



Universiteit  
Leiden

The Netherlands

## **The built-in selection bias of hazard ratios formalized using structural causal models**

Post, R.A.J.; Heuvel, E.R. van den; Putter, H.

### **Citation**

Post, R. A. J., Heuvel, E. R. van den, & Putter, H. (2024). The built-in selection bias of hazard ratios formalized using structural causal models. *Lifetime Data Analysis*, 30(2), 404-438.  
doi:10.1007/s10985-024-09617-y

Version: Publisher's Version

License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)

Downloaded from: <https://hdl.handle.net/1887/4239162>

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



# The built-in selection bias of hazard ratios formalized using structural causal models

Richard A. J. Post<sup>1</sup> · Edwin R. van den Heuvel<sup>1</sup> · Hein Putter<sup>2,3</sup>

Received: 29 October 2022 / Accepted: 10 January 2024  
© The Author(s) 2024

## Abstract

It is known that the hazard ratio lacks a useful causal interpretation. Even for data from a randomized controlled trial, the hazard ratio suffers from so-called built-in selection bias as, over time, the individuals at risk among the exposed and unexposed are no longer exchangeable. In this paper, we formalize how the expectation of the observed hazard ratio evolves and deviates from the causal effect of interest in the presence of heterogeneity of the hazard rate of unexposed individuals (frailty) and heterogeneity in effect (individual modification). For the case of effect heterogeneity, we define the causal hazard ratio. We show that the expected observed hazard ratio equals the ratio of expectations of the latent variables (frailty and modifier) conditionally on survival in the world with and without exposure, respectively. Examples with gamma, inverse Gaussian and compound Poisson distributed frailty and categorical (harming, beneficial or neutral) distributed effect modifiers are presented for illustration. This set of examples shows that an observed hazard ratio with a particular value can arise for all values of the causal hazard ratio. Therefore, the hazard ratio cannot be used as a measure of the causal effect without making untestable assumptions, stressing the importance of using more appropriate estimands, such as contrasts of the survival probabilities.

**Keywords** Causal inference · Causal hazard ratio · Selection bias · Individual effect modification · Heterogeneity of treatment effects

**Mathematics Subject Classification** 62D20 · 62N02

---

✉ Richard A. J. Post  
[r.a.j.post@tue.nl](mailto:r.a.j.post@tue.nl)

<sup>1</sup> Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands

<sup>2</sup> Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands

<sup>3</sup> Mathematical Institute, Leiden University, Leiden, The Netherlands

# 1 Introduction

When interested in time-to-event outcomes, ideally, one would like to know the hazard rates of an individual in the worlds with and without exposure. It is then standard practice to fit the observed hazard rates with a (time-invariant) Cox model (Cox 1972) to estimate the ratio of the expected hazard rates in both worlds. A decade ago, Hernán (2010) raised awareness that hazard ratios estimated from a randomized controlled trial (RCT) are unsuitable for causal inference. Firstly, the average hazard ratio could be uninformative as there will typically be time-varying hazard ratios. More importantly, even when period-specific hazard ratios are estimated, these can vary solely due to the loss of randomization over time by conditioning on survivors. The exposure assignment and risk factors become dependent when conditioning on individuals that survived  $t$ , i.e. survival time  $T \geq t$ , even if these risk factors are unrelated to the exposure (Aalen et al. 2015). As a result, effect measures based on hazard rates can suffer from non-collapsibility (Martinussen and Vansteelandt 2013; Aalen et al. 2015; Sjölander et al. 2016; Daniel et al. 2021).

In practice, the ratio of (partly) marginalized hazards, is estimated, that by the non-collapsibility, deviates from the conditional (causal) hazard ratio. This contrast is referred to as the built-in selection bias of hazard ratios as the bias results from conditioning on prior survival (Hernán 2010; Aalen et al. 2015; Sjölander et al. 2016; Stensrud et al. 2018; Young et al. 2020; Martinussen et al. 2020). This bias should not be confused with confounding bias that is absent when using data from an RCT (Didelez and Stensrud 2021). For exposure assignment  $A$ , and the potential survival time when the exposure is intervened on to  $a$  denoted by  $T^a$ , the expected observed hazard ratio from an RCT satisfies

$$\frac{\lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T \in [t, t+h] \mid T \geq t, A=a)}{\lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T \in [t, t+h] \mid T \geq t, A=0)} = \frac{\lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T^a \in [t, t+h] \mid T^a \geq t)}{\lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T^0 \in [t, t+h] \mid T^0 \geq t)} \quad (1)$$

(De Neve and Gerds 2020; Martinussen et al. 2020). The expected observed hazard ratio thus equals the ratio of hazard rates at time  $t$  for the potential outcomes of individuals from different populations; those for which  $T^a \geq t$  and those for which  $T^0 \geq t$ . As indicated before, these populations will typically not be exchangeable in other risk factors, implying that an effect found cannot be (solely) assigned to the exposure. The effect does thus not reflect how the hazard rate of an individual is affected by exposure. Only for cause-effect relations such that (1) is time-invariant, the estimand can be interpreted as  $\frac{\log(\mathbb{P}(T^a \geq t))}{\log(\mathbb{P}(T^0 \geq t))}$ . It has been recommended to use better interpretable estimands such as contrasts of quantiles, the restricted mean survival or survival probabilities of the potential outcomes respectively (Hernán 2010; Stensrud et al. 2018; Bartlett et al. 2020; Young et al. 2020), or the probabilistic index derived from the latter (De Neve and Gerds 2020). Alternatively, one can avoid interpretation issues by using accelerated failure time models (Hernán et al. 2005; Hernán 2010) or additive hazard models (Aalen et al. 2015; Martinussen et al. 2020).

Nevertheless, particularly in medical sciences, observed hazard ratios are still commonly presented by practitioners. In this paper, we formalize how the expected

observed hazard ratio deviates from the causal hazard ratio (as defined in Sect. 2), and thus quantify the built-in selection bias. To do so, we first present a general parameterization of cause-effect relations for time-to-event outcomes using a structural causal model in Sect. 2 and explain the loss of randomization over time by conditioning on survivors. We will limit ourselves to systems where the causal effect is appropriately described by a causal hazard ratio. The quantitative examples in which the hazard under no exposure varies among individuals, i.e. frailty, as presented in the literature (Aalen et al. 2015; Stensrud et al. 2017; Balan and Putter 2020) do fit in our framework, and we will formalize results for these examples. Additionally, we will extend these examples with causal effect heterogeneity, i.e. the causal effect on the hazard rate might vary between individuals (Stensrud et al. 2017). In Sect. 3, we define the causal hazard ratio and explain why this estimand is not identifiable from data. Practitioners instead compute the hazard ratio from data, and in Sect. 4 (which comprises the bulk of this paper) we derive what estimand is estimated: the survivor marginalized causal hazard ratio. This estimand describes a combination of the causal effect of interest and the difference in latent frailty- and modifying-features distribution between survivors in the exposed and unexposed universe. We point out exactly how this estimand deviates from the causal hazard ratio in the presence of frailty and effect heterogeneity. To develop understanding of how selection of frailty- and modifying-features affect the value of the estimand, we presented examples for systems in the presence of frailty (Sect. 4.1), effect heterogeneity (Sect. 4.2) or both (Sect. 4.3). In Sect. 5, we shortly discuss the implications of our results for the traditional Cox estimand. Finally, we present some concluding remarks in Sect. 6.

## 2 Notation

In this paper, probability distributions of factual and counterfactual outcomes are defined in terms of the potential outcome framework (Neyman 1990; Rubin 1974). Let  $T_i$  and  $A_i$  represent the (factual) stochastic outcome and exposure assignment level of individual  $i$ . Let  $T_i^a$  equal the potential outcome of individual  $i$  under an intervention of level  $a$  (counterfactual when  $A_i \neq a$ ). For those more familiar with the do-calculus,  $T^a$  is equivalent to  $T \mid do(A=a)$  as e.g. derived in (Pearl 2009, Equation 40) and (Bongers et al. 2021, Definition 8.6). Throughout this paper, we will assume causal consistency: if  $A_i = a$ , then  $T_i^a = T_i^{A_i} = T_i$ , implying that potential outcomes are independent of the assigned exposure levels of other individuals.

The hazard rate of  $T^a$  can vary among the individuals in the population of interest. We will parameterize this heterogeneity for hazards of  $T^0$  using a random variable  $U_{0i}$  that represents the frailty of individual  $i$  (see for example (Aalen et al. 2008, Chapter 6) or Balan and Putter (2020)). There can also be (relative) effect heterogeneity that we parameterize using the random variable  $U_{1i}$ , giving rise to an individual-specific hazard ratio. The hazard of the potential outcome  $T_i^a$  can be parameterized with a function that depends on  $U_{0i}$ ,  $U_{1i}$  and  $a$ . We describe cause-effect relations with a structural causal model (SCM) which is commonly used in the causal graphical literature, see e.g. (Pearl 2009, Chapter 1.4) and (Peters et al. 2018, Chapter 6), to

model observations. Instead, we include details on individual effect modifier  $U_1$  as well as the latent common cause of the outcomes  $U_0$ , to describe all the potential outcomes of an individual jointly. A SCM as presented in this paper is therefore a union of the SCM for observations ( $A=a$ ), and the so-called intervened SCMs for all possible  $do(A=a)$ . SCMs have been used before to describe hazards of potential outcomes (Hernán et al. 2000, 2001, 2005). Formulation of hazard rates of potential outcomes presented in the literature, e.g. by Aalen et al. (2015) and by Stensrud et al. (2017), naturally fit in this parameterization. However, as mentioned before, the dependence of  $T^a$  and  $T^0$  beyond shared frailty is typically not specified. The SCM consists of a joint probability distribution of  $(N_A, U_0, U_1, N_T)$  and a collection of structural assignments  $(f_A, f_\lambda)$  such that

$$\begin{aligned} A_i &:= f_A(N_{Ai}) \\ \lambda_i^a(t) &:= f_\lambda(t, U_{0i}, U_{1i}, a) \\ T_i^a &:= \min\{t>0: \exp(-\Lambda_i^a(t)) \leq N_{Ti}\}, \end{aligned} \quad (2)$$

where  $\forall t: f_\lambda(t, U_{0i}, U_{1i}, 0) \perp U_{1i} \mid U_{0i}$  (i.e. when  $a = 0$ ,  $f_\lambda$  does not contain  $U_1$ ),  $f_A$  is the inverse cumulative distribution function of  $A$ ,  $N_{Ai}, N_{Ti} \sim \text{Uni}[0, 1]$ ,  $\Lambda_i^a(t) = \int_0^t \lambda_i^a(s) ds$ , and  $U_{0i}, U_{1i} \perp N_{Ti}$  so that  $\lambda_i^a(t)$  equals the hazard rate of the potential outcome given  $U_0$  and  $U_1$ .

Note that the data generating mechanism is described by this SCM as  $T_i^{A_i} = T_i$ . If in SCM (2),

$$N_A \not\perp (U_0, U_1, N_T), \quad (3)$$

then there exists confounding as the distributions of  $(U_0, U_1, N_T)$  are not exchangeable between exposed and non-exposed individuals. However, in this work we focus on the distribution of data observed from a properly executed RCT, where by the randomization  $N_A \perp (U_0, U_1, N_T)$  so that there is no confounding. Note that  $\lambda_i^a(t)$  equals the hazard of the potential outcome of individual  $i$  under exposure  $a$ , i.e.

$$\lambda_i^a(t) = \lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T_i^a \in [t, t+h) \mid T_i^a \geq t, U_{0i}, U_{1i}), \quad (4)$$

and is thus a random variable when we consider an arbitrary individual. In this parameterization,  $U_0$  results in heterogeneity of the hazard under no exposure between individuals, and the presence of  $U_1$  results in heterogeneity of the effect of the exposure on the hazard between individuals. The SCM could be re-parameterized by including more details, e.g. measured risk factors, so that part of the unmeasured heterogeneity can be explained.

To understand how the so-called built-in selection bias is introduced, realize that  $T_i^a$  depends on  $a$  and the random variables  $U_{0i}, U_{1i}, N_{Ti}$  only, i.e.  $T_i^a := \min\{t>0: \exp(-\int_0^t f_\lambda(s, U_{0i}, U_{1i}, a) ds) \leq N_{Ti}\} = g(U_{0i}, U_{1i}, N_{Ti}, a)$ , for some function  $g$ . In an RCT,  $N_{Ai} \perp (U_{0i}, U_{1i}, N_{Ti})$ , so

$$T_i^a \perp A_i. \quad (5)$$

However, this independence might not hold conditionally on survival at time  $t$  since  $T_i := g(U_{0i}, U_{1i}, N_{Ti}, A_i)$ , so that  $A_i | T_i \geq t$  can inform on  $(U_{0i}, U_{1i}, N_{Ti})$  and thus on  $T_i^a$ , then

$$T_i^a \not\perp\!\!\!\perp A_i \mid T_i \geq t. \quad (6)$$

On the contrary,

$$T_i^a \perp\!\!\!\perp A_i \mid T_i^a \geq t, \quad (7)$$

as  $T_i^a := g(U_{0i}, U_{1i}, N_{Ti}, a)$ ,  $A_i | T_i^a \geq t$  does not inform on  $(U_{0i}, U_{1i}, N_{Ti})$ . In the literature, the dependence in (6) is often implicitly derived by recognizing that  $\{T \geq t\}$  is a collider that can thus open a back-door path between  $A$  and  $T^a$  (Aalen et al. 2015; Sjölander et al. 2016). The bias that results from conditioning on this collider is referred to as the built-in selection bias of the hazard ratio (Hernán 2010). This complicates causal inference which requires that the distribution of potential outcomes can be expressed in terms of the observed distribution.

In SCM (2), we did not restrict the distribution of  $U_0$  and  $U_1$  and only restricted  $f_\lambda$  and  $f_A$  to be properly defined hazard and inverse cumulative distribution functions respectively, so that the structural model is very general. It is important to realize that a SCM cannot be validated with data as it describes potential outcomes from different universes. For each individual the outcome can only be observed in one of the universes, and only the fit of the distribution of the outcomes in the factual world can be verified. In this work we focus on settings where the causal effect can be accurately described with a causal hazard ratio, which is defined in the next section. This will be the case when in SCM (2),

$$f_\lambda(t, U_{0i}, U_{1i}, a) = f_0(t, U_{0i})f_1(t, U_{1i}, a) \text{ and } f_1(t, U_{1i}, 0) = 1.$$

In the remainder of this manuscript we will restrict ourselves to cause-effect relations that meet this restriction.

### 3 The causal hazard ratio

If  $f_\lambda(t, U_{0i}, U_{1i}, a) = f_0(t, U_{0i})f_1(t, U_{1i}, a)$  and  $f_1(t, U_{1i}, 0) = 1$ , then the individual causal effect is described by  $f_1(t, U_{1i}, a)$ . The latter equals the ratio at time  $t$  of the hazard of an individual's potential outcome when exposed to level  $a$  and when not exposed, i.e.  $\frac{\lambda_i^a(t)}{\lambda_i^0(t)}$ . In the case of homogeneous effects,  $f_1(t, U_{1i}, a) = f_1(t, a)$  is equal for all individuals. In the case of heterogeneity of effects,  $f_1(t, U_{1i}, a)$  is the individual multiplicative causal effect. From a public health perspective, the ratio of the expected hazard rates in the world where everyone is exposed to  $a$  and in the world where all individuals are not exposed is of interest. This causal hazard ratio (CHR) of interest can be obtained as the ratio of the marginalized (over  $U_0$  and  $U_1$ ) conditional hazard rates in both worlds as presented in Definition 1.

**Definition 1 Causal hazard ratio** The causal hazard ratio (CHR) for cause-effect relations that can be parameterized with SCM 2 equals

$$\frac{\mathbb{E}[\lambda_i^a(t)]}{\mathbb{E}[\lambda_i^0(t)]} = \frac{\int \lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T^a \in [t, t+h] \mid T^a \geq t, U_0, U_1) dF_{U_0, U_1}}{\int \lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T^0 \in [t, t+h] \mid T^0 \geq t, U_0) dF_{U_0}}, \quad (8)$$

where we abbreviate the Lebesgue-Stieltjes integral of a function  $g$  with respect to probability law  $F_X$ , i.e.  $\int g(x) dF_X(x)$ , as  $\int g(X) dF_X$ .

When the parameterization of the cause-effect relations as SCM (2) would be known, the CHR can be expressed in terms of the distribution of the data generating mechanism as presented in Theorem 1.

**Theorem 1** *If the cause-effect relations of interest can be parameterized with SCM (2), and  $N_A \perp\!\!\!\perp U_0, U_1, N_T$  (no confounding), then*

$$\frac{\mathbb{E}[\lambda_i^a(t)]}{\mathbb{E}[\lambda_i^0(t)]} = \frac{\int \lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T \in [t, t+h] \mid T \geq t, U_0, U_1, A=a) dF_{U_0, U_1}}{\int \lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T \in [t, t+h] \mid T \geq t, U_0, A=0) dF_{U_0}}.$$

For an example, consider the commonly used frailty model where effect heterogeneity is absent, i.e.

$$\lambda_i^a(t) = U_{0i} \lambda_0(t) f_1(t, a).$$

The CHR equals the multiplicative effect that does not differ among individuals and equals  $f_1(t, a)$ . By applying Theorem 1, this CHR is indeed derived to equal

$$\frac{\int \lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T \in [t, t+h] \mid T \geq t, U_0, A=a) dF_{U_0}}{\int \lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T \in [t, t+h] \mid T \geq t, U_0, A=0) dF_{U_0}} = \frac{\lambda_0(t) \mathbb{E}[U_0] f_1(t, a)}{\lambda_0(t) \mathbb{E}[U_0]}.$$

It is important to note that  $f_1(t, a)$  deviates from the expected observed hazard ratio equal to

$$\frac{\lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T \in [t, t+h] \mid T \geq t, A=a)}{\lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T \in [t, t+h] \mid T \geq t, A=0)} = \frac{\lambda_0(t) \mathbb{E}[U_0 \mid T \geq t, A=a] f_1(t, a)}{\lambda_0(t) \mathbb{E}[U_0 \mid T \geq t, A=0]},$$

as we will elaborate on in Sect. 4.

In summary, it became clear that to derive the CHR from data, inference on the distribution of the latent frailty  $U_0$  and effect modifier  $U_1$  must be made. When their distributions are known, inference can be drawn from observed data, even when the parameters of the distributions are unknown. Software available to estimate frailty parameters are described by Balan and Putter (2020), and such methods could also be adapted to estimate the latent modifier distribution. However, in practice, the distributions of these latent variables are unknown. Even in the case without causal effect heterogeneity, it is impossible to distinguish the presence of frailty from a time-dependent causal effect (Balan and Putter 2020, Section 2.5). More precisely, different combinations of (varying) effect sizes and frailty distributions give rise to

the same marginal distribution. The same applies to combinations that also involve effect modifiers. In the case of clustered survival data (e.g. family data (Valberg et al. 2018)), at least theoretically, the shared frailty could be distinguished from violation of proportional hazards (Balan and Putter 2020). Reasoning along the same lines, individual frailty and marginal time-varying effects could only be derived from effect heterogeneity in the case of recurrent events with stationary distributions.

#### 4 Survivor marginalized causal hazard ratio

In Theorem 1, the actual CHR (see Definition 1) has been expressed in terms of the distributions of the observed data, and we concluded that these are not identifiable without making untestable assumptions on the distribution of  $(U_0, U_1)$ . Instead, practitioners often compute the hazard ratio from data, which expectation we refer to as the observed hazard ratio (OHR) and equals

$$\text{OHR}(t) = \frac{\lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T \in [t, t+h) \mid T \geq t, A=a)}{\lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T \in [t, t+h) \mid T \geq t, A=0)}. \quad (9)$$

To be precise, at time  $t$  the hazard rate can only be observed for non-censored individuals at that time ( $C(t) = 0$ ). However, in this work we will assume independent censoring, so that  $\mathbb{P}(T \mid T \geq t, A=a)$  is equal to  $\mathbb{P}(T \mid T \geq t, A=a, C(t)=0)$ .

To compare the OHR to the CHR that quantifies the causal effect of interest, the OHR should be expressed in terms of potential outcomes. For data from an RCT, by independence (7) and causal consistency, the OHR equals the survivor marginalized causal hazard ratio (SMCHR), i.e.

$$\text{SMCHR}(t) = \frac{\lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T^a \in [t, t+h) \mid T^a \geq t)}{\lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T^0 \in [t, t+h) \mid T^0 \geq t)}. \quad (10)$$

This SMCHR should not be confused with the ‘marginal causal hazard ratio’ defined by Martinussen et al. (2020) as

$$\frac{\lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T^a \in [t, t+h) \mid T^a \geq t, T^0 \geq t)}{\lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T^0 \in [t, t+h) \mid T^0 \geq t, T^a \geq t)}, \quad (11)$$

that could also be named the cross-world survivor marginalized causal hazard ratio and is not considered in this work.

We will study how the SMCHR (and thus the OHR from an RCT) differs from the CHR over time. By the law of total probability, the SMCHR in (10) equals

$$\frac{\lim_{h \rightarrow 0} \int h^{-1} \mathbb{P}(T^a \in [t, t+h) \mid T^a \geq t, U_0, U_1) dF_{U_0, U_1 \mid T^a \geq t}}{\lim_{h \rightarrow 0} \int h^{-1} \mathbb{P}(T^0 \in [t, t+h) \mid T^0 \geq t, U_0) dF_{U_0 \mid T^0 \geq t}}. \quad (12)$$

As the integration in the result of Theorem 1 is with respect to the population distribution of  $U_0$  and  $U_1$ , instead of those individuals for which  $T^a \geq t$  or  $T^0 \geq t$ , the SMCHR



deviates from the CHR, resulting in the built-in selection bias of the hazard (Hernán 2010; Aalen et al. 2015; Stensrud et al. 2018).

The problem induced for estimation of the CHR thus results from inference on a different estimand; the combined effect of the exposure of interest and the difference in latent frailty (and effect modification) distribution. To formalize how (10) deviates from the CHR that equals  $\frac{\mathbb{E}[f_0(t, U_0)f_1(t, U_1, a)]}{\mathbb{E}[f_0(t, U_0)]}$ , we focus on hazard functions that satisfy Condition 1 and do thus not have an infinite discontinuity.

### Condition 1 *Hazard without infinite discontinuity*

$$\forall t > 0: \exists \tilde{h} > 0 \text{ such that } \forall h^* \in (0, \tilde{h}): \mathbb{E}[f_0(t+h^*, U_0)f_1(t+h^*, U_1, a) \mid T^a \geq t] < \infty$$

The value of the SMCHR at time  $t$  is derived in Theorem 2 and can deviate from the CHR.

**Theorem 2** *If the cause-effect relations of interest can be parameterized with SCM (2), where*

$$\lambda_i^a(t) = f_0(t, U_{0i})f_1(t, U_{1i}, a),$$

*and Condition 1 applies, then*

$$\frac{\lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T^a \in [t, t+h) \mid T^a \geq t)}{\lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T^0 \in [t, t+h) \mid T^0 \geq t)} = \frac{\mathbb{E}[f_0(t, U_0)f_1(t, U_1, a) \mid T^a \geq t]}{\mathbb{E}[f_0(t, U_0) \mid T^0 \geq t]}.$$

From the proof presented in Appendix A.2, it becomes clear that the conditional expectations that determine the value of the SMCHR equal weighted means of  $f_0(t, u_0)f_1(t, u_1, a)$  and  $f_0(t, u_0)$  with weights  $\frac{\mathbb{P}(T^a \geq t \mid U_0 = u_0, U_1 = u_1)}{\mathbb{P}(T^a \geq t)}$  and  $\frac{\mathbb{P}(T^0 \geq t \mid U_0 = u_0)}{\mathbb{P}(T^0 \geq t)}$  respectively.

To develop our understanding of the difference between the SMCHR and the CHR, we will first continue to study the difference due to frailty and heterogeneity separately in the next two subsections. In the remainder of the section we present examples for cause-effect relations with effect heterogeneity in the presence of frailty, both for independent and dependent  $U_0$  and  $U_1$ . All programming codes used in the examples presented in this paper can be found online at <https://github.com/RAJP93/CHR>.

## 4.1 Causal effect homogeneity

In the case of homogeneous multiplicative causal effects on the hazard, i.e.  $f_1(t, U_{1i}, a) = f_1(t, a)$ , the ratio of the marginal hazard rates of individuals satisfying  $T_i^a \geq t$  and of those  $T_i^0 \geq t$  equals  $f_1(t, a)$  multiplied by a factor that depends on the difference in frailty distributions at time  $t$  in those two populations as derived in Corollary 1.

**Corollary 1** *If the cause-effect relations of interest can be parameterized with SCM (2), where*

$$\lambda_i^a(t) = f_0(t, U_{0i})f_1(t, a),$$

and condition 1 applies then

$$\frac{\lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T^a \in [t, t+h) \mid T^a \geq t)}{\lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T^0 \in [t, t+h) \mid T^0 \geq t)} = \frac{\mathbb{E}[f_0(t, U_0) \mid T^a \geq t]}{\mathbb{E}[f_0(t, U_0) \mid T^0 \geq t]} f_1(t, a).$$

As stated in the proof in Appendix A.3, the conditional expectation  $\mathbb{E}[f_0(t, U_0) \mid T^a \geq t]$  now equals a weighted mean of  $f_0(t, u_0)$ . The weights equal  $\frac{\mathbb{P}(T^a \geq t \mid U_0 = u_0)}{\mathbb{P}(T^a \geq t)}$  and over time increase for favourable values of  $U_0$ . If  $\forall t > 0: \Lambda^a(u_0, t) < \Lambda^0(u_0, t)$ , e.g. when  $\forall t > 0: f_1(t, a) < 1$ , then the weights  $\frac{\mathbb{P}(T^a \geq t \mid U_0 = u_0)}{\mathbb{P}(T^a \geq t)}$  increase slower than  $\frac{\mathbb{P}(T^0 \geq t \mid U_0 = u_0)}{\mathbb{P}(T^0 \geq t)}$  for favourable values of  $u_0$ , so that for all  $t > 0$ :

$$\frac{\mathbb{E}[f_0(t, U_0) \mid T^a \geq t]}{\mathbb{E}[f_0(t, U_0) \mid T^0 \geq t]} > 1. \quad (13)$$

Then, the SMCHR is larger than the CHR at all times. On the contrary, when  $\forall t > 0: \Lambda^a(u_0, t) > \Lambda^0(u_0, t)$ , then for all  $t > 0$ :

$$\frac{\mathbb{E}[f_0(t, U_0) \mid T^a \geq t]}{\mathbb{E}[f_0(t, U_0) \mid T^0 \geq t]} < 1, \quad (14)$$

and the SMCHR is larger than the CHR. An example of the latter was showcased by Stensrud et al. (2017), where a model with  $f_1(t, a) = 1.81^a$ ,  $f_0(u_0, t) = u_0 \lambda_0(t)$  and compound Poisson distributed frailty  $U_0$  could well explain the decrease of the effect of hormone replacement therapy on coronary heart disease in postmenopausal women over time as observed from an RCT by the Woman Health Initiative. Based on the same case study, Hernán (2010) explained that even when  $f_1(t, a)$  is time-invariant the SMCHR is time-varying, as we have formalized in Corollary 1, so that estimates can depend on the follow-up time.

For frailty models as presented by Aalen et al. (2015) and Stensrud et al. (2017), where  $f_0(t, U_{0i}) = U_{0i} \lambda_0(t)$ , it has been shown by Balan and Putter (2020) that  $\mathbb{E}[U_0 \mid T \geq t, A=a]$  can be expressed in terms of the Laplace transform of the frailty  $U_0$ . Reasoning along the same lines,  $\mathbb{E}[U_0 \mid T^a \geq t]$  is expressed in terms of the Laplace transform of the  $U_0$  in Lemma 1.

**Lemma 1** *If the cause-effect relations of interest can be parameterized with SCM (2), where*

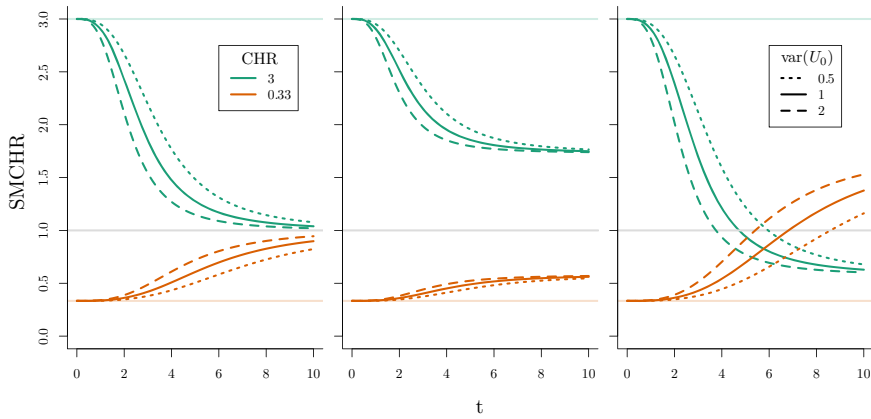
$$f_\lambda(t, U_{0i}, U_{1i}, a) = U_{0i} \lambda_0(t) f_1(t, a),$$

*then*

$$\mathbb{E}[U_0 \mid T^a \geq t] = -\frac{\mathcal{L}'_{U_0}(\int_0^t \lambda_0(s) f_1(s, a) ds)}{\mathcal{L}_{U_0}(\int_0^t \lambda_0(s) f_1(s, a) ds)}, \quad (15)$$

*where  $\mathcal{L}_{U_0}(c) = \mathbb{E}[\exp(-cU_0)]$  with derivative  $\mathcal{L}'_{U_0}(c)$ .*

As considered by (Balan and Putter 2020, Figure 5), we present examples with different frailty distributions. To illustrate the so-called selection bias, we consider a binary



**Fig. 1** SMCHR over time when  $\lambda_i^a(t) = U_{0i} t^2 \mu^a$  for  $\mu = 3$  (green) and  $\mu = \frac{1}{3}$  (orange), when  $U_0$  follows a gamma (left), inverse Gaussian (middle) or compound Poisson (right) distribution with variance 0.5 (dotted), 1 (solid) or 2 (dashed) respectively

exposure and let

$$\lambda_i^a(t) = U_{0i} \lambda_0(t) \mu^a,$$

where  $\lambda_0(t) = \frac{t^2}{20}$ ,  $\mathbb{E}[U_0] = 1$  and  $\text{var}(U_0) = \theta_0$  with  $U_0$  following a Gamma ( $\Gamma(\theta_0^{-1}, \theta_0)$ ), inverse Gaussian ( $\text{IG}(1, \theta_0^{-1})$ ) or compound Poisson ( $\text{CPoi}(3\theta_0^{-1}, \frac{1}{2}, \frac{2}{3}\theta_0)$ ) distribution respectively. The parameterizations, corresponding Laplace transforms and expressions for  $\mathbb{E}[U_0 | T^a \geq t]$  can be found in Appendix B. By applying Lemma 1,  $\mathbb{E}[U_0 | T^1 \geq t]$  and  $\mathbb{E}[U_0 | T^0 \geq t]$  can be derived. The expressions for these quantities are presented in Table 2 in Appendix C. The SMCHR then follows from Corollary 1 (as the conditional hazard is monotone increasing).

How the SMCHR deviates from the CHR (equal to  $\mu$ ) over time for  $\mu \in \{\frac{1}{3}, 3\}$ , and  $\theta_0 \in \{0.5, 1, 2\}$  is visualized in Fig. 1.

For both  $\mu = \frac{1}{3}$  and  $\mu = 3$  the selection of individuals that survive time  $t$  results in a SMCHR that evolves in the opposite direction of the causal effect, towards 1,  $\sqrt{\mu}$  and  $\sqrt{\mu^{-1}}$  respectively. For the case of a compound Poisson frailty, the logarithm of this latter limit is even opposite to the sign of the logarithm of the CHR due to the nonsusceptible individuals. For all types of frailty, the higher the variance of  $U_0$ , the larger the difference between the SMCHR and the CHR. For comparison we have also presented the survival curves of  $T^1$  and  $T^0$  in Fig. 9 in Appendix E for the setting where  $\theta_0 = 1$ . Note that for an RCT, by the independence in (5) and causal consistency,  $T^a$  follows the same distribution as the time-to-event for individuals exposed to  $a$  ( $T^a \stackrel{d}{=} T|A=a$ ).

## 4.2 Causal effect heterogeneity in the absence of frailty

Before we return to the general case presented in Theorem 2, let's consider the presence of effect heterogeneity in the absence of frailty, i.e.

$$f_{\lambda}(t, U_{0i}, U_{1i}, a) = \lambda_0(t) f_1(t, U_{1i}, a).$$

If the CHR,  $\mathbb{E}[f_1(t, U_{1i}, a)]$ , is equal for all  $t$ , the SMCHR is not, as over time the exposed individuals that 'benefit' more are more likely to survive. The effect of this selection on the SMCHR over time is formalized in Corollary 2.

**Corollary 2** *If the cause-effect relations of interest can be parameterized with SCM (2), where*

$$\lambda_i^a(t) = \lambda_0(t) f_1(t, U_{1i}, a),$$

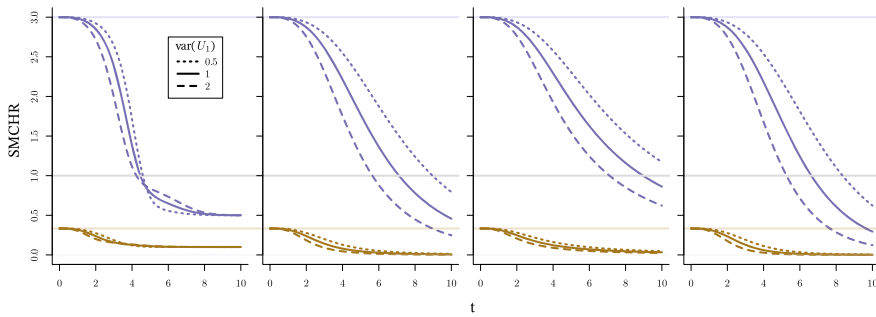
*and Condition 1 applies then*

$$\frac{\lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T^a \in [t, t+h] \mid T^a \geq t)}{\lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T^0 \in [t, t+h] \mid T^0 \geq t)} = \mathbb{E}[f_1(t, U_1, a) \mid T^a \geq t].$$

The SMCHR thus equals  $\mathbb{E}[f_1(t, U_1, a) \mid T^a \geq t]$ , which is smaller than  $\mathbb{E}[f_1(t, U_1, a)]$  as more weight is placed on lower values of  $f_1(t, u_1, a)$  that correspond to higher  $\mathbb{P}(T^a \geq t \mid U_1 = u_1)$ . Besides the selection of frailty factors, the selection of individual modifiers can thus also lead to selection bias of the estimated hazard ratio. For this hypothetical setting without frailty but with effect heterogeneity, the CHR at  $t$  is systematically lower (irrespective of whether the exposure is beneficial or harmful on average) than the SMCHR, so the exposure seems more 'beneficial' than it is. For a harming exposure, the resulting attenuation of the effect has only been explained due to the presence of frailty and not due to the presence of individual modifiers (Hernán 2010; Stensrud et al. 2017).

Similar to the examples presented in Sect. 4.1, we let  $\lambda_0 = \frac{t^2}{20}$ ,  $\mathbb{E}[U_1] = \mu$  and  $\text{var}(U_1) = \theta_1$  with  $U_1$  following a Gamma ( $\Gamma(\frac{\mu}{\theta_1}, \frac{\theta_1}{\mu})$ ), inverse Gaussian ( $\text{IG}(\mu, \frac{\mu^3}{\theta_1})$ ) or compound Poisson ( $\text{CPoi}(3\frac{\mu^2}{\theta_1}, \frac{1}{2}, \frac{2\theta_1}{3\mu})$ ) distribution respectively. By applying Lemma 1 for  $a = 1$  (since  $\lambda_i^1(t) = U_{1i}\lambda_0(t)$ ), we can derive  $\mathbb{E}[U_1 \mid T^1 \geq t]$ , which by Corollary 2 (as the conditional hazard is monotone increasing) equals the SMCHR, and is presented in Table 3 in Appendix C. Additionally, we derived  $\mathbb{E}[U_1 \mid T^1 \geq t]$  for a setting where the multiplicative hazard effect modifier  $U_1$  equals  $\mu_1$  ( $\leq 1$ , for individuals that benefit) with probability  $p_1$ ,  $\mu_2$  ( $\geq 1$ , for individuals that are harmed) with probability  $p_2$  or 1 (for individuals that are not affected). We define this distribution as the Benefit-Harm-Neutral,  $\text{BHN}(p_1, \mu_1, p_2, \mu_2)$ , distribution.

For  $\mathbb{E}[U_1] \in \{\frac{1}{3}, 3\}$ , and  $\theta_1 \in \{0.5, 1, 2\}$  the evolution of the conditional expectation is shown in Fig. 2 for all four effect-modifier distributions. For the BHN distribution, when  $\mathbb{E}[U_1]$  equals  $\frac{1}{3}$  and 3, we fix  $p_1 = 0.9$ ,  $\mu_1 = 0.1$  and  $p_1 = 0.05$ ,  $\mu_1 = 0.5$  respectively. Expressions for  $p_2$  and  $\mu_2$  such that  $\mathbb{E}[U_1] = \mu$  and  $\text{var}(U_1) = \theta_1$  can be found in Appendix B.4.



**Fig. 2** SMCHR over time when  $\lambda_i^a(t) = \frac{t^2}{20}(U_{1i})^a$  when  $U_1$  follows a BHN, gamma, inverse Gaussian or compound Poisson distribution (from left to right) with expectation 3 (blue) or  $\frac{1}{3}$  (brown) and variance 0.5 (dotted), 1 (solid) or 2 (dashed) respectively

When the exposure is in expectation harming ( $\mathbb{E}[U_1] = 3$ ), for all settings considered, there is a point in time that the SMCHR drops below 1. For the continuous distributions, the SMCHR won't stop decreasing. The decreases for the gamma and compound Poisson settings are very similar, while for the inverse Gaussian setting, this goes a bit slower. For the discrete setting, the SMCHR converges to  $\mu_1$  of 0.1 and 0.5, respectively. Again, as in the previous subsection, the higher the variability of the latent variable, the faster the SMCHR deviates from the CHR. Only for the discrete effect modifier, the lines cross for the different variances for  $\mathbb{E}[U_1] = 3$ , but this is the result of different fractions of individuals that are not affected by the exposure (as the mean and variance are coupled).

### 4.3 Causal effect heterogeneity in the presence of frailty

In the general case where effect heterogeneity and frailty are present, both heterogeneities affect the value of the SMCHR. By Theorem 2, the ratio evolves as  $\frac{\mathbb{E}[f_0(t, U_0)f_1(t, U_1, a)|T^a \geq t]}{\mathbb{E}[f_0(t, U_0)|T^0 \geq t]}$ . The numerator depends on the joint distribution of  $U_0$  and  $U_1$ . For illustration, we again consider a binary exposure and let

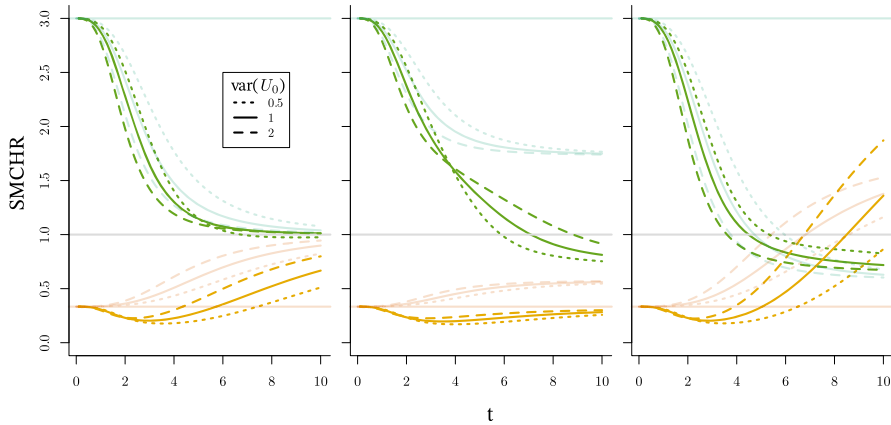
$$f_\lambda(t, U_{0i}, U_{1i}, a) = U_{0i}(U_{1i})^a \lambda_0(t) f_1(t, a),$$

such that the SMCHR equals  $\frac{\mathbb{E}[U_0 U_1 | T^1 \geq t]}{\mathbb{E}[U_0 | T^0 \geq t]} f_1(t, 1)$  and, by Lemma 1, can be derived from the Laplace transforms of  $U_0 U_1$  and  $U_0$  respectively.

#### 4.3.1 Independent $U_0$ and $U_1$

In the case of independence, the Laplace transform of the product equals  $\mathbb{E}[\mathcal{L}_{U_0}(c)U_1]$ , which generally does not adopt a tractable form. The case with a discrete effect modifier, introduced in Sect. 4.2, forms an exception. If  $U_1 \sim \text{BHN}(p_1, \mu_1, p_2, \mu_2)$ , then

$$\mathcal{L}_{U_0 U_1}(c) = p_1 \mathcal{L}_{U_0}(\mu_1 c) + p_2 \mathcal{L}_{U_0}(\mu_2 c) + (1 - p_1 - p_2) \mathcal{L}_{U_0}(c). \quad (16)$$



**Fig. 3** SMCHR over time when  $\lambda_i^a(t) = U_{0i}(U_{1i})^a \frac{t^2}{20}$  for a unit-variance BHN distributed  $U_1$  with  $\mathbb{E}[U_1] = 3$  (opaque green) and  $\mathbb{E}[U_1] = \frac{1}{3}$  (opaque orange) when  $U_0$  follows a gamma (left), inverse Gaussian (middle) or compound Poisson (right) distribution with variance 0.5 (dotted), 1 (solid) or 2 (dashed) respectively. For comparison, the lines presented in Fig. 1 are represented by transparent lines

We present the running example where  $f_1(t, a) = 1$  and  $\lambda_0 = \frac{t^2}{20}$ . As in Sect. 4.1,  $U_0$  follows a Gamma ( $\Gamma(\theta_0^{-1}, \theta_0)$ ), inverse Gaussian ( $\text{IG}(1, \theta_0^{-1})$ ) or compound Poisson ( $\text{CPois}(3\theta_0^{-1}, \frac{1}{2}, \frac{2}{3}\theta_0)$ ) distribution respectively. Moreover, the latent modifier  $U_1$  is independent of  $U_0$  and follows a unit-variance BHN distribution with mean  $\mu$ . The expressions for  $\mathbb{E}[U_0 U_1 \mid T^1 \geq t]$  are presented in Table 4 in Appendix C. As the  $\mathbb{E}[U_0 \mid T^0 \geq t]$  are independent of the  $U_1$  distribution these expectations are the same as presented in Table 2 in Appendix C. The SMCHR and its limit can be derived by applying Theorem 2 (as the conditional hazard is monotone increasing). Interestingly, for gamma frailty, as in the case without effect heterogeneity, this limit remains 1. For the inverse Gaussian frailty, selection of the effect modifier drastically changes the limit from  $\sqrt{\mathbb{E}[U_1]}$  to  $\sqrt{\mu_1}$ , which is always less or equal to 1. Finally, for compound Poisson frailty the limit changes from  $\sqrt{\mathbb{E}[U_1]^{-1}}$  to  $\frac{p_1}{\sqrt{\mu_1}} + \frac{p_2}{\sqrt{\mu_2}} + (1 - p_1 - p_2)$ .

The evolution of  $\frac{\mathbb{E}[U_0 U_1 \mid T^1 \geq t]}{\mathbb{E}[U_0 \mid T^0 \geq t]}$  over time is visualized in Fig. 3 for  $\theta_0 \in \{0.5, 1, 2\}$  and  $U_1 \sim \text{BHN}(0.9, 0.1, 0.03, 6.0)$ , such that  $\mathbb{E}[U_1] = \frac{1}{3}$  and  $\text{var}(U_1) = 1$ , and for  $U_1 \sim \text{BHN}(0.05, 0.5, 0.82, 3.5)$ , such that  $\mathbb{E}[U_1] = 3$  and  $\text{var}(U_1) = 1$ .

In the case the CHR is larger than one, the selection of less susceptible individuals (frailty) that are harmed less (effect modifier) in the exposed world, both cause the SMCHR to be smaller than the CHR. Then, the SMCHR decreases faster in the presence of effect heterogeneity. This explains the observation by Stensrud et al. (2017), “Interestingly, the magnitude of frailty bias is larger when a heterogeneous treatment effect is included”, for a simulation with frailty and random individual hazard ratios such that  $\mathbb{E}[\lambda_i^1(t)] = 1.81 > 1$ . For the gamma and compound Poisson frailty examples, this effect is relatively small as  $\mathbb{E}[U_1 U_0 \mid T^1 \geq t]$  is quite similar to  $\mathbb{E}[U_0 \mid T^1 \geq t]$  (presented in Table 4) for the selected  $p_1, \mu_1, p_2$  and  $\mu_2$ . However, for inverse Gaussian frailty, the SMCHR deviates much more from the CHR in the presence of effect

heterogeneity. In Fig. 7 in Appendix D, the evolution of the SMCHR is presented for a longer timescale, and the limits become apparent.

If the CHR is smaller than one, then the selection of less susceptible individuals (frailty) in the unexposed world and the selection of individuals that benefit more (effect modifier) in the exposed world have opposite effects on the SMCHR. For this case of discrete effect modifiers, the SMCHR first decreases by selecting individuals with more beneficial modifiers and later increases (above the CHR) when the frailty selection effect reveals. For the examples presented, the fraction  $p_1 = 0.9$  of the population with  $\mu_1 = 0.1$  are expected to survive so that over time the SMCHR will resemble the SMCHR in the absence of effect heterogeneity for this subpopulation (with the CHR equal to 0.1). The limit for gamma frailty is still one, so the SMCHR deviates less from the CHR due to the two opposed selection effects. The difference is strongly reduced for the inverse Gaussian frailty as the limit  $\sqrt{0.1}$  is close to the actual CHR. Finally, for the compound Poisson frailty, the SMCHR with effect heterogeneity crosses the SMCHR in the absence of effect heterogeneity as the frailty bias is larger for a CHR of 0.1 compared to one of  $\frac{1}{3}$ .

In summary, the bias for the CHR can further increase in the presence of effect heterogeneity, stressing the issues regarding the causal interpretation of OHRs (assuming no confounding). However, for beneficial exposures, the frailty bias can reduce in the presence of effect heterogeneity (e.g. inverse Gaussian frailty), illustrating that there might be settings where the SMCHR is close to the CHR.

#### 4.3.2 Dependent $U_0$ and $U_1$

In case the multiplicative effect of the exposure on the hazard of susceptible individuals is expected to be higher or lower than for less susceptible individuals, the distribution of  $U_0U_1$  will be less or more variable than when the latent variables are independent. Every bivariate joint distribution function,  $F_{(U_0, U_1)}$ , can be written using the marginal distribution functions and a copula  $C$  (Sklar 1959). As such,

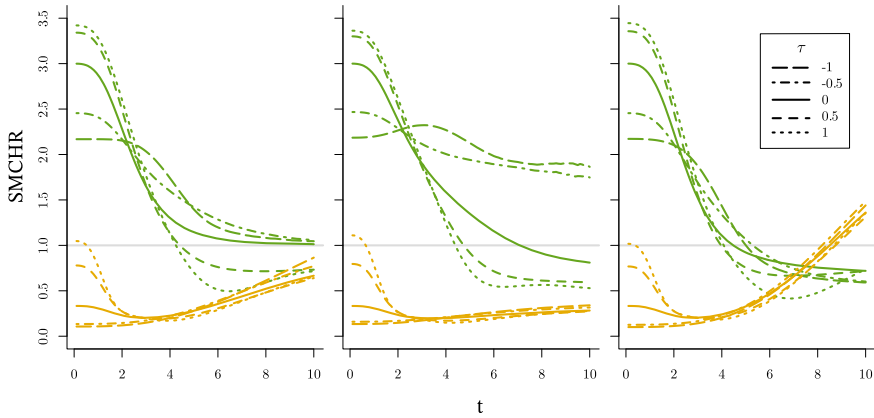
$$F_{(U_0, U_1)}(u_0, u_1) = C(F_{U_0}(u_0), F_{U_1}(u_1))$$

and the Kendall's  $\tau$  correlation coefficient of  $U_0$  and  $U_1$  can be written as a function of the copula (Nelsen 2006). To study how the dependence can affect the SMCHR for the setting presented in Fig. 3, we use a Gaussian copula

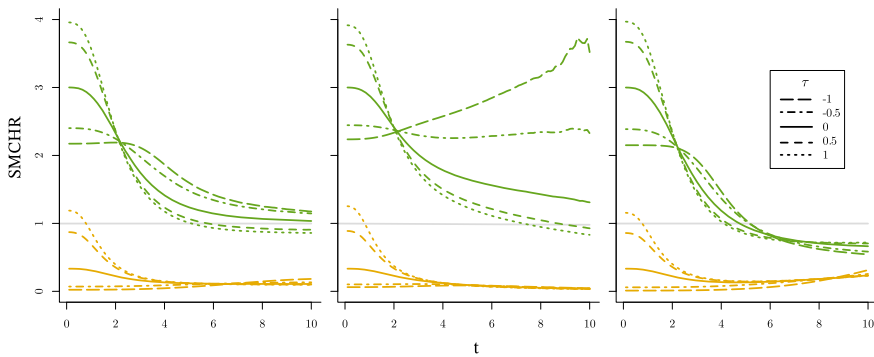
$$C(x, y) = \Phi_{2, \rho}(\Phi^{-1}(x), \Phi^{-1}(y)),$$

where  $\Phi$  and  $\Phi_{2, \rho}$  are the standard normal and bivariate normal with correlation  $\rho$  cumulative distribution functions, respectively. For  $\rho \in \{-1, \sin(-\frac{\pi}{4}), 0, \sin(\frac{\pi}{4}), 1\}$  (such that  $\tau \in \{-1, -\frac{1}{2}, 0, \frac{1}{2}, 1\}$ ) and  $\text{var}(U_0) = 1$ ,  $\mathbb{E}[U_0U_1 | T^1 \geq t]$  is derived empirically from simulations and are presented in Fig. 8 in Appendix D. The results were very similar when using a Frank, Clayton or Gumbel copula instead of the Gaussian copula. The SMCHRs are presented in Fig. 4.

Note that for  $\tau = 0$ , we recover the independent setting already shown in Fig. 3 that can be used for comparison. First of all, when  $U_0$  and  $U_1$  are dependent, the CHR



**Fig. 4** SMCHR over time for  $\lambda_i^a(t) = U_{0i}(U_{1i})^a \frac{t^2}{20}$ , for a unit-variance BHN distributed  $U_1$  with  $\mathbb{E}[U_1] = 3$  (green) or  $\mathbb{E}[U_1] = \frac{1}{3}$  (orange),  $U_0$  follows a gamma (left), inverse Gaussian (middle) or compound Poisson (right) distribution and the joint distribution of  $U_0$  and  $U_1$  follows from a Gaussian copula with varying Kendall's  $\tau$  correlation coefficients (see legend)



**Fig. 5** SMCHR over time for  $\lambda_i^a(t) = U_{0i}(U_{1i})^a \frac{t^2}{20}$ , for a unit-variance gamma distributed  $U_1$  with  $\mathbb{E}[U_1] = 3$  (green) or  $\mathbb{E}[U_1] = \frac{1}{3}$  (orange),  $U_0$  follows a gamma (left), inverse Gaussian (middle) or compound Poisson (right) distribution and the joint distribution of  $U_0$  and  $U_1$  follows from a Gaussian copula with varying Kendall's  $\tau$  correlation coefficients (see legend)

equals  $\mathbb{E}[U_1] + \text{cov}(U_0, U_1)$ . For CHRs greater than one, it becomes clear that the selection effect is more serious for cases with a high positive correlation between  $U_0$  and  $U_1$ . The stronger selection effect is due to the higher variability of  $U_0 U_1$ . For CHRs less than one, this trend is only true at short timescales, after which the frailty selection effect takes over since, for this example, for a large fraction of the individuals,  $p_1 = 0.9$ , the effect is the same ( $U_1 = 0.1$ ).

When we use a continuous gamma distributed  $U_1$  instead, the frailty selection effect is less apparent, as shown in Fig. 5.

So far, for CHRs larger than one, we have observed a monotonic SMCHR. However, in the case of strong dependence between  $U_0$  and  $U_1$  ( $|\tau| = 1$ ), for inverse Gaussian



**Table 1** Assuming no confounding, an OHR (at time  $t$ ) equal to a particular value  $x$  can occur for all values of a constant CHR as a result of selection of the frailty ( $U_0$ ) or modifier ( $U_1$ ) which might be dependent

		Cause	Presented examples
$x > 1$	$\text{CHR} > x$	Frailty or modifier selection	Figs. 1, 2 and 3
	$1 < \text{CHR} < x$	Dependence $U_0$ and $U_1$	Inverse Gaussian frailty - $\tau = 1$ (Figs. 4 and 5)
	$\text{CHR} < 1$	Frailty selection	Compound Poisson frailty (Figs. 1 and 3)
$x < 1$	$\text{CHR} > 1$	Frailty or modifier selection	Compound Poisson frailty (Figs. 1 and 3)
	$x < \text{CHR} < 1$	Modifier selection	Gamma distributed modifier (Fig. 5)
	$\text{CHR} < x$	Frailty selection	Figs. 1 and 3

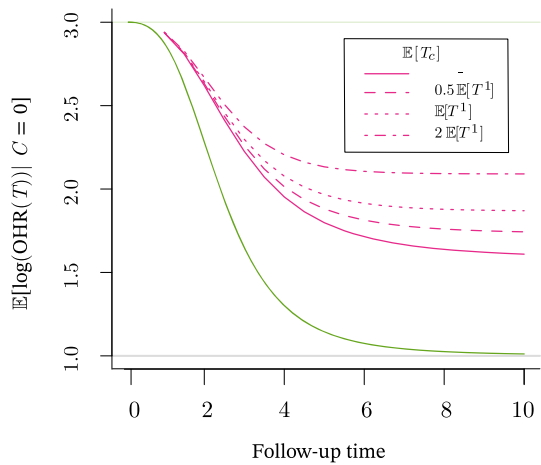
frailty,  $\mathbb{E}[U_0 \mid T^0 \geq t]$  decreases faster than  $\mathbb{E}[U_0 U_1 \mid T^1 \geq t]$  resulting in a non-monotonic trend for the SMCHR. For a Gamma distributed  $U_1$ , in the case of inverse Gaussian distributed  $U_0$  with  $\tau = 1$ , the SMCHR even equals a monotonic increasing function over time as shown in Fig. 5.

In Sect. 4, we have derived and applied Theorem 1 to several examples to illustrate the deviation of the SMCHR from the CHR. In summary, even when the CHR is constant, an OHR from an RCT equal to a particular value  $x$  (at time  $t$ ) can occur for different CHR values when the  $(U_0, U_1)$  distribution is unknown as summarized in Table 1.

## 5 Implications for the Cox model

We have demonstrated that in the presence of frailty and effect heterogeneity, even when the CHR is time-invariant, the SMCHR varies over time. Then, the proportional hazards assumption will not hold for an observed hazard ratio from an RCT (that is, with independent censoring, equal to the SMCHR as discussed at the start of Sect. 4). Despite the many options to deal with non-proportional hazards (see, e.g. (Therneau and Grambsch 2000, Section 6.5) or Bennett 1983; Hess 1994; Wei and Schaabel 2008), in the majority of epidemiological time-to-event studies, the misspecified traditional Cox's proportional hazard model is fitted. The logarithm of the Cox estimate can be interpreted as the logarithm of the OHR marginalized over the observed death times (Schemper et al. 2009), i.e.  $\mathbb{E}[\log(\text{OHR}(T)) \mid C = 0]$  for censoring indicator  $C$ . The logarithm of the Cox estimate obtained from an RCT thus equals a time-weighted average of the logarithm of the OHR. In the case of non-proportional hazards, even for independent censoring, the estimate is well-known to be affected by the censoring distribution. It differs from the average log hazard ratio  $\mathbb{E}[\log(\text{OHR}(T))]$  (Xu and O'Quigley 2000; Schemper et al. 2009; Boyd et al. 2012). Therefore, the bias of the Cox estimate, when the estimand is the CHR, will depend on the joint distribution of  $(U_0, U_1)$  as well as the censoring distribution. In most cases considered in Sect. 4, the deviation of the OHR from the CHR increased over time. For independent censoring,

**Fig. 6** Empirically obtained  $\exp(\mathbb{E}[\log(\text{OHR}(T)) \mid C = 0])$  for an increasing time to follow-up (solid pink), and cases with an additional exponentially distributed censoring time ( $T_C$ , dashed pink), when  $\lambda_i^a(t) = U_{0i}(U_{1i})^a \frac{t^2}{20}$ ,  $U_1 \sim \text{BHN}(0.05, 0.5, 0.82, 3.5)$  and  $U_0 \sim \Gamma(1, 1)$ . The SMCHR is also presented (green)



the probability of censoring increases over time, so the Cox estimate is closer to the OHR at short times. In Fig. 6, this is demonstrated for the gamma-frailty case ( $\text{var}(U_0) = 1$ , for which the SMCHR was presented in Fig. 3) by presenting empirically obtained  $\mathbb{E}[\log(\text{OHR}(T)) \mid C = 0]$  (with 1,000,000 replications) based on a varying follow-up time and loss to follow-up modelled with an exponential censoring-time distribution with varying means.

A time-varying OHR violates the proportional hazard assumption that can be verified when fitting a Cox model. When the assumption is not rejected in practice, the statistical test used is probably underpowered. In the presence of heterogeneity, only when the actual CHR would be time-varying, the OHR can be approximately constant when the selection effect and the change in CHR roughly cancel out (Stensrud et al. 2018; Stensrud and Hernán 2020). The data cannot be used to distinguish the latter case from the case with a constant CHR but no heterogeneity and, thus, no selection effect. Similarly, as mentioned at the end of Sect. 3, when the OHR would vary over time, we can never conclude whether this is the result of a time-varying causal effect or due to selection. However, the proportional hazard assumption would be violated in both cases, and a standard Cox model is inappropriate.

## 6 Discussion

In this paper, we have formalized how heterogeneity leads to deviation of the SMCHR (see Equation (10)) from the CHR of interest (see Definition 1) due to the selection of both the individual frailty factor ( $U_0$ ) and the individual effect modifier ( $U_1$ ). This work generalizes frailty examples presented in the literature (Hernán 2010; Aalen et al. 2015; Stensrud et al. 2017), by considering the possibility of multiplicative effect (on the hazard) heterogeneity that also results in non-exchangeability of exposed and unexposed individuals over time. As a result of the individual effect modifier ( $U_1$ ), the individuals that survive in the exposed groups are expected to benefit more or suffer less from the exposure. At the same time  $U_0 | T^1 \geq t$  will have a different distribution

than  $U_0|T^0 \geq t$ . When the CHR is larger than one, and  $U_0 \perp U_1$ , the selection effects act in the same direction. On the other hand, when the CHR is smaller than one and  $U_0 \perp U_1$ , the selection effects can act in opposite directions so that the SMCHR might be closer to the CHR than in the case without effect heterogeneity (see Fig. 3).

For data from an RCT, with independent censoring, the expected observed hazard ratio equals the studied SMCHR so that all results directly relate to this OHR. For observational data, the OHR does not equal the SMCHR due to confounding. However, when all confounders  $L$  are observed, i.e.  $T^a \perp A | L$ , one can study the conditional (on  $L$ ) OHR that in turn is equal to the conditional SMCHR. The presented theorems are valid while conditioning on  $L$ .

The intuition explained by Hernán (2010) suggests that an appropriate estimate of the SMCHR is expected to underestimate the actual effect size, while the sign of the logarithms of the SMCHR and the CHR are equal. However, we have shown that in the presence of effect heterogeneity, an SMCHR equal to a particular value  $x$  can occur both under  $\text{CHR} > 1$  as well as  $\text{CHR} < 1$  as summarized in Table 1. Therefore, OHRs from RCTs are not guaranteed to present a lower bound for the causal effect without making untestable assumptions on the  $(U_0, U_1)$  distribution. We have derived how the SMCHR will evolve due to the selection of frailty and effect modifiers in Theorem 1. However, in practice, only the evolution of the OHR can be found (assuming sufficient data is available). Even after assuming the absence of confounding (e.g. for an RCT), the CHR is non-identifiable without making (untestable) assumptions on the  $(U_0, U_1)$  distribution as discussed at the end of Sect. 3. We can thus not distinguish between a time-varying CHR without selection of  $U_0$  and  $U_1$  or a time-invariant CHR with selection, see e.g. Stensrud and Hernán (2020). Adjusting for other risk factors can lower the remaining variability of  $U_0$  and  $U_1$  so that the difference between the conditional OHR and CHR is reduced. Even for an RCT, it may thus help to focus on adjusted hazard ratios despite the absence of confounding. Nevertheless, adjusting for other risk factors will require more data and modelling decisions.

Finally, we want to remark that for cause-effect relations that cannot be described by SCM (2) with  $f_\lambda(t, U_{0i}, U_{1i}, a) = f_0(t, U_{0i})f_1(t, U_{1i}, a)$ , the CHR is not the appropriate measure to quantify the causal effect. Then, other causal hazard contrasts can be relevant that may or may not have an observable analog. For example, additive hazard models (when well-specified) do not suffer from the frailty selection as shown by Aalen et al. (2015), but these models will still suffer from latent modifier selection in the presence of effect heterogeneity ( $\mathbb{E}[U_1 | T^1 \geq t] > \mathbb{E}[U_1]$ ) as demonstrated in our companion paper (Post et al. 2024).

We hope that the discussed effect heterogeneity and formalization of the built-in selection bias of the OHR show the need to use more suitable estimands. As suggested by others, contrasts of the survival probabilities, the median, or the restricted mean survival time of potential outcomes are proper measures to quantify causal effects on time-to-event outcomes (Hernán 2010; Stensrud et al. 2018; Bartlett et al. 2020; Young et al. 2020). Modelling and estimating hazard rates can still be helpful for causal inference when the hazards are used to derive one of the appropriate causal estimands (Ryalen et al. 2018).

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## A Proofs

### A.1 Proof of theorem 1

**Proof**

$$\begin{aligned} \frac{\mathbb{E}[\lambda_i^a(t)]}{\mathbb{E}[\lambda_i^0(t)]} &= \frac{\mathbb{E}[f_0(t, U_{0i}) f_1(a, U_{1i}, t)]}{\mathbb{E}[f_0(t, U_{0i})]} \\ &= \frac{\int \lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T^a \in [t, t+h] \mid T^a \geq t, U_0, U_1) dF_{U_0, U_1}}{\int \lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T^0 \in [t, t+h] \mid T^0 \geq t, U_0) dF_{U_0}}, \end{aligned}$$

By randomization the independence in (7) applies and the CHR is equal to

$$\frac{\int \lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T^a \in [t, t+h] \mid T^a \geq t, U_0, U_1, A=a) dF_{U_0, U_1}}{\int \lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T^0 \in [t, t+h] \mid T^0 \geq t, U_0, A=0) dF_{U_0}}.$$

Finally, by causal consistency,

$$\frac{\mathbb{E}[\lambda_i^a(t)]}{\mathbb{E}[\lambda_i^0(t)]} = \frac{\int \lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T \in [t, t+h] \mid T \geq t, U_0, U_1, A=a) dF_{U_0, U_1}}{\int \lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T \in [t, t+h] \mid T \geq t, U_0, A=0) dF_{U_0}}. \quad \square$$

### A.2 Proof of theorem 2

**Proof** By the law of total probability,

$$\begin{aligned} \frac{\lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T^a \in [t, t+h] \mid T^a \geq t)}{\lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T^0 \in [t, t+h] \mid T^0 \geq t)} &= \\ \frac{\lim_{h \rightarrow 0} \int h^{-1} \mathbb{P}(T^a \in [t, t+h] \mid T^a \geq t, U_0, U_1) dF_{U_0, U_1 \mid T^a \geq t}}{\lim_{h \rightarrow 0} \int h^{-1} \mathbb{P}(T^0 \in [t, t+h] \mid T^0 \geq t, U_0) dF_{U_0 \mid T^0 \geq t}} \end{aligned}$$

First we focus on the integrand,

$$\begin{aligned} &h^{-1} \mathbb{P}(T^a \in [t, t+h] \mid T^a \geq t, U_0, U_1) \\ &= h^{-1} \frac{\mathbb{P}(T^a \geq t \mid U_0, U_1) - \mathbb{P}(T^a \geq t+h \mid U_0, U_1)}{\mathbb{P}(T^a \geq t \mid U_0, U_1)} \\ &= h^{-1} \left( 1 - \frac{\mathbb{P}(T^a \geq t+h \mid U_0, U_1)}{\mathbb{P}(T^a \geq t \mid U_0, U_1)} \right) \end{aligned}$$

$$\begin{aligned}
&= h^{-1} \left( 1 - \frac{\exp \left( - \int_0^{t+h} f_0(s, U_0) f_1(s, U_1, a) ds \right)}{\exp \left( - \int_0^t f_0(s, U_0) f_1(s, U_1, a) ds \right)} \right) \\
&= h^{-1} \left( 1 - \exp \left( - \int_t^{t+h} f_0(s, U_0) f_1(s, U_1, a) ds \right) \right)
\end{aligned}$$

For monotonic (increasing or decreasing) conditional hazard functions if  $h_2 < h_1$ , then

$$\begin{aligned}
&h_1^{-1} \left( 1 - \exp \left( - \int_t^{t+h_1} f_0(s, U_0) f_1(s, U_1, a) ds \right) \right) \\
&\leq h_2^{-1} \left( 1 - \exp \left( - \int_t^{t+h_2} f_0(s, U_0) f_1(s, U_1, a) ds \right) \right)
\end{aligned}$$

or

$$\begin{aligned}
&h_1^{-1} \left( 1 - \exp \left( - \int_t^{t+h_1} f_0(s, U_0) f_1(s, U_1, a) ds \right) \right) \\
&\geq h_2^{-1} \left( 1 - \exp \left( - \int_t^{t+h_2} f_0(s, U_0) f_1(s, U_1, a) ds \right) \right)
\end{aligned}$$

as the average integrated conditional hazard over the interval increases (or decreases). Moreover,

$$\lim_{h \rightarrow 0} h^{-1} \mathbb{P} \left( T^a \in [t, t+h) \mid T^a \geq t, U_0, U_1 \right) = f_0(s, U_0) f_1(s, U_1, a) \geq 0.$$

Then, the limit and integral can be interchanged by directly applying the monotone convergence theorem.

For non-monotone conditional hazard functions, when Condition 1 applies, for every  $t$ , there exist a  $\tilde{h}$  so that  $\forall h^* \in (0, \tilde{h})$ :  $\mathbb{E} \left[ f_0(t+h^*, U_0) f_1(t+h^*, U_1, a) \mid T^a \geq t \right] < \infty$ . Moreover, let  $t^* = \arg \max_{s \in (t, t+\tilde{h})} f_0(s, U_0) f_1(s, U_1, a)$ , so that for  $h < \tilde{h}$ :

$$h^{-1} \mathbb{P} \left( T^a \in [t, t+h) \mid T^a \geq t, U_0, U_1 \right) \leq h^{-1} \left( 1 - \exp \left( -h f_0(t^*, U_0) f_1(t^*, U_1, a) \right) \right).$$

Using the power series definition of the exponential function,

$$\begin{aligned}
&h^{-1} \mathbb{P} \left( T^a \in [t, t+h) \mid T^a \geq t, U_0, U_1 \right) \\
&\leq h^{-1} \left( 1 - \frac{1}{\sum_{k=0}^{\infty} h^k (f_0(t^*, U_0) f_1(t^*, U_1, a))^k \frac{1}{k!}} \right) \\
&= h^{-1} \frac{\sum_{k=1}^{\infty} h^k (f_0(t^*, U_0) f_1(t^*, U_1, a))^k \frac{1}{k!}}{\sum_{k=0}^{\infty} h^k (f_0(t^*, U_0) f_1(t^*, U_1, a))^k \frac{1}{k!}} \\
&= f_0(t^*, U_0) f_1(t^*, U_1, a) \frac{\sum_{k=1}^{\infty} h^{k-1} (f_0(t^*, U_0) f_1(t^*, U_1, a))^{k-1} \frac{1}{k!}}{\sum_{k=0}^{\infty} h^k (f_0(t^*, U_0) f_1(t^*, U_1, a))^k \frac{1}{k!}}
\end{aligned}$$

$$\begin{aligned}
 &= f_0(t^*, U_0) f_1(t^*, U_1, a) \frac{\sum_{k=0}^{\infty} h^k (f_0(t^*, U_0) f_1(t^*, U_1, a))^k \frac{1}{(k+1)!}}{\sum_{k=0}^{\infty} h^k (f_0(t^*, U_0) f_1(t^*, U_1, a))^k \frac{1}{k!}} \\
 &< f_0(t^*, U_0) f_1(t^*, U_1, a).
 \end{aligned}$$

Furthermore,  $\mathbb{E}[f_0(t^*, U_0) f_1(t^*, U_1, a) \mid T^a \geq t] < \infty$  when  $\forall h \in (0, \tilde{h})$ :  $\mathbb{E}[f_0(t + h, U_0) f_1(t + h, U_1, a) \mid T^a \geq t] < \infty$ . Then we can change the order of the limit and integral by application of the dominated convergence theorem and conclude,

$$\begin{aligned}
 &\frac{\lim_{h \rightarrow 0} \int h^{-1} \mathbb{P}(T^a \in [t, t + h) \mid T^a \geq t, U_0, U_1) dF_{U_0, U_1 \mid T^a \geq t}}{\lim_{h \rightarrow 0} \int h^{-1} \mathbb{P}(T^0 \in [t, t + h) \mid T^0 \geq t, U_0) dF_{U_0 \mid T^0 \geq t}} \\
 &= \frac{\mathbb{E}[f_0(t, U_0) f_1(t, U_1, a) \mid T^a \geq t]}{\mathbb{E}[f_0(t, U_0) \mid T^0 \geq t]}.
 \end{aligned}$$

Applying Bayes rule, we obtain

$$\begin{aligned}
 &\mathbb{E}[f_0(t, U_0) f_1(t, U_1, a) \mid T^a \geq t] \\
 &= \int f_0(t, U_0) f_1(t, U_1, a) dF_{(U_0, U_1) \mid T^a \geq t} \\
 &= \int f_0(t, U_0) f_1(t, U_1, a) \frac{\mathbb{P}(T^a \geq t \mid U_0, U_1)}{\mathbb{P}(T^a \geq t)} dF_{(U_0, U_1)} \\
 &= \int f_0(t, U_0) f_1(t, U_1, a) \frac{\exp\left(-\int_0^t f_0(s, U_0) f_1(s, U_1, a) ds\right)}{\int \exp\left(-\int_0^t f_0(s, U_0) f_1(s, U_1, a) ds\right) dF_{(U_0, U_1)}} dF_{(U_0, U_1)}.
 \end{aligned}$$

Furthermore,

$$\begin{aligned}
 &\mathbb{E}[f_0(t, U_0) \mid T^a \geq t] \\
 &= \int f_0(t, U_0) dF_{U_0 \mid T^a \geq t} \\
 &= \int f_0(t, U_0) \frac{\mathbb{P}(T^a \geq t \mid U_0)}{\mathbb{P}(T^a \geq t)} dF_{U_0} \\
 &= \int f_0(t, U_0) \frac{\int \exp\left(-\int_0^t f_0(s, U_0) f_1(s, U_1, a) ds\right) dF_{U_1 \mid U_0}}{\int \exp\left(-\int_0^t f_0(s, U_0) f_1(s, U_1, a) ds\right) dF_{(U_0, U_1)}} dF_{U_0},
 \end{aligned}$$

such that for  $a = 0$ ,

$$\mathbb{E}[f_0(t, U_0) \mid T^0 \geq t] = \int f_0(t, U_0) \frac{\exp\left(-\int_0^t f_0(s, U_0) ds\right)}{\int \exp\left(-\int_0^t f_0(s, U_0) ds\right) dF_{U_0}} dF_{U_0}.$$

The ratio of both gives us the result.  $\square$

### A.3 Proof of corollary 1

**Proof** By Theorem 2,

$$\frac{\lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T^a \in [t, t+h] \mid T^a \geq t)}{\lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T^0 \in [t, t+h] \mid T^0 \geq t)} = \frac{\mathbb{E}[f_0(t, U_0) f_1(t, U_1, a) \mid T^a \geq t]}{\mathbb{E}[f_0(t, U_0) \mid T^0 \geq t]}.$$

and equals

$$\frac{\int f_0(t, U_0) f_1(t, U_1, a) \frac{\exp(-\Lambda^a(t, U_0, U_1))}{\int \exp(-\Lambda^a(t, U_0, U_1)) dF_{(U_0, U_1)}} dF_{(U_0, U_1)}}{\left( \int f_0(t, U_0) \frac{\exp(-\Lambda^0(t, U_0))}{\int \exp(-\Lambda^0(t, U_0)) dF_{U_0}} dF_{U_0} \right)^{-1}},$$

where  $\Lambda^a(t, u_0, u_1) = \int_0^t f_0(s, u_0) f_1(s, u_1, a) ds$  and thus  $\Lambda^0(t, u_0) = \int_0^t f_0(s, u_0) ds$ .

As  $U_1$  is degenerate,

$$\begin{aligned} & \frac{\int f_0(t, U_0) \frac{\exp(-\int_0^t f_0(s, U_0) f_1(s, a) ds)}{\int \exp(-\int_0^t f_0(s, U_0) f_1(s, a) ds) dF_{U_0}} dF_{U_0}}{\left( \int f_0(t, U_0) \frac{\exp(-\int_0^t f_0(s, U_0) ds)}{\int \exp(-\int_0^t f_0(s, U_0) ds) dF_{U_0}} dF_{U_0} \right)^{-1}} f_1(t, a) \\ &= \frac{\mathbb{E}[f_0(t, U_0) \mid T^a \geq t]}{\mathbb{E}[f_0(t, U_0) \mid T^0 \geq t]} f_1(t, a). \end{aligned}$$

□

### A.4 Proof of lemma 1

**Proof** By Bayes rule, the probability density of  $U_0$  given  $T^a \geq t$ ,  $f(u_0 \mid T^a \geq t)$  equals

$$\begin{aligned} f_{U_0 \mid T^a \geq t}(u_0) &= \frac{\mathbb{P}(T^a \geq t \mid U_0 = u_0) f(u_0)}{\int \mathbb{P}(T^a \geq t \mid U_0) dF_{U_0}} \\ &= \frac{\exp(-u_0 \int_0^t \lambda_0(s) f_1(s, a) ds) f(u_0)}{\int \exp(-U_0 \int_0^t \lambda_0(s) f_1(s, a) ds) dF_{U_0}}. \end{aligned}$$

So that the Laplace transform of  $U_0 \mid T^a \geq t$  can be written as

$$\begin{aligned} \mathcal{L}_{U_0 \mid T^a \geq t}(c) &= \mathbb{E}[\exp(-U_0 c) \mid T^a \geq t] \\ &= \int \exp(-U_0 c) dF_{U_0 \mid T^a \geq t}(u_0) \end{aligned}$$

$$\begin{aligned}
 &= \int \exp(-u_0 c) \frac{\exp\left(-u_0 \int_0^t \lambda_0(s) f_1(s, a) ds\right) f(u_0)}{\int \exp\left(-U_0 \int_0^t \lambda_0(s) f_1(s, a) ds\right) dF_{U_0}} du_0 \\
 &= \int \frac{\exp\left(-u_0(c + \int_0^t \lambda_0(s) f_1(s, a) ds)\right) f(u_0)}{\int \exp\left(-U_0 \int_0^t \lambda_0(s) f_1(s, a) ds\right) dF_{U_0}} du_0 \\
 &= \frac{\mathbb{E}\left[\exp\left(-U_0(c + \int_0^t \lambda_0(s) f_1(s, a) ds)\right)\right]}{\mathbb{E}\left[\exp\left(-U_0 \int_0^t \lambda_0(s) f_1(s, a) ds\right)\right]} \\
 &= \frac{\mathcal{L}_{U_0}(c + \int_0^t \lambda_0(s) f_1(s, a) ds)}{\mathcal{L}_{U_0}(\int_0^t \lambda_0(s) f_1(s, a) ds)}.
 \end{aligned}$$

Since for a random variable  $X$ ,  $\mathbb{E}[X] = -\mathcal{L}'_X(0)$ ,

$$\mathbb{E}[U_0 \mid T^a \geq t] = -\mathcal{L}'_{U_0|T^a \geq t}(0) = -\frac{\mathcal{L}'_{U_0}\left(\int_0^t \lambda_0(s) f_1(s, a) ds\right)}{\mathcal{L}_{U_0}\left(\int_0^t \lambda_0(s) f_1(s, a) ds\right)}.$$

□

## A.5 Proof of corollary 2

**Proof** By Theorem 2,

$$\frac{\lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T^a \in [t, t+h] \mid T^a \geq t)}{\lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T^0 \in [t, t+h] \mid T^0 \geq t)} = \frac{\mathbb{E}[f_0(t, U_0) f_1(t, U_1) \mid T^a \geq t]}{\mathbb{E}[f_0(t, U_0) \mid T^0 \geq t]}.$$

and equals

$$\begin{aligned}
 &\int f_0(t, U_0) f_1(t, U_1, a) \frac{\exp(-\Lambda^a(t, U_0, U_1))}{\int \exp(-\Lambda^a(t, U_0, U_1)) dF_{(U_0, U_1)}} dF_{(U_0, U_1)} \\
 &\left( \int f_0(t, U_0) \frac{\exp(-\Lambda^0(t, U_0))}{\int \exp(-\Lambda^0(t, U_0)) dF_{U_0}} dF_{U_0} \right)^{-1},
 \end{aligned}$$

where  $\Lambda^a(t, u_0, u_1) = \int_0^t f_0(s, u_0) f_1(s, u_1, a) ds$  and thus  $\Lambda^0(t, u_0) = \int_0^t f_0(s, u_0) ds$ . As now  $U_0$  is degenerate,

$$\int f_1(t, U_1, a) \frac{\exp(-\Lambda^a(t, U_1))}{\int \exp(-\Lambda^a(t, U_1)) dF_{U_1}} dF_{U_1} = \mathbb{E}[f_1(t, U_1, a) \mid T^a \geq t],$$

where  $\Lambda^a(t, u_1) = \int_0^t \lambda_0(s) f_1(s, u_1, a) ds$ .

□



## B Laplace transforms

### B.1 Gamma

If  $X \sim \Gamma(k, \theta)$ , then  $\mathbb{E}[X] = k\theta$ ,  $\text{var}[X] = k\theta^2$ ,

$$\begin{aligned}\mathcal{L}_X(c) &= (1 + \theta c)^{-k}, \\ \mathcal{L}'_X(c) &= -\frac{\theta k}{\theta c + 1} \mathcal{L}_X(c),\end{aligned}$$

and

$$\mathcal{L}''_X(c) = \frac{\theta^2 k(k+1)}{(\theta c + 1)^2} \mathcal{L}_X(c).$$

When  $f_\lambda(t, U_{0i}, U_{1i}, a) = U_{0i} \lambda_0(t) f_1(t, a)$  and  $U_0 \sim \Gamma(k, \theta)$ , by Lemma 1,

$$\mathbb{E}[U_0 \mid T^a \geq t] = \frac{\theta k}{\theta \Lambda^a(t) + 1}, \quad (17)$$

where  $\Lambda^a(t) = \int_0^t \lambda_0(s) f_1(s, a) ds$ . If  $\mathbb{E}[U_0] = \mu$  and  $\text{var}(U_0) = \theta_0$ , then  $k = \frac{\mu^2}{\theta_0}$ ,  $\theta = \frac{\theta_0}{\mu}$  and as such

$$\mathbb{E}[U_0 \mid T^a \geq t] = \frac{\mu}{\frac{\theta_0}{\mu} \Lambda^a(t) + 1}. \quad (18)$$

In particular, when  $k = \theta^{-1}$ , then  $\mathbb{E}[U_0] = 1$ ,  $\text{var}(U_0) = \theta_0$  and

$$\mathbb{E}[U_0 \mid T^a \geq t] = (\theta_0 \Lambda^a(t) + 1)^{-1}. \quad (19)$$

### B.2 Inverse Gaussian

If  $X \sim \text{IG}(\mu, \lambda)$ , then  $\mathbb{E}[X] = \mu$ ,  $\text{var}[X] = \frac{\mu^3}{\lambda}$ ,

$$\begin{aligned}\mathcal{L}_X(c) &= \exp\left(\frac{\lambda}{\mu} \left(1 - \sqrt{1 + \frac{2\mu^2 c}{\lambda}}\right)\right), \\ \mathcal{L}'_X(c) &= -\frac{\mu}{\sqrt{\frac{2\mu^2 c}{\lambda} + 1}} \mathcal{L}_X(c),\end{aligned}$$

and

$$\mathcal{L}''_X(c) = \left( \mu^2 \frac{\lambda}{\lambda + 2\mu^2 c} + \frac{\mu}{\lambda \left(\frac{2\mu^2 c}{\lambda} + 1\right)^{\frac{3}{2}}} \right) \mathcal{L}_X(c).$$

When  $f_\lambda(t, U_{0i}, U_{1i}, a) = U_{0i} \lambda_0(t) f_1(t, a)$  and  $U_0 \sim \text{IG}(\mu, \lambda)$ , by Lemma 1,

$$\mathbb{E}[U_0 \mid T^a \geq t] = \frac{\mu}{\sqrt{2\Lambda^a(t) \frac{\mu^2}{\lambda} + 1}}, \quad (20)$$

where  $\Lambda^a(t) = \int_0^t \lambda_0(s) f_1(s, a) ds$ . If  $\text{var}(U_0) = \theta_0$ , then  $\lambda = \frac{\mu^3}{\theta_0}$  and as such

$$\mathbb{E}[U_0 \mid T^a \geq t] = \frac{\mu}{\sqrt{2\Lambda^a(t) \frac{\theta_0}{\mu} + 1}}. \quad (21)$$

In particular when  $\mu = 1$  and  $\lambda = \theta_0^{-1}$ , then  $\mathbb{E}[U_0] = 1$ ,  $\text{var}(U_0) = \theta_0$  and

$$\mathbb{E}[U_0 \mid T^a \geq t] = (2\theta_0 \Lambda^a(t) + 1)^{-\frac{1}{2}}. \quad (22)$$

### B.3 Compound poisson

If  $X \sim \text{CPoi}(\rho, \eta, \nu)$ , then  $X = \sum_{i=1}^N Y_i$ , where  $N \sim \text{Poi}(\rho)$ , and  $Y \sim \Gamma(\eta, \nu)$ ,  $\mathbb{E}[X] = \rho\eta\nu$ ,  $\text{var}[X] = \rho\eta\nu^2 + \eta^2\nu^2\rho$ ,

$$\begin{aligned} \mathcal{L}_X(c) &= \exp\left(\rho \left(\left(\frac{\nu^{-1}}{\nu^{-1} + c}\right)^\eta - 1\right)\right), \\ \mathcal{L}'_X(c) &= -\rho\eta\nu \left(\frac{1}{c\nu + 1}\right)^{\eta+1} \mathcal{L}_X(c), \end{aligned}$$

and

$$\mathcal{L}''_X(c) = \eta\nu^2\rho \left(\frac{1}{c\nu + 1}\right)^{\eta+2} \left(\eta\rho \left(\frac{1}{c\nu + 1}\right)^\eta + \eta + 1\right) \mathcal{L}_X(c).$$

When  $f_\lambda(t, U_{0i}, U_{1i}, a) = U_{0i} \lambda_0(t) f_1(t, a)$  and  $U_0 \sim \text{CPoi}(\rho, \eta, \nu)$ , by Lemma 1,

$$\mathbb{E}[U_0 \mid T^a \geq t] = \rho\eta\nu \left(\frac{1}{\Lambda^a(t)\nu + 1}\right)^{\eta+1}, \quad (23)$$

where  $\Lambda^a(t) = \int_0^t \lambda_0(s) f_1(s, a) ds$ . Let  $\mathbb{E}[U_0] = \mu$  and  $\text{var}(U_0) = \theta_0$ , then  $\rho = \frac{\mu^2(1+\eta)}{\eta\theta_0}$ ,  $\nu = \frac{\theta_0}{\mu(1+\eta)}$  and as such

$$\mathbb{E}[U_0 \mid T^a \geq t] = \mu \left(\frac{1}{\Lambda^a(t) \frac{\theta_0}{\mu(1+\eta)} + 1}\right)^{\eta+1}. \quad (24)$$

In particular, when  $\eta = \frac{1}{2}$ ,  $\rho = 3\theta_0^{-1}$  and  $\nu = \frac{2}{3}\theta_0$ , then  $\mathbb{E}[U_0] = 1$ ,  $\text{var}(U_0) = \theta_0$  and

$$\mathbb{E}[U_0 \mid T^a \geq t] = (1 + \Lambda^a(t) \frac{2}{3}\theta_0)^{-\frac{3}{2}}. \quad (25)$$

## B.4 Discrete

Let  $\mathbb{P}(X = \mu_i) = p_i$  for  $i > 0$  and  $i \leq k$ ,

$$\mathcal{L}_X(c) = \sum_{i=1}^n p_i \exp(-c\mu_i),$$

$$\mathcal{L}'_X(c) = \sum_{i=1}^n -\mu_i p_i \exp(-c\mu_i),$$

and

$$\mathcal{L}''_X(c) = \sum_{i=1}^n \mu_i^2 p_i \exp(-c\mu_i).$$

Furthermore, if  $Y \sim F$ , then

$$\mathcal{L}_{XY}(c) = \sum_{i=1}^n p_i \mathcal{L}_Y(c\mu_i),$$

$$\mathcal{L}'_{XY}(c) = \sum_{i=1}^n p_i \mu_i \mathcal{L}'_Y(c\mu_i),$$

and

$$\mathcal{L}''_{XY}(c) = \sum_{i=1}^n p_i \mu_i^2 \mathcal{L}''_Y(c\mu_i).$$

Let  $U_1$  equal  $\mu_1$ ,  $\mu_2$  or 1 with probability  $p_1$ ,  $p_2$  and  $1 - p_1 - p_2$  respectively. If

$$p_2 = \frac{(\mu - \mu_1 p_1 + p_1 - 1)^2}{\mu^2 - 2\mu - \mu_1^2 p_1 + 2\mu_1 p_1 - p_1 + \theta_1 + 1},$$

and

$$\mu_2 = \frac{\mu^2 - \mu - \mu_1^2 p_1 + \mu_1 p_1 + \theta_1}{\mu - \mu_1 p_1 + p_1 - 1},$$

such that  $p_2 \in [0, 1]$  and  $\mu_2 \geq 1$ , then  $\mathbb{E}[U_1] = \mu$  and  $\text{var}(U_1) = \theta_1$ .

When,  $f_\lambda(t, U_{0i}, U_{1i}, a) = \lambda_0(t)(U_{1i})^a f_1(t, a)$ , by Lemma 1,  $\mathbb{E}[U_1 \mid T^1 \geq t]$  equals

$$\frac{\mu_1 p_1 \exp(-\Lambda^1(t)\mu_1) + \mu_2 p_2 \exp(-\Lambda^1(t)\mu_2) + (1 - p_1 - p_2) \exp(-\Lambda^1(t))}{p_1 \exp(-\Lambda^1(t)\mu_1) + p_2 \exp(-\Lambda^1(t)\mu_2) + (1 - p_1 - p_2) \exp(-\Lambda^1(t))},$$

where  $\Lambda^1(t) = \int_0^t \lambda_0(s) f_1(s, 1) ds$ .

Furthermore, when  $f_\lambda(t, U_{0i}, U_{1i}, a) = U_{0i} \lambda_0(t) (U_{1i})^a f_1(t, a)$ , and  $U_0$  and  $U_1$  are independent, by Lemma 1,  $\mathbb{E}[U_0 U_1 \mid T^1 \geq t]$  equals

$$\frac{p_1 \mu_1 \mathcal{L}'_{U_0}(\Lambda^1(t) \mu_1) + p_2 \mu_2 \mathcal{L}'_{U_0}(\Lambda^1(t) \mu_2) + (1 - p_1 - p_2) \mathcal{L}'_{U_0}(\Lambda^1(t))}{p_1 \mathcal{L}_{U_0}(\Lambda^1(t) \mu_1) + p_2 \mathcal{L}_{U_0}(\Lambda^1(t) \mu_2) + (1 - p_1 - p_2) \mathcal{L}_{U_0}(\Lambda^1(t))}.$$

## C Supplementary tables

**Table 2** Conditional expectations and resulting SMCHR when the CHR equals  $\mu$  for different frailty distributions such that  $\mathbb{E}[U_0] = 1$  and  $\text{var}(U_0) = \theta_0$  (in absence of effect modification)

	$U_0 \sim \Gamma(\theta_0^{-1}, \theta_0)$	$IG(1, \theta_0^{-1})$	$CPoi\left(3\theta_0^{-1}, \frac{1}{2}, \frac{2}{3}\theta_0\right)$
$\mathbb{E}[U_0 \mid T^1 \geq t]$	$\frac{60}{60 + \mu t^3 \theta_0}$	$\sqrt{\frac{30}{30 + \mu t^3 \theta_0}}$	$90^{\frac{3}{2}} (90 + \mu t^3 \theta_0)^{-\frac{3}{2}}$
$\mathbb{E}[U_0 \mid T^0 \geq t]$	$\frac{60}{60 + t^3 \theta_0}$	$\sqrt{\frac{30}{30 + t^3 \theta_0}}$	$90^{\frac{3}{2}} (90 + t^3 \theta_0)^{-\frac{3}{2}}$
$\frac{\lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T^1 \in (t, t+h) \mid T^1 \geq t)}{\lim_{h \rightarrow 0} h^{-1} \mathbb{P}(T^0 \in [t, t+h) \mid T^0 \geq t)}$	$1 + \frac{60(\mu-1)}{60 + \mu t^3 \theta_0}$	$\mu \frac{\sqrt{30 + t^3 \theta_0}}{\sqrt{30 + \mu t^3 \theta_0}}$	$\mu \left(\frac{90 + t^3 \theta_0}{90 + \mu t^3 \theta_0}\right)^{\frac{3}{2}}$

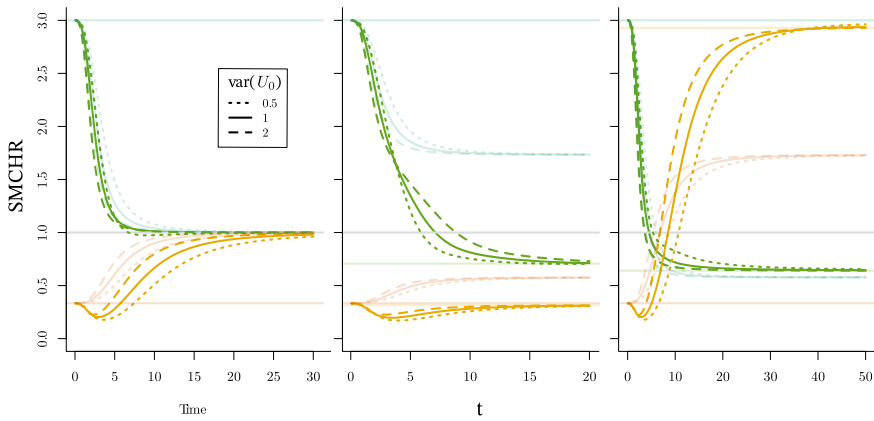
**Table 3** Conditional expectation of the individual effect modifier  $U_1$  when the CHR equals  $\mu$  for different modifier distributions such that  $\mathbb{E}[U_1] = \mu$  and  $\text{var}(U_0) = \theta_1$  (in absence of frailty)

$\mathbb{E}[U_1 \mid T^{-1} \geq t]$	$U_1 \sim$		
	$\Gamma\left(\frac{\mu}{\theta_1}, \frac{\theta_1}{\mu}\right)$	$\text{IG}\left(\mu, \frac{\mu^3}{\theta_1}\right)$	$\text{CPoi}\left(3\frac{\mu^2}{\theta_1}, \frac{1}{2}, \frac{2\theta_1}{3\mu}\right)$
			$\text{BHN}(p_1, \mu_1, p_2, \mu_2)$
	$\frac{60\mu^2}{\theta_1 t^3 + 60\mu}$	$\frac{\mu\sqrt{30\mu}}{\sqrt{t^3\theta_1 + 30\mu}}$	$\mu\left(\frac{\theta_1 t^3}{90\mu} + 1\right)^{-\frac{3}{2}}$
			$\frac{p_1\mu_1 + p_2\mu_2 \exp\left(\frac{-t^3(\mu_2 - \mu_1)}{60}\right) + (1 - p_1 - p_2) \exp\left(\frac{-t^3(1 - \mu_1)}{60}\right)}{p_1 + p_2 \exp\left(\frac{-t^3(\mu_2 - \mu_1)}{60}\right) + (1 - p_1 - p_2) \exp\left(\frac{-t^3(1 - \mu_1)}{60}\right)}$

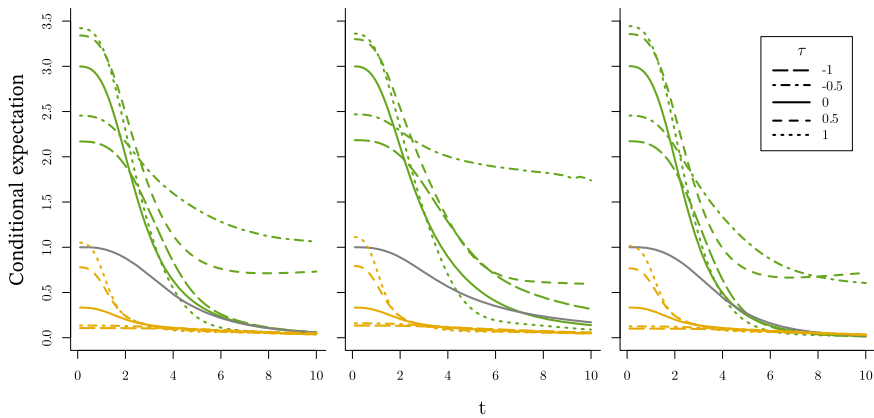
**Table 4** Conditional expectation  $\mathbb{E}[U_0 U_1 \mid T^1 \geq t]$  and its limiting forms when  $U_1$  follows a BHN( $p_1, \mu_1, p_2, \mu_2$ ) distribution for different unit-expectations frailty distributions with  $\text{var}(U_0) = \theta_0$  while  $U_1 \perp U_0$ . Here  $p_3 = 1 - p_1 - p_2$  and  $\mu_3 = 1$ . For comparison  $\mathbb{E}[U_0 \mid T^0 \geq t]$ , as shown before in Table 2, is also presented

$U_0 \sim$	$\mathbb{E}[U_1 U_0 \mid T^1 \geq t]$	$\mathbb{E}[U_0 \mid T^0 \geq t]$
$\Gamma(\theta_0^{-1}, \theta_0)$	$\frac{\sum_{i=1}^3 p_i \mu_i \left(1 + \frac{\theta_0 t^3}{60} \mu_i\right)^{-(1+\theta_0^{-1})}}{\sum_{i=1}^3 p_i \left(1 + \frac{\theta_0 t^3}{60} \mu_i\right)^{-\theta_0^{-1}}}$ $= (\theta_0 \frac{t^3}{60})^{-1} + o\left(t^{-3}\right)$	$\frac{60}{60+t^3 \theta_0}$
$\text{IG}(1, \theta_0^{-1})$	$\frac{\sum_{i=1}^3 p_i \mu_i \left(2 \theta_0 \frac{t^3}{60} \mu_i\right)^{-\frac{1}{2}} \exp\left(\theta_0^{-1} \left(1 - \sqrt{1 + \theta_0 2 \frac{t^3}{60} \mu_i}\right)\right)}{\sum_{i=1}^3 p_i \exp\left(\theta_0^{-1} \left(1 - \sqrt{1 + \theta_0 2 \frac{t^3}{60} \mu_i}\right)\right)}$ $= \sqrt{\mu_1} \sqrt{\frac{30}{t^3 \theta_0}} + o\left(t^{-\frac{3}{2}}\right)$	$\sqrt{\frac{30}{30+t^3 \theta_0}}$
$\text{CPoi}\left(3\theta_0^{-1}, \frac{1}{2}, \frac{2}{3}\theta_0\right)$	$\frac{\sum_{i=1}^3 p_i \mu_i \left(\frac{3}{3+2\theta_0 \mu_i} \frac{t^3}{60}\right)^{\frac{3}{2}} \exp\left(3\theta_0^{-1} \sqrt{\frac{3}{3+2\theta_0 \mu_i} \frac{t^3}{60}} - 1\right)}{\sum_{i=1}^3 p_i \exp\left(3\theta_0^{-1} \sqrt{\frac{3}{3+2\theta_0 \mu_i} \frac{t^3}{60}} - 1\right)}$ $= 90 \frac{3}{2} \left(\theta_0 t^3\right)^{-\frac{3}{2}} \sum_{i=1}^3 \frac{p_i}{\sqrt{\mu_i}} + o\left(t^{-\frac{9}{2}}\right)$	$90 \frac{3}{2} \left(90 + t^3 \theta_0\right)^{-\frac{3}{2}}$

## D Supplementary Figures



**Fig. 7** SMCHR over time when  $\lambda_i^a(t) = U_{0i}(U_{1i})^a \frac{t^2}{20}$  for a unit-variance BHN distributed  $U_1$  with  $\mathbb{E}[U_1] = 3$  (opaque green) and  $\mathbb{E}[U_1] = \frac{1}{3}$  (opaque orange) when  $U_0$  follows a gamma (left), inverse Gaussian (middle) or compound Poisson (right) distribution with variance 0.5 (dotted), 1 (solid) or 2 (dashed) respectively. For comparison, the lines presented in Fig. 1 as well as the limits of the SMCHR are represented by transparent lines



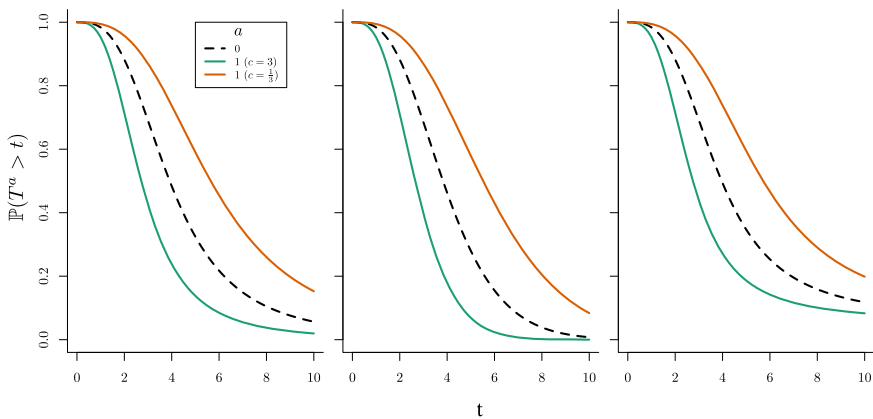
**Fig. 8**  $\mathbb{E}[U_0 U_1 \mid T^1 \geq t]$  over time for  $\lambda_i^a(t) = U_{0i}(U_{1i})^a \frac{t^2}{20}$ , for a unit-variance BHN distributed  $U_1$  with  $\mathbb{E}[U_1] = 3$  (green) or  $\mathbb{E}[U_1] = \frac{1}{3}$  (orange),  $U_0$  follows a gamma (left), inverse Gaussian (middle) or compound Poisson (right) distribution and the joint distribution of  $U_0$  and  $U_1$  follows from a Gaussian copula with varying Kendall's  $\tau$  correlation coefficients (see legend). Furthermore, the evolution of  $\mathbb{E}[U_0 \mid T^0 \geq t]$  is presented (grey)

## E Survival curves examples

The survival curves of  $T^0$  and  $T^1$  for the examples presented in this paper can be expressed in terms of the Laplace transforms presented in the previous section as shown in Table 5, where  $\Lambda^0(t) = \frac{t^3}{60}$ . For the example where  $U_0 \not\perp U_1$ , the survival curve for  $T^1$  is obtained empirically from simulation. For data from a RCT,  $T^a \stackrel{d}{=} T \mid A=a$ .

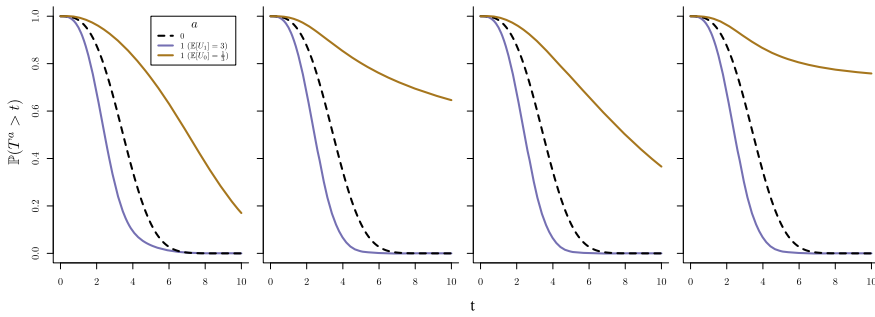
**Table 5** Survival curves for  $T^0$  and  $T^1$  for different  $\lambda_i^a(t)$  where  $\Lambda^0(t) = \frac{t^3}{60}$

$\lambda_i^a(t)$	$S_{T^0}(t)$	$S_{T^1}(t)$
$U_{0i} \frac{t^2}{20} \mu^a$	$\mathcal{L}_{U_0}(\Lambda^0(t))$	$\mathcal{L}_{U_0}(\mu \Lambda^0(t))$
$\frac{t^2}{20} (U_{1i})^a$	$\exp(-\Lambda^0(t))$	$\mathcal{L}_{U_1}(\Lambda^0(t))$
$U_{0i} \frac{t^2}{20} (U_{1i})^a$	$\mathcal{L}_{U_0}(\Lambda^0(t))$	$\mathcal{L}_{U_0 U_1}(\Lambda^0(t))$

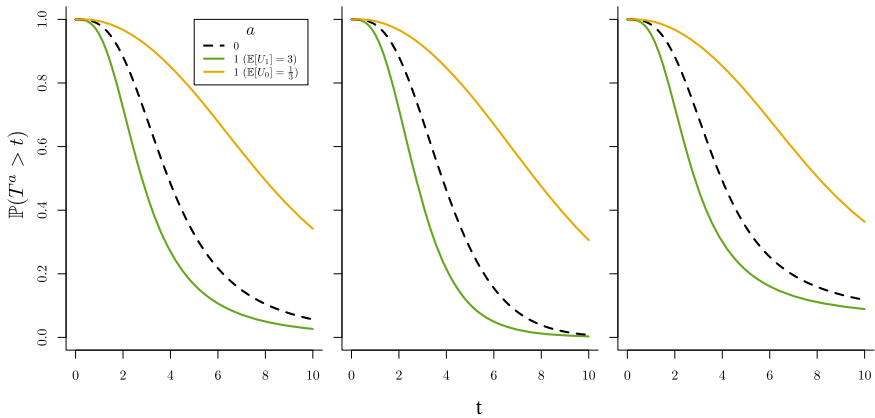


**Fig. 9** Survival curves for  $T^0$  (black), and  $T^1$  (coloured) when  $\lambda_i^a(t) = U_{0i} \frac{t^2}{20} \mu^a$  for  $\mu = 3$  (green) and  $\mu = \frac{1}{3}$  (orange), when  $U_0$  follows a gamma, inverse Gaussian or compound Poisson distribution (from left to right) with variance 1. The corresponding SMCHR curves were presented in Fig. 1

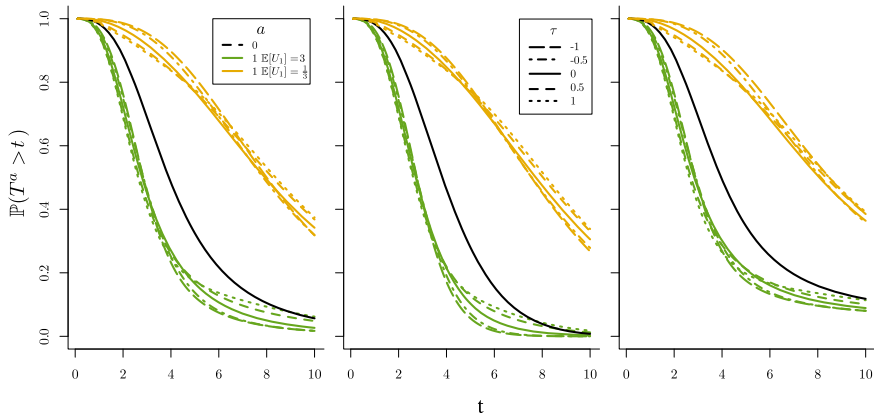




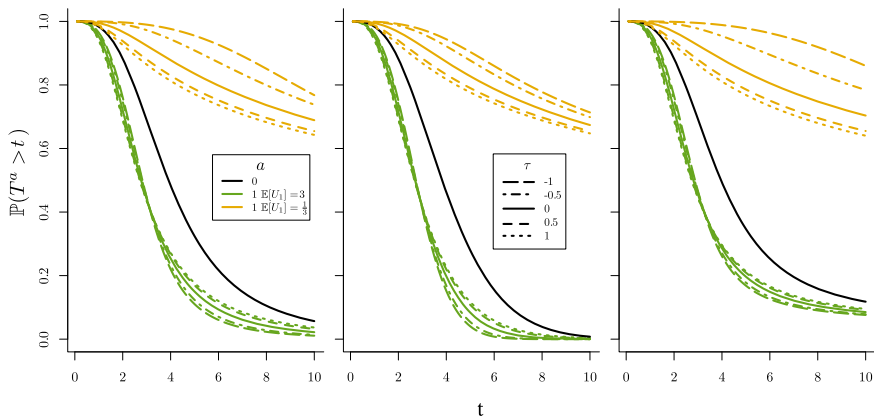
**Fig. 10** Survival curves for  $T^0$  (black), and  $T^1$  (coloured) when  $\lambda_i^a(t) = U_{0i} \frac{t^2}{20} (U_{1i})^a$ , when  $U_1$  follows a BHN, gamma, inverse Gaussian or compound Poisson distribution (from left to right) with expectation 3 (blue) or  $\frac{1}{3}$  (brown) and variance 1. The corresponding SMCHR curves were presented in Fig. 2



**Fig. 11** Survival curves for  $T^0$  (black), and  $T^1$  (coloured) when  $\lambda_i^a(t) = U_{0i} (U_{1i})^a \frac{t^2}{20}$  for a unit-variance BHN distributed  $U_1$  with  $\mathbb{E}[U_1] = 3$  (green) and  $\mathbb{E}[U_1] = \frac{1}{3}$  (orange) when  $U_0$  follows a gamma, inverse Gaussian or compound Poisson distribution (from left to right) with variance 1. The corresponding SMCHR curves were presented in Fig. 3



**Fig. 12** Survival curves for  $T^0$  (black), and  $T^1$  (coloured) when  $\lambda_i^a(t) = U_{0i}(U_{1i})^a \frac{t^2}{20}$ , for a unit-variance BHN distributed  $U_1$  with  $\mathbb{E}[U_1] = 3$  (green) or  $\mathbb{E}[U_1] = \frac{1}{3}$  (orange), when  $U_0$  follows a gamma, inverse Gaussian or compound Poisson distribution (from left to right) with variance 1 and the joint distribution of  $U_0$  and  $U_1$  follows from a Gaussian copula with varying Kendall's  $\tau$ . The corresponding SMCHR curves were presented in Fig. 4



**Fig. 13** Survival curves for  $T^0$  (black), and  $T^1$  (coloured) when  $\lambda_i^a(t) = U_{0i}(U_{1i})^a \frac{t^2}{20}$ , for a unit-variance gamma distributed  $U_1$  with  $\mathbb{E}[U_1] = 3$  (green) or  $\mathbb{E}[U_1] = \frac{1}{3}$  (orange) when  $U_0$  follows a gamma, inverse Gaussian or compound Poisson distribution (from left to right) with variance 1. and the joint distribution of  $U_0$  and  $U_1$  follows from a Gaussian copula with varying Kendall's  $\tau$ . The corresponding SMCHR curves were presented in Fig. 5

## References

- Aalen OO, Borgan Ø, Gjessing HK (2008) Survival and Event History Analysis, 1st edn. Springer, New York
- Aalen OO, Cook RJ, Røysland K (2015) Does Cox analysis of a randomized survival study yield a causal treatment effect? Lifetime Data Anal 21(4):579–593
- Balan TA, Putter H (2020) A tutorial on frailty models. Stat Methods Med Res 29(11):3424–3454

- Bartlett JW, Morris TP, Stensrud MJ, Daniel RM, Vansteelandt SK, Burman CF (2020) The hazards of period specific and weighted hazard ratios. *Stat Biopharm Res* 12(4):518–519
- Bennett S (1983) Analysis of survival data by the proportional odds model. *Stat Med* 2(2):273–277
- Bongers S, Forré P, Peters J, Mooij JM (2021) Foundations of structural causal models with cycles and latent variables. *Ann Stat* 49(5):2885–2915
- Boyd AP, Kittelson JM, Gillen DL (2012) Estimation of treatment effect under non-proportional hazards and conditionally independent censoring. *Stat Med* 31(28):3504–3515
- Cox DR (1972) Regression models and life-tables. *J Roy Stat Soc B* 34(2):187–220
- Daniel R, Zhang J, Farewell D (2021) Making apples from oranges: Comparing noncollapsible effect estimators and their standard errors after adjustment for different covariate sets. *Biom J* 63(3):528–557
- De Neve J, Gerds TA (2020) On the interpretation of the hazard ratio in Cox regression. *Biom J* 62(3):742–750
- Didelez V, Stensrud MJ (2021) On the logic of collapsibility for causal effect measures. *Biometrical Journal*
- Hernán MA (2010) The hazards of hazard ratios. *Epidemiology* 21(1):13–15
- Hernán MA, Brumback B, Robins JM (2000) Marginal Structural Models to Estimate the Causal Effect of Zidovudine on the Survival of HIV-Positive Men. *Epidemiology* 11(5)
- Hernán MA, Brumback B, Robins JM (2001) Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *J Am Stat Assoc* 96(454):440–448
- Hernán MA, Cole SR, Margolick J, Cohen M, Robins JM (2005) Structural accelerated failure time models for survival analysis in studies with time-varying treatments. *Pharmacoepidemiol Drug Saf* 14(7):477–491
- Hess KR (1994) Assessing time-by-covariate interactions in proportional hazards regression models using cubic spline functions. *Stat Med* 13(10):1045–1062
- Martinussen T, Vansteelandt S (2013) On collapsibility and confounding bias in Cox and Aalen regression models. *Lifetime Data Anal* 19(3):279–296
- Martinussen T, Vansteelandt S, Andersen P (2020) Subtleties in the interpretation of hazard contrasts. *Lifetime Data Anal* 26(4):833–855
- Nelsen RB (2006) *An Introduction to Copulas*, 2nd edn. Springer
- Neyman J (1990) On the Application of Probability Theory to Agricultural Experiments. *Essay on Principles*. Section 9. *Stat Sci* 5(4):465–472
- Pearl J (2009) *Causality: Models, reasoning, and inference*, 2nd edn. Cambridge University Press, Cambridge
- Peters J, Janzing D, Schölkopf B (2018) *Elements of causal inference: foundations and learning algorithms*. The MIT Press, Cambridge, Massachusetts
- Post RAJ, van den Heuvel ER, Putter H (2024) Bias of the additive hazard model in the presence of causal effect heterogeneity. *Lifetime Data Anal*. <https://doi.org/10.1007/s10985-024-09616-z>
- Rubin DB (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol* 66(5):688–701
- Ryalen PC, Stensrud MJ, Røysland K (2018) Transforming cumulative hazard estimates. *Biometrika* 105(4):905–916
- Schemper M, Wakounig S, Heinze G (2009) The estimation of average hazard ratios by weighted cox regression. *Stat Med* 28(19):2473–2489
- Sjölander A, Dahlqvist E, Zetterqvist J (2016) A note on the noncollapsibility of rate differences and rate ratios. *Epidemiology* 27(3):356–359
- Sklar A (1959) Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de statistique de l'Université de Paris* 8:229–231
- Stensrud MJ, Hernán MA (2020) Why test for proportional hazards? *JAMA* 323(14):1401–1402
- Stensrud MJ, Valberg M, Røysland K, Aalen OO (2017) Exploring selection bias by causal frailty models: The magnitude matters. *Epidemiology* 28(3):379–386
- Stensrud MJ, Aalen JM, Aalen OO, Valberg M (2018) Limitations of hazard ratios in clinical trials. *Eur Heart J* 40(17):1378–1383
- Therneau TM, Grambsch PM (2000) *Modeling Survival Data: Extending the Cox Model*, 1st edn. Springer
- Valberg M, Stensrud MJ, Aalen OO (2018) The surprising implications of familial association in disease risk. *BMC Public Health* 18(1):1–9
- Wei G, Schaibel DE (2008) Estimating cumulative treatment effects in the presence of nonproportional hazards. *Biometrics* 64(3):724–732

Xu R, O'Quigley J (2000) Estimating average regression effect under non-proportional hazards. *Biostatistics* 1(4):423–439

Young JG, Stensrud MJ, Tchetgen Tchetgen EJ, Hernán MA (2020) A causal framework for classical statistical estimands in failure-time settings with competing events. *Stat Med* 39(8):1199–1236

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.