE under model condition M4b)

According to model condition M4b) we have $P(X \geq \tau) \geq \delta > 0$ for a real number δ. We consider $n^{-1}\big[\ell(\widehat{A}_n, \widehat{\beta}_n) - \ell(\widetilde{A}_n, \beta_0)\big]$

$$\begin{aligned} &\leq \frac{1}{n}\sum_{i=1}^{n}\left[\int_0^\tau \log\left(e^{\hat{\beta}_n^T Z_i(t)} n\Delta\widehat{A}_n(t)\mathcal{G}'\left(\int_0^t e^{\hat{\beta}_n^T Z_i(u)} d\widehat{A}_n(u)\right)\right)\mathbb{1}(C_i \geq t) dN_i(t)\right. \\ &\qquad\qquad - \frac{1}{n}\sum_{i=1}^{n}\mathcal{G}\left(\int_0^{\tau\wedge X_i} e^{\hat{\beta}_n^T Z_i(t)} d\widehat{A}_n(t)\right) + O_p(1) \\ &\overset{i)}{\leq} \frac{1}{n}\sum_{i=1}^{n}\left[\int_0^\tau \log\left(e^{M} n\Delta\widehat{A}_n(t)\mathcal{G}'\left(\int_0^t e^{M} d\widehat{A}_n(u)\right)\right)\mathbb{1}(C_i \geq t) dN_i(t)\right. \\ &\qquad\qquad - \frac{1}{n}\sum_{i=1}^{n}\mathcal{G}\big(e^{-M}\widehat{A}_n(X_i\wedge\tau)\big) + O_p(1) \end{aligned}$$

$$
\begin{aligned}
&\overset{ii)}{\leq} \quad \left(\int_0^\tau \mathbb{P}_n dN(t)\right) \log\left(\int_0^\tau e^M \mathcal{G}'\big(e^M \widehat{A}_n(t)\big)\, d\widehat{A}_n(t)\right) \\
&\qquad\qquad\qquad\qquad -\frac{1}{n}\sum_{i:X_i\geq\tau} \mathcal{G}\big(e^{-M}\widehat{A}_n(\tau)\big) + O_p(1) \\
&\leq \quad \mathbb{P}_n \mathbb{1}\{X \leq \tau, \Delta\varepsilon = 1\} \log\left(\tau\, e^M \widehat{A}_n(\tau)\, \mathcal{G}'\big(e^M \widehat{A}_n(\tau)\big)\right) \\
&\qquad\qquad\qquad\qquad -\mathbb{P}_n \mathbb{1}\{X \geq \tau\} \mathcal{G}\big(e^{-M}\widehat{A}_n(\tau)\big) + O_p(1) \\
&= \quad \left\{\mathbb{P}_n \mathbb{1}\{X \leq \tau, \Delta\varepsilon = 1\} \left[\frac{\log \mathcal{G}'\big(e^M \widehat{A}_n(\tau)\big)}{\mathcal{G}\big(e^{-M}\widehat{A}_n(\tau)\big)} + \frac{\log\big(e^M \widehat{A}_n(\tau)\big)}{\mathcal{G}\big(e^{-M}\widehat{A}_n(\tau)\big)}\right] - \mathbb{P}_n \mathbb{1}\{X \geq \tau\}\right\} \\
&\qquad\qquad\qquad\qquad\qquad\qquad \times \mathcal{G}\big(e^{-M}\widehat{A}_n(\tau)\big) + O_p(1).
\end{aligned}
$$

The arguments for $i)$ and $ii)$ are the same as in Section B.2.2. For n sufficiently large, $\mathbb{P}_n \mathbb{1}\{X \geq \tau\}$ will be arbitrary close to $P(X \geq \tau) > \delta$. Therefore, if the sequence $\widehat{A}_n(\tau)$ diverged, $n^{-1}\big[\ell(\widehat{A}_n, \widehat{\beta}_n) - \ell(\widetilde{A}_n, \beta_0)\big]$ would go to $-\infty$, which contradicts the definition of $(\widehat{A}_n, \widehat{\beta}_n)$ as a maximum likelihood estimator. $\square$

For model condition M4b) it is sufficient to verify that $\log(ax_n)/\mathcal{G}(a^{-1}x_n) \to 0$ and $\log \mathcal{G}'(ax_n)/\mathcal{G}(a^{-1}x_n) \to 0$ for $x_n \to \infty$. Condition M4b) is satisfied for example for the Box-Cox transformation models with $\rho > 0$. This can easily be checked: For x_n sufficiently large we have

$$
\frac{\log(ax_n)}{\mathcal{G}(a^{-1}x_n)} = \rho \cdot \frac{\log(ax_n)}{\{(1 + a^{-1}x_n)^\rho - 1\}} \leq \rho \cdot \frac{\log(ax_n)}{(a^{-1}x_n)^\rho} = \rho\, a^\rho \cdot \frac{\log(x_n)}{(x_n)^\rho} + o(1) = o(1),
$$

and with l'Hospital's rule we obtain:

$$
\begin{aligned}
\lim_{n\to\infty} \frac{\log \mathcal{G}'(ax_n)}{\mathcal{G}(a^{-1}x_n)} &= \lim_{n\to\infty} \rho\,(\rho - 1) \cdot \frac{\log\,(1 + ax_n)}{(1 + a^{-1}x_n)^\rho - 1} \\
&= \lim_{n\to\infty} (\rho - 1) \cdot \frac{a}{(1 + ax_n)(1 + a^{-1}x_n)^{\rho-1}} = 0.
\end{aligned}
$$

B.3 Consistency

Setting with administrative censoring. The log-likelihood function presented as (4) in Section 3.2, has a second representation,

$$\begin{aligned}
\ell^a(\beta, A_0) &= \sum_{i=1}^n \Big[\int_0^\tau \log\left(e^{\beta^T Z_i(t)}\alpha_0(t)\mathcal{G}'\Big(\int_0^t e^{\beta^T Z_i(u)} dA_0(u)\Big)\right) \mathbb{1}(C_i \geq t) Y_i(t) dN_i(t)\Big] \\
&\quad - \sum_{i=1}^n \mathcal{G}\left(\int_0^{X_i} e^{\beta^T Z_i(t)} dA_0(t)\right) \\
&\quad - \sum_{i:\Delta_i\varepsilon_i=2}\left[\int_{X_i}^\tau \mathbb{1}(C_i \geq t) e^{\beta^T Z_i(t)} \mathcal{G}'\left(\int_0^t Y_i(u) e^{\beta^T Z_i(u)} dA_0(u)\right) dA_0(t)\right]. \qquad (6)
\end{aligned}$$

Maximizing the discretized log-likelihood function with respect to jump sizes it is obtained that $\left[n\widehat{A}_n^0\{X_j\}\right]^{-1} = \Phi_n(X_j, \widehat{A}_n^0, \widehat{\beta}_n)$ with

$$\begin{aligned}
\Phi_n^a(s, \widehat{A}_n^0, \widehat{\beta}_n) &= n^{-1}\sum_{k=1}^n \mathcal{G}'\left(\int_0^\tau \mathbb{1}(C_i \geq t) Y_i(t) e^{\hat{\beta}_n^T Z_i(t)} d\widehat{A}_n^0(t)\right) e^{\hat{\beta}_n^T Z_i(s)} Y_i^a(s) \\
&\qquad - n^{-1}\sum_{k=1}^n \int_s^\tau \frac{Y_k^a(s) e^{\widehat{\beta}_n^T Z_k(s)} \mathcal{G}''\big(\int_0^t e^{\widehat{\beta}_n^T Z_k(u)} d\widehat{A}_n^0(u)\big)}{\mathcal{G}'\big(\int_0^t e^{\widehat{\beta}_n^T Z_k(u)} d\widehat{A}_n^0(u)\big)} \mathbb{1}(C_k \geq t) dN_k(t) \\
&= \frac{1}{n}\sum_{k=1}^n \mathbb{1}\big((X_k \wedge \tau) \geq X_j\big) e^{\widehat{\beta}_n^T Z_k(X_j)} \mathcal{G}'\left(\int_0^{X_k} e^{\widehat{\beta}_n^T Z_k(t)} d\widehat{A}_n^0(t)\right) \\
&\quad - \frac{1}{n}\sum_{k=1}^n \mathbb{1}(X_k \geq X_j) e^{\widehat{\beta}_n^T Z_k(X_j)} \int_{X_j}^\tau \frac{\mathcal{G}''\big(\int_0^t e^{\widehat{\beta}_n^T Z_j(u)} d\widehat{A}_n^0(u)\big)}{\mathcal{G}'\big(\int_0^t e^{\widehat{\beta}_n^T Z_j(u)} d\widehat{A}_n^0(u)\big)} \mathbb{1}(C_k \geq t) dN_k(t) \\
&\quad + \frac{1}{n}\sum_{k:\Delta_k\varepsilon_k=2} \mathbb{1}(X_k \leq X_j)\mathbb{1}(C_k \geq X_j) e^{\widehat{\beta}_n^T Z_k(X_j)} \mathcal{G}'\left(\int_0^{X_j} e^{\widehat{\beta}_n^T Z_k(t)} d\widehat{A}_n^0(t)\right) \\
&\quad + \frac{1}{n}\sum_{k:\Delta_k\varepsilon_k=2} \mathbb{1}(X_k \leq X_j) e^{\widehat{\beta}_n^T Z_k(X_j)} \int_{X_k}^\tau \mathbb{1}(C_k \geq t) e^{\widehat{\beta}_n^T Z_k(t)} \mathcal{G}''\left(\int_0^t e^{\widehat{\beta}_n^T Z_k(u)} d\widehat{A}_n^0(u)\right) d\widehat{A}_n^0(t),
\end{aligned}$$

We define the Kullback-Leibler distance

$$\begin{aligned}
\mathcal{K}_a(\theta_0,\theta_*) &= E_{F_{\theta_0},G}\left[\log\left\{\left(\frac{f_{\theta_0}(X)}{f_{\theta_*}(X)}\right)^{\Delta\mathbb{1}(\varepsilon=1)}\left(\frac{S_{\theta_0}(C\wedge\tau)}{S_{\theta_*}(C\wedge\tau)}\right)^{1-\Delta}\left(\frac{S_{\theta_0}(C\wedge\tau)}{S_{\theta_*}(C\wedge\tau)}\right)^{\Delta\mathbb{1}(\varepsilon\neq1)}\right\}\right] \\
&= \int E_{\theta_0}\left[\log\left\{\left(\frac{f_{\theta_0}(X)}{f_{\theta_*}(X)}\right)^{\Delta\mathbb{1}(\varepsilon=1)}\left(\frac{S_{\theta_0}(c\wedge\tau)}{S_{\theta_*}(c\wedge\tau)}\right)^{1-\Delta\mathbb{1}(\varepsilon=1)}\right\}\right]dG(c). \qquad (7)
\end{aligned}$$

and it is immediately obtained that

$$\begin{aligned}
-\mathcal{K}_a(\theta_0,\theta_*) &\le \int\log\left[\int\Delta\mathbb{1}(\varepsilon=1)\left(\frac{f_{\theta_*}(x)}{f_{\theta_0}(x)}\right)\big(dP(x,\varepsilon=1)+dP(x,\varepsilon=2)\big)\right. \\
&\qquad\left.+\int\big(1-\Delta\mathbb{1}(\varepsilon=1)\big)\left(\frac{S_{\theta_*}(c)}{S_{\theta_0}(c)}\right)\big(dP(x,\varepsilon=1)+dP(x,\varepsilon=2)\big)\right]dG(c) \\
&= \int\log\left[\int\mathbb{1}(T\le c)\left(\frac{f_{\theta_*}(x)}{f_{\theta_0}(x)}\right)f_{\theta_0}(x)\,dx\right. \\
&\qquad\left.+\int\big(1-\Delta\mathbb{1}(\varepsilon=1)\big)\big(dP(x,\varepsilon=1)+dP(x,\varepsilon=2)\big)\left(\frac{S_{\theta_*}(c)}{S_{\theta_0}(c)}\right)\right]dG(c) \\
&= \int\log\left[\int_0^c f_{\theta_*}(x)dx+\left\{1-\int_0^c f_{\theta_0}(x)dx\right\}\left(\frac{S_{\theta_*}(c)}{S_{\theta_0}(c)}\right)\right]dG(c) \\
&= \int\log\left[\big(1-S_{\theta_*}(c)\big)+\left\{1-\big(1-S_{\theta_0}(c)\big)\right\}\left(\frac{S_{\theta_*}(c)}{S_{\theta_0}(c)}\right)\right]dG(c) \\
&= \int\log\left[\big(1-S_{\theta_*}(c)\big)+S_{\theta_*}(c)\right]dG(c)=0,
\end{aligned}$$

which means that the Kullback-Leibler information is nonnegative. With the Helly-Bray Lemma we obtain that for every sequence $\widehat{A}_n$ there is a subsequence $\widehat{A}_{n_k}$ and $A_*\in\Theta$ with $\widehat{A}_{n_k}\to A_*$. Together with model condition M1) we obtain sequential compactness for Θ, which means that for every sequence $\widehat{\theta}_n$ there is a subsequence $\widehat{\theta}_{n_k}$ and $\theta_*\in\Theta$ with $\widehat{\theta}_{n_k}\to\theta_*$. From the strong law of large numbers (SLLN) and because $\widehat{\theta}_n=(\widehat{\beta}_n,\widehat{A}^0_n)$ maximizes the likelihood function, we obtain

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n f_{\widehat{\theta}_{n_k}}(X_i)S_{\widehat{\theta}_{n_k}}(C_i\wedge\tau)^{1-\Delta\mathbb{1}(\varepsilon_i=1)}=E_{\theta_0}\left[f_{\theta_*}(X)S_{\theta_*}(C\wedge\tau)^{1-\Delta\mathbb{1}(\varepsilon=1)}\right]$$

$$\geq E_{\theta_0}\big[f_{\theta_0}(X)S_{\theta_0}(C\wedge\tau)^{1-\Delta\mathbb{1}(\varepsilon=1)}\big] = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\big[f_{\theta_0}(X)S_{\theta_0}(C\wedge\tau)^{1-\Delta\mathbb{1}(\varepsilon=1)}\big].$$

This means that the Kullback-Leibler distance is nonpositive and takes the value zero if $f_{\theta_*}(X)S_{\theta_*}(C\wedge\tau)^{1-\Delta\mathbb{1}(\varepsilon=1)} = f_{\theta_0}(X)S_{\theta_0}(C\wedge\tau)^{1-\Delta\mathbb{1}(\varepsilon=1)}$. We finally ascertain that the Kullback-Leibler information takes the value zero if and only if the limit of the maximum likelihood estimator is the true parameter. By Jensen's inequality

$$E_{F_{\theta_0},G}\left[\log\left\{\left(\frac{f_{\theta_0}(X)}{f_{\theta_*}(X)}\right)^{\Delta\mathbb{1}(\varepsilon=1)}\left(\frac{S_{\theta_0}(C\wedge\tau)}{S_{\theta_*}(C\wedge\tau)}\right)^{1-\Delta\mathbb{1}(\varepsilon=1)}\right\}\right]\geq 0$$
$$\int\int\log\left\{\left(\frac{f_{\theta_0}(X)}{f_{\theta_*}(X)}\right)^{\Delta\mathbb{1}(\varepsilon=1)}\left(\frac{S_{\theta_0}(c\wedge\tau)}{S_{\theta_*}(c\wedge\tau)}\right)^{1-\Delta\mathbb{1}(\varepsilon=1)}\right\}dP^{\theta_0}(x)\,G_c(c).$$

Therefore, $\mathcal{K}_a(\theta_0,\theta_*)$ is equal to zero if and only if the above term takes the value zero for $P^{G_c}-$almost all $C\in[0,\tau]$. Applying integration by parts and other manipulations for the inner integral of (7) it is obtained that for every $c\in[0,\tau]$

$$\begin{aligned}
&\int\log\left\{\left(\frac{f_{\theta_0}(X)}{f_{\theta_*}(X)}\right)^{\Delta\mathbb{1}(\varepsilon=1)}\left(\frac{S_{\theta_0}(c\wedge\tau)}{S_{\theta_*}(c\wedge\tau)}\right)^{1-\Delta\mathbb{1}(\varepsilon=1)}\right\}dP^{\theta_0}(x)\\
&\quad= \int\Delta\mathbb{1}(\varepsilon=1)\,\log\left(\frac{f_{\theta_0}(x)}{f_{\theta_*}(x)}\right)dP^{\theta_0}(x)\\
&\qquad\qquad+\int\big(1-\Delta\mathbb{1}(\varepsilon=1)\big)dP^{\theta_0}(x)\cdot\log\left(\frac{S_{\theta_0}(c\wedge\tau))}{S_{\theta_*}(c\wedge\tau)}\right)\\
&\quad= \int_0^{c\wedge\tau}\log\left(\frac{\alpha_{\theta_0}(x)\exp\big(-A_{\theta_0}(x)\big)}{\alpha_{\theta_*}(x)\exp\big(-A_{\theta_*}(x)}\right)f^1_{\theta_0}(x)dx\\
&\qquad\qquad+\left(\int dP^{\theta_0}(x)-\int_0^{c\wedge\tau}f_{\theta_0}(x)dx\right)\cdot\log\left(\frac{S_{\theta_0}(c\wedge\tau)}{S_{\theta_*}(c\wedge\tau)}\right)\\
&\quad= \int_0^{c\wedge\tau}\big(\log\alpha_{\theta_0}(x)-\log\alpha_{\theta_*}(x)\big)f_{\theta_0}(x)dx+\int_0^{C\wedge\tau}\big(A_{\theta_*}(x)-A_{\theta_0}(x)\big)f_{\theta_0}(x)dx\\
&\qquad\qquad+\big(A_{\theta_*}(c\wedge\tau)-A_{\theta_0}(c\wedge\tau)\big)S_{\theta_0}(c\wedge\tau)
\end{aligned}$$

$$
\begin{aligned}
= & \int_0^{c\wedge\tau} \big(\log\alpha_{\theta_0}(x) - \log\alpha_{\theta_*}(x)\big)\alpha_{\theta_0}(x)S_{\theta_0}(x)dx \\
& \qquad + \big(A_{\theta_*}(c\wedge\tau) - A_{\theta_0}(c\wedge\tau)\big)\big(1 - S_{\theta_0}(c\wedge\tau)\big) \\
& \qquad\qquad - \int_0^{c\wedge\tau} \big(\alpha_{\theta_*}(x) - \alpha_{\theta_0}(x)\big)\big(1 - S_{\theta_0}(x)\big)dx \\
& \qquad + \big(A_{\theta_*}(c\wedge\tau) - A_{\theta_0}(c\wedge\tau)\big)S_{\theta_0}(c\wedge\tau) \\
= & \int_0^{C\wedge\tau} \big(\log\alpha_{\theta_0}(x) - \log\alpha_{\theta_*}(x)\big)\alpha_{\theta_0}(x)S_{\theta_0}(x)dx \\
& \qquad\qquad + \int_0^{C\wedge\tau} \big(\alpha_{\theta_*}(x) - \alpha_{\theta_0}(x)\big)S_{\theta_0}(x)dx \\
= & \int_0^{C\wedge\tau} \left\{\frac{\alpha_{\theta_*}(x)}{\alpha_{\theta_0}(x)} - 1 - \log\left(\frac{\alpha_{\theta_*}(x)}{\alpha_{\theta_0}(x)}\right)\right\}\alpha_{\theta_0}(x)S_{\theta_0}(x)dx.
\end{aligned}
$$

As $(y-1) \geq \log y \;\; \forall y \in \mathbb{R}^+$, this means that $\mathcal{K}_a(\theta_0, \theta_*) = 0$

$$
\begin{aligned}
\Leftrightarrow \log\left(\frac{\alpha_{\theta_*}(x)}{\alpha_{\theta_0}(x)}\right) &= \frac{\alpha_{\theta_*}(x)}{\alpha_{\theta_0}(x)} - 1 \quad [P^{\theta_0}] \;\; \forall x \leq c\wedge\tau \\
\Leftrightarrow \quad \alpha_{\theta_*}(x) &= \alpha_{\theta_0}(x) \qquad\qquad [P^{\theta_0}] \;\; \forall x \leq c\wedge\tau.
\end{aligned}
$$

From the monotonicity of $\mathcal{G}$ and from model condition $M2$) we obtain that

$$
\begin{aligned}
\alpha_{\theta_*}(t)e^{\beta_*^T Z(t)}\mathcal{G}'\left(e^{\beta_*^T Z(s)}dA_{\theta_*}(s)\right) &= \alpha_{\theta_0}(t)e^{\beta_0^T Z(t)}\mathcal{G}'\left(e^{\beta_0^T Z(s)}dA_{\theta_0}(s)\right) \\
\Rightarrow \; \mathcal{G}\left(\int_0^t e^{\beta_*^T Z(s)}dA_{\theta_*}(s)\right) &= \mathcal{G}\left(\int_0^t e^{\beta_0^T Z(s)}dA_0(s)\right)
\end{aligned}
$$

$$
\Rightarrow \int_0^t e^{\beta_*^T Z(s)}dA_{\theta_*}(s) = \int_0^t e^{\beta_0^T Z(s)}dA_0(s) \quad \forall t \in [0,\tau]
$$

and therefore $\beta_* = \beta_0$ and $A_{\theta_*} = A_{\theta_0}$. From this we conclude that $\widehat{\theta}_{n_k} \to \theta_* = \theta_0$.

Setting with independent right censoring. Maximizing the log-likelihood function (2) in Section 3.2 it is obtained that $\left[n\widehat{A}_n^0\{X_j\}\right]^{-1} = \Phi_n(X_j, \widehat{A}_n^0, \widehat{\beta}_n)$, with

$$
\begin{aligned}
\Phi_n(X_j,\widehat{A}_n^0,\widehat{\beta}_n) &= \frac{1}{n}\sum_{k=1}^{n}\mathbb{1}\big((X_k\wedge\tau)\geq X_j\big)\,e^{\widehat{\beta}_n^T Z_k(X_j)}\,\mathcal{G}'\left(\int_0^{X_k} e^{\widehat{\beta}_n^T Z_k(t)}d\widehat{A}_n^0(t)\right)\\
&-\frac{1}{n}\sum_{k=1}^{n}\mathbb{1}(X_k\geq X_j)e^{\widehat{\beta}_n^T Z_k(X_j)}\int_{X_j}^{\tau}\frac{\mathcal{G}''\big(\int_0^t e^{\widehat{\beta}_n^T Z_j(u)}d\widehat{A}_n^0(u)\big)}{\mathcal{G}'\big(\int_0^t e^{\widehat{\beta}_n^T Z_j(u)}d\widehat{A}_n^0(u)\big)}\mathbb{1}(C_k\geq t)dN_k(t)\\
&+\frac{1}{n}\sum_{k:\Delta_k\varepsilon_k\neq 1}\mathbb{1}(X_k\leq X_j)w_k^*(X_j)e^{\widehat{\beta}_n^T Z_k(X_j)}\,\mathcal{G}'\left(\int_0^{X_j} e^{\widehat{\beta}_n^T Z_k(t)}d\widehat{A}_n^0(t)\right)\\
&+\frac{1}{n}\sum_{k:\Delta_k\varepsilon_k\neq 1}\mathbb{1}(X_k\leq X_j)e^{\widehat{\beta}_n^T Z_k(X_j)}\int_{X_k}^{\tau}w_k^*(t)e^{\widehat{\beta}_n^T Z_k(t)}\,\mathcal{G}''\left(\int_0^{t} e^{\widehat{\beta}_n^T Z_k(u)}d\widehat{A}_n^0(u)\right)d\widehat{A}_n^0(t).
\end{aligned}
$$

Asymptotic equivalence of $\Phi_n^a(X_j,\widehat{A}_n^0,\widehat{\beta}_n)$ and $\Phi_n(X_j,\widehat{A}_n^0,\widehat{\beta}_n)$. is obtained by

$$
\begin{aligned}
&\big|\Phi_n^a(X_j,\widehat{A}_n^0,\widehat{\beta}_n)-\Phi_n(X_j,\widehat{A}_n^0,\widehat{\beta}_n)\big|\\
&=\Bigg|\frac{1}{n}\sum_{k:\Delta_k\varepsilon_k\neq 1}\mathbb{1}(X_k\leq X_j)w_k^*(X_j)e^{\widehat{\beta}_n^T Z_k(X_j)}\,\mathcal{G}'\left(\int_0^{X_j} e^{\widehat{\beta}_n^T Z_k(t)}d\widehat{A}_n^0(t)\right)\\
&\quad+\frac{1}{n}\sum_{k:\Delta_k\varepsilon_k\neq 1}\mathbb{1}(X_k\leq X_j)e^{\widehat{\beta}_n^T Z_k(X_j)}\int_{X_k}^{\tau}w_k^*(t)e^{\widehat{\beta}_n^T Z_k(t)}\,\mathcal{G}''\left(\int_0^{t} e^{\widehat{\beta}_n^T Z_k(u)}d\widehat{A}_n^0(u)\right)d\widehat{A}_n^0(t)\\
&\quad-\frac{1}{n}\sum_{k:\Delta_k\varepsilon_k\neq 1}\mathbb{1}(X_k\leq X_j)\mathbb{1}(C_k\geq X_j)e^{\widehat{\beta}_n^T Z_k(X_j)}\,\mathcal{G}'\left(\int_0^{X_j} e^{\widehat{\beta}_n^T Z_k(t)}d\widehat{A}_n^0(t)\right)\\
&\quad-\frac{1}{n}\sum_{k:\Delta_k\varepsilon_k\neq 1}\mathbb{1}(X_k\leq X_j)e^{\widehat{\beta}_n^T Z_k(X_j)}\\
&\qquad\qquad\times\int_{X_k}^{\tau}\mathbb{1}(C_k\geq t)e^{\widehat{\beta}_n^T Z_k(t)}\,\mathcal{G}''\left(\int_0^{t} e^{\widehat{\beta}_n^T Z_k(u)}d\widehat{A}_n^0(u)\right)d\widehat{A}_n^0(t)\Bigg|\\
&\leq\Bigg|e^M\sup_{y\leq\tau e^M\widehat{A}_n^0(\tau)}\mathcal{G}'(y)\cdot\frac{1}{n}\sum_{k:\Delta_k\varepsilon_k\neq 1}\mathbb{1}(X_k\leq X_j)\big|w_k^*(X_j)-\mathbb{1}(C_k\geq X_j)\big|\Bigg|\\
&\qquad\qquad+O_p(1)\cdot\sup_{t\in[0,\tau]}\big|w_k^*(t)-\mathbb{1}(C_k\geq t)\big|=o_p(1).
\end{aligned}
$$

Asymptotic equivalence of the Kullback-Leibler distances. Comparing the Kullback-Leibler distance for the competing risk models with administrative censoring and with independent right censoring, equivalence is obtained by

$$
\begin{aligned}
\mathcal{K}(\theta_0,\theta_*) - \mathcal{K}_a(\theta_0,\theta_*) = \int E_{\theta_0}\Big[&\mathbb{1}(T \le c, \varepsilon = 2) \\
&\times \log\left\{\left(\frac{\exp\big(-\int_X^\tau \widetilde{w}(t) dA_{\theta_0}(t)\big)}{\exp\big(-\int_X^\tau \widetilde{w}(t) dA_{\theta_*}(t)} \Big/ \frac{\exp\big(-\int_X^\tau \mathbb{1}(c \ge t) dA_{\theta_0}(t)\big)}{\exp\big(-\int_X^\tau \mathbb{1}(c \ge t) dA_{\theta_*}(t)}\right)\right\}\Big] \, dP^C(c) \\
\le \quad \int_0^\tau \int_0^\tau & E_{\theta_0}\big[\widetilde{w}(t) - \mathbb{1}(c \ge t)\big] d\big(A_{\theta_*} - A_{\theta_0}\big)(t) \, dP^C(c) = o_p(1).
\end{aligned}
$$

Consistency of the parameter estimates can therefore be concluded with the same arguments as for the setting with administrative censoring.

B.4 Weak convergence and variance estimation

Proof of Lemma 3. By the definition of $(\widehat{\theta}, \widehat{w}_n)$

$$
\begin{aligned}
\sqrt{n}\big(\Psi(\widehat{\theta}_n, \widetilde{w}) - \Psi(\theta_0, \widetilde{w})\big) &= \sqrt{n}\big(\Psi(\widehat{\theta}_n, \widetilde{w}) - \Psi_n(\widehat{\theta}_n, \widehat{w}_n)\big) + o_p^*(1) \\
&= \sqrt{n}\big(\Psi(\widehat{\theta}_n, \widehat{w}) - \Psi_n(\widehat{\theta}_n, \widehat{w})\big) + o_p^*(1) \\
&= \sqrt{n}\big(\Psi(\theta_0, \widehat{w}) - \Psi_n(\theta_0, \widehat{w})\big) + o_p^*(1 + \sqrt{n}\|\widehat{\theta}_n - \theta_0\|) \\
&= -\sqrt{n}\big((\Psi_n - \Psi)(\theta_0, \widetilde{w}) + (\Psi_n - \Psi)(\theta_0, \widehat{w}) - (\Psi_n - \Psi)(\theta_0, \widetilde{w})\big) \\
&\qquad + o_p^*(1 + \sqrt{n}\|\widehat{\theta}_n - \theta_0\|). \qquad (8)
\end{aligned}
$$

The rest of the proof is along the lines of Van der Vaart and Wellner (1996). Since the derivative $\dot{\Psi}_0$ is continuously invertible, there is a constant $c > 0$ such that

$$
\|\dot{\Psi}_0^{\widetilde{w}}(\theta - \theta_0)\| \ge c\|\theta - \theta_0\| \quad \forall \theta, \theta_0
$$

and together with the differentiability of Ψ, this yields

$$
\|\Psi(\theta, \widetilde{w}) - \Psi(\theta_0, \widetilde{w})\| \ge c\|\theta - \theta_0\| + o_p(\|\theta - \theta_0\|).
$$

Applying this to (8) we obtain

$$
\sqrt{n}\|\widehat{\theta}_n - \theta_0\|\big(c + o_p(1)\big) \le O_p(1) + o_p^*(1 + \sqrt{n}\|\widehat{\theta}_n - \theta_0\|),
$$

which means that $\widehat{\theta}_n$ is $\sqrt{n}$-consistent for θ_0 with regard to $\|\cdot\|$. Replacing the left side in (8) by

$$\sqrt{n}\,\dot{\Psi}_0^{\widetilde{w}}(\widehat{\theta}_n - \theta_0) + o_p^*(\sqrt{n}\|\widehat{\theta}_n - \theta_0\|),$$

(5) is obtained. Weak convergence of $\sqrt{n}(\widehat{\theta} - \theta_0)$ is then obtained by the contiunuity of $\left[\dot{\Psi}_0^{\widetilde{w}}\right]^{-1}$ and by the continuous mapping theorem. □

In case of iid observations Lemma 3 can be applied with $\Psi_n^w(\theta, h) = \mathbb{P}_n\,\psi(\theta, h, w)$ and with $\Psi^w(\theta, h) = \mathcal{P}\,\psi(\theta, h, w)$ and $\sqrt{n}(\Psi_n - \Psi)(\theta) = \{\mathbb{G}_n\psi(\theta, h, w) : h \in \mathcal{H}\}$ is the empirical process indexed by the class of functions $\{\psi(\theta, h, w) : h \in \mathcal{H}\}$. Then condition $a)$ reduces to

$$\left\|\mathbb{G}_n\big(\psi(\widehat{\theta}_n, h, w) - \psi(\widehat{\theta}_0, h, w)\big)\right\|_{\mathcal{H}} = o_p^*\big(1 + \|\widehat{\theta}_n - \theta_0\|\big). \tag{9}$$

Lemma 2 (Van der Vaart and Wellner (1996), Lemma 3.3.5)**.** *Suppose that the class of functions*

$$\Big\{\psi(\theta, h, w) - \psi(\theta_0, h, w) : \|\theta - \theta_0\| < \delta, h \in \mathcal{H}\Big\}$$

is P-Donsker for some $\delta > 0$ and that

$$\sup_{h\in\mathcal{H}}\mathcal{P}\big\{\psi(\theta, h, w) - \psi(\theta_0, h, w)\big\}^2 \to 0, \quad \theta \to \theta_0.$$

If $\hat{\theta}_n$ converges in outer probability to θ_0, then condition $a)$ is satisified.

To verify condition $a)$ of Lemma 3 it can be argued that for the setting with independent right-censoring,

$$\big\{\psi(\theta, h, \widehat{w}_n) - \psi(\theta_0, h, \widehat{w}_n) : \|\theta - \theta_0\| < \delta, h \in \mathcal{H}\big\}$$

and for the setting with administrative censoring

$$\big\{\psi(\theta, h, \mathbb{1}(L < t \leq C)) - \psi(\theta_0, h, \mathbb{1}(L < t \leq C) : \|\theta - \theta_0\| < \delta, h \in \mathcal{H}\big\}$$

are Donsker classes because its components are Donsker classes. Note that our weights are Donsker classes as their components are indicator functions and product limit estimators. Arguments as in Parner (1998) and Kosorok (2008), presented as Donsker preservation theorems on pp. 172 et seq., are directly applicable. The second condition of Lemma 2 is valid as pointwise convergence can be strengthened to $\mathcal{L}^2$ convergence by dominated convergence.

Score operator for the model with administrative censoring. We define the empirical score operator for the model with administrative censoring by $\Psi_n^a(h,\beta,A) = \Psi_{n1}^a(h_1,\beta,A) + \Psi_{n2}^a(h_2,\beta,A)$ with $h=(h_1,h_2)\in\mathcal{H}$,

$$\begin{aligned}
\Psi_{n1}^a(h_1,\beta,A) \;=\;& h_1^T\,\mathbb{P}_n\left\{\int_0^\tau \left(\xi(t,A,\beta)\int_0^t e^{\beta^T Z(s)}Z(s)dA(s)+Z(t)\right)Y(t)dN^a(t)\right.\\
&\left. -\;\mathcal{G}'\left(\int_0^\tau Y(t)e^{\beta^T Z(t)}dA(t)\right)\int_0^\tau Y(t)e^{\beta^T Z(t)}Z(t)dA(t)\right\}, \qquad (10)\\
\Psi_{n2}^a(h_2,\beta,A) \;=\;& \mathbb{P}_n\left\{\int_0^\tau h_2(t)dN^a(t)+\int_0^\tau \xi(t,A,\beta)\left(\int_0^t Y^a(s)e^{\beta^T Z(s)}h_2(s)dA(s)\right)dN^a(t)\right.\\
&\left. -\;\mathcal{G}'\left(\int_0^\tau Y^a(t)e^{\beta^T Z(u)}dA(u)\right)\int_0^\tau Y^a(t)e^{\beta^T Z(t)}h_2(t)dA(t)\right\}, \qquad (11)
\end{aligned}$$

and

$$\xi(t,\beta,A)=\mathcal{G}''\left(\int_0^t e^{\beta^T Z(u)}dA(u)\right)\Big/\mathcal{G}'\left(\int_0^t e^{\beta^T Z(u)}dA(u)\right).$$

An alternative representation of the score is obtained by taking the Gateaux derivatives of (6), to obtain

$$\begin{aligned}
\Psi_{n1}^a(h_1,\beta,A) \;=\;& h_1^T\,\mathbb{P}_n\left\{\int_0^\tau \xi(t,A,\beta)\int_0^t Z(s)e^{\beta Z(s)}dA(s)dN(t)+\int_0^\tau Z(t)dN(t)\right.\\
&-\mathcal{G}'\left(\int_0^X e^{\beta^T Z(t)}dA_0(t)\right)\int_0^X e^{\beta^T Z(t)}Z(t)dA_0(t)\\
&-\int_X^\tau \mathbb{1}(C\ge t,\Delta\varepsilon=2)e^{\beta^T Z(t)}Z(t)\mathcal{G}'\left(\int_0^t e^{\beta^T Z(u)}dA(u)\right)dA(t)\\
&-\int_X^\tau \mathbb{1}(C\ge t,\Delta\varepsilon=2)e^{\beta^T Z(t)}\left(\int_0^t e^{\beta^T Z(s)}Z(s)dA(s)\right)\\
&\left.\times\mathcal{G}''\left(\int_0^t e^{\beta^T Z(s)}dA(s)\right)dA(t)\right\}. \qquad (12)
\end{aligned}$$

and

$$
\begin{aligned}
\Psi_{n2}^{a}(h_2,\beta,A) = \mathbb{P}_n \Bigg\{ & \int_0^\tau \tilde{q}(t)\mathbb{1}(C\geq t)dN(t) \\
& + \int_0^\tau \xi(t,A,\beta)\left(\int_0^t e^{\beta^T Z(s)}\tilde{q}(s)dA(s)\right)\mathbb{1}(C\geq t)dN(t) \\
& - \int_0^X e^{\beta^T Z(t)}\tilde{q}(t)dA(t)\mathcal{G}'\left(\int_0^X e^{\beta^T Z(t)}dA(t)\right) \\
& - \int_X^\tau \mathbb{1}(C\geq t,\Delta\varepsilon=2)Y(t)e^{\beta^T Z(t)}\tilde{q}(t)\mathcal{G}'\left(\int_0^t e^{\beta^T Z(u)}dA(u)\right)dA(t) \\
& - \int_X^\tau \mathbb{1}(C\geq t,\Delta\varepsilon=2)Y(t)e^{\beta^T Z(t)}\left(\int_0^t e^{\beta^T Z(s)}\tilde{q}(s)dA(s)\right) \\
& \qquad \mathcal{G}''\left(\int_0^t e^{\beta^T Z(u)}dA(u)\right)dA(t)\Bigg\}, \qquad (13)
\end{aligned}
$$

The limiting version $\Psi^a(h,\beta,A)=\Psi_1^a(h_1,\beta,A)+\Psi_2^a(h_2,\beta,A)$ is obtained by substituting the empirical measure with the probability measure $\mathcal{P}$. Defining a path $\big(\beta-\beta_0, A(\cdot)-A_0(\cdot)\big) = \big(\beta_0 g_1, \int_0^{(\cdot)} g_2(s)dA_0(s)\big)$ for $g=(g_1,g_2)\in\mathcal{H}$, we derive $\dot{\Psi}_0^a(h,g)$ by taking the Gateaux derivatives

$$
\begin{aligned}
\dot{\Psi}_{11}(h_1,g_1) &= h_1^T\mathcal{I}_{\beta_0}(\beta-\beta_0) \\
\dot{\Psi}_{12}(h_1,g_2) &= \int h_1^T\mathcal{A}^* d(A-A_0)(t) \\
\dot{\Psi}_{21}(h_2,g_1) &= \mathcal{A}[h_2](\beta-\beta_0) \\
\dot{\Psi}_{22}(h_2,g_2) &= \int \big(p(t)h_2(t)+\mathcal{K}[h_2(t)]\big)d(A-A_0)(t),
\end{aligned}
$$

specifically

$$
\begin{aligned}
\mathcal{I}_{\beta_0} &= E\left[\int_0^\tau \xi(t,\beta_0,A_0)\int_0^t e^{\beta_0^T Z(u)} Z(u)Z(u)^T dA_0(u) Y^a(t) dN^a(t)\right] \\
&\quad + E\left[\int_0^\tau \xi'(t,\beta_0,A_0)\left\{\int_0^t e^{\beta_0^T Z(u)} Z(u) dA_0(u)\right\}^{\otimes 2} Y(t) dN^a(t)\right] \\
&\quad - E\left[\mathcal{G}'\left(\int_0^\tau e^{\beta_0^T Z(t)} dA_0(t)\right)\int_0^\tau Y(t) e^{\beta_0^T Z(t)} Z(t)Z(t)^T dA_0(t)\right] \\
&\quad - E\left[\mathcal{G}''\left(\int_0^\tau e^{\beta_0^T Z(t)} dA_0(t)\right)\left\{\int_0^\tau Y^a(t) e^{\beta_0^T Z(t)} Z(t) dA_0(t)\right\}^{\otimes 2}\right] \\
&= -E\left[\int_0^\tau \mathcal{G}'\left(\int_0^t e^{\beta_0^T Z(u)} dA_0(u)\right) Y^a(t) e^{\beta_0^T Z(t)} Z(t)Z(t)^T dA_0(t)\right] \\
&\quad + E\left[\int_0^\tau \xi'(t,\beta_0,A_0)\left\{\int_0^t e^{\beta_0^T Z(u)} Z(u) dA_0(u)\right\}^{\otimes 2} Y^a(t) dN^a(t)\right] \\
&\quad - E\left[\mathcal{G}''\left(\int_0^\tau Y^a(t) e^{\beta_0^T Z(t)} dA_0(t)\right)\left\{\int_0^\tau Y^a(t) e^{\beta_0^T Z(t)} Z(t) dA_0(t)\right\}^{\otimes 2}\right],
\end{aligned}
$$

$$
\begin{aligned}
\int_0^\tau h_1^T \mathcal{A}^*(t) g(t) dA_0(t) &= E\Bigg[h^T\Bigg\{\int_0^\tau \xi'(t,A,\beta)\left(\int_0^t Z(s) e^{\beta^T Z(s)} dA(s)\right) \\
&\qquad\qquad \times e^{\beta_0^T Z(t)} g_2(t)\alpha_0(t) Y(t) dN^a(t) \\
&\quad + \int_0^\tau \xi(t,A,\beta)\left(\int_0^t e^{\beta^T Z(s)} Z(s) g_2(s) dA(s)\right) dN^a(t)
\end{aligned}
$$

$$
\begin{aligned}
&- \int_0^\tau Y^a(t) e^{\beta^T Z(t)} h_2(t) dA(t)\, \mathcal{G}'\left(\int_0^\tau e^{\beta^T Z(t)} dA(t)\right) \\
&\qquad\qquad \int_0^\tau Y^a(t) e^{\beta^T Z(t)} Z(t) dA(t) \\
&- \mathcal{G}'\left(\int_0^\tau e^{\beta^T Z(t)} dA(t)\right)\int_0^\tau Y^a(t) e^{\beta^T Z(t)} Z(t) h_2(t) dA(t)\Bigg\}\Bigg],
\end{aligned}
$$

$$
\begin{aligned}
A[\int h_2 dA_0] \;=\; & E\left[\int_0^\tau \xi'(t,\beta_0,A_0)\int_0^t Y^a(s)e^{\beta_0^T Z(s)}h_2(s)dA_0(s)\int_0^t Y^a(s)e^{\beta_0^T Z(s)}Z(s)dA_0(s)\right] \\
& + E\left[\int_0^\tau \xi'(t,\beta_0,A_0)\int_0^t Y^a(s)e^{\beta_0^T Z(s)}h_2(s)Z(s)dA_0(s)\right] \\
& - E\left[\int_0^\tau Y^a(t)e^{\beta_0^T Z(t)}h_2(t)dA_0(t)\left(\int_0^\tau Y^a(t)e^{\beta_0^T Z(t)}Z(t)dA_0(t)\right)\widetilde{\xi}(\beta_0,A_0)\right] \\
& - E\left[\left(\int_0^\tau Y^a(s)e^{\beta_0^T Z(s)}h_2(s)Z(s)dA_0(s)\right)\mathcal{G}'\left(\int_0^\tau e^{\beta_0^T Z(t)}dA_0(t)\right)\right], \\
p(t) \;=\; & E\left\{Y^a(t)e^{\beta_0^T Z(t)}\int_t^{\tau\wedge C}\xi(s,\beta_0,A_0)dN(s)\right\} \\
& - E\left\{Y^a(t)e^{\beta_0^T Z(t)}\mathcal{G}'\left(\int_0^\tau e^{\beta_0^T Z(u)}dA_0(u)\right)\right\} \\
= \;& - E\left\{Y^a(t)e^{\beta_0^T Z(t)}\mathcal{G}'\left(\int_0^t e^{\beta_0^T Z(u)}dA_0(u)\right)\right\}, \\
\mathcal{K}[h_2] \;=\; & - E\left\{Y^a(t)e^{\beta_0^T Z(t)}\widetilde{\xi}(\beta_0,A_0)\int_0^\tau Y^a(s)e^{\beta_0^T Z(s)}h_2(s)dA_0(s)\right\} \\
& + E\left\{Y^a(t)e^{\beta_0^T Z(t)}\int_t^{\tau\wedge C}\xi'(s,\beta_0,A_0)\int_0^s Y^a(u)e^{\beta_0^T Z(u)}h_2(u)dA_0(u)dN(s)\right\}
\end{aligned}
$$

with

$$
\widetilde{\xi}(\beta_0,A_0) = \mathcal{G}''\left(\int_0^\tau e^{\beta^T Z(t)}dA_0(t)\right).
$$

From this $\Psi^a(\theta,h)$ is Gateaux differentiable in a neighborhood of θ_0 with continuous derivatives. Continuous Gateaux differentiability can be strengthened to continuous Frechét differentiability with an additional continuity condition such as in proposition 1, page 455 by Bickel et al. (1993). As in Parner (1998) it can be argued that for some $\varepsilon > 0$

$$
\sup\left\{\left\|\frac{\partial}{\partial t}\Psi(\theta_0+th)\right\|_p : \|h\|_p \le 1,\ |t| < \varepsilon\right\} < \infty.
$$

Alternatively Lemma 2 from Kosorok et al (2004) may be applied.

For directions $h = (h_1, h_2)$ and $g = (g_1, g_2)$, with $h_1, g_1 \in \mathbb{R}^d$ and $h_2, g_2 \in \mathcal{D}[0,\tau]$ and a path $\big(\beta - \beta_0, A(\cdot) - A_0(\cdot)\big) = \big(\beta_0 g_1, \int_0^{(\cdot)} g_2(s) dA_0(s)\big)$, defining a ring homomorphism, $\dot\Psi_0(h,g)$ has a representation

$$\big((\beta-\beta_0, A-A_0), h\big) \;\mapsto\; \begin{pmatrix} \dot\Psi_{11} & \dot\Psi_{12} \\ \dot\Psi_{21} & \dot\Psi_{22} \end{pmatrix} \left[\begin{pmatrix} h_1 \\ h_2 \end{pmatrix}, \begin{pmatrix} \beta - \beta_0 \\ A - A_0 \end{pmatrix} \right].$$

Invertibility of $\dot\Psi_0^a$ is first establised for the model with administrative censoring: By Doob decomposition of $N^a(t)$ and integration by parts we obtain a representation of the scores $\Psi_1^a(h_1, \beta, A)$ and $\Psi_2^a(h_2, \beta, A)$ as martingale integrals,

$$\begin{aligned}
\Psi_1^a(h_1,\beta,A) = h_1^T\,\mathcal{P}\bigg\{ \int_0^\tau \bigg[& \mathcal{G}''\left(\int_0^t e^{\beta^T Z(u)} dA_0(u)\right)\left(\int_0^t Z(u) e^{\beta^T Z(u)} dA_0(u)\right) \\
& + Z(t)\mathcal{G}'\left(\int_0^t e^{\beta^T Z(u)} dA_0(u)\right)\bigg] Y^a(t) e^{\beta^T Z(t)} dA_0(t) \\
& + \int_0^\tau \left[\xi(t,A,\beta) \int_0^t Z(u) e^{\beta^T Z(u)} dA_0(u) + Z(t)\right] \mathbb{1}(t \le C) dM(t) \\
& - \mathcal{G}'\left(\int_0^\tau e^{\beta^T Z(t)} dA_0(t)\right) \int_0^\tau Y^a(t) e^{\beta^T Z(t)} Z(t) dA_0(t) \bigg\} \\
= \;\; & h_1^T\,\mathcal{P} \int_0^\tau \left[\xi(t,A_0,\beta) \int_0^t h_1^T Z(u) e^{\beta^T Z(u)} dA_0(u) + h_1^T Z(t)\right] \mathbb{1}(t \le C) dM(t)
\end{aligned}$$

and with a similar calculation

$$\Psi_2^a(h_2,\beta,A) = \mathcal{P} \int_0^\tau \left[\xi(t,A,\beta) \int_0^t Y^a(u) h_2(u) e^{\beta^T Z(u)} dA_0(u) + h_2(u)\right] \mathbb{1}(t \le C) dM(t).$$

As in Parner (1998) and in Kosorok (2008), we apply a decomposition

$$\sigma_1(h) = \begin{pmatrix} \mathcal{I}_{\beta_0} & 0 \\ 0 & p(\cdot) \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}, \qquad \sigma_2 = \dot\Psi_0^a - \sigma_1 = \begin{pmatrix} 0 & \dot\Psi_{12} \\ \dot\Psi_{21} & \mathcal{K} \end{pmatrix}.$$

The following conclusion from the open mapping theorem is then utilized, e.g. van der Vaart and Wellner (1996), van der Vaart (1998) and Parner (1998), Kosorok (2008):

Lemma 3. *Let $\mathbb{B}_1$ and $\mathbb{B}_2$ denote Banach spaces with a continuously invertible isomorphism from one to the other. Let $V : \mathbb{B}_1 \mapsto \mathbb{B}_2$ denote an operator such that the difference between V and some invertible operator $\mathcal{I}$ is nothing more than a compact operator $\mathcal{K}$ then V is onto $\Leftrightarrow$ V is one to one $\Leftrightarrow$ V has an inverse V^{-1}.*

From Banach inverse theorem we obtain that if $\dot{\Psi}_0^a$ is a continuous linear operator with inverse $\left[\dot{\Psi}_0^a\right]^{-1}$, then $\left[\dot{\Psi}_0^a\right]^{-1}$ is continuous. This means that it is sufficient to argue that *i)* σ_1 is invertible, *ii)* σ_2 is compact *iii)* $\dot{\Psi}_0^a$ is one-to-one, linear and continuous.

i) For the invertibility of σ_1 it needs to be ascertained that $\dot{\Psi}_{11}$ and $p(t)$ are continuously invertible. The operator $p(t)$ is invertible because $p(t) \neq 0 \, \forall t \in [0, \tau]$. By the fundamental Fisher information property it is obtained for $h_1 \in \mathbb{R}^d$ that

$$
\begin{aligned}
-\, h_1^T \dot{\Psi}_{11} h_1 = \Psi_1^a(h_1, \beta_0, A_0)^2 & \\
= \quad & E\left\{ \int_0^\tau h_1^T \left[\xi(t, A_0, \beta) \int_0^t Z(u) e^{\beta^T Z(u)} dA_0(u) + Z(t) \right] \mathbb{1}(t \leq C) dM(t) \right\}^2 \\
= \quad & E \int_0^\tau \left\{ h_1^T \left(\frac{\mathcal{G}''\left(\int_0^t e^{\beta_0^T Z(u)} dA_0(u) \right)}{\mathcal{G}'\left(\int_0^t e^{\beta_0^T Z(u)} dA_0(u) \right)} \int_0^t Z(u) e^{\beta^T Z(u)} dA_0(u) + Z(t) \right) \right\}^2 \\
& \qquad \times \mathbb{1}(t \leq C) Y(t) e^{\beta^T Z(t)} G' \left(\int_0^t e^{\beta_0^T Z(u)} dA_0(u) \right) dA_0(t)
\end{aligned}
$$

and we conclude that $-\dot{\Psi}_{11}(h_1, h_1) > 0$ for $h_1 \neq 0$ under model condition $M6$) with h_2 being set to zero. Invertibility is then implied by the positive definiteness of $\dot{\Psi}_{11}$.

ii) $\dot{\Psi}_{21}$ is compact because of its finite dimensional range. $\dot{\Psi}_{12}$ and $\mathcal{K}$ are integral operators with continuous kernel and therefore compact. This is a consequence of the

Arzelà Ascoli theorem.

iii) As a consequence of the fundamental Fisher information property we obtain $-\dot{\Psi}_0^a(h,h) = \left[\Psi_1(h_1,\beta_0,A_0) + \Psi_2(h_2,\beta_0,A_0)\right]^2$ and the martingale representation of the score implies

$$
\begin{aligned}
-\dot{\Psi}_0^a(h,h) &= \left[\Psi_1^a(h_1,\beta_0,A_0) + \Psi_2^a(h_2,\beta_0,A_0)\right]^2 \\
&= E\left\{\int_0^\tau \left[\left(h_1^T Z(t) + h_2(t)\right) + \xi(t,A_0,\beta_0)\int_0^t e^{\beta_0^T Z(u)}\left(h_1^T Z(u) + h_2(u)\right)dA_0(u)\right]\right. \\
&\qquad\qquad \left.\times \mathbb{1}(t \le C)dM(t)\right\}^2 \\
&= E\int_0^\tau \left[\left(h_1^T Z(t) + h_2(t)\right) + \xi(t,A_0,\beta_0)\int_0^t e^{\beta_0^T Z(u)}\left(h_1^T Z(u) + h_2(u)\right)dA_0(u)\right]^2 \\
&\qquad\qquad \times Y^a(t) e^{\beta_0^T Z(t)} G'\left(\int_0^t e^{\beta_0^T Z(u)} dA_0(u)\right) dA_0(t).
\end{aligned}
$$

This means that under model condition M6), $\dot{\Psi}_0^a(h,h) > 0$ for every $h \in \mathcal{H}$, $h \neq 0$ and therefore the operator $\dot{\Psi}_0^a$ is one-to-one. With Banach inverse theorem we conclude that $\dot{\Psi}_0^a$ is continuously invertible under model condition M6).

An alternative way to establish continuous invertibility of $\dot{\Psi}^a$ is to ascertain that Ψ_{11} and $V = \Psi_{22} - \Psi_{21}\Psi_{11}^{-1}\Psi_{12}$ are continuously invertible, as

$$
\begin{pmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{pmatrix}
\begin{pmatrix} \Psi_{11}^{-1} & -\Psi_{11}^{-1}\Psi_{12} \\ 0 & I \end{pmatrix}
=
\begin{pmatrix} I & \Psi_{21}\Psi_{11}^{-1} \\ 0 & V \end{pmatrix},
$$

(Vaart and Wellner, 1996, and van der Vaart, 1998).

Invertibility of $\dot{\Psi}_{22}$. The operator can be decomposed as $\dot{\Psi}_{22} = \mathcal{I} + \mathcal{K}$. The operator $\mathcal{I}[h_2] = ph_2$ is invertible because $p(t) \neq 0$ for all $t \in [0,\tau]$. The integral operator $\mathcal{K} = \mathcal{K}[h_2]$ is compact. By the Fisher information property it is ascertained that

$$-\dot{\Psi}_{22}(h_2,h_2) = \left[\Psi_2^a(h_2,\beta_0,A_0)\right]^2$$

$$
\begin{aligned}
&= E\left\{\int_0^\tau \left[\xi(t,A,\beta)\int_0^t Y^a(u)h_2(u)e^{\beta^T Z(u)}dA_0(u)+h_2(t)\right]\mathbb{1}(t\le C)dM(t)\right\}^2 \\
&= E\int_0^\tau \left(\frac{\mathcal{G}''\left(\int_0^t e^{\beta_0^T Z(u)}dA_0(u)\right)}{\mathcal{G}'\left(\int_0^t e^{\beta_0^T Z(u)}dA_0(u)\right)}\int_0^t Y^a(u)h_2(u)e^{\beta^T Z(u)}dA_0(u)+h_2(t)\right)^2 \\
&\qquad \times\mathbb{1}(t\le C)Y(t)e^{\beta^T Z(t)}G'\left(\int_0^t e^{\beta_0^T Z(u)}dA_0(u)\right)dA_0(t).
\end{aligned}
$$

From this we obtain that under model condition M6), $-\dot\Psi_{22}(h_2,h_2)>0$ for $h\neq 0$, which implies that the operator is one-to-one.

Invertibility of $V=\dot\Psi_{22}-\dot\Psi_{21}\dot\Psi_{11}^{-1}\dot\Psi_{12}$ For the decomposition $V=\mathcal{I}+\mathcal{K}+\widetilde{\mathcal{K}}$ it can be argued that $\mathcal{I}$ is invertible, $\mathcal{K}$ is compact and $\widetilde{K}$ is compact because it is a composition of a continuous operator $-\dot\Psi_{21}\dot\Psi_{22}$ and a compact operator $\dot\Psi_{21}$. If the efficient information matrix for the finite dimensional parameter is nonsingular, it can be concluded with Lemma 25.92 of van der Vaart (1998) that V is one to one.

Invertibility of $\dot\Psi_0$. Taking the Gateux derivatives of the log-likelihood function represented as (3), we obtain the score operator $\Psi_n(h,\beta,A)=\Psi_{n1}(h_1,\beta,A)+\Psi_{n2}(h_2,\beta,A)$ with $h=(h_1,h_2)\in\mathcal{H}$,

$$
\begin{aligned}
\Psi_{n1}(h_1,\beta,A) &= h_1^T\mathbb{P}_n\left\{\int_0^\tau q(t)\mathbb{1}(C\ge t)dN(t)\right. \\
&\quad +\int_0^\tau \xi(t,A,\beta)\left(\int_0^t e^{\beta^T Z(s)}\tilde q(s)dA(s)\right)\mathbb{1}(C\ge t)dN(t) \\
&\quad -\int_0^X e^{\beta^T Z(t)}\tilde q(t)dA(t)\mathcal{G}'\left(\int_0^{X_i} e^{\beta^T Z(t)}dA(t)\right) \\
&\quad -\int_X^\tau \mathbb{1}(\Delta\varepsilon=2)w^*(t)e^{\beta^T Z(t)}\tilde q(t)\mathcal{G}'\left(\int_0^t e^{\beta^T Z(u)}dA(u)\right)dA(t)
\end{aligned}
$$

$$- \int_X^\tau \mathbb{1}(\Delta\varepsilon = 2) w^*(t) e^{\beta^T Z(t)} \left(\int_0^t e^{\beta^T Z(s)} \tilde{q}(s) dA(s) \right)$$

$$\mathcal{G}'' \left(\int_0^t e^{\beta^T Z(u)} dA(u) \right) dA(t) \Bigg\},$$

$$\begin{aligned}
\Psi_{n2}(h_1, \beta, A) \quad = \quad & \int_0^\tau \xi(t, A, \beta) \int_0^t Z(s) e^{\beta^T Z(s)} dA(s) dN(t) + \int_0^\tau Z(t) dN(t) \\
& - \mathcal{G}' \left(\int_0^X e^{\beta^T Z(t)} dA_0(t) \right) \int_0^X e^{\beta^T Z(t)} Z(t) dA_0(t) \\
& - \int_X^\tau \mathbb{1}(\Delta\varepsilon = 2) w^*(t) e^{\beta^T Z(t)} Z(t) \, \mathcal{G}' \left(\int_0^t e^{\beta^T Z(u)} dA(u) \right) dA(t) \\
& - \int_X^\tau \mathbb{1}(\Delta\varepsilon = 2) w^*(t) e^{\beta^T Z(t)} \left(\int_0^t e^{\beta^T Z(s)} Z(s) dA(s) \right) \\
& \qquad \mathcal{G}'' \left(\int_0^t e^{\beta^T Z(s)} dA(s) \right) dA(t)
\end{aligned}$$

and with

$$\xi(t, \beta, A) = \mathcal{G}'' \left(\int_0^t e^{\beta^T Z(u)} dA(u) \right) \Big/ \mathcal{G}' \left(\int_0^t e^{\beta^T Z(u)} dA(u) \right).$$

We thereby only need to consider those terms that are related to the contribution of the competing risk events. The scores are asymptotically equivalent

$$\begin{aligned}
\Leftrightarrow \qquad & \frac{1}{n} \sum_{i:\, \Delta_i = 1, \varepsilon_i \neq 1} \left\{ \int_X^\tau \mathbb{1}(\Delta\varepsilon = 2) \mathbb{1}(C_i \geq t) e^{\beta^T Z(t)} \tilde{q}(t) \mathcal{G}' \left(\int_0^t e^{\beta^T Z(u)} dA(u) \right) dA(t) \right. \\
& + \int_X^\tau \mathbb{1}(\Delta\varepsilon = 2) \mathbb{1}(C_i \geq t) e^{\beta^T Z(t)} \left(\int_0^t e^{\beta^T Z(s)} \tilde{q}(s) dA(s) \right) \\
& \qquad \left. \mathcal{G}'' \left(\int_0^t e^{\beta^T Z(u)} dA(u) \right) dA(t) \right\}
\end{aligned}$$

$$
\begin{aligned}
\stackrel{!}{=} \; & \frac{1}{n} \sum_{i:\, \Delta_i = 1, \varepsilon_i \neq 1} \Bigg\{ \int_X^\tau \mathbb{1}(\Delta\varepsilon = 2) w^*(t) e^{\beta^T Z(t)} \tilde{q}(t) \mathcal{G}' \left(\int_0^t e^{\beta^T Z(u)} dA(u) \right) dA(t) \\
& + \int_X^\tau \mathbb{1}(\Delta\varepsilon = 2) w^*(t) e^{\beta^T Z(t)} \left(\int_0^t e^{\beta^T Z(s)} \tilde{q}(s) dA(s) \right) \\
& \qquad \mathcal{G}'' \left(\int_0^t e^{\beta^T Z(u)} dA(u) \right) dA(t) \Bigg\},
\end{aligned}
$$

and

$$
\begin{aligned}
& \frac{1}{n} \sum_{i:\, \Delta_i = 1, \varepsilon_i \neq 1} \Bigg\{ \int_X^\tau \mathbb{1}(\Delta\varepsilon = 2) \mathbb{1}(C_i \geq t)\, e^{\beta^T Z(t)} Z(t)\, \mathcal{G}' \left(\int_0^t e^{\beta^T Z(u)} dA(u) \right) dA(t) \\
& + \int_X^\tau \mathbb{1}(\Delta\varepsilon = 2) \mathbb{1}(C_i \geq t)\, e^{\beta^T Z(t)} \left(\int_0^t e^{\beta^T Z(s)} \tilde{q}(s) dA(s) \right) \\
& \qquad \mathcal{G}'' \left(\int_0^t e^{\beta^T Z(u)} dA(u) \right) dA(t) \Bigg\} \\
\stackrel{!}{=} \; & \frac{1}{n} \sum_{i:\, \Delta_i = 1, \varepsilon_i \neq 1} \Bigg\{ \int_X^\tau \mathbb{1}(\Delta\varepsilon = 2) w^*(t) e^{\beta^T Z(t)} Z(t)\, \mathcal{G}' \left(\int_0^t e^{\beta^T Z(u)} dA(u) \right) dA(t) \\
& + \int_X^\tau \mathbb{1}(\Delta\varepsilon = 2) w^*(t) e^{\beta^T Z(t)} \left(\int_0^t e^{\beta^T Z(s)} \tilde{q}(s) dA(s) \right) \\
& \qquad \mathcal{G}'' \left(\int_0^t e^{\beta^T Z(u)} dA(u) \right) dA(t) \Bigg\} + o_p(t).
\end{aligned}
$$

This is easily obtained by the means of the Glivenko-Cantelli theorem. The continuous invertibility for the model with independent right censoring then follows from $\dot{\Psi}_0 = \dot{\Psi}_0^a + o_p(1)$.

B.5 BMT data set

B.5.1 Pp plots for the subdistribution

We apply a modified version of a graphical method for model selection proposed by Cheng Wei and Ying (1997). We define a sequence of transformed event times by $Y(t) = \log \mathcal{G}^{-1}\big(A(t)\big) = \beta^T Z + \log A_0(t)$. The theoretical subdistribution of this

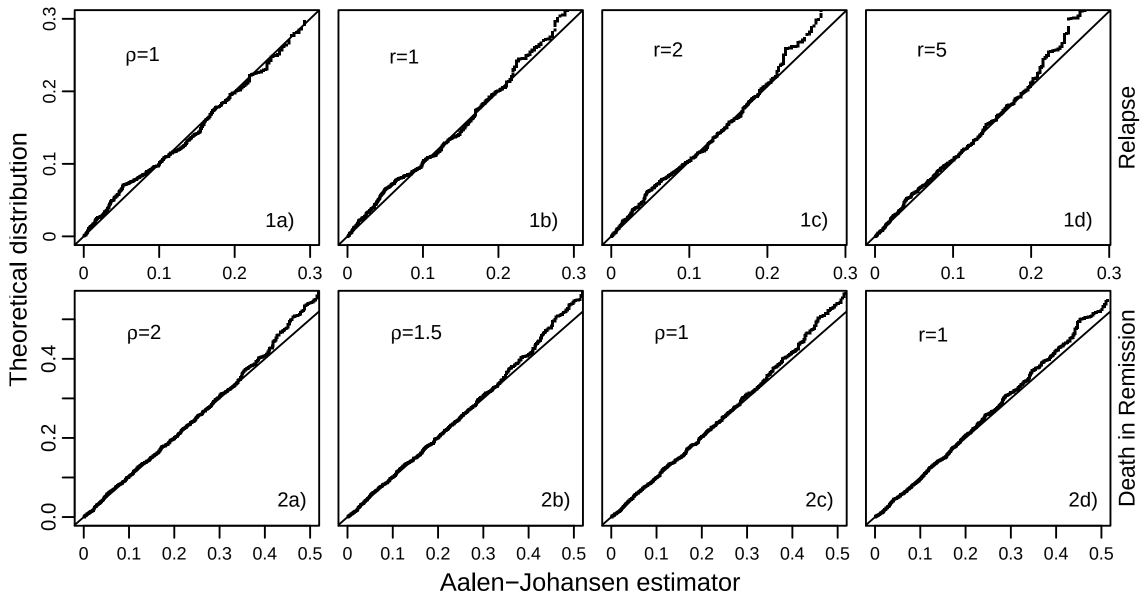

Figure 3: Pp-plots for the bmt data set. First row for outcome relapse: $1a$) Fine-Gray model, $1b$) Proportional odds model, $1c), 1d$) Logarithmic transformation model with $r = 2$ and $r = 5$. Second row for outcome death in remission: $2a, b$) Box-Cox transformation models with $\rho = 2$ and $\rho = 1.5$, $2c$) Fine-Gray model, $2d$) Proportional odds model.

sequence is defined by $P\big(Y(T) \leq t, \varepsilon = j\big) = P\big(F_j(T) \leq 1 - e^{-\mathcal{G}(e^t)}, \varepsilon = j\big)$ with a range on $[0, p_j]$ for the event types $j \in \{1, 2\}$ and $p_j := P(\varepsilon = j)$. Defining $P\big(T \leq F_j^{-1}(1 - e^{-\mathcal{G}(e^t)}), \varepsilon = 1\big) = \big(1 - e^{-\mathcal{G}(e^t)}\big) \wedge p_j$, this can be plotted against the Aalen-Johansen estimator for $(Y_i, \Delta_i, \varepsilon_i)$. The method is slightly sensitive to an increased variability of the Aalen-Johansen estimator for diminishing sample sizes. We therefore displayed the plots for sets corresponding to approximately 95 percent of individuals in figure 3. The pp-plots suggest that all considered models fit reasonably well and that for none of the considered models a significant lack of fit was apparent.

B.5.2 Fine Gray model with time interaction terms

Time interaction terms can be incorporated in the proportional hazards model to accommodate nonproportionality over time and to improve the fit of the model. This is an option in particular if the interpretation of the parameter estimates in terms of hazard ratios is desired. It may be difficult, however, to synthesize the different parameter estimates and corresponding time interaction terms and to reach a conclusion for the overall effect on the hazard function. Transformation models may be preferable from the point of view that the direction of the covariate effects are immediately interpretable.

We fitted a Fine-Gray model with first and second order interaction terms for both endpoints for the bmt dataset. The subdistribution hazard is thereby assumed to take a form $\alpha(t) = \alpha_0(t)\exp(\beta_1 Z + \beta_2 Zt + \beta_3 Zt^2)$. As displayed in table 7 we obtain a completely overfitted model with 27 distorted values for the estimated covariates. Further difficulties in interpretation arise from the fact that the covariates and its corresponding interaction terms have in many cases opposing directions.

After backward selection, only a small number of parameters are associated with time dependent covariates. For most of the covariates the direction of the parameter and the corresponding interaction terms are opposed.

For relapse the significant covariates after backward selection were CML, wtime, sint, sadv and mism. Significant first order interaction terms we obtained from the covariates karn, sint, sadv and match. Significant second order interaction terms were obtained from CML and from karn.

Result from the Fine-Gray model with time interaction terms

Relapse

	$\widehat{\beta}$	$SE(\widehat{\beta})$	$p-$value
AML	-0.162	0.256	0.530
CML	-0.792	0.160	0.008
wtime	-0.008	0.004	0.008
sex	0.227	0.199	0.250
karn	-0.363	0.249	0.150
sint	1.652	0.316	0.000
sadv	2.470	0.296	0.000
match	0.198	0.254	0.440
mism	-0.917	0.717	0.200
AML$*t$	-0.031	0.041	0.450
CML$*t$	-0.036	0.043	0.400
wtime$*t$	0.001	0.000	0.140
sex$*t$	-0.043	0.030	0.160
karn$*t$	0.115	0.047	0.013
sint$*t$	-0.132	0.045	0.004
sadv$*t$	-0.121	0.046	0.008
match$*t$	-0.075	0.044	0.086
mism$*t^2$	-0.051	0.130	0.690
AML$*t^2$	0.001	0.001	0.330
CML$*t^2$	0.001	0.001	0.130
wtime$*t^2$	0.000	0.000	0.250
sex$*t^2$	0.001	0.001	0.420
karn$*t^2$	-0.002	0.001	0.018
sint$*t^2$	0.002	0.001	0.030
sadv$*t^2$	0.001	0.001	0.230
match$*t^2$	0.001	0.001	0.280
mism$*t^2$	0.000	0.002	0.930

Dead in Remission

	$\widehat{\beta}$	$SE(\widehat{\beta})$	$p-$value
AML	-0.254	0.234	0.280
CML	0.045	0.191	0.810
wtime	0.006	0.002	0.004
sex	0.056	0.136	0.680
karn	-0.539	0.171	0.002
sint	0.283	0.164	0.084
sadv	0.474	0.213	0.026
match	0.686	0.150	0.000
mism	0.620	0.238	0.009
AML$*t$	0.053	0.060	0.380
CML$*t$	0.079	0.051	0.120
wtime$*t$	0.000	0.000	0.260
sex$*t$	-0.025	0.033	0.450
karn$*t$	0.030	0.042	0.470
sint$*t$	-0.009	0.038	0.810
sadv$*t$	-0.010	0.052	0.850
match$*t$	-0.007	0.036	0.850
mism$*t$	0.188	0.068	0.006
AML$*t^2$	0.000	0.002	0.930
CML$*t^2$	0.000	0.001	0.000
wtime$*t^2$	0.000	0.000	0.110
sex$*t^2$	0.001	0.001	0.290
karn$*t^2$	0.000	0.001	0.590
sint$*t^2$	0.001	0.001	0.000
sadv$*t^2$	0.000	0.001	0.950
match$*t^2$	0.000	0.001	0.660
mism$*t^2$	-0.008	0.003	0.014

After Backward selection

Relapse

	$\widehat{\beta}$	$SE(\widehat{\beta})$	$p-$value
CML	-0.792	0.160	0.000
wtime	-0.008	0.004	0.020
sint	1.217	0.230	0.000
sadv	2.312	0.210	0.000
mism	-1.280	0.374	0.001
karn$*t$	0.076	0.028	0.006
sint$*t$	-0.039	0.014	0.006
sadv$*t$	-0.083	0.021	0.000
match$*t$	-0.030	0.014	0.031
CML$*t^2$	0.001	0.000	0.010
karn$*t^2$	-0.002	0.001	0.000

Dead in Remission

	$\widehat{\beta}$	$SE(\widehat{\beta})$	$p-$value
AML	-0.306	0.137	0.025
CML	-0.007	0.002	0.000
karn	-0.435	0.107	0.000
sint	0.285	0.108	0.008
sadv	0.429	0.130	0.001
match	0.688	0.098	0.000
mism	0.601	0.229	0.009
AML$*t$	0.057	0.026	0.039
CML$*t$	0.076	0.022	0.001
mism$*t$	0.197	0.064	0.002
mism$*t^2$	-0.008	0.003	0.008

Table 7: Results for the Fine-Gray model with time interaction terms: Estimates for outcome relapse are diplayed on the left hand side, for outcome death in remission on the right hand side. Table at the bottom: Results after backward selection.

B.6 Simulation settings

Model	(β_{11}, β_{12})	a	b	τ	type 1	type 2	cens.
Fine-Gray	$(1, -0.2)$	–	$[\tau]$	4	35	59	6
Fine-Gray	$(1, -0.2)$	1	2	1.9	29	45	26
Fine-Gray	$(1, -0.2)$	0.5	1	0.95	22	31	47
prop.odds	$(1, -0.2)$	–	$[\tau]$	3.5	55	39	6
prop.odds	$(1, -0.2)$	0.75	1.5	1.4	48	26	26
prop.odds	$(1, -0.2)$	0.5	3.5	2.5	52	31	17

Table 8: Simulation studies for the weighted NPMLE for the choice of parameter $(\beta_{11}, \beta_{12}) = (1.0, -0.2)$. Censorings were uniformly distributed on $[a, b]$ with values of a, b as listed above and with $b = [\tau]$ indicating that there was only administrative censoring at τ. Columns 6-8: number of the events of interest (type 1), competing risk events (type 2) and censorings (cens.) in %.

			Bias and Standard errors					Coverage	
Size	$[a,b]$	parameter	Bias	%	SE	SEE_1	SEE_2	Cov_1	Cov_2
50	$[\tau]$	β_{11}	0.076	7.6	0.336	0.288	0.310	0.922	0.950
		β_{12}	−0.013	6.3	0.278	0.257	0.267	0.920	0.943
		$A_n(\tau/4)$	−0.008	4.0	0.075	0.071	0.072	0.893	0.895
		$A_n(\tau/2)$	−0.010	3.4	0.096	0.091	0.092	0.901	0.909
		$A_n(\tau)$	−0.012	3.3	0.106	0.101	0.103	0.906	0.920
200	$[\tau]$	β_{11}	0.016	1.6	0.142	0.136	0.140	0.937	0.949
		β_{12}	−0.006	3.2	0.123	0.121	0.123	0.934	0.936
		$A_n(\tau/4)$	−0.001	0.3	0.035	0.036	0.037	0.947	0.954
		$A_n(\tau/2)$	0.001	0.3	0.046	0.047	0.047	0.943	0.941
		$A_n(\tau)$	0.000	0.1	0.051	0.052	0.052	0.950	0.950
500	$[\tau]$	β_{11}	0.006	0.6	0.087	0.086	0.087	0.941	0.949
		β_{12}	0.000	0.2	0.077	0.076	0.076	0.953	0.952
		$A_n(\tau/4)$	−0.001	0.6	0.023	0.023	0.023	0.944	0.944
		$A_n(\tau/2)$	−0.001	0.4	0.029	0.029	0.029	0.950	0.951
		$A_n(\tau)$	−0.001	0.3	0.032	0.033	0.033	0.953	0.953
50	$[1,2]$	β_{11}	0.072	7.2	0.371	0.314	0.341	0.919	0.953
		β_{12}	−0.003	1.5	0.313	0.283	0.298	0.931	0.952
		$A_n(\tau/4)$	−0.005	4.4	0.053	0.049	0.050	0.864	0.874
		$A_n(\tau/2)$	−0.010	4.8	0.075	0.070	0.072	0.882	0.892
		$A_n(\tau)$	−0.018	6.1	0.107	0.099	0.101	0.869	0.879
200	$[1,2]$	β_{11}	0.022	2.2	0.160	0.149	0.153	0.938	0.939
		β_{12}	−0.010	4.9	0.133	0.132	0.135	0.945	0.947
		$A_n(\tau/4)$	−0.003	2.2	0.026	0.025	0.025	0.922	0.923
		$A_n(\tau/2)$	−0.003	1.3	0.037	0.036	0.037	0.930	0.931
		$A_n(\tau)$	−0.002	0.7	0.054	0.054	0.053	0.941	0.939
500	$[1,2]$	β_{11}	0.005	0.5	0.093	0.093	0.094	0.953	0.958
		β_{12}	−0.004	2.1	0.084	0.083	0.084	0.951	0.956
		$A_n(\tau/4)$	0.000	0.3	0.016	0.016	0.016	0.946	0.947
		$A_n(\tau/2)$	0.000	0.1	0.024	0.023	0.023	0.936	0.939
		$A_n(\tau)$	0.000	0.0	0.034	0.034	0.034	0.948	0.949
50	$[0.5,1]$	β_{11}	0.104	10.4	0.438	0.370	0.406	0.911	0.968
		β_{12}	−0.018	9.2	0.380	0.328	0.353	0.922	0.951
		$A_n(\tau/4)$	−0.007	10.6	0.050	0.036	0.037	0.823	0.838
		$A_n(\tau/2)$	−0.009	7.7	0.054	0.050	0.051	0.854	0.875
		$A_n(\tau)$	−0.018	8.9	0.088	0.079	0.082	0.845	0.850
200	$[0.5,1]$	β_{11}	0.018	1.8	0.184	0.172	0.176	0.928	0.944
		β_{12}	−0.008	3.8	0.162	0.157	0.159	0.949	0.949
		$A_n(\tau/4)$	−0.001	2.0	0.018	0.018	0.018	0.930	0.934
		$A_n(\tau/2)$	−0.003	2.3	0.028	0.027	0.027	0.920	0.922
		$A_n(\tau)$	−0.005	2.4	0.045	0.044	0.044	0.922	0.927
500	$[0.5,1]$	β_{11}	0.007	0.7	0.105	0.107	0.108	0.953	0.960
		β_{12}	−0.006	2.8	0.100	0.098	0.098	0.944	0.950
		$A_n(\tau/4)$	0.000	0.3	0.011	0.011	0.011	0.934	0.935
		$A_n(\tau/2)$	−0.001	0.4	0.017	0.017	0.017	0.949	0.947
		$A_n(\tau)$	−0.001	0.3	0.028	0.029	0.029	0.945	0.941

Table 9: Simulation studies for the Fine-Gray model, parameter of interest $(\beta_{11}, \beta_{12}) = (1, -0.2)$. Censorings uniformly distributed on $[a, b]$, with $[\tau]$ 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. $\text{Cov}_1, \text{Cov}_2$) 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	$[\tau]$	β_{11}	0.026	2.6	0.321	0.308	0.323	0.936	0.956
		β_{12}	−0.008	4.0	0.291	0.275	0.284	0.938	0.951
		$A_n(\tau/4)$	−0.001	0.2	0.265	0.246	0.251	0.906	0.915
		$A_n(\tau/2)$	0.021	1.9	0.387	0.355	0.364	0.922	0.926
		$A_n(\tau)$	0.016	1.2	0.449	0.418	0.429	0.918	0.925
200	$[\tau]$	β_{11}	0.012	1.2	0.156	0.153	0.155	0.946	0.953
		β_{12}	−0.002	1.1	0.137	0.134	0.135	0.942	0.945
		$A_n(\tau/4)$	−0.002	0.2	0.123	0.122	0.122	0.927	0.930
		$A_n(\tau/2)$	0.010	0.9	0.173	0.174	0.175	0.943	0.941
		$A_n(\tau)$	0.009	0.7	0.204	0.205	0.206	0.943	0.945
500	$[\tau]$	β_{11}	0.005	0.5	0.098	0.096	0.097	0.943	0.947
		β_{12}	−0.003	1.3	0.089	0.085	0.085	0.937	0.938
		$A_n(\tau/4)$	−0.001	0.2	0.075	0.077	0.077	0.949	0.949
		$A_n(\tau/2)$	−0.001	0.1	0.106	0.109	0.109	0.954	0.956
		$A_n(\tau)$	−0.002	0.1	0.126	0.128	0.128	0.954	0.955
50	$[0.75, 1.5]$	β_{11}	0.039	3.9	0.353	0.323	0.340	0.923	0.956
		β_{12}	−0.007	3.3	0.295	0.287	0.298	0.935	0.962
		$A_n(\tau/4)$	−0.011	2.7	0.142	0.132	0.134	0.886	0.892
		$A_n(\tau/2)$	−0.004	0.6	0.241	0.215	0.220	0.905	0.906
		$A_n(\tau)$	0.021	2.0	0.460	0.398	0.403	0.873	0.870
200	$[0.75, 1.5]$	β_{11}	0.005	0.5	0.161	0.159	0.161	0.951	0.956
		β_{12}	−0.013	6.5	0.141	0.140	0.141	0.942	0.949
		$A_n(\tau/4)$	−0.005	1.3	0.065	0.066	0.067	0.943	0.943
		$A_n(\tau/2)$	−0.005	0.7	0.108	0.107	0.107	0.935	0.940
		$A_n(\tau)$	0.002	0.2	0.183	0.186	0.186	0.931	0.931
500	$[0.75, 1.5]$	β_{11}	0.007	0.7	0.104	0.100	0.101	0.943	0.947
		β_{12}	0.001	0.5	0.092	0.088	0.088	0.933	0.935
		$A_n(\tau/4)$	−0.002	0.5	0.044	0.042	0.042	0.937	0.938
		$A_n(\tau/2)$	−0.001	0.2	0.068	0.068	0.068	0.942	0.942
		$A_n(\tau)$	0.006	0.6	0.117	0.117	0.117	0.947	0.947
50	$[0.5, 3.5]$	β_{11}	0.021	2.1	0.341	0.316	0.330	0.929	0.952
		β_{12}	−0.012	6.0	0.301	0.276	0.290	0.926	0.949
		$A_n(\tau/4)$	0.013	2.0	0.212	0.203	0.208	0.936	0.944
		$A_n(\tau/2)$	0.038	3.9	0.331	0.323	0.331	0.939	0.939
		$A_n(\tau)$	0.058	4.7	0.490	0.468	0.474	0.921	0.929
200	$[0.5, 3.5]$	β_{11}	0.009	0.9	0.160	0.156	0.158	0.948	0.955
		β_{12}	0.003	1.7	0.141	0.138	0.139	0.949	0.949
		$A_n(\tau/4)$	−0.002	0.4	0.097	0.099	0.100	0.945	0.948
		$A_n(\tau/2)$	−0.002	0.2	0.148	0.154	0.154	0.945	0.944
		$A_n(\tau)$	−0.001	0.1	0.210	0.219	0.216	0.944	0.941
500	$[0.5, 3.5]$	β_{11}	0.008	0.8	0.103	0.099	0.099	0.936	0.942
		β_{12}	−0.004	1.9	0.088	0.087	0.087	0.948	0.949
		$A_n(\tau/4)$	0.000	0.0	0.064	0.063	0.063	0.952	0.953
		$A_n(\tau/2)$	−0.003	0.3	0.096	0.097	0.097	0.945	0.946
		$A_n(\tau)$	−0.002	0.2	0.138	0.137	0.135	0.941	0.936

Table 10: Simulation studies for the proportional odds model, parameter of interest $(\beta_{11}, \beta_{12}) = (1, -0.2)$. Censorings uniformly distributed on $[a, b]$, with $[\tau]$ 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. $\mathrm{Cov}_1, \mathrm{Cov}_2$) coverage probability of 0.95 confidence intervals for the variance estimators.

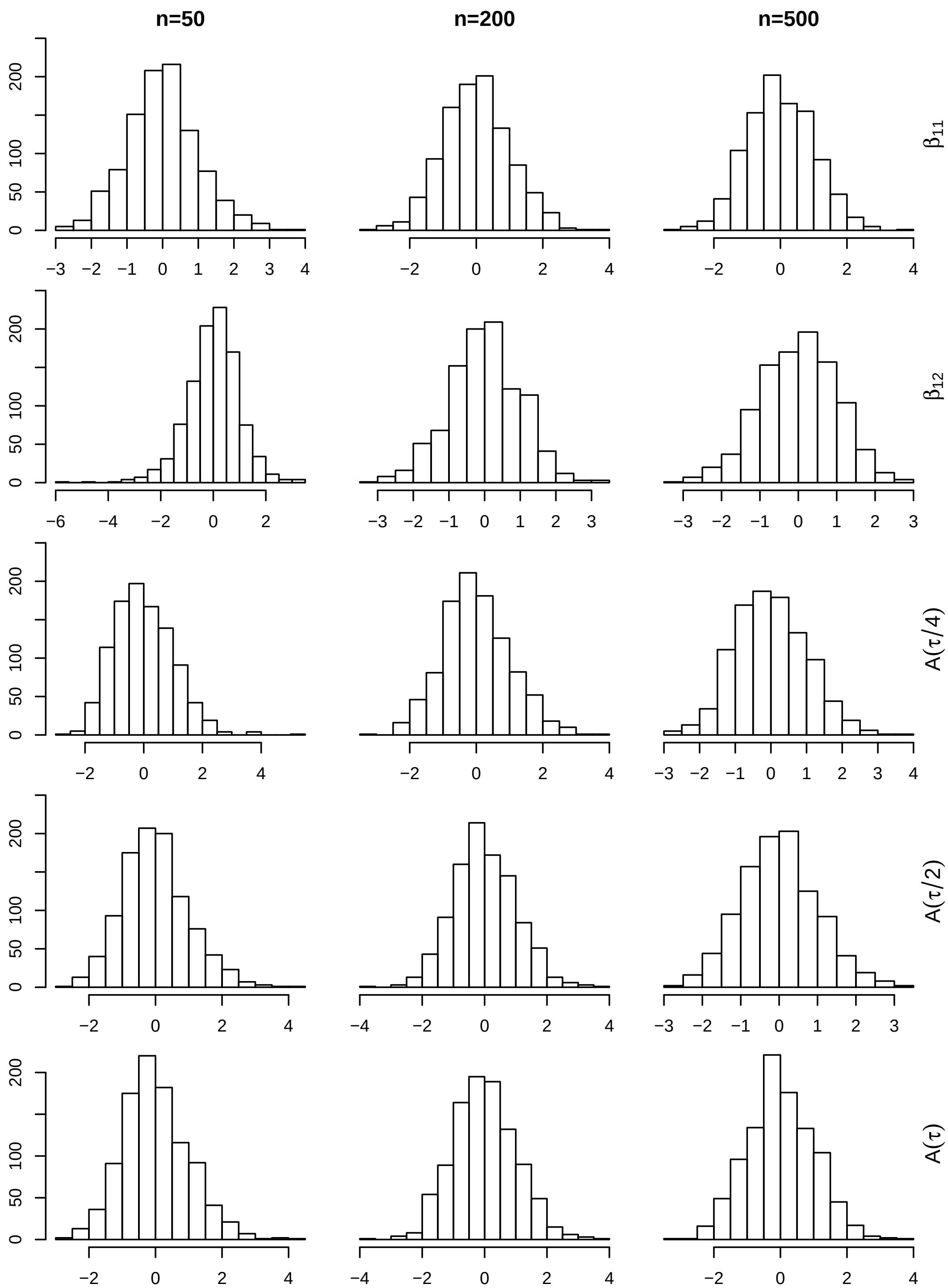

Figure 4: Simulation study for the Fine-Gray model with parameter of interest $(\beta_{11}, \beta_{12}) = (0.5, -0.5)$ and censorings only occuring for individuals at the endpoint of study $\tau = 4$. Histograms for the standardized parameter estimates, rows from top to bottom: 1) β_{11}, 2)β_{12}, 3) $A(\tau/4)$, 4) $A(\tau/2)$, 5) $A(\tau)$. On average 5% censorings and 33% events of interest were generated for this setting.

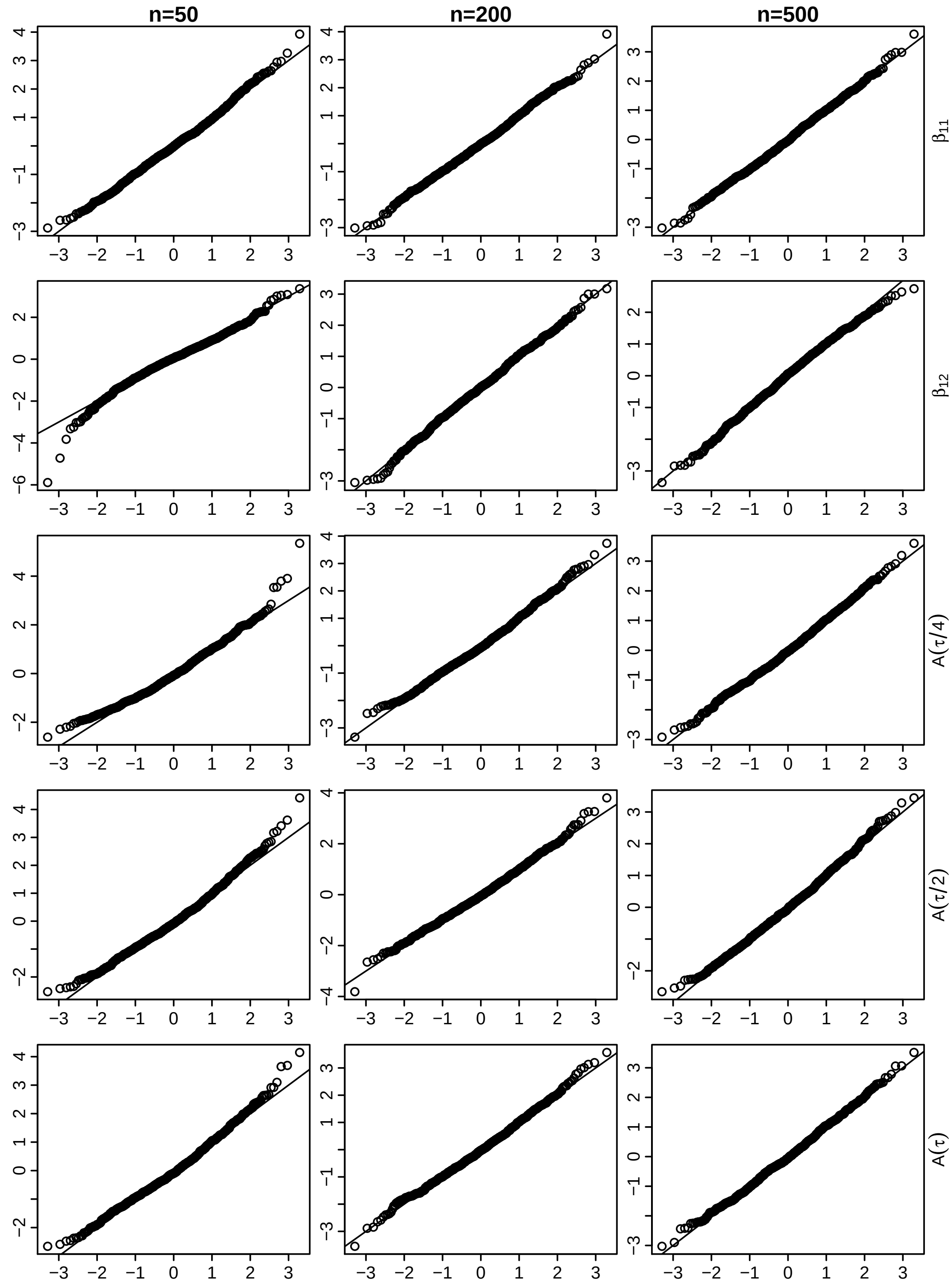

Figure 5: Simulation study for the Fine-Gray model with parameter of interest $(\beta_{11}, \beta_{12}) = (0.5, -0.5)$ and censorings only occuring for individuals at the endpoint of study $\tau = 4$. Qq plots for the standardized parameter estimates, rows from top to bottom: 1) β_{11}, 2)β_{12}, 3) $A(\tau/4)$, 4) $A(\tau/2)$, 5) $A(\tau)$. On average 5% censorings and 33% events of interest were generated for this setting.

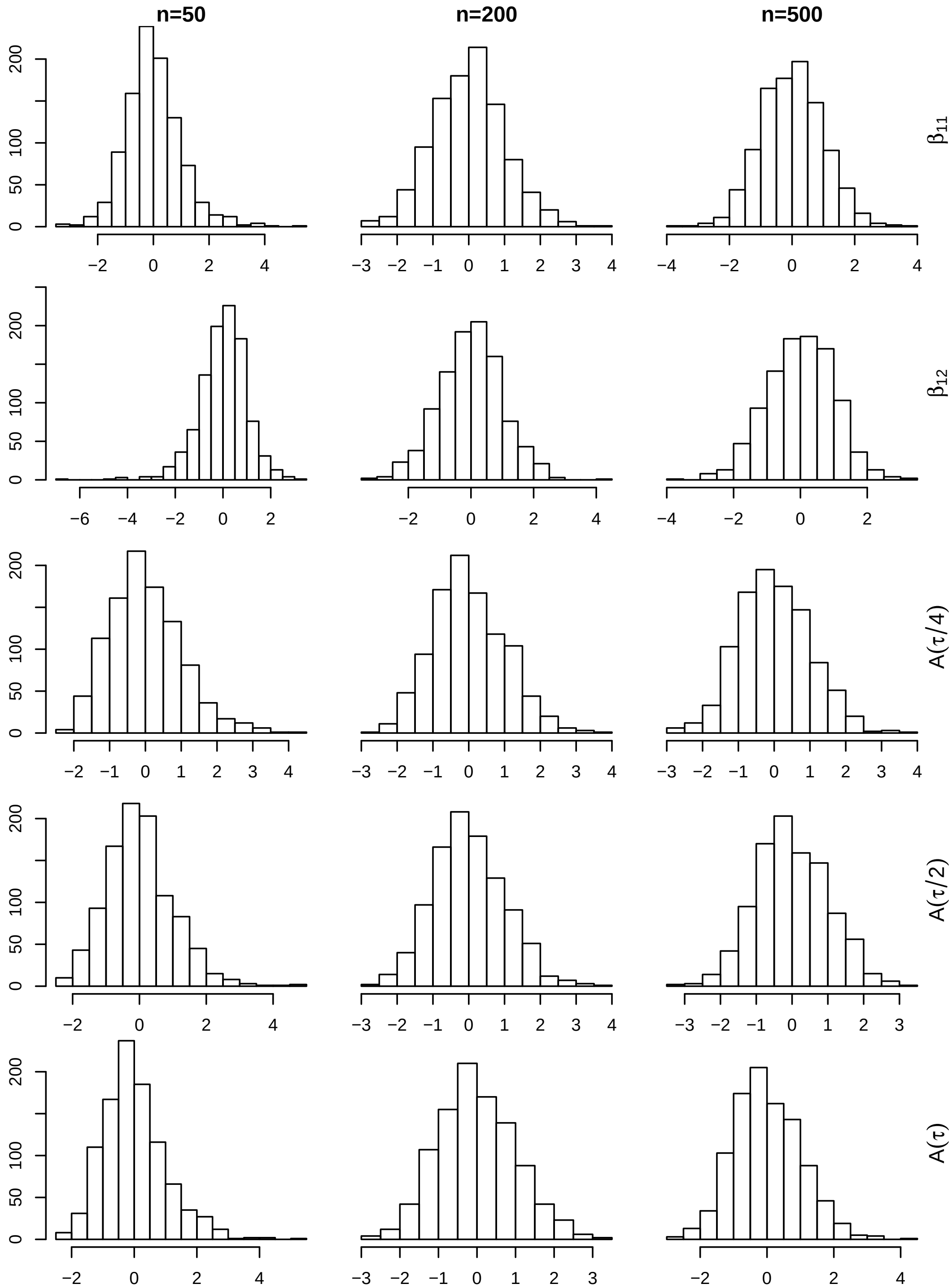

Figure 6: Simulation study for the Fine-Gray model with parameter of interest $(\beta_{11}, \beta_{12}) = (0.5, -0.5)$, censorings uniformly distributed on $[1, 2]$. Histograms for the standardized parameter estimates, rows from top to bottom: 1) β_{11}, 2)β_{12}, 3) $A(\tau/4)$, 4) $A(\tau/2)$, 5) $A(\tau)$. On average 25% censorings and 26% events of interest were generated for this setting.

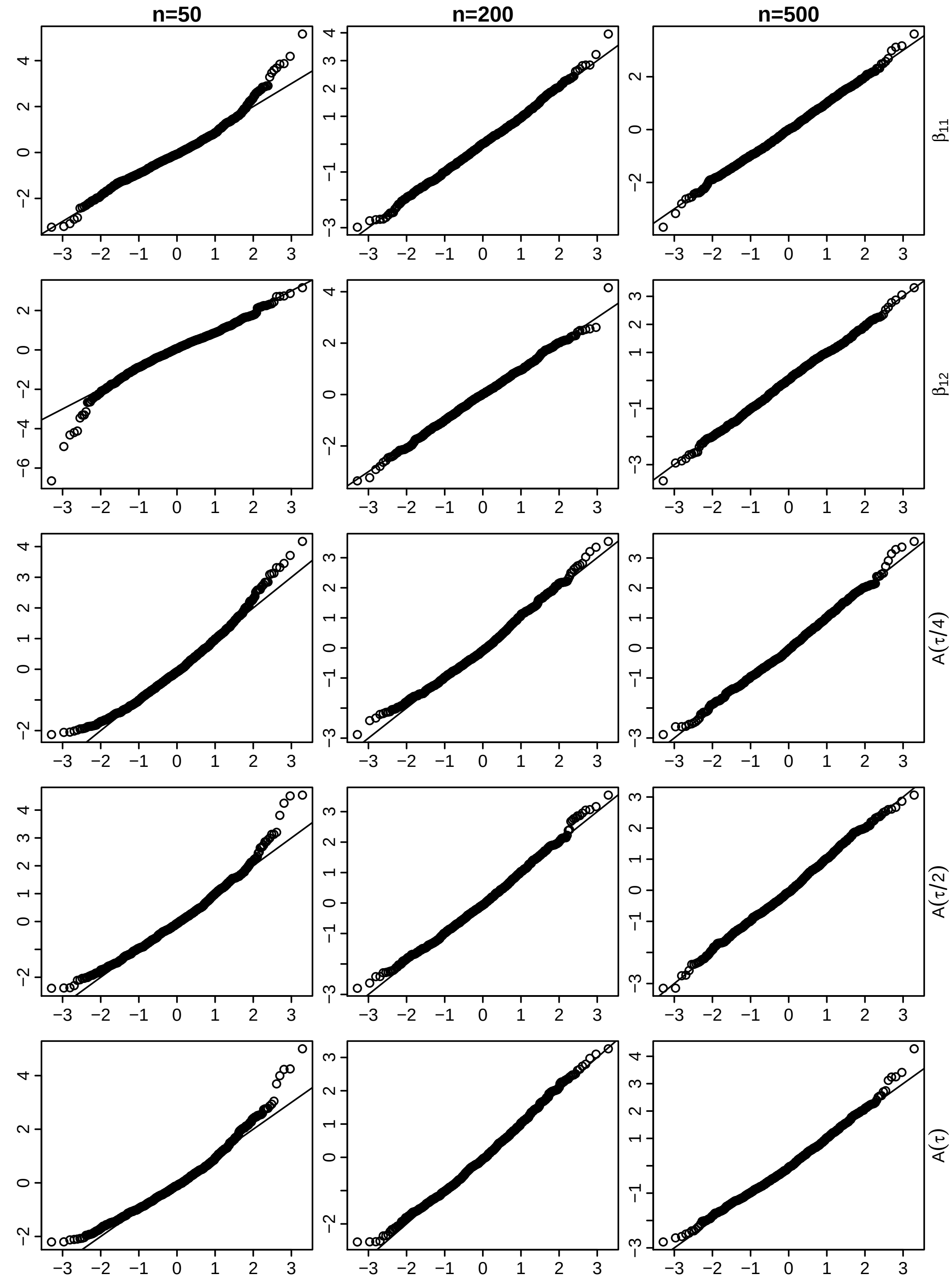

Figure 7: Simultion study for the Fine-Gray model with parameter of interest $(\beta_{11}, \beta_{12}) = (0.5, -0.5)$, censorings uniformly distributed on $[1, 2]$. On average 25% censorings and 26% events of interest were generated for this setting. Qq plots for the standardized parameter estimates, rows from top to bottom: 1) β_{11}, 2)β_{12}, 3) $A(\tau/4)$, 4) $A(\tau/2)$, 5) $A(\tau)$

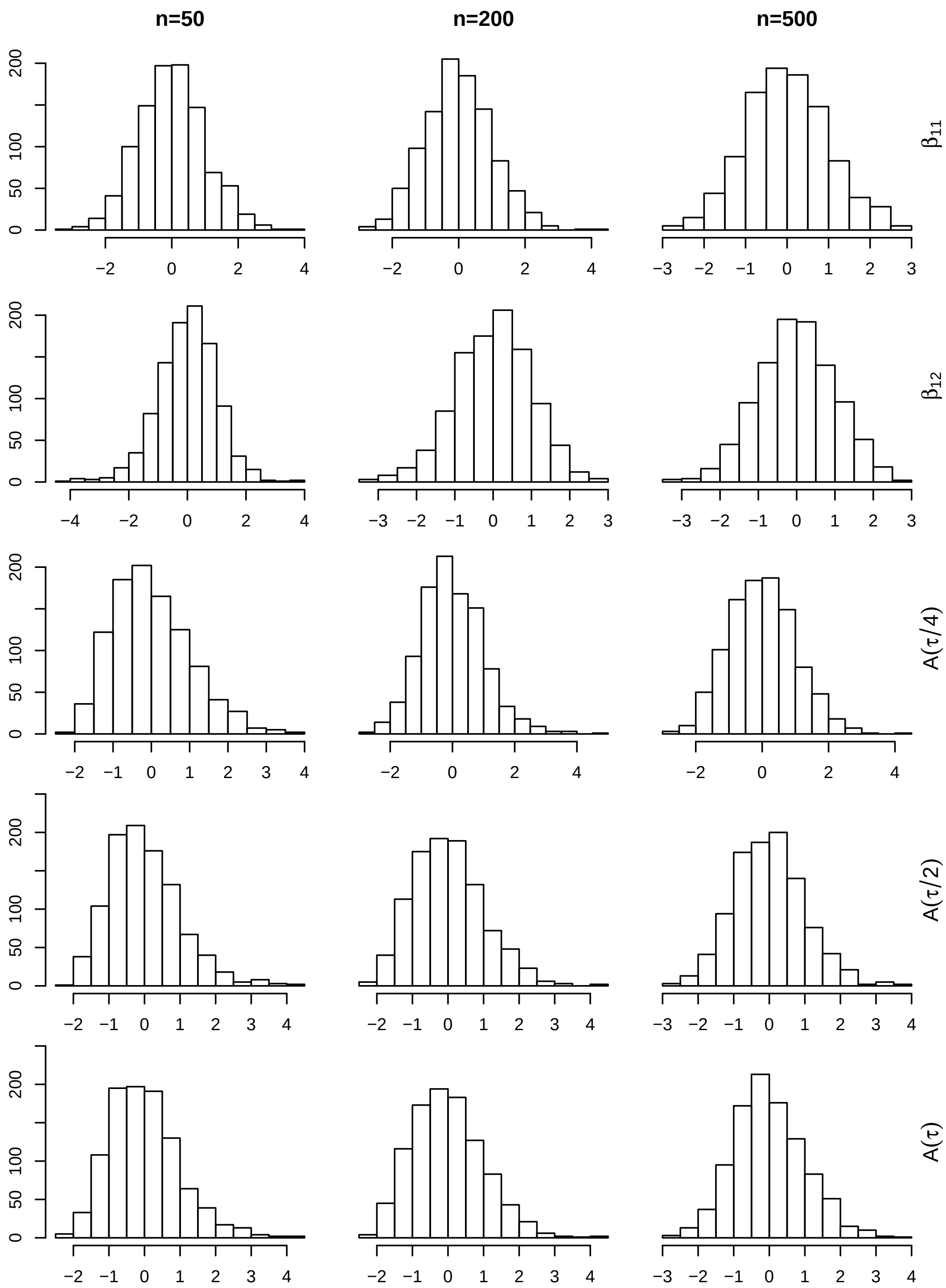

Figure 8: Simulation study for the proportional odds model with parameter of interest $(\beta_{11}, \beta_{12}) = (0.5, -0.5)$ and censorings only occuring for individuals at the endpoint of study $\tau = 3.5$. Histograms for the standardized parameter estimates, rows from top to bottom: 1) β_{11}, 2)β_{12}, 3) $A(\tau/4)$, 4) $A(\tau/2)$, 5) $A(\tau)$. On average 5% of censorings and 56% events of interest were generated for this setting.

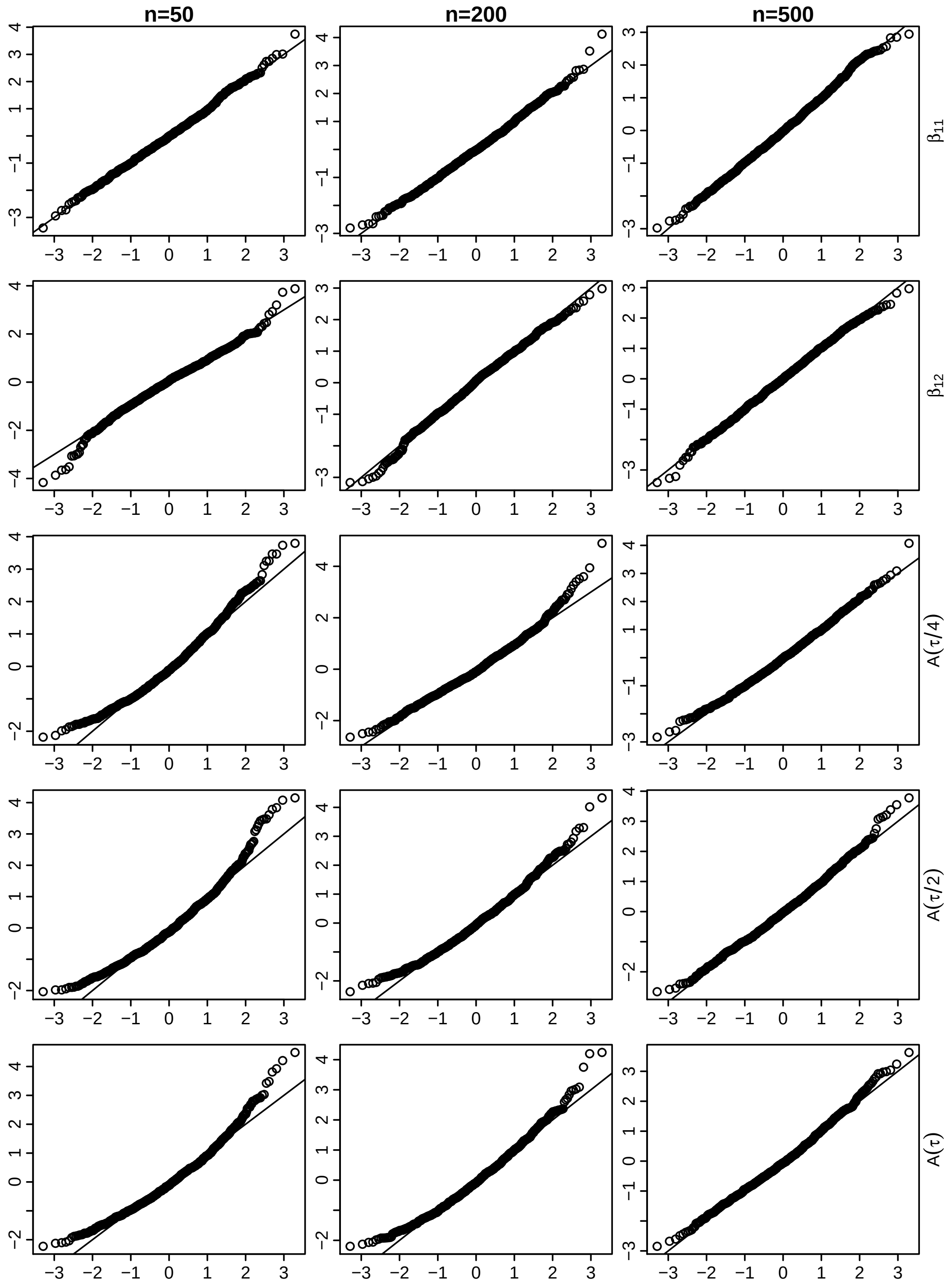

Figure 9: Simulation study for the proportional odds model with parameter of interest $(\beta_{11}, \beta_{12}) = (0.5, -0.5)$ and censorings only occuring for individuals at the endpoint of study $\tau = 3.5$. Qq plots for the standardized parameter estimates, rows from top to bottom: 1) β_{11}, 2)β_{12}, 3) $A(\tau/4)$, 4) $A(\tau/2)$, 5) $A(\tau)$. On average 5% of censorings and 56% events of interest were generated for this setting.

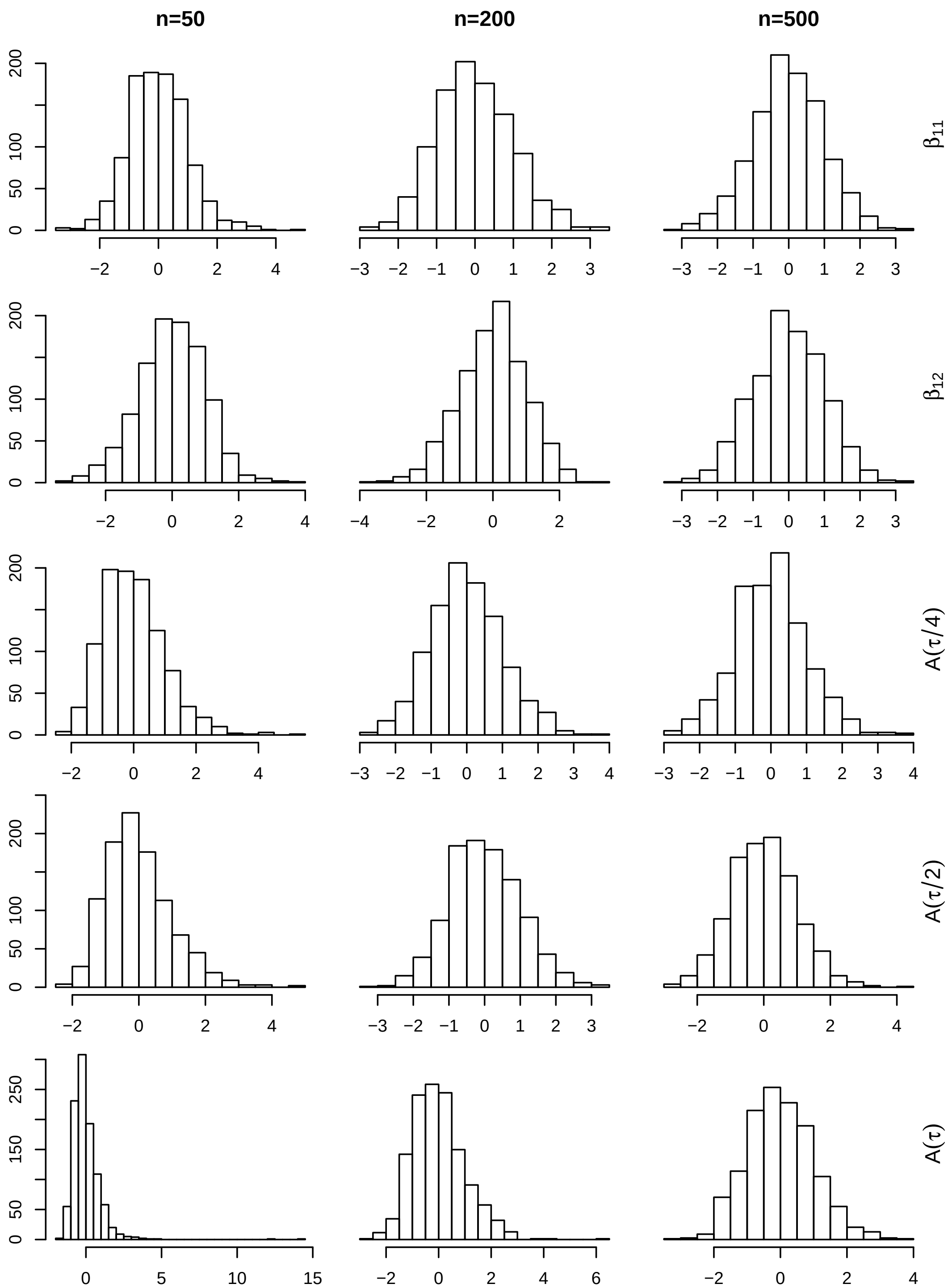

Figure 10: Simulation study for the proportional odds model with parameter of interest $(\beta_{11}, \beta_{12}) = (0.5, -0.5)$, censorings uniformly distributed on $[0.75, 1.5]$. Histograms for the standardized parameter estimates, rows from top to bottom: 1) β_{11}, 2)β_{12}, 3) $A(\tau/4)$, 4) $A(\tau/2)$, 5) $A(\tau)$. On average 24% censorings and 48% events of interest were generated for this setting.

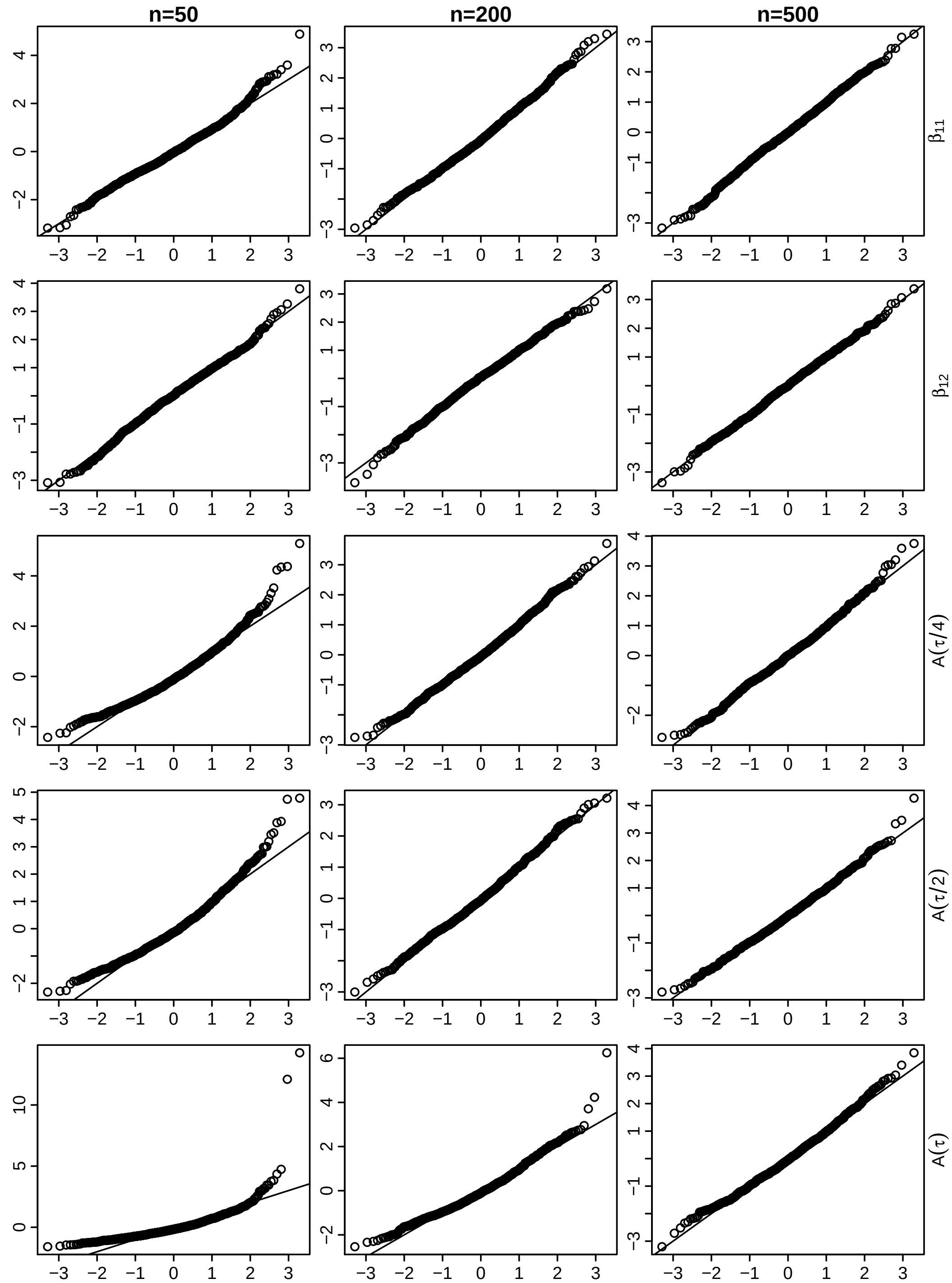

Figure 11: Simulation study for the proportional odds model with parameter of interest $(\beta_{11}, \beta_{12}) = (0.5, -0.5)$, censorings uniformly distributed on $[0.75, 1.5]$. Qq plots for the standardized parameter estimates, rows from top to bottom: 1) β_{11}, 2)β_{12}, 3) $A(\tau/4)$, 4) $A(\tau/2)$, 5) $A(\tau)$. On average 24% censorings and 48% events of interest were generated for this setting.