Universiteit Leiden
The Netherlands

# Statistical modelling of competing risks with incomplete data: with applications to allogeneic stem cell transplantation

Bonneville, E.F.

**Citation**
Bonneville, E. F. (2025, July 2). *Statistical modelling of competing risks with incomplete data: with applications to allogeneic stem cell transplantation*. Retrieved from https://hdl.handle.net/1887/4252266

| | |
|---|---|
| Version: | Publisher's Version |
| License: | Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden |
| Downloaded from: | https://hdl.handle.net/1887/4252266 |

**Note:** To cite this publication please use the final published version (if applicable).

# Chapter 6

# Why you should avoid using multiple Fine–Gray models: insights from (attempts at) simulating proportional subdistribution hazards data



Chapter based on: **Bonneville, E. F.**, de Wreede, L. C. and Putter, H. (2024) Why you should avoid using multiple Fine–Gray models: insights from (attempts at) simulating proportional subdistribution hazards data. *Journal of the Royal Statistical Society Series A: Statistics in Society*, qnae056. DOI: 10.1093/jrsssa/qnae056.

## Abstract

Studies considering competing risks will often aim to estimate the cumulative incidence functions conditional on an individual's baseline characteristics. While the Fine–Gray subdistribution hazard model is tailor-made for analysing only one of the competing events, it may still be used in settings where multiple competing events are of scientific interest, where it is specified for each cause in turn. In this work, we provide an overview of data-generating mechanisms where proportional subdistribution hazards hold for at least one cause. We use these to motivate why the use of multiple Fine–Gray models should be avoided in favour of better alternatives such as cause-specific hazard models.

## 6.1 Introduction

Competing risks are ubiquitous across medical studies, where patients can experience one of several distinct events, such as death due to different causes. One of the typical aims for a study considering competing risks is to estimate the model-based cumulative probabilities of experiencing one or more of the competing events by a certain time for specific patients (the predicted cumulative incidence functions). Presently, given a set of baseline covariates and outcome data, cumulative incidence functions are generally estimated using either the Fine–Gray subdistribution hazard model (Fine and Gray, 1999), or by fitting and thereafter combining cause-specific Cox proportional hazards models for each event (Putter *et al.*, 2007). The former is often the tool of choice when developing prognostic models for a single event of interest, as it does not require explicitly specifying models for the competing events. It is also arguably simpler to externally validate, requiring only cumulative subdistribution baseline hazard estimates at relevant timepoints together with the estimated regression coefficients (instead of the full cause-specific cumulative hazards for all events).

Even when only one of the endpoints is of primary interest, multiple authors have argued in favour of a more holistic approach to competing risks analyses, suggesting that *all* events should be studied together (Andersen *et al.*, 2012; Gerds *et al.*, 2012; Latouche *et al.*, 2013). For example, as emphasised by Latouche *et al.* (2007), the effect of a given covariate on the cumulative incidence of one event should not be considered in isolation from its effect on the cumulative incidence of competing events. Consider patients with a malignant haematological disease, that are at risk of (competing) disease relapse and non-relapse mortality after undergoing an allogeneic stem cell transplantation (alloSCT). A less intensive pre-transplantation conditioning regimen may not be able to sufficiently control the disease compared with a more intensive regimen (i.e. increased risk of relapse), but it will be less toxic for the patient (i.e. reduced risk of non-relapse mortality)—see Shimoni *et al.* (2016) for an example of opposing effects of conditioning regimen on the cumulative incidence functions of relapse and non-relapse mortality.

In a context where more than one of the competing events are of (possibly equal) interest, one may opt to fit a Fine–Gray model for *each* competing event in turn. Since these models are fitted independently, it is possible that the sum of the estimated cumulative incidence functions (the total failure probability, TFP) for given covariate values at certain timepoints exceeds 1. This was recently illustrated by Austin *et al.* (2021) with data considering cardiovascular and non-cardiovascular death, where the TFP exceeded 1 for 5% of patients at 5 years after hospital admission. This known issue of the TFP exceeding 1 occurs partly as a result of at least one of the specified Fine–Gray models being incorrect (Beyersmann *et al.*, 2012). That is, the assumption of proportional subdistribution hazards fails to hold for at least one of the models.

A useful starting point for further understanding these issues when using multiple Fine–Gray models is to consider a simplified context with two competing risks, and suppose that the Fine–Gray model has been correctly specified for one event (cause 1). The objective of this article is to outline the implications (i.e. the implied assumptions) of specifying a Fine–Gray model for cause 1 on the cumulative incidence function for cause 2. To do so, we provide an overview of data-generating mechanisms (DGMs) where the Fine–Gray model is correctly specified for at least cause 1. For these DGMs, we (a) give example specifications (e.g. choice of distributions), (b) refer to instances where they have been used across the methodological competing risks literature (e.g. as part of simulation studies), if at all, and (c) touch upon potential difficulties from a simulator's perspective. While these DGMs are particularly relevant for methodological researchers aiming to simulate competing risks data under different assumptions, they provide additional insights for applied researchers seeking to motivate their choice of analysis method(s) in a more principled way. In the discussion, we therefore reflect on what the characteristics of the outlined DGMs imply for the use of multiple Fine–Gray models in practice, and argue in favour of cause-specific hazard models for cumulative incidence prediction.

## 6.2 Competing risks and the Fine–Gray model

In a competing risks setting, we assume that individuals can experience only one of $K$ distinct events or, phrased differently, that only the first event of interest is observed. We denote the failure time as $T$, and the competing event indicator as $D \in \{1, ..., K\}$. In practice, individuals are subject to a right-censoring time $C$ (generally assumed independent of $T$ and $D$), and we thus only observe realisations of $\tilde{T} = \min(T, D)$ and $\tilde{D} = I(T \leq C)D$. The cause-specific hazard for the $k^{\text{th}}$ event is defined as

$$h_k(t) = \lim_{\Delta t \downarrow 0} \frac{P(t \leq T < t + \Delta t, D = k \mid T \geq t)}{\Delta t}.$$

These hazards fully define the event-free survival function,

$$P(T > t) = \exp\left( - \sum_{k=1}^{K} \int_0^t h_k(u)\, \mathrm{d}u \right) = \exp\left( - \sum_{k=1}^{K} H_k(t) \right),$$

assuming the distribution of $T$ is continuous, and $H_k(t)$ is the cause-specific cumulative hazard function for the $k^{\text{th}}$ event. The cause-specific cumulative incidence function is then defined as

$$F_k(t) = P(T \leq t, D = k) = \int_0^t h_k(u) S(u-)\, \mathrm{d}u,$$

where $S(u-)$ is the event-free survival probability just prior to $u$.

A relevant question is whether we can model $F_k(t)$ directly, without needing to model all cause-specific hazards. In order to do so, the idea is to specify a hazard for the $k^{\text{th}}$ event, $\lambda_k(t)$, that satisfies

$$P(T \leq t, D = k) = 1 - \exp\left(-\int_0^t \lambda_k(u)\,\mathrm{d}u\right),$$

analogously to the standard single-event survival setting. Rearranging the above yields

$$\begin{aligned}\lambda_k(t) &= \frac{-\,\mathrm{d}\log\{1 - F_k(t)\}}{\mathrm{d}t}, \\ &= \frac{\mathrm{d}F_k(t)}{\mathrm{d}t} \times \{1 - F_k(t)\}^{-1},\end{aligned}$$

which is the commonly known expression for the subdistribution hazard. It can also be written as $\lambda_k(t) = f_k(t)/\{1 - F_k(t)\}$, where $f_k(t) = \mathrm{d}F_k(t)/\,\mathrm{d}t$ is referred to as the subdensity function (Gray, 1988). $F_k(t)$ is not a true distribution function since $F_k(\infty) = P(D = k) < 1$, and is hence known as a 'subdistribution' function. The cause-specific hazard can also be written in terms of the subdensity function, as $h_k(t) = f_k(t)/S(t)$. Thus, the cause-specific hazard conditions on being event-free by $t$, while the subdistribution hazard conditions on not having failed by event $k$ by $t$.

The Fine–Gray model is a semiparametric model that assumes proportionality on the subdistribution hazard scale. Using covariate vector $\mathbf{Z}$, the Fine–Gray model for cause $k$ can be written as

$$\lambda_k(t \,|\, \mathbf{Z}) = \lambda_{k0}(t) \exp(\boldsymbol{\beta}_k^\top \mathbf{Z}),$$

with $\lambda_{k0}(t)$ being the subdistribution baseline hazard function and $\boldsymbol{\beta}_k$ representing the effects of covariates $\mathbf{Z}$ on the subdistribution hazard. The cumulative incidence function for the $k^{\text{th}}$ event can then be written as

$$F_k(t \,|\, \mathbf{Z}) = 1 - \exp\left\{-\exp(\boldsymbol{\beta}_k^\top \mathbf{Z}) \int_0^t \lambda_{k0}(u)\,\mathrm{d}u\right\},$$

which corresponds to modelling the cumulative incidence function with a complementary log-log transformation. Furthermore, let $F_{k0}(t) = 1 - \exp(-\int_0^t \lambda_{k0}(u)\,\mathrm{d}u)$ be the baseline cumulative incidence function, i.e. the cumulative incidence when $\mathbf{Z} = 0$. We can then also write the Fine–Gray model as

$$1 - F_k(t \,|\, \mathbf{Z}) = \{1 - F_{k0}(t)\}^{\exp(\boldsymbol{\beta}_k^\top \mathbf{Z})},$$

which is a similar relation to that of the survival functions in a Cox model.

## 6.3 Data-generating mechanisms

For the sake of simplicity, we restrict ourselves to $K = 2$ competing events, a single time-constant covariate $X$, and assume that cause 1 is of primary interest. Results can be generalised to more complex settings. We let $h_k(t \mid X)$, $\lambda_k(t \mid X)$ and $F_k(t \mid X) = P(T \leq t, D = k \mid X)$, respectively, denote the cause-specific hazard, subdistribution hazard and cumulative incidence function for cause $k$, conditional on $X$. Furthermore, we let $\beta_k$ and $\gamma_k$ represent the effect of $X$ on the subdistribution hazard and cause-specific hazard of cause $k$, respectively. In what follows, we present DGMs for which a Fine–Gray model correctly holds for cause 1. When illustrating the different DGMs, $X$ is assumed to be binary.

In essence, the task is to specify a joint density $f(T, D \mid X)$ where the Fine–Gray model is correctly specified for cause 1. With $S(T \mid X) = \exp\{-H_1(T \mid X) - H_2(T \mid X)\}$, we have

$$\begin{aligned} f(T, D \mid X) &= \{h_1(T \mid X)S(T \mid X)\}^{I(D=1)}\{h_2(T \mid X)S(T \mid X)\}^{I(D=2)}S(T \mid X)^{1-I(D=1)-I(D=2)}, \\ &= f_1(T \mid X)^{I(D=1)} f_2(T \mid X)^{I(D=2)}\{1 - F_1(T \mid X) - F_2(T \mid X)\}^{1-I(D=1)-I(D=2)}, \end{aligned}$$

as written by Andersen and Ravn (2023). Since the Fine–Gray model provides an expression for $F_1(t \mid X)$, we need only think about what assumptions to make regarding cause 2. Additionally, the hazard functions comprising the above density should fulfil various restrictions, which are outlined in the work of Haller and Ulm (2014). Namely,

1. All hazard functions must be non-negative for all time points $t > 0$.
2. The cause-specific and subdistribution hazards (for the event of interest) should be identical before the occurrence of the first competing event. Therefore, $h_1(t \mid X) = \lambda_1(t \mid X)$ at $t = 0$.
3. $F_1(t \mid X)$ must converge to $P(D = 1 \mid X)$ as $t \to \infty$. If $P(D = 1 \mid X) < 1$, this in turn implies that $\lim_{t\to\infty} \lambda_1(t \mid X) = 0$ and even that the cumulative subdistribution hazard $\Lambda_1(t \mid X)$ should not go to infinity for $t \to \infty$.

### 6.3.1 Using the reduction factor

A first approach to specifying $f(T, D \mid X)$ is to make assumptions about all cause-specific hazard functions. That is, we would like to select a set of cause-specific hazard functions for which proportionality holds on the subdistribution hazard of event 1. To do so, we can make use of the link between $h_1(t \mid X)$ and $\lambda_1(t \mid X)$, which is given by

$$\lambda_1(t \mid X) = h_1(t \mid X)\frac{S(t \mid X)}{1 - F_1(t \mid X)}, \tag{6.1}$$

with the latter expression referred to as the *reduction factor* by Putter *et al.* (2020). In the book by Beyersmann *et al.* (2012) (Equation 5.3.9), this has also been written as

$$h_1(t \mid X) = \lambda_1(t \mid X) \left\{ 1 + \frac{F_2(t \mid X)}{S(t \mid X)} \right\},$$

which holds since $S(t \mid X) = 1 - \sum_{k=1}^{2} F_k(t \mid X)$. The above expressions allow to simulate data by specifying $\lambda_1(t \mid X)$ and one of $h_1(t \mid X)$, $h_2(t \mid X)$, or $h_1(t \mid X) + h_2(t \mid X)$, and thereafter deriving the implied cause-specific hazard(s). With both cause-specific hazards being defined, one should be able simulate using standard methods, i.e. with latent times or using the all-cause hazard function (Beyersmann *et al.*, 2009).

#### 6.3.1.1 Specifying $\lambda_1(t \mid X)$ and $h_2(t \mid X)$

Since we assume that the Fine–Gray model is correctly specified for $\lambda_1(t \mid X)$, we can express our assumptions regarding cause 2 for instance by specifying a model for $h_2(t \mid X)$, which can be any hazard-based regression model (e.g. cause-specific Cox, additive hazards, or other), and derive the implied $h_1(t \mid X)$. By re-arranging Equation 6.1 and thereafter integrating with respect to $t$, we can write

$$\overbrace{h_1(t \mid X) \exp\{-H_1(t \mid X) - H_2(T \mid X)\}}^{f_1(t \mid X)} = \lambda_1(t \mid X) \exp\{-\Lambda_1(t \mid X)\},$$

$$\exp\{-H_1(t \mid X)\} = 1 - \int_0^t \lambda_1(u \mid X) \exp\{-\Lambda_1(u \mid X) + H_2(u \mid X)\} \, \mathrm{d}u.$$

It then follows that, given choices of $\lambda_1(t \mid X)$ and $h_2(t \mid X)$, the implied cause-specific hazard for event 1 is given by

$$h_1(t \mid X) = \frac{\lambda_1(t \mid X) \exp\{-\Lambda_1(t \mid X) \ + H_2(t \mid X)\}}{1 - \int_0^t \lambda_1(u \mid X) \exp\{-\Lambda_1(u \mid X) \ + H_2(u \mid X)\} \, \mathrm{d}u}. \tag{6.2}$$

The above expression has the advantage of naturally ensuring that $\lambda_1(t \mid X) = h_1(t \mid X)$ at $t = 0$. However, depending on the choices of $\lambda_1(t \mid X)$ and $h_2(t \mid X)$, the implied $h_1(t \mid X)$ may become negative at certain timepoints. Specifically, this occurs when the integral in the denominator (corresponding to $1 - \exp\{-H_1(t \mid X)\}$, the 'net risk' for cause 1) in Equation 6.2 exceeds 1. Another way of looking at these potentially negative hazard values is to express the event-free survival $S(t \mid X)$ in terms of $\lambda_1(t \mid X)$ and $h_2(t \mid X)$, as

$$S(t \mid X) = \left[ 1 - \int_0^t \lambda_1(u \mid X) \exp\{-\Lambda_1(u \mid X) \ + H_2(u \mid X)\} \, \mathrm{d}u \right] \times \exp\{-H_2(t \mid X)\}. \tag{6.3}$$

Therefore, when the implied net risk of cause 1 exceeds one (generally as a result of excessively large cause-specific hazard for cause 2), the event-free survival itself become negative, meaning that $1 - S(t \mid X)$ (the TFP) exceeds 1.

To illustrate this DGM, we specify

$$\lambda_1(t \mid X) = \nu_1 e^{\kappa_1 t} \exp(\beta_1 X),$$

where $\kappa_1$ and $\nu_1$ respectively are the shape and rate parameters of a Gompertz baseline hazard. Fixing $\beta_1 = 0.5$, and choosing a negative shape $\kappa_1 = -2$, and rate $\nu_1 = 0.5$, we have that $P(D = 1 \mid X) = 1 - [1 - \{1 - \exp(\nu_1/\kappa_1)\}]^{\exp(\beta_1 X)}$. This asymptote of the Gompertz cumulative distribution function, which is less than 1 when the shape parameter is negative, has made it an attractive choice of distribution in work investigating direct parametric modelling of cumulative incidence functions (Jeong and Fine, 2006). For cause 2, we assume cause-specific proportional hazards as

$$h_2(t \mid X) = a_2 b_2 t^{a_2 - 1} \exp(\gamma_2 X),$$

where $a_2$ and $b_2$ are respectively the shape and rate parameters of a Weibull baseline hazard. We fix $\{a_2, b_2, \gamma_2\} = \{0.5, 1.25, 0.25\}$, and derive $h_1(t \mid X)$ using Equation 6.2. Figure 6.1 depicts the true (obtained with numerical integration) stacked cumulative incidence functions and cause-specific hazards (conditional on $X = 1$) for both causes, as well as the implied subdistribution hazard ratios. We see that this mechanism (when $X = 1$) is only properly defined prior to $t \approx 3.20$, after which $h_1(t \mid X = 1)$ is negative. The corresponding cumulative incidence functions demonstrate that beyond this timepoint, there is no probability space left to fill, as the TFP has already reached one. Additionally, panel C from Figure 6.1 emphasises that proportional subdistribution hazards do not hold for cause 2.

Making appropriate use of this approach for simulation purposes means paying attention to the fact that the choice $h_2(t \mid X)$ will have to respect the remaining probability space left over by the proportional subdistribution hazards structure assumed by cause 1. Practically speaking, this means specifying a $h_2(t \mid X)$ such that the implied event-free survival in Equation 6.3 does not become negative. One may also choose to set a maximum follow-up time, before which all hazards behave appropriately for all $X$ and the TFP is smaller than 1. In Figure 6.1, this could be achieved by setting a maximum follow-up time smaller or equal to 3.20 (or adjusting the parameters of the hazards function in order to allow a larger maximum follow-up time). An example use of this DGM is found in the work of Lambert *et al.* (2017), as part of a simulation study assessing the performance of a proposed flexible parametric approach for modelling the subdistribution hazard of one event. Both $\lambda_1(t \mid X)$ and $h_2(t \mid X)$ assumed proportional hazards with mixture Weibull baseline hazards, and the maximum (simulated) follow-up time was set to 5 years.

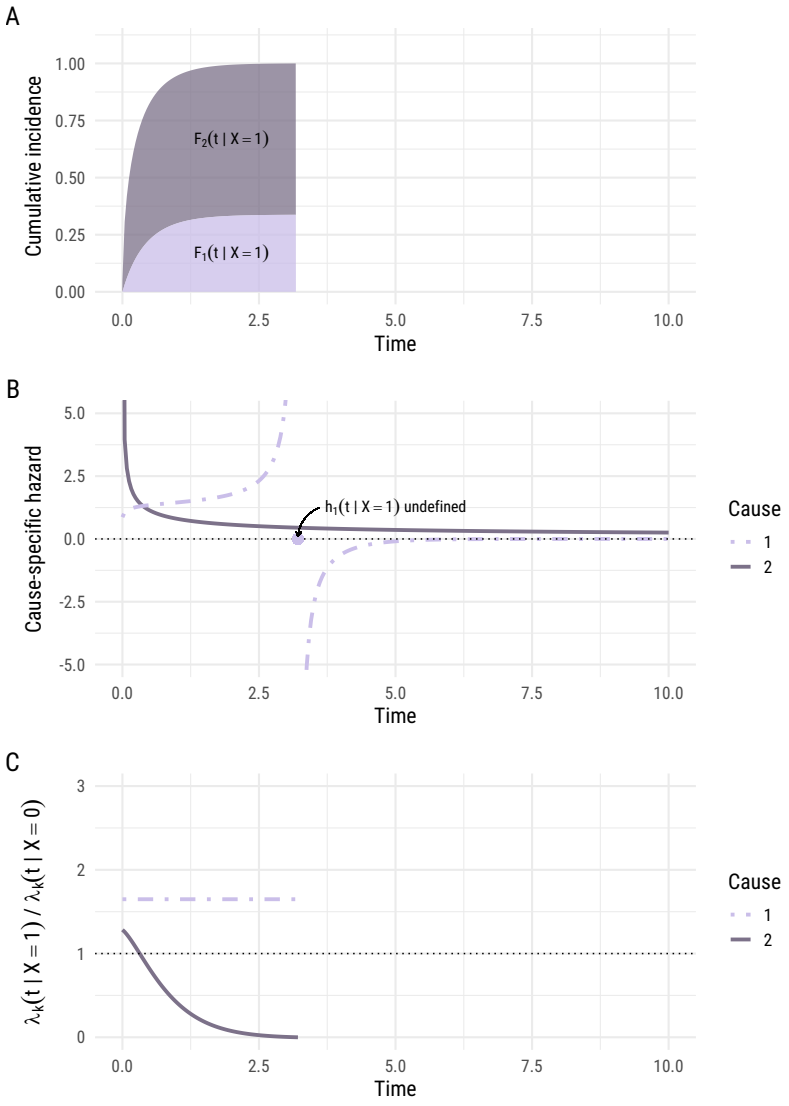


Figure 6.1: True stacked cumulative incidence functions (panel A) and cause-specific hazards (panel B) conditional on $X = 1$, and the subdistribution hazard ratios (panel C) for both causes, under DGM described in Section 6.3.1.1. This DGM assumes a Fine–Gray model for cause 1 with Gompertz baseline subdistribution hazard, and cause-specific Cox model for cause 2 with a Weibull baseline hazard.

#### 6.3.1.2 Specifying $\lambda_1(t \mid X)$ and $h_1(t \mid X)$

If we instead choose to specify both the subdistribution and cause-specific hazards for event 1, the cause-specific hazard for cause 2 can be derived by re-arranging Equation 6.1 as

$$h_2(t \mid X) = \lambda_1(t \mid X) - h_1(t \mid X) - \frac{\mathrm{d}}{\mathrm{d}t} \log\left\{ \frac{\lambda_1(t \mid X)}{h_1(t \mid X)} \right\}. \tag{6.4}$$

While a Gompertz baseline hazard could again be specified for $\lambda_1(t \mid X)$, it is important to specify $h_1(t \mid X)$ such that $h_1(t \mid X) = \lambda_1(t \mid X)$ at $t = 0$. From a simulator's point of view, it means paying attention to the fact that due to the form of the reduction factor (a time-dependent weight), proportionality generally cannot hold for both $\lambda_1(t \mid X)$ and $h_1(t \mid X)$ simultaneously. If the Fine–Gray model holds for $\lambda_1(t \mid X)$, the cause-specific hazard ratio $h_1(t \mid X = 1)/h_1(t \mid X = 0)$ should usually be time-dependent.

An exception to the above is found in the work of Saadati *et al.* (2018). There, a DGM is presented where $h_2(t \mid X)$ is chosen such that proportionality holds on both the subdistribution and cause-specific hazard scale for cause 1. The chosen $h_2(t \mid X)$ is based on setting the reduction factor to 1 for all covariate patterns and timepoints, which also implies that the covariate effects on the subdistribution and cause-specific hazard for event 1 need to be equal to each other.

Similarly to Equation 6.2, the implied $h_2(t \mid X)$ when specifying $\lambda_1(t \mid X)$ and $h_1(t \mid X)$ may also become negative. To understand an instance of where this can occur, note that we can write the all-cause cumulative hazard as a function of $\lambda_1(t \mid X)$ and $h_1(t \mid X)$ using Equation 6.1, as

$$-\log\{S(t \mid X)\} = \frac{\lambda_1(t \mid X) \exp\{-\Lambda_1(t \mid X)\}}{h_1(t \mid X)}.$$

As an example, we can use the same Gompertz parametrisation for $\lambda_1(t \mid X)$ as in the previous sub-section, namely $\{\kappa_1, \nu_1, \beta_1\} = \{-2, 0.5, 0.5\}$. Suppose now we also decide to use a Gompertz baseline hazard for $h_1(t \mid X)$, using the same parametrisation (i.e. base rate also equal to $\nu_1$ and $\beta_1 = \gamma_1$, ensuring $\lambda_1(t \mid X) = h_1(t \mid X)$ at $t = 0$), but instead setting the shape parameter to 2 (exponentially increasing). By solving $-\log\{S(t \mid X)\} - H_1(t \mid X) = 0$, one can find the timepoint at which $H_1(t \mid X)$ starts to exceed the all-cause cumulative hazard. Prior to this timepoint, $H_2(t \mid X)$ will be forced to decrease in order to maintain $-\log\{S(t \mid X)\} = H_1(t \mid X) + H_2(t \mid X)$, implying negative $h_2(t \mid X)$. Note also that this is not the fault of the Gompertz distribution: it is perfectly possible to simulate competing risks data with baseline cause-specific Gompertz hazards.

From a simulator's point of view, a DGM based on directly specified $\lambda_1(t \mid X)$ and $h_1(t \mid X)$ is rather tedious to implement given (a) the restriction that $\lambda_1(t \mid X) = h_1(t \mid X)$

at $t = 0$ for all $X$, (b) the (generally) time-dependent nature of the cause-specific hazard ratio for event 1. Even when $\beta_1 = \gamma_1$ (same covariate effects on cause-specific and subdistribution hazard of cause 1), which is unlikely to be the case in practice unless there are relatively few cause 2 failures, specifying an adequate $h_1(t \mid X)$ is not very flexible. As part of work on simulating proportional subdistribution hazard data with time-varying effects, Haller and Ulm (2014) provide an example of simulating from a DGM based on specifying both $h_1(t \mid X)$ and $\lambda_1(t \mid X)$. There, $h_1(t \mid X)$ is chosen to be time constant, with rate equal to $\lambda_1(t \mid X)$ at $t = 0$.

#### 6.3.1.3 Specifying $\lambda_1(t \mid X)$ and a model for the all-cause hazard

The reduction factor could also form the basis for a DGM if a model is specified for the all-cause hazard $\sum_{k=1}^{K} H_k(t \mid X) = -\log\{S(t \mid X)\}$, together with the Fine–Gray model for $\lambda_1(t \mid X)$. One can derive the implied $h_1(t \mid X)$ from Equation 6.1, and subtract it from the all-cause hazard to obtain $h_2(t \mid X)$. Proportional hazards on the all-cause scale however typically implies that the cause-specific hazards will be non-proportional. Note that this DGM requires that $-\log\{S(t \mid X)\} > \Lambda_1(t \mid X)$ at all timepoints and for all $X$. That is, that the all-cause cumulative hazard is always greater than the cumulative subdistribution hazard of event 1, otherwise the implied cause-specific hazard for cause 2 is forced to be negative. When simulating from this DGM, this could for example occur if the specified covariate effects differ substantially between the subdistribution hazard model for cause 1 and the all-cause model (e.g. pushing $\Lambda_1(t \mid X)$ above $-\log\{S(t \mid X)\}$ for $X = 1$). Nevertheless, as long as precautions are taken when specifying $-\log\{S(t \mid X)\}$, this DGM again represents a valid way to specify $f(T, D \mid X)$ such that proportional subdistribution hazards hold for cause 1. To the best of our knowledge, this mechanism has not been used in articles simulating proportional subdistribution hazards data.

### 6.3.2 Squeezing

Instead of specifying the various hazard functions, we can also work with the cumulative incidence functions directly. Recall that the Fine–Gray model for cause 1 can be expressed as

$$1 - F_1(t \mid X) = \{1 - F_{10}(t)\}^{\exp(\beta_1 X)}.$$

The idea is now to specify $F_{10}(t)$ directly. Since $F_k(\infty) = P(D = k) < 1$, we have to first pick some proper cumulative distribution $\tilde{F}_{10}(t)$ (e.g. exponential or Weibull cumulative distribution function, CDF) with $\lim_{t\to\infty} \tilde{F}_{10}(t) = 1$ and scale it down by a factor $0 < p < 1$ (since $p = 0$ or $p = 1$ would imply no competing risks). This leaves

$$1 - F_1(t \mid X) = \left[1 - p\{\tilde{F}_{10}(t)\}\right]^{\exp(\beta_1 X)}.$$

Note that $\lim_{t\to\infty} F_{10}(t) = p$, and $p_1(x) = P(D = 1 \mid X) = 1 - (1-p)^{\exp(\beta_1 X)}$. The probability of experiencing cause 2 therefore needs to be 'squeezed' into the remaining probability space $p_2(x) = P(D = 2 \mid X) = 1 - P(D = 1 \mid X) = (1-p)^{\exp(\beta_1 X)}$. Since $p_2(x)$ is determined by $p_1(x)$, it is guaranteed that $p_1(x) + p_2(x) = 1$. The second cumulative incidence function takes the form

$$P(T \leq t, D = 2 \mid X) = P(T \leq t \mid D = 2, X)P(D = 2 \mid X),$$

where $P(T \leq t \mid D = 2, X)$ can be chosen to be any standard CDF, which is then scaled down by $P(D = 2 \mid X)$. When simulating using this DGM, it is convenient to first generate the competing event indicator, and thereafter draw event times conditional on this indicator e.g. for event 1, drawing from $P(T \leq t \mid D = 1, X)$. For more details, see section 5.3.6 of Beyersmann *et al.* (2012). This DGM is arguably the most commonly used approach to simulate proportional subdistribution hazard data, as it ensures the TFP remains below or equal to 1. Multiple simulation studies, along with the original article proposing the Fine–Gray model, have simulated data in this way (Austin *et al.*, 2021; Bellach *et al.*, 2019; Fine and Gray, 1999; Saadati *et al.*, 2018).

To illustrate this mechanism, we use Weibull-type distribution functions and set

$$\tilde{F}_{10}(t) = 1 - \exp(-b_1 t^{a_1}),$$
$$P(T \leq t \mid D = 2, X) = 1 - \exp\{-b_2 t^{a_2} \exp(\beta_2^* X)\},$$

with $\{a_1, b_1, \beta_1, p\} = \{1.25, 1, 0.5, 0.2\}$ and $\{a_2, b_2, \beta_2^*\} = \{1.5, 1, 0.5\}$. $\beta_2^*$ is denoted as such since it is not a subdistribution log hazard ratio, but instead denotes the effect of $X$ on $P(T \leq t \mid D = 2, X)$. Figure 6.2 shows the baseline hazards and hazard ratios ($X = 1$ relative to $X = 0$) over time for the cause-specific and subdistribution hazards of both events. For this DGM, these functions are arguably more interesting to show, since they are only implicitly specified e.g. $h_1(t \mid X)$ is obtained by dividing $\mathrm{d}F_1(t \mid X)/\,\mathrm{d}t$ by $1 - F_1(t \mid X) - F_2(t \mid X)$. Note that in panels B and D (and also in panel C in Figure 6.1), no logarithmic transformation was applied to the y-axis (as would normally be the case for hazard ratio plots) due to the cause-specific and subdistribution hazard ratios for cause 2 going to zero as $t \to \infty$.

This DGM provides a clear picture on why one may choose to avoid Fine–Gray models for more than one cause. As other authors have similarly noted (Beyersmann *et al.*, 2012), a Fine–Gray model being specified for cause 1 effectively constrains the remaining probability space available to cause 2 to a maximum of $1 - P(D = 1 \mid X)$. Equivalently, this is a constraint on $\Lambda_2(t \mid X) = -\log\{1 - F_2(t \mid X)\}$, the cumulative subdistribution hazard for cause 2 conditional on $X$. In the case of a binary $X$, this translates to the coefficient of a Fine–Gray model for the competing cause needing to be *determined* by the Fine–Gray model for cause 1. Explictly, we can express the

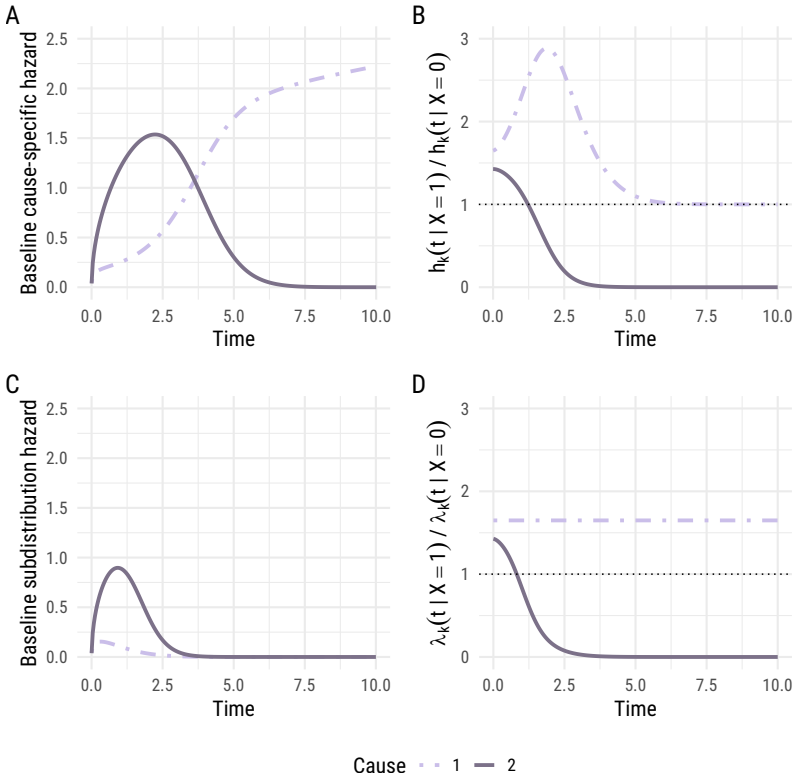


Figure 6.2: Baseline hazards (panels A and C) and hazard ratios $X = 1$ relative to $X = 0$ (panels B and D) over time for the cause-specific and subdistribution hazards of both events, based on the 'squeezing' DGM.

cumulative subdistribution hazard ratio for cause 2 as $t \to \infty$ for this DGM as

$$\begin{aligned}\lim_{t\to\infty} \frac{\Lambda_2(t \mid X=1)}{\Lambda_2(t \mid X=0)} &= \frac{-\log\{1-P(D=2 \mid X=1)\}}{-\log\{1-P(D=2 \mid X=0)\}}, \\ &= \frac{\log\{P(D=1 \mid X=1)\}}{\log\{P(D=1 \mid X=0)\}}, \\ &= \frac{\log\{1-(1-p)^{\exp(\beta_1)}\}}{\log(p)},\end{aligned}$$

by using $P(D=2 \mid X) = 1-\exp\{-\lim_{t\to\infty}\Lambda_2(t \mid X)\}$ and $P(D=2 \mid X) = 1-P(D=1 \mid X)$. Therefore, the cumulative subdistribution hazard ratio for cause 2 as $t \to \infty$ should be completely determined by $\beta_1$ and $p$. Since running a Fine–Gray model for the second cause does not account for this restriction, it is typically misspecified, leading to issues such as the TFP exceeding 1. As shown in Figure 6.2, the extent of this misspecification can be alarming: the true subdistribution hazard ratio for the competing cause is severely time-dependent, for which the time-averaged subdistribution hazard ratio (see Grambauer *et al.*, 2010) obtained from a Fine–Gray model is perhaps a suboptimal summary.

### 6.3.3 Two Fine–Gray models

The previous sub-section may suggest that it is impossible for proportional subdistribution hazards to hold for more than one competing event. In fact, if we choose to also directly specify the cumulative incidence for cause 2 in the same style as in the 'squeezing' mechanism (instead of having it determined by cause 1), we can achieve proportional subdistribution hazards for both events. Suppose we have

$$\begin{aligned}F_1(t \mid X) &= 1 - \left[1 - p_{10}\{\tilde{F}_{10}(t)\}\right]^{\exp(\beta_1 X)}, \\ F_2(t \mid X) &= 1 - \left[1 - p_{20}\{\tilde{F}_{20}(t)\}\right]^{\exp(\beta_2 X)}.\end{aligned}$$

If we let $P(D=k \mid X) = 1-(1-p_{k0})^{\exp(\beta_k X)} = p_k(x)$, then we can write that as $t \to \infty$, $\text{TFP}(x) = p_1(x) + p_2(x)$. To determine the event indicator when simulating, we would draw $u \sim \mathcal{U}(0,1)$, and set

$$D = \begin{cases} 1, & \text{if } u \le p_1(x), \\ 2, & \text{if } p_1(x) < u \le p_2(x). \end{cases}$$

The event times can then be drawn as in the preceding subsection, by inverting $P(T \le t \mid D=k, X)$. Those with $u > p_1(x) + p_2(x)$ are technically considered 'cured', that is, never at risk of any of the competing events. To illustrate this DGM, we set

$$\begin{aligned}\tilde{F}_{10}(t) &= 1 - \exp(-b_1 t^{a_1}), \\ \tilde{F}_{20}(t) &= 1 - \exp(-b_2 t^{a_2}),\end{aligned}$$

with $\{a_1, b_1, \beta_1, p_{10}\} = \{0.75, 1, 0.5, 0.25\}$ and $\{a_2, b_2, \beta_2, p_{20}\} = \{0.75, 1, 0.5, 0.5\}$. Figure 6.3 shows the baseline cumulative incidence functions, and those conditional on $X = 1$. The chosen baseline cumulative incidence functions and subdistribution hazard ratios result in the TFP exceeding 1 from $t \approx 3$ when $X = 1$. While this mechanism can be flexible when baseline hazard rates are small and covariate effects are modest, it by design cannot guarantee a TFP $\leq 1$ for any timepoint. Indeed, if $X$ is not binary, and instead continuous and unbounded, the TFP is guaranteed to exceed 1 (though for small $t$, $X$ will need to be very extreme for this to occur) (Austin *et al.*, 2021).

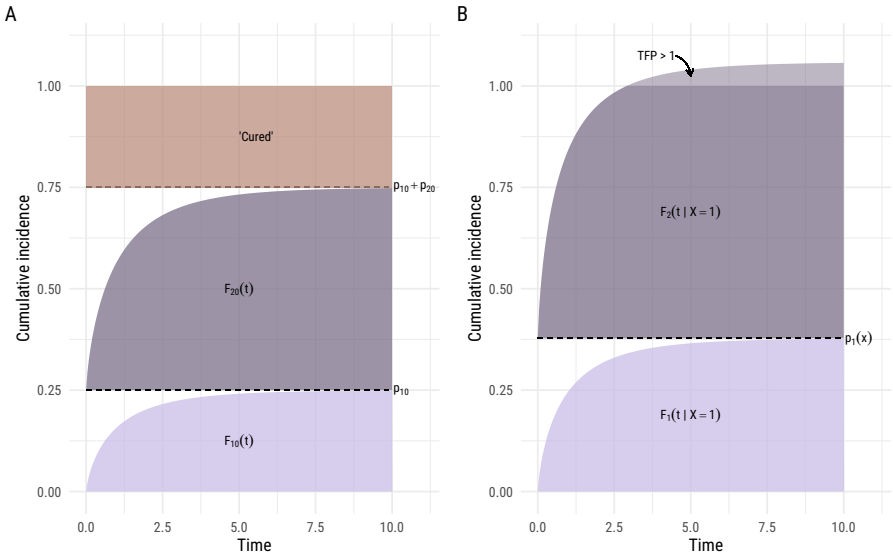


Figure 6.3: Stacked cumulative incidence functions conditional on $X = 0$ (baseline, panel A) and those conditional on $X = 1$ (panel B), for a DGM in which proportional subdistribution hazards can hold for both causes.

Two other useful implementations of (variants of) this DGM are found in the simulations of Mao and Lin (2017), and Mozumder *et al.* (2018)—both investigating the performance of different approaches (semiparametric and flexible parametric, respectively) for direct modelling of the cumulative incidence functions. Mao and Lin directly specify Gompertz cumulative subdistribution hazards $\Lambda_k(t \mid X)$ for both events, and thereafter invert $P(T \leq t \mid D = k, X)$ analogously using $P(T \leq t, D = k \mid X) = 1 - \exp\{-\Lambda_k(t \mid X)\}$ and its limit as $t \to \infty$. Mozumbder et al. specify mixture Weibull baseline subdistribution hazards for both events, and then derive the implied cause-specific hazards using the reduction factor, and use these for simulating.

An important point is that both of the above approaches specify a maximum follow-up

time in their simulations. Indeed, as pointed out by Latouche *et al.* (2013), 'it is possible that such models may hold over restricted time ranges, which has practical implications for studies with limited longitudinal follow-up', with 'such models' referring to multiple proportional subdistribution hazard models. Note also that specifying a maximum follow-up time $\tau$ means that the $1 - F_1(\tau \mid X) - F_2(\tau \mid X)$ proportion of individuals that did not experience the event by $\tau$ are simply considered as censored at $\tau$.

## 6.4 Discussion

In this work, we have outlined various ways of specifying a joint density $f(T, D \mid X)$ in which a Fine–Gray model for cause 1 is correctly specified. The goal was to outline the possible assumptions that can be made regarding the (cumulative incidence of) cause 2, given that a Fine–Gray model is correctly specified for cause 1. From a simulator's perspective, all DGMs are fundamentally aiming to do the same thing, which is to fill up the probability space leftover from the assumed Fine–Gray model for cause 1 i.e. $1 - F_1(t \mid X)$. The 'squeezing' DGM does this in the most explicit way, by making sure cause 2 fills up all of the remaining space, in turn ensuring that the TFP remains below or equal to 1 for all $X$ and at all timepoints. This makes it the approach with the fewest restrictions when simulating data for which proportional subdistribution hazards holds for one cause only. Using the reduction factor in contrast can be quite inflexible, possibly producing negative cause-specific hazard values without great care in choices of parameter values and distributions. Note also that the described DGMs can be readily adapted in order to simulate under link functions other than the complementary log-log, discussed for example by Gerds *et al.* (2012). More generally, there is no DGM for which proportional subdistribution hazards holds simultaneously for both causes, unless one assumes finite follow-up and a bounded covariate space.

Suppose that we have a dataset where proportional subdistribution hazards perfectly hold for both causes up to some maximum follow-up time, e.g. a simulated dataset, for which we know the DGM. Here, fitting a Fine–Gray model for each cause in turn (i.e. using the exact models the data were generated from) can still lead to the TFP exceeding 1. This can occur more generally in finite samples, since the Fine–Gray models for each cause are estimated separately. As a result, alternative approaches have been developed (based on the complete data likelihood) to facilitate simultaneous modelling of all cumulative incidence functions, while incorporating this TFP constraint (Mao and Lin, 2017; Shi *et al.*, 2013) The parametric approach suggested by Shi *et al.* (2013) actually does so by explicitly incorporating the 'squeezing' of the second cause (i.e. that covariates effects on cause 2 depend on the asymptote of the cumulative incidence function for cause 1) into the likelihood. Note also that modelling the cause-specific hazards and combining these to obtain predicted cumulative incidence functions ensures that the TFP remains below 1 for any timepoint and covariate

combination (Austin *et al.*, 2021). The cause-specific approach also extends naturally (i.e. without TFP issues or tedious algebra) to settings with more than two competing events.

In applied settings, proportionality assumptions for either event on any of cause-specific and subdistribution hazard scales, such as those made in the outlined DGMs, will never hold exactly. Using alloSCT data, Gerds *et al.* (2012) compared the performance of cause-specific hazard and Fine–Gray models (as well as other transformation models) for predicting competing events relapse and non-relapse mortality. Both predictive accuracy (based on cross-validated Brier score) and individual predictions were similar for both approaches. Wolbers *et al.* (2009) reported that the cause-specific and Fine–Gray approaches showed comparable calibration when predicting coronary heart disease, though they did not consider calibration of the competing event. Kantidakis *et al.* (2023) also reported similar predictive performance of the cause-specific and Fine–Gray approaches when applied on a dataset of patients with extremity soft-tissue sarcoma. This was as part of a broader comparison with machine learning techniques, with the goal of predicting competing events disease progression and death.

One explanation for this comparable performance is that both models make use of their non-parametric baseline hazard to compensate, to some extent, for misspecified (i.e. non-proportional) covariate effects—allowing them to still predict the cumulative incidence functions fairly accurately (noted in Shi *et al.*, 2013 for Fine–Gray models). Differences in performance may also be modest in settings with a shorter follow-up period or when there is heavy censoring: the time-dependent weight relating the cause-specific and subdistribution hazards (the reduction factor) is less influential earlier in time. Nevertheless, there are situations in which one would expect (multiple) Fine–Gray models to underperform with respect to cause-specific hazard approaches. First, when the estimated TFP exceeds 1 in a non-negligible proportion of patients, as in the example by Austin *et al.* (2021), the predictions for one or more of the competing events must by definition be partly miscalibrated (due to risk overestimation). Second, when the true cumulative incidence curves for an event (e.g. relapse probabilities for two conditioning regimens) cross each other, the predicted curves by the Fine–Gray model will not be allowed to cross. Poythress *et al.* (2020) therefore suggest to compare predicted curves to their non-parametric counterparts as a possible diagnostic, and note that cause-specific hazards models are much more suited to capturing more complex cumulative incidence function shapes. For example, the cause-specific hazards could be modelled using flexible parametric approaches, which naturally accommodate time-varying effects (Hinchliffe and Lambert, 2013; Kipourou *et al.*, 2019).

In conclusion, the described DGMs outline the variety of ways in which proportional subdistribution hazards could hold for at least one of two event types. In terms of cumulative incidence prediction for both causes, we argue that cause-specific hazard models should be preferred over multiple Fine–Gray models, as they (a) by design ensure that the TFP does not exceed 1, (b) are able to capture complex shapes for the

cumulative incidence functions (although in the Fine–Gray context, one could technically include time by covariate interactions), and (c) additionally provide inference on the cause-specific hazards, which are the 'natural building blocks for competing risks modeling' (Saadati *et al.*, 2018). The {riskRegression} R package in particular provides useful functions for developing and validating prediction models based on cause-specific hazards (Ozenne *et al.*, 2017). While using a Fine–Gray model for one cause only may still be defendable (e.g. for prediction purposes, when other causes are truly a nuisance), it does go against the holistic approach to competing risks analyses described in the introduction, where all causes should ideally be studied together. Cause-specific hazard models, which are often misunderstood to be less suitable for prediction compared to Fine–Gray models (see e.g. D'Amico *et al.*, 2018), should perhaps also be the preferred approach also in settings where predicting a single cause is of interest. When the main goal is simultaneous inference on the cumulative incidence functions, the proposed semiparametric approach by Mao and Lin (2017) is a promising alternative to multiple Fine–Gray models, as it (a) provides more efficient inference, (b) allows the use of different link functions (e.g. accommodates non-proportional hazards, and allows odds ratio interpretation of parameters) for different events, (c) does not need to model the censoring distribution. For inference at specific timepoints, one may also consider to specify models using pseudovalues as the outcome variable (Klein and Andersen, 2005).

## Supplementary materials

The R code to reproduce the figures for the described DGMs is available at https://github.com/survival-lumc/FineGrayDGM.