Supplemental material for "Estimation and inference for semi-competing risks based on data from a nested case-control study"

Ina Jazić

Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA

inj466@mail.harvard.edu

Stephanie Lee

Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA

Sebastien Haneuse

Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA

Introduction

This document is organized as follows:

Appendix A: Additional details on the illness-death model, including score functions for the illness-death model under the Weibull specification of the baseline hazards, as well as for the spline-based specification of the baseline hazards under both the unpenalized and penalized log-likelihoods

Appendix B: Additional details on asymptotics, including a proof of the unbiasedness of the unpenalized score functions, as well as arguments for the asymptotic properties of the penalized log-likelihood.

Appendix C: Detailed specifications of the simulation settings described in the main manuscript

Appendix D: Additional results for the simulation study described in the main manuscript

Appendix E: Descriptive statistics and additional results from the analysis of the CIBMTR stem cell transplantation data

A Details on the illness-death model

A.1 Classes of observations under the illness-death model

Table 1: Y and δ observed for each event (non-terminal and terminal) across the four designated classes of observations in semi-competing risks data

		NT		T	
		Y_1	δ_1	Y_2	δ_2
1)	Both events	T_1	1	T_2	1
2)	Terminal only	T_2	0	T_2	1
3)	Non-terminal only	T_1	1	C	0
4)	Neither event	C	0	C	0

A.2 Derivation of observed data likelihood for illness-death model

The rates of transition among the three states in the illness-death model are dictated by three intensity, or hazard, functions, formally represented below and illustrated in the accompanying figure:

$$
\begin{aligned}
\lambda_1(t_1) &= \lim_{\Delta\to 0} \frac{1}{\Delta} P(T_1 \in [t_1, t_1+\Delta) | T_1 \geq t_1, T_2 \geq t_1), \quad t_1 > 0 \\
\lambda_2(t_2) &= \lim_{\Delta\to 0} \frac{1}{\Delta} P(T_2 \in [t_2, t_2+\Delta) | T_1 \geq t_2, T_2 \geq t_2), \quad t_2 > 0 \\
\lambda_3(t_2|t_1) &= \lim_{\Delta\to 0} \frac{1}{\Delta} P(T_2 \in [t_2, t_2+\Delta) | T_1 = t_1, T_2 \geq t_2), \quad t_2 > t_1
\end{aligned}
$$

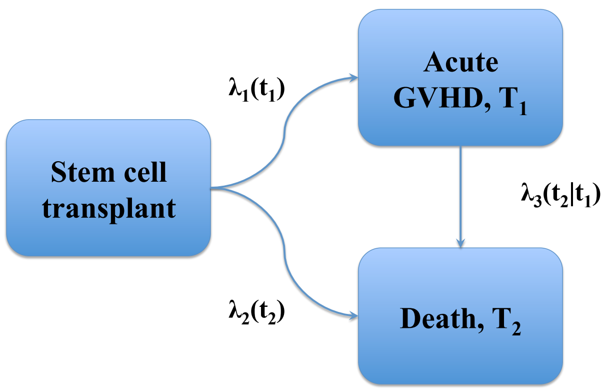


Figure 1: Flowchart representation of the illness-death model for acute graft-versus-host disease (aGVHD) following hematopoietic stem cell transplantation.

Note that in this document, we adopt the semi-Markov assumption for λ_3 (where the relevant timescale is the sojourn time) – however, this observed data likelihood can be written out for other assumptions for λ_3. A Markov specification would replace $t_2 - t_1$ with simply t_2 in the expression for λ_3 below. The joint density of the two survival times is

$$f(t_1, t_2) = \exp\left[-\int_0^{t_1} \{\lambda_1(s) + \lambda_2(s)\} ds\right] \lambda_1(t_1) \times \exp\left\{-\int_0^{t_2 - t_1} \lambda_3(s) ds\right\} \lambda_3(t_2 - t_1)$$

Define the hazards:

$$\begin{aligned}
\lambda_1(t_1) &= \gamma\lambda_1^*(t_1) = \gamma\lambda_{01}(t_1)\exp(\boldsymbol{\beta}_1'\mathbf{X}_1) \\
\lambda_2(t_2) &= \gamma\lambda_2^*(t_2) = \gamma\lambda_{02}(t_2)\exp(\boldsymbol{\beta}_2'\mathbf{X}_2) \\
\lambda_3(t_2 - t_1) &= \gamma\lambda_3^*(t_2 - t_1) = \gamma\lambda_{03}(t_2 - t_1)\exp(\boldsymbol{\beta}_3'\mathbf{X}_3)
\end{aligned}$$

Then, we can write out the likelihood contribution for each class of observations in Table A.1 based on the joint density:

1. $f_1(y_1, y_2|\gamma) = \exp[-\gamma\{\Lambda_1^*(y_1) + \Lambda_2^*(y_1) + \Lambda_3^*(y_2 - y_1)\}]\{\gamma\lambda_1^*(y_1)\}\{\gamma\lambda_3^*(y_2 - y_1)\}$
2. $f_2(y_1, y_2|\gamma) = \exp[-\gamma\{\Lambda_1^*(y_1) + \Lambda_2^*(y_1)\}]\gamma\lambda_2^*(y_1)$
3. $f_3(y_1, y_2|\gamma) = \exp[-\gamma\{\Lambda_1^*(y_1) + \Lambda_2^*(y_1) + \Lambda_3^*(y_2 - y_1)\}]\gamma\lambda_1^*(y_1)$
4. $f_4(y_1, y_2|\gamma) = \exp[-\gamma\{\Lambda_1^*(y_1) + \Lambda_2^*(y_1)\}]$

Note that $y_1 = y_2$ for classes 2 and 4 (where the patient does not experience the non-terminal event), which allows the likelihood contribution to be written in terms of y_1. Next, we integrate the frailty γ out of each of these likelihood contributions.
Let $A = \Lambda_1^*(y_1) + \Lambda_2^*(y_1) + \Lambda_3^*(y_2 - y_1)$ and let $A_{eq} = \Lambda_1^*(y_1) + \Lambda_2^*(y_1)$.

1. (Both events)

$$\begin{aligned}
f_1(y_1, y_2) &= \lambda_1^*(y_1)\lambda_3^*(y_2 - y_1)\int_0^\infty \gamma^2 \exp(-\gamma A)\frac{1}{\Gamma(1/\theta)}\gamma^{\frac{1}{\theta}-1}\left(\frac{1}{\theta}\right)^{\frac{1}{\theta}}\exp\left(-\frac{\gamma}{\theta}\right)d\gamma \\
&= \lambda_1^*(y_1)\lambda_3^*(y_2 - y_1)\left(\frac{\frac{1}{\theta}}{\frac{1}{\theta} + A}\right)^{\frac{1}{\theta}}\int_0^\infty \gamma^2\frac{1}{\Gamma(1/\theta)}\gamma^{\frac{1}{\theta}-1}\left(\frac{1}{\theta} + A\right)^{\frac{1}{\theta}}\exp\left\{-\gamma\left(\frac{1}{\theta} + A\right)\right\}d\gamma
\end{aligned}$$

The integral gives the second moment of a $Gamma\left(\frac{1}{\theta}, \frac{1}{\theta} + A\right)$ random variable. $E[\gamma^2] = Var[\gamma] + (E[\gamma])^2 = \frac{1/\theta}{(1/\theta + A)^2} + \left(\frac{1/\theta}{1/\theta + A}\right)^2 = (1 + \theta)(1 + \theta A)^{-2}$.

$$\begin{aligned}
&= \lambda_1^*(y_1)\lambda_3^*(y_2 - y_1)(1 + \theta A)^{-\frac{1}{\theta}}(1 + \theta)(1 + \theta A)^{-2} \\
&= \lambda_1^*(y_1)\lambda_3^*(y_2 - y_1)(1 + \theta)[1 + \theta\{\Lambda_1^*(y_1) + \Lambda_2^*(y_1) + \Lambda_3^*(y_2 - y_1)\}]^{-\frac{1}{\theta}-2}
\end{aligned}$$

2. (Terminal only) Recall $y_1 = y_2$ in this case.

$$f_2(y_1, y_2) = \lambda_2^*(y_1) \int_0^\infty \gamma \exp\left(-\gamma A_{eq}\right) \frac{1}{\Gamma(1/\theta)} \gamma^{\frac{1}{\theta}-1} \left(\frac{1}{\theta}\right)^{\frac{1}{\theta}} \exp\left(-\frac{\gamma}{\theta}\right) d\gamma$$
$$= \lambda_2^*(y_1)(1+\theta A_{eq})^{-\frac{1}{\theta}} \int_0^\infty \gamma \frac{1}{\Gamma(1/\theta)} \gamma^{\frac{1}{\theta}-1} \left(\frac{1}{\theta} + A_{eq}\right)^{\frac{1}{\theta}} \exp\left\{-\gamma \left(\frac{1}{\theta} + A_{eq}\right)\right\} d\gamma$$

The integral gives the mean of a $Gamma\left(\frac{1}{\theta}, \frac{1}{\theta} + A_{eq}\right)$ random variable, which is $(1+\theta A_{eq})^{-1}$.

$$= \lambda_2^*(y_1)[1 + \theta\{\Lambda_1^*(y_1) + \Lambda_2^*(y_1) + \Lambda_3^*(y_2 - y_1)\}]^{-\frac{1}{\theta}-1}$$

3. (Non-terminal only)

$$f_3(y_1, y_2) = \lambda_1^*(y_1) \int_0^\infty \gamma \exp\left(-\gamma A\right) \frac{1}{\Gamma(1/\theta)} \gamma^{\frac{1}{\theta}-1} \left(\frac{1}{\theta}\right)^{\frac{1}{\theta}} \exp\left(-\frac{\gamma}{\theta}\right) d\gamma$$
$$= \lambda_1^*(y_1)(1+\theta A)^{-\frac{1}{\theta}} \int_0^\infty \gamma \frac{1}{\Gamma(1/\theta)} \gamma^{\frac{1}{\theta}-1} \left(\frac{1}{\theta} + A\right)^{\frac{1}{\theta}} \exp\left\{-\gamma \left(\frac{1}{\theta} + A\right)\right\} d\gamma$$

The integral gives the mean of a $Gamma\left(\frac{1}{\theta}, \frac{1}{\theta} + A\right)$ random variable, which is $(1+\theta A)^{-1}$.

$$= \lambda_1^*(y_1)[1 + \theta\{\Lambda_1^*(y_1) + \Lambda_2^*(y_1) + \Lambda_3^*(y_2 - y_1)\}]^{-\frac{1}{\theta}-1}$$

4. (Neither event) Recall $y_1 = y_2$ in this case.

$$f_4(y_1, y_2) = \int_0^\infty \exp\left(-\gamma A_{eq}\right) \frac{1}{\Gamma(1/\theta)} \gamma^{\frac{1}{\theta}-1} \left(\frac{1}{\theta}\right)^{\frac{1}{\theta}} \exp\left(-\frac{\gamma}{\theta}\right) d\gamma$$
$$= (1+\theta A_{eq})^{-\frac{1}{\theta}} \int_0^\infty \frac{1}{\Gamma(1/\theta)} \gamma^{\frac{1}{\theta}-1} \left(\frac{1}{\theta} + A_{eq}\right)^{\frac{1}{\theta}} \exp\left\{-\gamma \left(\frac{1}{\theta} + A_{eq}\right)\right\} d\gamma$$

This is the integral of a $Gamma\left(\frac{1}{\theta}, \frac{1}{\theta} + A_{eq}\right)$ pdf on its whole range, which is 1.

$$= [1 + \theta\{\Lambda_1^*(y_1) + \Lambda_2^*(y_1) + \Lambda_3^*(y_2 - y_1)\}]^{-\frac{1}{\theta}}$$

Then, we can write out the full likelihood. Let $A_i = \Lambda_1^*(y_{i1}) + \Lambda_2^*(y_{i1}) + \Lambda_3^*(y_{i2} - y_{i1})$ for individual

i.

$$\mathcal{L}(\boldsymbol{\xi}) = \prod_{i=1}^{n} f_1(y_{i1}, y_{i2})^{\delta_{i1}\delta_{i2}} f_2(y_{i1}, y_{i2})^{(1-\delta_{i1})\delta_{i2}} f_3(y_{i1}, y_{i2})^{\delta_{i1}(1-\delta_{i2})} f_4(y_{i1}, y_{i2})^{(1-\delta_{i1})(1-\delta_{i2})}$$
$$= \prod_{i=1}^{n} \lambda_1^*(y_{i1})^{\delta_{i1}} \lambda_2^*(y_{i2})^{(1-\delta_{i1})\delta_{i2}} \lambda_3^*(y_{i2} - y_{i1})^{\delta_{i1}\delta_{i2}} (1+\theta)^{\delta_{i1}\delta_{i2}} (1+\theta A_i)^{-(\frac{1}{\theta}+\delta_{i1}+\delta_{i2})}$$

A.3 Specification of weighted illness-death model: Weibull

One specification we adopt for the baseline hazards is a Weibull distribution, such that:

$$\lambda_{0g}(t) = \phi_{g1}\phi_{g2}t^{\phi_{g1}-1}; \quad \Lambda_{0g}(t) = \phi_{g2}t^{\phi_{g1}}$$
$$\log \lambda_{0g}(t) = \log \phi_{g1} + \log \phi_{g2} + (\phi_{g1} - 1)\log t$$

Let $h = \log \theta, a_g = \log \phi_{g1}$ and $k_g = \log \phi_{g2}$.

$$A_i = y_{i1}^{\exp a_1} e^{k_1 + \boldsymbol{\beta}_1' \mathbf{X}_{1i}} + y_{i1}^{\exp a_2} e^{k_2 + \boldsymbol{\beta}_2' \mathbf{X}_{2i}} + \{(y_{i2} - y_{i1})^{\exp a_3}\} e^{k_3 + \boldsymbol{\beta}_3' \mathbf{X}_{3i}}$$

The parameters of interest are $\boldsymbol{\xi} = (h, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, k_1, a_1, k_2, a_2, k_3, a_3)$, which we solve for using `nleqslv` in `R` on the basis of the score equations $\mathcal{U}^w(\boldsymbol{\xi}) = \sum_{i=1}^N \mathbf{U}_i^w(\boldsymbol{\xi}) = \sum_{i=1}^N R_i \pi_i^{-1} \mathbf{U}_i(\boldsymbol{\xi})$, with $\mathbf{U}_i(\boldsymbol{\xi}) = \partial \ell_i(\boldsymbol{\xi}) / \partial \boldsymbol{\xi}$ listed below:

$$U_{i,1}(\boldsymbol{\xi}) = \frac{\delta \ell_i(\boldsymbol{\xi})}{\delta h} = e^h \left\{ \frac{\delta_{i1}\delta_{i2}}{1+e^h} + \frac{\log(1+e^h A_i)}{e^{2h}} - \left(e^{-h} + \delta_{i1} + \delta_{i2}\right) \frac{A_i}{1+e^h A_i} \right\}$$
$$\mathbf{U}_{i,2}(\boldsymbol{\xi}) = \frac{\delta \ell_i(\boldsymbol{\xi})}{\delta \boldsymbol{\beta}_1} = \delta_{i1}\mathbf{X}_{1i} - \left(e^{-h} + \delta_{i1} + \delta_{i2}\right) \frac{y_{i1}^{\exp a_1} e^{h+k_1+\boldsymbol{\beta}_1'\mathbf{X}_{1i}} \mathbf{X}_{1i}}{1+e^h A_i}$$
$$\mathbf{U}_{i,3}(\boldsymbol{\xi}) = \frac{\delta \ell_i(\boldsymbol{\xi})}{\delta \boldsymbol{\beta}_2} = (1-\delta_{i1})\delta_{i2}\mathbf{X}_{2i} - \left(e^{-h} + \delta_{i1} + \delta_{i2}\right) \frac{y_{i1}^{\exp a_2} e^{h+k_2+\boldsymbol{\beta}_2'\mathbf{X}_{2i}} \mathbf{X}_{2i}}{1+e^h A_i}$$
$$\mathbf{U}_{i,4}(\boldsymbol{\xi}) = \frac{\delta \ell_i(\boldsymbol{\xi})}{\delta \boldsymbol{\beta}_3} = \delta_{i1}\delta_{i2}\mathbf{X}_{3i} - \left(e^{-h} + \delta_{i1} + \delta_{i2}\right) \frac{\{(y_{i2} - y_{i1})^{\exp a_3}\} e^{h+k_3+\boldsymbol{\beta}_3'\mathbf{X}_{3i}} \mathbf{X}_{3i}}{1+e^h A_i}$$
$$U_{i,5}(\boldsymbol{\xi}) = \frac{\delta \ell_i(\boldsymbol{\xi})}{\delta k_1} = e^{k_1} \left\{ \delta_{i1} e^{-k_1} - \left(e^{-h} + \delta_{i1} + \delta_{i2}\right) \frac{y_{i1}^{\exp a_1} e^{h+\boldsymbol{\beta}_1'\mathbf{X}_{1i}}}{1+e^h A_i} \right\}$$
$$U_{i,6}(\boldsymbol{\xi}) = \frac{\delta \ell_i(\boldsymbol{\xi})}{\delta a_1} = e^{a_1} \left\{ \delta_{i1}\left(e^{-a_1} + \log y_{i1}\right) - \left(e^{-h} + \delta_{i1} + \delta_{i2}\right) \frac{y_{i1}^{\exp a_1} (\log y_{i1}) e^{h+k_1+\boldsymbol{\beta}_1'\mathbf{X}_{1i}}}{1+e^h A_i} \right\}$$
$$U_{i,7}(\boldsymbol{\xi}) = \frac{\delta \ell_i(\boldsymbol{\xi})}{\delta k_2} = e^{k_2} \left\{ (1-\delta_{i1})\delta_{i2} e^{-k_2} - \left(e^{-h} + \delta_{i1} + \delta_{i2}\right) \frac{y_{i1}^{\exp a_2} e^{h+\boldsymbol{\beta}_2'\mathbf{X}_{2i}}}{1+e^h A_i} \right\}$$

$$U_{i,8}(\boldsymbol{\xi}) = \frac{\delta \ell_i(\boldsymbol{\xi})}{\delta a_2} = e^{a_2}\left\{(1-\delta_{i1})\delta_{i2}\left(e^{-a_2}+\log y_{i1}\right) - \left(e^{-h}+\delta_{i1}+\delta_{i2}\right)\frac{y_{i1}^{\exp a_2}(\log y_{i1})e^{h+k_2+\boldsymbol{\beta}_2'\mathbf{X}_{2i}}}{1+e^h A_i}\right\}$$

$$U_{i,9}(\boldsymbol{\xi}) = \frac{\delta \ell_i(\boldsymbol{\xi})}{\delta k_3} = e^{k_3}\left\{\delta_{i1}\delta_{i2}e^{-k_3} - \left(e^{-h}+\delta_{i1}+\delta_{i2}\right)\frac{(y_{i2}-y_{i1})^{\exp a_3}e^{h+\boldsymbol{\beta}_3'\mathbf{X}_{3i}}}{1+e^h A_i}\right\}$$

For $U_{i,10}(\boldsymbol{\xi})$, we adjust the notation slightly. Computing A_i in this case gives errors since $\log(y_{i2} - y_{i1})$ is calculated when $\delta_{i1} = 0$. To fix this, we can break the log-likelihood back into the sum of contributions from each of the four classes of observations. Define A_i as before, and $A_{i,eq} = y_{i1}^{\exp a_1}e^{k_1+\boldsymbol{\beta}_1'\mathbf{X}_{i1}} + y_{i1}^{\exp a_2}e^{k_2+\boldsymbol{\beta}_2'\mathbf{X}_{2i}}$ (that is, the value of A_i when $y_{i1} = y_{i2}$). Note that $A_{i,eq}$ is not a function of a_3.

$$\begin{aligned}
\ell_w(\boldsymbol{\xi}) = \sum_{i=1}^{n}\frac{1}{\pi_i}\Big[&\delta_{i1}\delta_{i2}\Big\{a_1+k_1+(e^{a_1}-1)\log y_{i1}+\boldsymbol{\beta}_1'\mathbf{X}_{1i}+a_3+k_3+(e^{a_3}-1)\log(y_{i2}-y_{i1}) \\
&\quad +\boldsymbol{\beta}_3'\mathbf{X}_{3i} - \left(e^{-h}+2\right)\log(1+e^h A_i)\Big\} \\
&+(1-\delta_{i1})\delta_{i2}\left\{a_2+k_2+(e^{a_2}-1)\log y_{i1}+\boldsymbol{\beta}_2'\mathbf{X}_{2i}-\left(e^{-h}+1\right)\log(1+e^h A_{i,eq})\right\} \\
&+\delta_{i1}(1-\delta_{i2})\left\{a_1+k_1+(e^{a_1}-1)\log y_{i1}+\boldsymbol{\beta}_1'\mathbf{X}_{1i}-\left(e^{-h}+1\right)\log(1+e^h A_i)\right\} \\
&+(1-\delta_{i1})(1-\delta_{i2})\left\{-e^{-h}\log(1+e^h A_{i,eq})\right\}\Big]
\end{aligned}$$

When finding $U_{i,10}(\boldsymbol{\xi})$, we only need to account for log-likelihood contributions from patients who experienced a non-terminal event (members of classes 1 and 3).

$$\begin{aligned}
U_{i,10}(\boldsymbol{\xi}) &= \frac{\delta \ell_i(\boldsymbol{\xi})}{\delta a_3} \\
&= e^{a_3}\Bigg[\delta_{i1}\delta_{i2}\left\{e^{-a_3}+\log(y_{i2}-y_{i1}) - \left(e^{-h}+2\right)\frac{(y_{i2}-y_{i1})^{\exp a_3}\log(y_{i2}-y_{i1})e^{h+k_3+\boldsymbol{\beta}_3'\mathbf{X}_{3i}}}{1+e^h A_i}\right\} \\
&\qquad +\delta_{i1}(1-\delta_{i2})\left\{-\left(e^{-h}+1\right)\frac{(y_{i2}-y_{i1})^{\exp a_3}\log(y_{i2}-y_{i1})e^{h+k_3+\boldsymbol{\beta}_3'\mathbf{X}_{3i}}}{1+e^h A_i}\right\}\Bigg]
\end{aligned}$$

A.4 Specification of weighted illness-death model: B-splines

We also model the log-baseline hazard (BH) functions using splines. That is, for $g = 1, 2, 3$, we define

$$\log\lambda_{0g}(t) = \sum_{k=1}^{K_g} B_{gk}(t)\phi_{gk};\, \lambda_{0g}(t) = \exp\left\{\sum_{k=1}^{K_g} B_{gk}(t)\phi_{gk}\right\};\, \Lambda_{0g}(t) = \int_0^t \exp\left\{\sum_{k=1}^{K_g} B_{gk}(s)\phi_{gk}\right\}ds$$

B_{gk} is the k^{th} basis function of the B-spline based on data informing BH g (here, we use cubic splines). K_g is the number of spline basis functions used for BH g, which corresponds to (# internal knots + 3). Let $\boldsymbol{\phi}_g = (\phi_{g1}, \ldots, \phi_{gK_g})'$, a $(K_g \times 1)$ vector of spline coefficients that are parameters of interest in our model. Operationally, the spline basis functions are computed using the `splines2` package in `R`, which assigns internal knots based on quantiles of the specified data - we used the event times for each transition. Knot placement was determined by y_1 among observations for which $\delta_1 = 1$ for BH 1, y_1 among observations for which $\delta_1 = 0$ and $\delta_2 = 1$ for BH 2 (note that for these observations, $y_1 = y_2$), and $y_2 - y_1$ among observations for which $\delta_1 = 1$ and $\delta_2 = 1$ for BH 3. The lower boundary knot was set at 0 for all hazards, and the upper boundary knot was set at $\max(y_1) + 0.01$ (for BH 1 and BH 2) or $\max(y_2 - y_1) + 0.01$ (for BH 3). To evaluate the splines at non-event times, the `predict` function in `splines2` was used. Let $h = \log\theta$ and let

$$A_i = \left[\int_0^{y_{i1}} \exp\left\{\sum_{k=1}^{K_1} B_{1k}(s)\phi_{1k}\right\} ds\right] e^{\boldsymbol{\beta}_1'\mathbf{x}_{1i}} + \left[\int_0^{y_{i1}} \exp\left\{\sum_{k=1}^{K_2} B_{2k}(s)\phi_{2k}\right\} ds\right] e^{\boldsymbol{\beta}_2'\mathbf{x}_{2i}} + \left[\int_0^{y_{i2}-y_{i1}} \exp\left\{\sum_{k=1}^{K_3} B_{3k}(s)\phi_{3k}\right\} ds\right] e^{\boldsymbol{\beta}_3'\mathbf{x}_{3i}}$$

Our parameters of interest are $\boldsymbol{\xi} = (h, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\phi}_3)$, which we solve for using on the basis of the score equations:

$$\mathcal{U}^w(\boldsymbol{\xi}) = \sum_{i=1}^{N} \mathbf{U}_i^w(\boldsymbol{\xi}) = \sum_{i=1}^{N} R_i \pi_i^{-1} \mathbf{U}_i(\boldsymbol{\xi})$$

with $\mathbf{U}_i(\boldsymbol{\xi}) = \frac{\partial}{\partial \boldsymbol{\xi}} \ell_i(\boldsymbol{\xi})$ listed below.

$$U_{i,1}(\boldsymbol{\xi}) = \frac{\partial \ell_i(\boldsymbol{\xi})}{\partial h} = e^h \left\{ \frac{\delta_{i1}\delta_{i2}}{1+e^h} + \frac{\log(1+e^h A_i)}{e^{2h}} - \left(e^{-h} + \delta_{i1} + \delta_{i2}\right) \frac{A_i}{1+e^h A_i} \right\}$$

$$\mathbf{U}_{i,2}(\boldsymbol{\xi}) = \frac{\partial \ell_i(\boldsymbol{\xi})}{\partial \boldsymbol{\beta_1}} = \delta_{i1}\mathbf{x}_i - \left(e^{-h} + \delta_{i1} + \delta_{i2}\right) \frac{\left[\int_0^{y_{i1}} \exp\left\{\sum_{k=1}^{K_1} B_{1k}(s)\phi_{1k}\right\} ds\right] e^{h+\boldsymbol{\beta}_1'\mathbf{x}_i}\mathbf{x}_i}{1+e^h A_i}$$

$$\mathbf{U}_{i,3}(\boldsymbol{\xi}) = \frac{\partial \ell_i(\boldsymbol{\xi})}{\partial \boldsymbol{\beta_2}} = (1-\delta_{i1})\delta_{i2}\mathbf{x}_i - \left(e^{-h} + \delta_{i1} + \delta_{i2}\right) \frac{\left[\int_0^{y_{i1}} \exp\left\{\sum_{k=1}^{K_2} B_{2k}(s)\phi_{2k}\right\} ds\right] e^{h+\boldsymbol{\beta}_2'\mathbf{x}_i}\mathbf{x}_i}{1+e^h A_i}$$

$$\mathbf{U}_{i,4}(\boldsymbol{\xi}) = \frac{\partial \ell_i(\boldsymbol{\xi})}{\partial \boldsymbol{\beta_3}} = \delta_{i1}\delta_{i2}\mathbf{x}_i - \left(e^{-h} + \delta_{i1} + \delta_{i2}\right) \frac{\left[\int_0^{y_{i2}-y_{i1}} \exp\left\{\sum_{k=1}^{K_3} B_{3k}(s)\phi_{3k}\right\} ds\right] e^{h+\boldsymbol{\beta}_3'\mathbf{x}_i}\mathbf{x}_i}{1+e^h A_i}$$

$$U_{i,5l}(\boldsymbol{\xi}) = \frac{\partial \ell_i(\boldsymbol{\xi})}{\partial \phi_{1l}} = \delta_{i1}B_{1l}(y_{i1}) - \left(e^{-h} + \delta_{i1} + \delta_{i2}\right) \frac{\left[\int_0^{y_{i1}} \exp\left\{\sum_{k=1}^{K_1} B_{1k}(s)\phi_{1k}\right\} B_{1l}(s) ds\right] e^{h+\boldsymbol{\beta}_1'\mathbf{x}_i}}{1+e^h A_i}$$

$$U_{i,6l}(\boldsymbol{\xi}) = \frac{\partial \ell_i(\boldsymbol{\xi})}{\partial \phi_{2l}}$$
$$= (1-\delta_{i1})\delta_{i2} B_{2l}(y_{i1}) - \left(e^{-h} + \delta_{i1} + \delta_{i2}\right) \frac{\left[\int_0^{y_{i1}} \exp\left\{\sum_{k=1}^{K_2} B_{2k}(s)\phi_{2k}\right\} B_{2l}(s) ds\right] e^{h+\boldsymbol{\beta}_2'\mathbf{x}_i}}{1+e^h A_i}$$

For $\mathbf{U}_{i,7}(\boldsymbol{\xi})$, we adjust the notation slightly. Computing A_i in this case gives errors since $\log(y_{i2}-y_{i1})$ is calculated when $\delta_{i1} = 0$. To fix this, we can break the log-likelihood back into the sum of contributions from each of the four classes of observations. Define A_i as before, and

$$A_{i,eq} = \left[\int_0^{y_{i1}} \exp\left\{\sum_{k=1}^{K_1} B_{1k}(s)\phi_{1k}\right\} ds\right] e^{\boldsymbol{\beta}_1'\mathbf{x}_{1i}} + \left[\int_0^{y_{i1}} \exp\left\{\sum_{k=1}^{K_2} B_{2k}(s)\phi_{2k}\right\} ds\right] e^{\boldsymbol{\beta}_2'\mathbf{x}_{2i}}$$

(that is, the value of A_i when $y_{i1} = y_{i2}$).

$$\begin{aligned}\ell_i(\boldsymbol{\xi}) = \delta_{i1}\delta_{i2} &\left\{\log\lambda_{01}(y_{i1}) + \boldsymbol{\beta}_1'\mathbf{x}_i + \log\lambda_{03}(y_{i2}-y_{i1}) + \boldsymbol{\beta}_3'\mathbf{x}_i + \log(1+e^h) - \left(e^{-h}+2\right)\log(1+e^h A_i)\right\} \\ &+ (1-\delta_{i1})\delta_{i2}\left\{\log\lambda_{02}(y_{i1}) + \boldsymbol{\beta}_2'\mathbf{x}_i - \left(e^{-h}+1\right)\log(1+e^h A_{i,eq})\right\} \\ &+ \delta_{i1}(1-\delta_{i2})\left\{\log\lambda_{01}(y_{i1}) + \boldsymbol{\beta}_1'\mathbf{x}_i - \left(e^{-h}+1\right)\log(1+e^h A_i)\right\} \\ &+ (1-\delta_{i1})(1-\delta_{i2})\left\{-e^{-h}\log(1+e^h A_{i,eq})\right\}\end{aligned}$$

When finding $\mathbf{U}_7$, only take derivatives with respect to classes 1 and 3 (people who experienced a non-terminal event).

$$U_{i,7l}(\boldsymbol{\xi}) = \frac{\partial \ell_i(\boldsymbol{\xi})}{\partial \phi_{3l}}$$
$$= \delta_{i1}\delta_{i2}\left(M_{3l}(y_{i2}-y_{i1}) - \left(e^{-h}+2\right)\frac{\left[\int_0^{y_{i2}-y_{i1}} \exp\left\{\sum_{k=1}^{K_3} B_{3k}(s)\phi_{3k}\right\} B_{3l}(s)ds\right] e^{h+\boldsymbol{\beta}_3'\mathbf{x}_i}}{1+e^h A_i}\right)$$
$$+ \delta_{i1}(1-\delta_{i2})\left(-\left(e^{-h}+1\right)\frac{\left[\int_0^{y_{i2}-y_{i1}} \exp\left\{\sum_{k=1}^{K_3} B_{3k}(s)\phi_{3k}\right\} B_{3l}(s)ds\right] e^{h+\boldsymbol{\beta}_3'\mathbf{x}_i}}{1+e^h A_i}\right)$$

In our implementation, the score equations are programmed in `Rcpp`, while the maximizer is obtained using `nleqslv` in `R`. In practice, our approach to selecting starting values involved initially fitting a flexible Bayesian model using `SemiCompRisks`. Starting values for the spline coefficients were set to be those that produced the smallest squared difference from the estimated baseline hazard from the Bayesian model on a fine grid. This implementation can currently be found on GitHub at `https://github.com/ijazic/FreqIDSpline` and will be included in an upcoming release of the `SemiCompRisks` package.

A.5 Penalized likelihood for flexible spline-based illness-death model

To ensure smoothness in the estimated log-BH functions, penalty terms may be placed on the norms of their second derivatives. The resulting penalized weighted log-likelihood has the following form:

$$\ell^{wp}(\boldsymbol{\xi}) = \ell^{w}(\boldsymbol{\xi}) - \sum_{g=1}^{3} \kappa_g \int (\log \lambda_{0g})''^2(u)du$$

The corresponding score functions are

$$\mathcal{U}^{wp}(\boldsymbol{\xi}) = \sum_{i=1}^{N} \mathbf{U}_i^{w}(\boldsymbol{\xi}) - \mathbf{U}^{p}(\boldsymbol{\xi}) = \sum_{i=1}^{N} R_i \pi_i^{-1} \mathbf{U}_i(\boldsymbol{\xi}) - \mathbf{U}^{p}(\boldsymbol{\xi})$$

with $\mathbf{U}_i(\boldsymbol{\xi}) = \frac{\partial}{\partial \boldsymbol{\xi}} \ell_i(\boldsymbol{\xi})$ listed in the previous section, and $\mathbf{U}^p(\boldsymbol{\xi}) = \frac{\partial}{\partial \boldsymbol{\xi}} \sum_{g=1}^{3} \kappa_g \int (\log \lambda_{0g})''^2(u)du$ listed below. Note that $\mathbf{U}^p(\boldsymbol{\xi})$ is external to the likelihood and thus cannot be broken down into individual-level contributions. Since only the $\boldsymbol{\phi}$ parameters appear in the penalty terms, all elements of $\mathbf{U}^p(\boldsymbol{\xi})$ are 0 except:

$$U_{5l}^{p}(\boldsymbol{\xi}) = \frac{\partial}{\partial \phi_{1l}} \sum_{g=1}^{3} \kappa_g \int (\log \lambda_{0g})''^2(u)du = \kappa_1 \left\{ 2 \sum_{k=1}^{K_1} \phi_{1k} \int B''_{1k}(u) B''_{1l}(u)du \right\}$$

$$U_{6l}^{p}(\boldsymbol{\xi}) = \frac{\partial}{\partial \phi_{2l}} \sum_{g=1}^{3} \kappa_g \int (\log \lambda_{0g})''^2(u)du = \kappa_2 \left\{ 2 \sum_{k=1}^{K_2} \phi_{2k} \int B''_{2k}(u) B''_{2l}(u)du \right\}$$

$$U_{7l}^{p}(\boldsymbol{\xi}) = \frac{\partial}{\partial \phi_{3l}} \sum_{g=1}^{3} \kappa_g \int (\log \lambda_{0g})''^2(u)du = \kappa_3 \left\{ 2 \sum_{k=1}^{K_3} \phi_{3k} \int B''_{3k}(u) B''_{3l}(u)du \right\}$$

Below is the derivation of $U_{5l}^{p}(\boldsymbol{\xi})$. First, we can express the full penalty term in terms of partial second derivatives of the B-spline basis functions (which can be calculated using `splines2`):

$$\begin{aligned} \int (\log \lambda_{01})''^2(u)du &= \int \left[\frac{\delta^2}{\delta u^2} \left\{ \sum_{k=1}^{K_1} B_{1k}(u)\phi_{1k} \right\} \right]^2 du = \int \left[\sum_{k=1}^{K_1} \left\{ \frac{\delta^2}{\delta u^2} B_{1k}(u)\phi_{1k} \right\} \right]^2 du \\ &= \int \left\{ \sum_{k=1}^{K_1} B''_{1k}(u)\phi_{1k} \right\}^2 du \end{aligned}$$

Then, the derivative follows (in convenient form, since these integrals do not need to be updated as the parameter values change):

$$\frac{\partial}{\partial \phi_{1l}} \left\{ \int (\log \lambda_{01})''^2(u) du \right\} = \frac{\partial}{\partial \phi_{1l}} \int \left\{ \sum_{k=1}^{K_1} B''_{1k}(u)\phi_{1k} \right\}^2 du = \int \frac{\partial}{\partial \phi_{1l}} \left\{ \sum_{k=1}^{K_1} B''_{1k}(u)\phi_{1k} \right\}^2 du$$
$$= \int 2 \left\{ \sum_{k=1}^{K_1} B''_{1k}(u)\phi_{1k} \right\} B''_{1l}(u) du = 2 \sum_{k=1}^{K_1} \phi_{1k} \int B''_{1k}(u) B''_{1l}(u) du$$

Let $\hat{\boldsymbol{\xi}}$ be the solution to the above score equations (the maximum penalized likelihood estimator) and let $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \kappa_3)'$. The smoothing parameters can be chosen by maximizing an approximation to the standard cross-validation score, as in Joly *et al.* (1998)(1):

$$\bar{V}(\boldsymbol{\kappa}) = \ell^w(\hat{\boldsymbol{\xi}}) + \text{trace}\left[\hat{\mathbf{H}}^{-1}\hat{\mathbf{I}}\right]$$

where $\hat{\mathbf{H}} = -\frac{\partial^2}{\partial \boldsymbol{\xi}} \ell^{wp}(\boldsymbol{\xi})|_{\hat{\boldsymbol{\xi}}}$ and $\hat{\mathbf{I}} = \sum_{i=1}^{N} \mathbf{U}_i^w(\boldsymbol{\xi}) \mathbf{U}_i^w(\boldsymbol{\xi})^T|_{\hat{\boldsymbol{\xi}}}$. $\boldsymbol{\kappa}$ can be found via a golden section search with respect to this cross-validation criterion.

B Further details on asymptotics

B.1 Proof of unbiased score functions

Assuming the model is correctly specified, we have that:

$$\begin{aligned}
E_{\mathbf{Y},\mathbf{R}}[\mathcal{U}^w(\boldsymbol{\xi})] &= \sum_{i=1}^{N} E_{\mathbf{Y},\mathbf{R}}\left[\frac{R_i}{P(R_i=1|\mathbf{Y}_i)}\mathbf{U}_i(\boldsymbol{\xi})\right] \\
&= \sum_{i=1}^{N} E_{\mathbf{Y}}\left[E_{\mathbf{R}|\mathbf{Y}}\left[\frac{R_i}{P(R_i=1|\mathbf{Y}_i)}\mathbf{U}_i(\boldsymbol{\xi})\right]\right] \\
&= \sum_{i=1}^{N} E_{\mathbf{Y}}\left[\frac{E_{\mathbf{R}|\mathbf{Y}}[R_i]}{P(R_i=1|\mathbf{Y}_i)}\mathbf{U}_i(\boldsymbol{\xi})\right] \\
&= \sum_{i=1}^{N} E_{\mathbf{Y}}\left[\frac{P(R_i=1|\mathbf{Y}_i)}{P(R_i=1|\mathbf{Y}_i)}\mathbf{U}_i(\boldsymbol{\xi})\right] \;=\; E_{\mathbf{Y}}\left[\mathcal{U}(\boldsymbol{\xi})\right] \;=\; \mathbf{0}.
\end{aligned}$$

where $\mathcal{U}(\boldsymbol{\xi}) = \sum_{i=1}^{N}\mathbf{U}_i(\boldsymbol{\xi})$ are the complete-data score functions.

B.2 Asymptotics for the penalized log-likelihood

The penalized log-likelihood we consider is

$$\ell^{wp}(\boldsymbol{\xi}) = \ell^{w}(\boldsymbol{\xi}) - \sum_{g=1}^{3}\kappa_g\int(\log\lambda_{0g})''^2(u)du.$$

Note, $\boldsymbol{\xi}$ is a vector of length p_ξ=1 + $\sum_{g=1}^{3}p_g$ + $\sum_{g=1}^{3}K_g$ where p_g is the length of $\mathbf{X}_g$ and K_g is the number of basis functions in the B-spline specification of $\log\lambda_{0g}(t)$. Since the penalty function is quadratic in the spline coefficient vector, one can rewrite the penalized log-likelihood as

$$\ell^{wp}(\boldsymbol{\xi}) = \ell^{w}(\boldsymbol{\xi}) - \sum_{g=1}^{3}\kappa_g\boldsymbol{\phi}_g^T\mathbf{K}_g\boldsymbol{\phi}_g$$

where $\boldsymbol{K}_g$, $g\in\{1,2,3\}$, are non-negative $K_g\times K_g$ definite matrices, that are functions of t_1 and t_2. Following Gray (1994)(2) and Berhane and Weissfeld (2003)(3), one can perform a Taylor series expansion to reveal that

$$\sqrt{N}(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi}_0) = N\left(\boldsymbol{J}+\tilde{\kappa}\tilde{\mathbf{K}}\right)^{-1}N^{-1/2}\mathcal{U}^w(\boldsymbol{\xi}_0) + o_p(1)$$

where $\boldsymbol{\xi}_0$ is the true value of the parameter vector, $\boldsymbol{J} = E_{\mathbf{Y},\mathbf{R}}\left[\partial\mathbf{U}^w(\boldsymbol{\xi})/\partial\boldsymbol{\xi}\right]|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0}$ is the information matrix corresponding to the weighted unpenalized log-likelihood and $\mathcal{U}^w(\boldsymbol{\xi})$ is the weighted unpenalized score vector. Furthermore, $\tilde{\mathbf{K}}$ is a $p_\xi\times p_\xi$ block-diagonal matrix with zero rows and columns for components of $\boldsymbol{\xi}$ that are not penalized and blocks $\boldsymbol{K}_g$ for components that are, while $\tilde{\kappa}$ is a $p_\xi \times p_\xi$ diagonal matrix with zero's and κ_g corresponding to $\tilde{\mathbf{K}}$. From the Central Limit Theorem, we then have that $\sqrt{N}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0)$ is asymptotically normal with mean $\mathbf{0}$ and variance given as the limit of $N\mathbf{V}$ where

$$\mathbf{V} = \left(\boldsymbol{J} + \tilde{\kappa}\tilde{\mathbf{K}}\right)^{-1} \boldsymbol{\Gamma} \left(\boldsymbol{J} + \tilde{\kappa}\tilde{\mathbf{K}}\right)^{-1}$$

which follows from the asymptotic normality of $\mathcal{U}^w(\boldsymbol{\xi})$. Note, this result assumes K_g remains fixed as $N \to \infty$ [2].

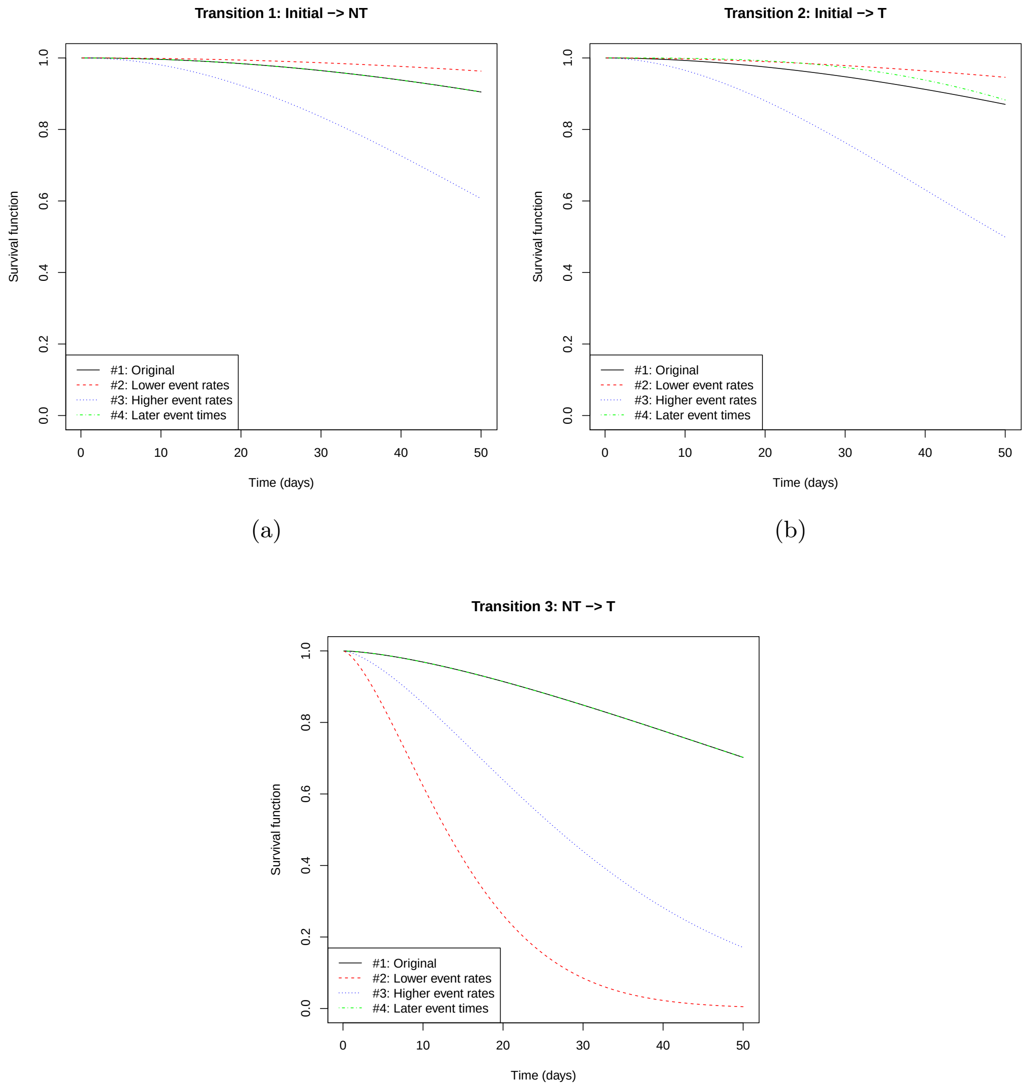
C Detailed specification of simulation settings

Table 2: True values of regression coefficients $\boldsymbol{\beta}$

	Hazard 1	Hazard 2	Hazard 3
X_A	1	-1	-0.5
X_B	0.5	-0.5	-0.5
X_C	0.5	0.5	0.5

Table 3: Weibull baseline hazard settings, using parameterization $\lambda_{0g}(t) = \phi_{g1}\phi_{g2}t^{\phi_{g1}-1}$

		Hazard 1	Hazard 2	Hazard 3
#1: Original	ϕ_{g1}	2	1.85	1.5
	ϕ_{g2}	4×10^{-5}	1×10^{-4}	1×10^{-2}
#2: Lower event rates	ϕ_{g1}	2	1.85	1.5
	ϕ_{g2}	1.5×10^{-5}	4×10^{-5}	1.5×10^{-2}
#3: Higher event rates	ϕ_{g1}	2	1.85	1.5
	ϕ_{g2}	2×10^{-4}	5×10^{-4}	5×10^{-3}
#4: Later event times	ϕ_{g1}	2	3	1.5
	ϕ_{g2}	4×10^{-5}	1×10^{-6}	1×10^{-2}



(a) (b) (c)

Figure 2: Baseline survival function plots for the three transitions in the illness-death model, under the Weibull settings above. Note that the “Original” and “Later event times” settings have identical baseline survival functions for transitions 1 and 3. These figures do not take into account the uniform censoring.

D Additional simulation results

Table 4: All iterations whose exponentiated estimates of $\log\theta$ were > 6 median absolute deviations (MAD) away from the median estimate were removed; the number of iterations removed (out of 10,000) under all scenarios displayed in the main manuscript are identified here. The number of iterations that would have been filtered out based on > 4 MAD and > 5 MAD rules are displayed as well.

		Table 1 Scenario 1		Table 2 Scenario 1		Table 3 Scenario 2		Table 4 Scenario 3	Table 4 Scenario 4
		$m = 1$	$m = 3$	$m = 1$	$m = 3$	$m = 1$	$m = 3$	$m = 1$	$m = 1$
NT-NCC	> 4 MAD	828	307	240	236	329	211	309	525
	> 5 MAD	95	276	155	87	191	138	138	488
	> 6 MAD	77	14	115	63	119	90	71	6
T-NCC	> 4 MAD	270	153	30	114	168	130	–	–
	> 5 MAD	219	85	7	0	49	34	–	–
	> 6 MAD	0	76	3	0	10	8	–	–
NT-SNCC	> 4 MAD	136	105	112	81	147	119	–	–
	> 5 MAD	76	46	65	42	41	29	–	–
	> 6 MAD	70	38	0	0	13	8	–	–
T-SNCC	> 4 MAD	147	118	104	90	36	35	–	–
	> 5 MAD	93	60	56	48	9	11	–	–
	> 6 MAD	89	52	0	0	1	3	–	–

Table 5: Results for $\beta_{B1} - \beta_{B3}$ and $\beta_{C1} - \beta_{C3}$ under simulation scenario #1 with θ=1, based on 10,000 simulated datasets. Shown are percent bias, empirical standard errors, standard error estimates and coverage for Wald-based 95% confidence intervals (CS: conservative sandwich estimator; PR: perturbation resampling estimator), for four designs that vary by the index event (NT: non-terminal; T: terminal) and whether the supplementation of non-index cases was used. Also shown is relative uncertainty, defined as the ratio of the (empirical) standard error for estimates based on a given NCC design to that of the estimates from an analysis of the full cohort.

Parameter/	$m=1$							$m=3$						
Design			SE		Cov					SE		Cov		
	% Bias	Emp	CS	PR	CS	PR	Rel Unc	% Bias	Emp	CS	PR	CS	PR	Rel Unc
β_{B1}														
NT-NCC	1.5	0.087	0.086	0.086	0.95	0.96	1.62	1.3	0.064	0.064	0.064	0.95	0.95	1.20
NT-SNCC	0.3	0.070	0.070	0.069	0.95	0.94	1.30	0.2	0.058	0.058	0.056	0.95	0.93	1.09
T-NCC	-0.3	0.088	0.088	0.086	0.95	0.94	1.65	0.3	0.065	0.064	0.061	0.95	0.93	1.21
T-SNCC	0.1	0.072	0.071	0.070	0.95	0.95	1.35	0.2	0.059	0.059	0.057	0.95	0.94	1.11
β_{B2}														
NT-NCC	2.8	0.360	0.324	0.313	0.94	0.93	4.84	-0.0	0.232	0.209	0.196	0.94	0.92	3.11
NT-SNCC	0.2	0.087	0.087	0.086	0.95	0.94	1.17	0.1	0.078	0.078	0.076	0.95	0.95	1.05
T-NCC	0.4	0.090	0.089	0.088	0.95	0.95	1.21	0.2	0.078	0.079	0.077	0.95	0.95	1.05
T-SNCC	0.3	0.089	0.088	0.087	0.95	0.94	1.20	0.2	0.078	0.078	0.077	0.95	0.95	1.05
β_{B3}														
NT-NCC	-0.8	0.088	0.088	0.093	0.94	0.96	1.11	-0.4	0.082	0.082	0.084	0.95	0.96	1.04
NT-SNCC	-0.2	0.080	0.079	0.080	0.95	0.95	1.00	-0.2	0.080	0.079	0.080	0.95	0.95	1.00
T-NCC	-0.7	0.122	0.120	0.117	0.94	0.93	1.54	0.0	0.092	0.091	0.087	0.95	0.93	1.16
T-SNCC	-0.2	0.079	0.079	0.079	0.95	0.95	1.00	-0.2	0.079	0.079	0.080	0.95	0.95	1.00
β_{C1}														
NT-NCC	3.9	0.091	0.087	0.095	0.93	0.97	1.76	2.6	0.066	0.064	0.069	0.93	0.96	1.27
NT-SNCC	0.0	0.068	0.068	0.069	0.95	0.96	1.32	0.2	0.056	0.056	0.058	0.95	0.96	1.08
T-NCC	-0.4	0.087	0.086	0.086	0.95	0.95	1.69	0.0	0.063	0.063	0.061	0.95	0.95	1.22
T-SNCC	0.2	0.069	0.070	0.070	0.95	0.96	1.34	0.1	0.057	0.057	0.058	0.95	0.96	1.10
β_{C2}														
NT-NCC	6.6	0.344	0.311	0.306	0.94	0.94	4.88	4.6	0.220	0.200	0.189	0.94	0.93	3.12
NT-SNCC	0.3	0.083	0.083	0.083	0.95	0.95	1.18	0.2	0.074	0.074	0.074	0.95	0.95	1.05
T-NCC	-0.2	0.087	0.086	0.086	0.95	0.95	1.23	0.1	0.075	0.075	0.074	0.95	0.95	1.06
T-SNCC	0.4	0.085	0.085	0.084	0.95	0.94	1.20	0.2	0.074	0.074	0.074	0.95	0.96	1.05
β_{C3}														
NT-NCC	3.8	0.098	0.091	0.109	0.90	0.98	1.27	2.7	0.086	0.083	0.093	0.93	0.97	1.10
NT-SNCC	-0.1	0.078	0.078	0.081	0.95	0.97	1.00	-0.0	0.078	0.078	0.082	0.95	0.97	1.00
T-NCC	-1.1	0.121	0.118	0.118	0.94	0.94	1.56	-0.1	0.090	0.090	0.089	0.95	0.95	1.16
T-SNCC	-0.1	0.078	0.078	0.081	0.95	0.97	1.00	-0.0	0.078	0.078	0.082	0.95	0.96	1.00

Table 6: Results for $\beta_{B1} - \beta_{B3}$ and $\beta_{C1} - \beta_{C3}$ under simulation scenario #2 with θ=1, based on 10,000 simulated datasets. Shown are percent bias, empirical standard errors, standard error estimates and coverage for Wald-based 95% confidence intervals (CS: conservative sandwich estimator; PR: perturbation resampling estimator), for four designs that vary by the index event (NT: non-terminal; T: terminal) and whether the supplementation of non-index cases was used. Also shown is relative uncertainty, defined as the ratio of the (empirical) standard error for estimates based on a given NCC design to that of the estimates from an analysis of the full cohort.

Parameter/	$m = 1$							$m = 3$						
Design			SE		Cov					SE		Cov		
	% Bias	Emp	CS	PR	CS	PR	Rel Unc	% Bias	Emp	CS	PR	CS	PR	Rel Unc
β_{B1}														
NT-NCC	2.0	0.135	0.139	0.129	0.96	0.96	1.86	1.5	0.094	0.097	0.095	0.95	0.97	1.29
NT-SNCC	0.3	0.101	0.100	0.100	0.95	0.95	1.39	0.1	0.082	0.081	0.081	0.95	0.94	1.13
T-NCC	-0.4	0.149	0.149	0.144	0.94	0.93	2.05	-0.1	0.101	0.101	0.098	0.95	0.95	1.39
T-SNCC	0.1	0.103	0.100	0.099	0.94	0.94	1.38	-0.2	0.082	0.081	0.081	0.95	0.94	1.10
β_{B2}														
NT-NCC	13.4	0.715	0.594	0.582	0.94	0.94	6.91	5.3	0.408	0.369	0.353	0.94	0.93	3.95
NT-SNCC	0.4	0.123	0.123	0.123	0.95	0.95	1.19	0.6	0.110	0.110	0.109	0.95	0.94	1.06
T-NCC	0.7	0.123	0.124	0.123	0.95	0.94	1.19	0.7	0.109	0.110	0.109	0.95	0.95	1.06
T-SNCC	0.3	0.126	0.124	0.123	0.94	0.94	1.20	0.6	0.110	0.110	0.109	0.95	0.95	1.05
β_{B3}														
NT-NCC	-1.5	0.136	0.137	0.138	0.94	0.98	1.23	-0.9	0.120	0.120	0.125	0.94	0.97	1.08
NT-SNCC	-0.8	0.112	0.110	0.112	0.95	0.96	1.00	-0.7	0.112	0.110	0.113	0.95	0.96	1.00
T-NCC	-3.0	0.216	0.205	0.193	0.93	0.91	1.95	-1.5	0.151	0.148	0.143	0.94	0.94	1.36
T-SNCC	-0.5	0.111	0.111	0.112	0.95	0.96	1.01	-0.5	0.110	0.110	0.113	0.95	0.96	1.00
β_{C1}														
NT-NCC	7.0	0.143	0.141	0.135	0.95	0.97	2.04	4.2	0.096	0.098	0.099	0.95	0.97	1.37
NT-SNCC	0.5	0.098	0.097	0.098	0.95	0.95	1.40	0.3	0.079	0.079	0.079	0.95	0.95	1.13
T-NCC	-1.1	0.147	0.145	0.142	0.94	0.93	2.11	-0.3	0.098	0.099	0.096	0.95	0.95	1.40
T-SNCC	0.8	0.098	0.097	0.098	0.95	0.95	1.41	0.7	0.078	0.079	0.079	0.95	0.95	1.12
β_{C2}														
NT-NCC	14.8	0.648	0.572	0.556	0.93	0.92	6.57	5.6	0.392	0.357	0.343	0.94	0.94	3.97
NT-SNCC	0.1	0.119	0.118	0.118	0.95	0.95	1.21	0.2	0.105	0.104	0.104	0.95	0.95	1.07
T-NCC	-1.0	0.120	0.121	0.119	0.95	0.96	1.22	-0.3	0.105	0.105	0.104	0.95	0.95	1.07
T-SNCC	-0.5	0.117	0.118	0.118	0.96	0.96	1.20	-0.5	0.105	0.104	0.104	0.95	0.95	1.07
β_{C3}														
NT-NCC	6.0	0.159	0.146	0.159	0.90	0.98	1.43	3.8	0.132	0.125	0.140	0.92	0.97	1.19
NT-SNCC	-0.9	0.112	0.109	0.114	0.94	0.96	1.01	-0.7	0.112	0.109	0.114	0.94	0.96	1.00
T-NCC	-5.6	0.213	0.204	0.195	0.93	0.92	1.92	-2.4	0.150	0.147	0.144	0.94	0.95	1.35
T-SNCC	0.1	0.110	0.109	0.114	0.94	0.96	1.01	0.1	0.110	0.109	0.114	0.94	0.96	1.00

Table 7: Results for $\beta_{A1} - \beta_{A3}$ and θ under simulation scenario #1 with θ=3, based on 10,000 simulated datasets. Shown are percent bias, empirical standard errors, standard error estimates and coverage for Wald-based 95% confidence intervals (CS: conservative sandwich estimator; PR: perturbation resampling estimator), for four designs that vary by the index event (NT: non-terminal; T: terminal) and whether the supplementation of non-index cases was used. Also shown is relative uncertainty, defined as the ratio of the (empirical) standard error for estimates based on a given NCC design to that of the estimates from an analysis of the full cohort.

Parameter/	$m = 1$							$m = 3$						
Design			SE		Cov					SE		Cov		
	% Bias	Emp	CS	PR	CS	PR	Rel Unc	% Bias	Emp	CS	PR	CS	PR	Rel Unc
β_{A1}														
NT-NCC	2.8	0.153	0.134	0.135	0.95	0.97	2.22	2.1	0.105	0.098	0.097	0.94	0.95	1.52
NT-SNCC	0.3	0.092	0.092	0.091	0.95	0.95	1.34	0.3	0.076	0.076	0.074	0.95	0.95	1.10
T-NCC	0.1	0.111	0.110	0.109	0.95	0.94	1.61	0.2	0.083	0.082	0.079	0.95	0.93	1.20
T-SNCC	0.2	0.092	0.092	0.090	0.95	0.94	1.34	0.2	0.076	0.076	0.074	0.95	0.94	1.11
β_{A2}														
NT-NCC	9.4	0.722	0.539	0.523	0.93	0.92	6.08	2.1	0.401	0.350	0.333	0.93	0.91	3.38
NT-SNCC	0.5	0.132	0.133	0.132	0.95	0.95	1.11	0.4	0.123	0.122	0.122	0.95	0.95	1.03
T-NCC	0.7	0.135	0.135	0.135	0.95	0.95	1.14	0.5	0.123	0.123	0.122	0.95	0.95	1.04
T-SNCC	0.5	0.133	0.133	0.132	0.95	0.95	1.12	0.5	0.123	0.122	0.121	0.95	0.95	1.03
β_{A3}														
NT-NCC	-3.0	0.138	0.126	0.122	0.95	0.95	1.50	-2.3	0.107	0.106	0.106	0.95	0.95	1.16
NT-SNCC	0.5	0.094	0.093	0.093	0.95	0.94	1.02	0.4	0.093	0.092	0.093	0.95	0.94	1.01
T-NCC	0.8	0.139	0.139	0.135	0.95	0.94	1.51	0.4	0.108	0.107	0.103	0.95	0.94	1.17
T-SNCC	0.5	0.094	0.093	0.093	0.95	0.94	1.02	0.4	0.093	0.092	0.092	0.95	0.94	1.00
θ														
NT-NCC	23.1	0.632	0.591	0.714	0.69	0.84	1.80	13.9	0.527	0.452	0.500	0.74	0.84	1.50
NT-SNCC	-0.2	0.366	0.377	0.367	0.95	0.95	1.04	0.2	0.356	0.358	0.357	0.95	0.95	1.02
T-NCC	-1.2	0.466	0.488	0.467	0.95	0.95	1.33	-0.0	0.385	0.390	0.379	0.95	0.95	1.10
T-SNCC	-0.2	0.367	0.377	0.367	0.95	0.95	1.05	0.2	0.355	0.358	0.358	0.95	0.96	1.01

Table 8: Results for $\beta_{B1} - \beta_{B3}$ and $\beta_{C1} - \beta_{C3}$ under simulation scenario #1 with θ=3, based on 10,000 simulated datasets. Shown are percent bias, empirical standard errors, standard error estimates and coverage for Wald-based 95% confidence intervals (CS: conservative sandwich estimator; PR: perturbation resampling estimator), for four designs that vary by the index event (NT: non-terminal; T: terminal) and whether the supplementation of non-index cases was used. Also shown is relative uncertainty, defined as the ratio of the (empirical) standard error for estimates based on a given NCC design to that of the estimates from an analysis of the full cohort.

Parameter/	$m=1$							$m=3$						
Design			SE		Cov					SE		Cov		
	% Bias	Emp	CS	PR	CS	PR	Rel Unc	% Bias	Emp	CS	PR	CS	PR	Rel Unc
β_{B1}														
NT-NCC	2.3	0.164	0.146	0.139	0.97	0.96	2.42	1.3	0.103	0.100	0.098	0.96	0.95	1.53
NT-SNCC	0.1	0.090	0.091	0.090	0.95	0.94	1.33	-0.1	0.075	0.075	0.073	0.95	0.94	1.11
T-NCC	-0.0	0.110	0.110	0.108	0.95	0.95	1.62	0.1	0.081	0.082	0.078	0.96	0.94	1.20
T-SNCC	0.1	0.090	0.091	0.089	0.95	0.95	1.33	0.1	0.074	0.075	0.073	0.95	0.94	1.10
β_{B2}														
NT-NCC	5.7	0.535	0.447	0.422	0.93	0.92	5.94	1.8	0.320	0.284	0.271	0.94	0.93	3.56
NT-SNCC	0.8	0.107	0.107	0.106	0.95	0.94	1.19	0.8	0.095	0.095	0.093	0.95	0.95	1.06
T-NCC	0.8	0.108	0.108	0.107	0.95	0.95	1.20	0.7	0.095	0.095	0.093	0.95	0.95	1.05
T-SNCC	0.8	0.107	0.107	0.106	0.95	0.94	1.19	0.7	0.095	0.095	0.093	0.95	0.95	1.05
β_{B3}														
NT-NCC	-0.7	0.145	0.133	0.130	0.96	0.96	1.53	-0.6	0.109	0.108	0.109	0.95	0.95	1.15
NT-SNCC	-0.1	0.096	0.095	0.095	0.95	0.94	1.01	-0.1	0.096	0.094	0.094	0.95	0.94	1.01
T-NCC	-0.2	0.142	0.141	0.137	0.95	0.94	1.49	0.0	0.110	0.109	0.105	0.95	0.94	1.15
T-SNCC	-0.1	0.097	0.095	0.095	0.95	0.94	1.02	-0.1	0.095	0.094	0.095	0.95	0.95	1.00
β_{C1}														
NT-NCC	7.1	0.157	0.143	0.142	0.96	0.97	2.35	3.8	0.099	0.098	0.097	0.95	0.95	1.49
NT-SNCC	-0.1	0.089	0.088	0.087	0.95	0.95	1.33	0.1	0.073	0.073	0.072	0.95	0.95	1.09
T-NCC	-0.3	0.107	0.107	0.106	0.95	0.95	1.60	0.0	0.080	0.080	0.077	0.95	0.94	1.19
T-SNCC	-0.0	0.088	0.088	0.087	0.95	0.95	1.32	0.1	0.073	0.073	0.072	0.95	0.96	1.09
β_{C2}														
NT-NCC	12.2	0.509	0.430	0.412	0.94	0.93	6.00	5.5	0.300	0.272	0.260	0.94	0.94	3.54
NT-SNCC	-0.0	0.103	0.102	0.102	0.95	0.95	1.21	0.2	0.090	0.090	0.089	0.95	0.94	1.06
T-NCC	-0.2	0.104	0.104	0.104	0.95	0.95	1.22	-0.0	0.091	0.091	0.089	0.95	0.95	1.07
T-SNCC	0.0	0.102	0.102	0.101	0.95	0.95	1.20	0.0	0.090	0.090	0.089	0.95	0.95	1.06
β_{C3}														
NT-NCC	7.6	0.139	0.130	0.137	0.96	0.98	1.51	4.3	0.105	0.105	0.108	0.95	0.96	1.15
NT-SNCC	0.1	0.093	0.092	0.092	0.95	0.95	1.01	0.3	0.092	0.091	0.091	0.95	0.96	1.00
T-NCC	-0.2	0.138	0.137	0.135	0.95	0.94	1.50	0.1	0.106	0.105	0.102	0.94	0.95	1.16
T-SNCC	0.2	0.093	0.092	0.092	0.95	0.95	1.01	0.3	0.092	0.091	0.091	0.95	0.95	1.00

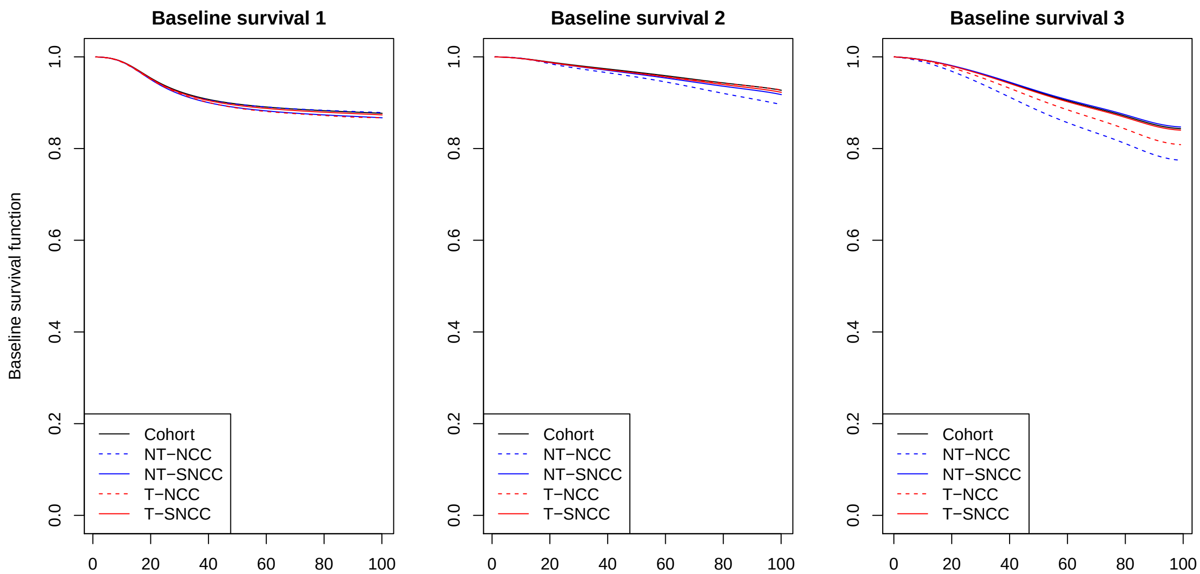
E Additional material related to the CIBMTR dataset

Table 9: Descriptive statistics for the CIBMTR cohort of stem cell transplantation patients

				Observed outcome category, %			
		N	%	Both aGVHD & death	aGVHD & cens for death	Death without aGVHD	Cens for both
Total subjects		8838		5.1	12.6	10.0	72.4
Gender	Male	4891	55.3	5.3	13.3	9.0	72.4
	Female	3947	44.7	4.7	11.7	11.0	72.6
Age, years	<20	1666	18.9	3.0	12.5	7.0	77.6
	20-39	2892	32.7	5.4	13.4	9.4	71.8
	40-59	3750	42.4	5.6	12.4	11.3	70.7
	≥60	530	6.0	5.5	9.1	12.8	72.6
Race	Caucasian	7754	87.8	5.0	12.7	9.9	72.4
	African-American	378	4.3	7.7	16.1	13.5	62.7
	Asian	478	5.4	4.2	9.0	7.1	79.7
	Other	228	2.6	4.8	7.9	11.8	75.4
HLA compatibility	Identical sibling	3665	41.5	3.2	10.1	7.6	79.1
	8/8	3696	41.8	5.8	13.8	10.6	69.8
	7/8	1477	16.7	7.9	15.6	14.1	62.4
Karnofsky score	<90%	2650	30.0	6.6	12.0	15.2	66.2
	90-100%	6188	70.0	4.4	12.8	7.7	75.1
Diagnosis	AML	4528	51.2	4.2	11.6	10.6	73.6
	ALL	1875	21.2	5.1	12.6	10.6	71.7
	CML	1397	15.8	6.1	15.1	7.1	71.7
	MDS	1038	11.7	7.3	13.3	9.6	69.7
Disease status	Early	4524	51.2	4.1	12.3	6.2	77.4
	Intermediate	2107	23.8	4.9	13.3	9.2	72.6
	Advanced	2207	25.0	7.2	12.4	18.4	62.1
Year of transplant	Before 2007	5406	61.2	5.9	12.4	10.7	71.1
	2007 or later	3432	38.8	3.8	12.9	8.6	74.7
Graft type	Bone marrow	3072	34.8	4.2	12.1	10.7	72.9
	Peripheral blood	5766	65.2	5.5	12.8	9.5	72.2
Conditioning intensity	Myeloablative	7055	79.8	5.2	13.1	9.5	72.2
	Non-myeloablative	1783	20.2	4.4	10.6	11.7	73.3
GVHD prophylaxis	Tac/MTX	3333	37.7	4.1	12.9	8.9	74.1
	Tac	1591	18.0	6.2	14.3	11.3	68.2
	CSA/MTX	2610	29.5	5.0	11.9	9.2	73.9
	CSA	834	9.4	7.7	12.7	11.6	68.0
	Ex vivo TCD/$CD34^{+}$	470	5.3	4.3	7.4	13.6	74.7
In vivo T-cell depletion	No	6877	77.8	5.3	13.3	9.1	72.2
	Yes	1961	22.2	4.3	9.8	12.6	73.2

Table 10: Hazard ratio estimates and Wald-based 95% confidence intervals for the spline-based illness-death model fit to the full CIBMTR cohort.

	Hazard 1	Hazard 2	Hazard 3
Female	0.86 (0.77,0.96)	1.19 (1.03,1.37)	1.08 (0.88,1.32)
Age			
<20	0.91 (0.76,1.08)	0.67 (0.53,0.86)	0.59 (0.41,0.85)
20-39	1.00	1.00	1.00
40-59	1.01 (0.89,1.14)	1.32 (1.11,1.56)	1.25 (1.00,1.57)
≥60	0.94 (0.73,1.21)	1.63 (1.20,2.20)	2.28 (1.40,3.70)
African-American	1.47 (1.16,1.87)	1.59 (1.15,2.21)	1.46 (0.95,2.23)
HLA compatibility			
Identical sibling	1.00	1.00	1.00
8/8	1.94 (1.66,2.26)	1.53 (1.24,1.89)	1.37 (1.06,1.77)
7/8	2.59 (2.14,3.13)	2.10 (1.60,2.77)	1.63 (1.22,2.18)
Karnofsky score ≥90%	0.87 (0.77,1.00)	0.56 (0.47,0.66)	0.69 (0.55,0.85)
Diagnosis			
AML	1.00	1.00	1.00
ALL	1.12 (0.96,1.30)	1.35 (1.11,1.64)	1.27 (0.96,1.69)
CML	1.45 (1.23,1.70)	0.96 (0.75,1.22)	1.21 (0.90,1.63)
MDS	1.41 (1.18,1.69)	0.71 (0.56,0.91)	1.44 (1.06,1.95)
Disease status			
Early	1.00	1.00	1.00
Intermediate	1.04 (0.91,1.19)	1.43 (1.18,1.74)	1.16 (0.89,1.51)
Advanced	1.15 (1.00,1.33)	3.12 (2.59,3.77)	1.49 (1.15,1.92)
2007 or later	0.85 (0.75,0.97)	0.82 (0.69,0.98)	0.61 (0.48,0.78)
Bone marrow graft	1.29 (1.12,1.48)	0.74 (0.62,0.88)	1.26 (0.98,1.61)
Myeloablative conditioning	0.63 (0.54,0.74)	1.05 (0.86,1.28)	0.88 (0.64,1.22)
GVHD prophylaxis			
Tac/MTX	1.00	1.00	1.00
Tac	1.31 (1.13,1.52)	1.28 (1.05,1.57)	1.45 (1.08,1.94)
CsA/MTX	1.08 (0.93,1.25)	1.35 (1.10,1.65)	1.39 (1.06,1.84)
CsA	1.69 (1.37,2.09)	1.56 (1.18,2.05)	1.87 (1.32,2.66)
Ex vivo TCD/CD34$^+$ selection	0.64 (0.47,0.87)	1.49 (1.08,2.05)	1.79 (1.03,3.13)
In vivo T-cell depletion	0.64 (0.55,0.74)	1.16 (0.98,1.38)	1.13 (0.86,1.47)
θ		0.62 (0.17,2.24)	



(a) (b) (c)

Figure 3: Estimates of the baseline survival functions from the weighted illness-death model fits (log baseline-hazard modeled with splines) for the four NCC/SNCC studies based on the CIBMTR data, as well as the cohort data.

References

[1] Joly P, Commenges D and Letenneur L. A penalized likelihood approach for arbitrarily censored and truncated data: application to age-specific incidence of dementia. *Biometrics* 1998; 54(1): 185–194.

[2] Gray RJ. Spline-based test in survival analysis. *Biometrics* 1994; 50(3): 640–652.

[3] Berhane K and Weissfeld LA. Inference in spline-based models for multiple time-to-event data, with applications to a breast cancer prevention trial. *Biometrics* 2003; 59(4): 859–868.