Check for updates

Biometrical Journal

WILEY-VCH

Biometrical Journal

RESEARCH ARTICLE OPEN ACCESS

Revisiting Hazard Ratios: Can We Define Causal Estimands for Time-Dependent Treatment Effects?

Dominic Edelmann

Division of Biostatistics, German Cancer Research Center, Heidelberg, Germany

Correspondence: Dominic Edelmann (dominic.edelmann@dkfz.de)

Received: 20 July 2025 | **Revised:** 10 November 2025 | **Accepted:** 14 November 2025

ABSTRACT

In this paper, some aspects concerning the causal interpretation of hazard contrasts are revisited. It is first investigated, in which sense the hazard ratio constitutes a causal effect. It is demonstrated that the hazard ratio at a timepoint t represents a causal effect for the population at baseline, but in general not for any population at risk at time t. Moreover, the scenario is studied, in which the survival curves coincide up to some timepoint t and then separate. This investigation provides valuable insight both on the causal interpretation of the conventional hazard ratio and on properties of the recently proposed causal hazard ratio. The findings suggest that, without making further assumptions, there is in general no meaningful estimand for a treatment effect at time $t > 0$. It is therefore advocated to develop alternative estimands grounded in medically plausible assumptions about the joint distribution of counterfactual survival times.

1 | Introduction

The Cox proportional hazards regression (Cox 1972) remains by far the most widely used model for time-to-event outcomes, and the hazard ratio has consequently become one of the most frequently reported effect measures in medical research. The very term "effect measure" implicitly suggests a causal interpretation, and it is common practice to treat changes in the hazard ratio over time as if they directly reflected time-dependent treatment effects (Jachno et al. 2022; Fay and Li 2024). Such interpretations have long been recognized as problematic (Keiding 1999). More recent work (Hernán 2010; Martinussen et al. 2020; Martinussen 2022) has sharpened these concerns, showing that even in a randomized trial, a change in the hazard ratio does not necessarily reflect a change in the treatment effect. This work is motivated from the article by Beyersmann et al. (2025).

Fifteen years after Hernán's landmark paper, "The hazards of hazard ratios," (Hernán 2010) the article by Beyersmann et al. (2025)—and the special issue surrounding it—may well mark a turning point for the inference of time-dependent effects. While until now, the literature has mainly focused on describing the pitfalls of interpreting the hazard ratio as a causal treatment effect, Beyersmann et al. (2025) proposed a more pragmatic course, offering guidance for a cautious yet meaningful causal interpretation of hazard contrasts.

Most authors agree that on one hand the hazard ratio represents—at least in a certain sense—a causal effect and on the other hand can lead to counterintuitive conclusions when naively interpreted as treatment effect at a certain time point. Yet there is still substantial confusion and apparent dissonance in the literature. One reason for this confusion is that researchers often fail to distinguish properly between (1) whether the hazard ratio mathematically represents a causal effect and (2) whether it admits a meaningful or intuitive causal interpretation. This conflation of mathematical and semantic issues leads to misunderstandings. Indeed, some papers do not even define the term "causal effect," making any strict logical analysis impossible. In this paper, such confusion is avoided

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Biometrical Journal* published by Wiley-VCH GmbH.

by separating mathematical results from their interpretation as clearly as possible. The mathematical setup for our discussion is described in the following paragraph.

1.1 | Mathematical Setup

Assumptions 1. Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $T_0 : \Omega \to (0, \infty)$ and $T_1 : \Omega \to (0, \infty)$ denote two measurable functions on $(\Omega, \mathcal{F})$ representing counterfactual survival outcomes. T_0 and T_1 are both assumed to be continuous with density functions $f_0(t)$ and $f_1(t)$, respectively. For $i = 1, 2$, the cumulative distribution functions is denoted by $F_i(t)$, the survival functions by $S_i(t) = 1 - F_i(t)$ and the hazard functions by

$$h_i(t) = \lim_{\Delta t \searrow 0} \frac{P(t < T_i \leq t + \Delta t | T_i \geq t)}{\Delta t} = \frac{f_i(t)}{S_i(t)}.$$

Moreover, it is assumed that there is an interval $[0, \tau]$ such that

$$f_0(t) > 0, f_1(t) > 0 \text{ for all } t \in [0, \tau].$$

The (conventional) **hazard ratio** (HR) between P_{T_1} and P_{T_0} at time $t \in [0, \tau]$ then exists and is given by

$$\text{HR}(t) = \frac{h_1(t)}{h_0(t)}. \tag{1}$$

Avoiding discussions about identifiability and estimation, we will take an "all-knowing" population perspective in this paper, meaning that the joint distribution of (T_0, T_1) is fully known. For this reason, no assumptions on randomization of treatment are required and the introduction of a treatment variable is omitted. Yet it makes sense to read this paper with a situation of a randomized trial in mind allowing for identification of the hazard ratio in Equation (1).

1.2 | Outline

In Section 2, we revisit the question in which sense the hazard ratio represents causal effect offering a simple and concise solution: for any $t \in [0, \tau]$, HR(t) is a causal effect for the full population at baseline ($t = 0$), but not for a meaningful subpopulation at risk at time t. Consequently, interpretation of HR(t) as time-dependent treatment effect is questionable.

In Section 3, we take a closer look at the case where the survival curves coincide up to some timepoint u. Although often dismissed as trivial, this setting provides valuable insight both on the causal interpretation of the conventional hazard ratio and on properties of the recently proposed causal hazard ratio (Martinussen et al. 2020; Martinussen 2022).

The findings of these sections suggest that—without additional structure—there is no generally useful or coherent definition of a treatment effect at time $t > 0$ and that assumptions on the dependence structure of T_0 and T_1 are required to enable inference for time-dependent treatment effects. The paper concludes with an outlook on future steps to establish statistical methods to infer time-dependent treatment effects in Sections 4 an 5.

Throughout this comment, the hazard ratio is exemplarily treated, all results similarly hold for the hazard difference or other time-dependent hazard contrasts; the described issues are not related to the non-collapsibility of the hazard ratio (Aalen et al. 2015).

2 | Causal Effect Properties of the Hazard Ratio

2.1 | Motivation

There is substantial confusion in the literature about the question in which sense the hazard ratio HR(t) in Equation (1) represents a causal effect. In this section, a simple and concise answer to this question is provided: for any $t > 0$, HR(t) is a causal effect for the full population at baseline ($t = 0$), but it is not a causal effect for any subpopulation

$$A_t \subseteq \{T_0 \geq t, T_1 \geq t\}.$$

The subsets A_t in turn are the only sensible subsets to consider when defining a treatment effect for a population at time t; there appears to be no reasonable way to define a treatment effect at time t for populations in which one (or both) counterfactual survival times can be smaller than t.

The subsequent arguments rely on the notion of a population causal effect introduced in Definition 2.2. This definition formalizes the characterization of such an effect as "any contrast between functionals of the marginal distributions of counterfactual outcomes" (Hernán and Robins 2010, p. 7). The reader is reminded that the general mathematical setup for this article has already been provided in Section 1.1.

2.2 | Mathematical Formulation

Definition 2.1. For any probability measure Q on a measurable space $(\Omega, \mathcal{F})$, let Q_X denote the induced probability measure, that is, $Q_X(A) := Q(X^{-1}(A))$ for $A \in \mathcal{F}$.

Definition 2.2. Let $\mathcal{P}$ denote a set of probability measures on a measurable space $(\Omega, \mathcal{F})$. Let Y_0 and Y_1 denote two measurable functions denoting counterfactual outcomes on $(\Omega, \mathcal{F})$ with $P_{Y_0}, P_{Y_1} \in \mathcal{P}$. Let $\phi : \mathcal{P} \to \mathbb{R}$ denote some functional and f denote a contrast function $f : \mathbb{R}^2 \to \mathbb{R}$. Then

$$\text{CE}(P_{Y_1}, P_{Y_0}) = f(\phi(P_{Y_1}), \phi(P_{Y_0}))$$

is called a **population causal effect** for comparing Y_1 and Y_0 under the probability distribution P.

The restriction to real-valued functionals and contrast functions is made for reasons of simplicity. In Appendix A, a proposition for a more general definition of causal effects is attempted.

Remark 2.3. The attentive reader may have realized that the precise definition of the term *contrast function* has been omitted. To avoid unnecessary technical discussions, we avoid an exact definition of this term. For the following results, one only requires the following properties:

15214036, 2025, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bimj.70100 by Martha Krebsforschungszentrum - Deutsches Krebsforschungszentrum (DKFZ) , Wiley Online Library on [02/03/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

1. The difference $f(x,y) = x - y$ and the ratio $f(x,y) = x/y$ are contrast functions.
2. For each contrast function f, there is an element $c_0 \in \{0,1\}$, such that $f(x,x) = c_0$ for all $x \in \mathbb{R}$.

Proposition 2.4. *Under Assumptions 1, for each $t \in [0,\tau]$, the hazard ratio $HR(t)$ is a population causal effect for comparing T_0 and T_1 under P.*

Proof. Let $\mathcal{P}$ be a suitably chosen set (e.g., $\mathcal{P} = \{P_{T_0}, P_{T_1}\}$). For probability measures Q with density function g and cumulative distribution function G, set

$$\phi(Q) = \frac{g(t)}{1 - G(t)} \qquad f(x,y) = x/y. \qquad \square$$

Proposition 2.5. *Under Assumption 1, for $t \in (0,\tau]$, $HR(t)$ is in general **not** a population causal effect for comparing T_0 and T_1 under $P(\cdot|A_t)$ for any subset*

$$A_t \subseteq \{T_0 \geq t, T_1 \geq t\}$$

with $P(A_t) > 0$.

Proof. The statement of Proposition 2.5 can be shown by providing a simple counterexample. To avoid redundancy, Example 1 in Section 3 is used. Set $t > u$. For each

$$A_t \subseteq \{T_0 \geq t, T_1 \geq t\},$$

the distributions of T_1 and T_0 under A_t are equal implying that any causal effect for comparing T_0 and T_1 under $P(\cdot|A_t)$ is 0 or 1 (cf. Remark 2.3). On the other hand, $0 < \mathrm{HR}(t) < 1$. $\square$

To simplify notation, we will often refer to a population causal effect under $P(\cdot|A_t)$ as a *causal effect for the population A_t*, a notion that we have already informally used before.

Corollary 2.6. *The (conventional) hazard ratio $HR(t)$ is in general not equal to the **causal hazard ratio** $HR_C(t)$ defined as (Martinussen et al. 2020; Martinussen 2022)*

$$HR_C(t) = \lim_{\Delta t \searrow 0} \frac{P(t < T_1 \leq t + \Delta t | T_0 \geq t, T_1 \geq t)}{P(t < T_0 \leq t + \Delta t | T_0 \geq t, T_1 \geq t)}.$$

2.3 | Interpretation and Connection to Literature

We have established that $\mathrm{HR}(t)$ is a causal effect for the population at baseline and consequently represents a population treatment effect in the sense that it is a contrast between the distributions of the counterfactuals T_0 and T_1. A useful conclusion is that, if $HR(t) \neq 1$ for some $t \in [0,\tau]$, then the distributions of T_0 and T_1 are not equal.

On the other hand, if one would like to interpret a causal effect measure as a *population treatment effect at time t*, it appears natural to require that the population on which the corresponding causal effect is defined is at risk at time t and that both counterfactual survival times occur after t. $\mathrm{HR}(t)$ is in general not a causal effect of a population guaranteed to be at risk at time t and hence should not be naively interpreted as a treatment effect a time t.

Moreover providing an intuitive interpretation of the causal effect $\mathrm{HR}(t)$ in the general case appears very difficult, particularly when $\mathrm{HR}(t)$ is treated as a stand-alone measure. A better explanation may be possible when contrasting the hazard functions in a functional sense, as proposed by Beyersmann et al. (2025).

While this section offers a strikingly simple perspective on the subtleties concerning the causality of the hazard ratio, the core findings of this section are of course not entirely new. Hernán and Robins (2010), e.g., clearly stated that the hazard ratio is a causal effect, but its "causal interpretation is not straightforward", providing an example where $\mathrm{HR}(t) < 1$ indicates a beneficial treatment effect, while treatment is in fact harmful. Likewise, the conclusion by Martinussen (2022) that "the hazard ratio obtained cannot be given a causal interpretation when interpreted as a hazard ratio" is closely connected to our finding that $\mathrm{HR}(t)$ is not a causal effect at time t (cf. Corollary 2.6).

3 | The Curious Case of Equal Survival Curves

3.1 | Motivation

Many articles on the causal interpretation of the hazard ratio devote a brief remark to the setting, where the survival curves coincide up to a time u and then diverge. The prevailing view is that $\mathrm{HR}(t)$ does have causal interpretation in this case and that pitfalls such as highlighted in Section 2 are irrelevant. For instance, Hernán and Robins (2010, p. 225) argued that the "built-in selection bias of hazard ratios does not happen if the survival curves are the same in the treated and the untreated", and Beyersmann et al. (2025) concurred that interpreting the treatment is "ineffective on one time interval and beneficial on the other" is "certainly correct" in this setting.

In what follows we revisit this scenario in depth and show that it is far more intricate than these brief summary comments suggest. The literature typically restricts—explicitly or implicitly—to the trivial scenario

$$\{T_0 \geq t\} = \{T_1 \geq t\} \text{ for all } t \leq u,$$

and neglects the general case

$$P(T_0 \geq t) = P(T_1 \geq t) \text{ for all } t \leq u.$$

Using concrete counter-examples, it is shown that when the latter equality holds, $\mathrm{HR}(t)$ is not always a causal effect for a subset of $\{T_0 \geq t, T_1 \geq t\}$. Moreover, it is demonstrated that the causal hazard ratio $\mathrm{HR}_C(t)$ (or any other causal effect for a subpopulation of $\{T_0 \geq t, T_1 \geq t\}$) is not necessarily a satisfying stand-alone measure for the treatment affect at time t. In particular, there exist settings where $\mathrm{HR}_C(t) = 1$ for all $t > 0$, but the corresponding survival curves $S_0(t)$ and $S_1(t)$ differ for all values $t > u$. Even more there exists examples, where the causal hazard ratio implies a benefit of treatment, i.e., $\mathrm{HR}_C(t) < 1$ for $t \geq u$, but the survival curves suggest that treatment is harmful, i.e., $S_0(t) > S_1(t)$ for

15214036, 2025, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bimj.70100 by Martha Krebsforschungszentrum - Deutsches Krebsforschungszentrum (DKFZ) , Wiley Online Library on [02/03/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

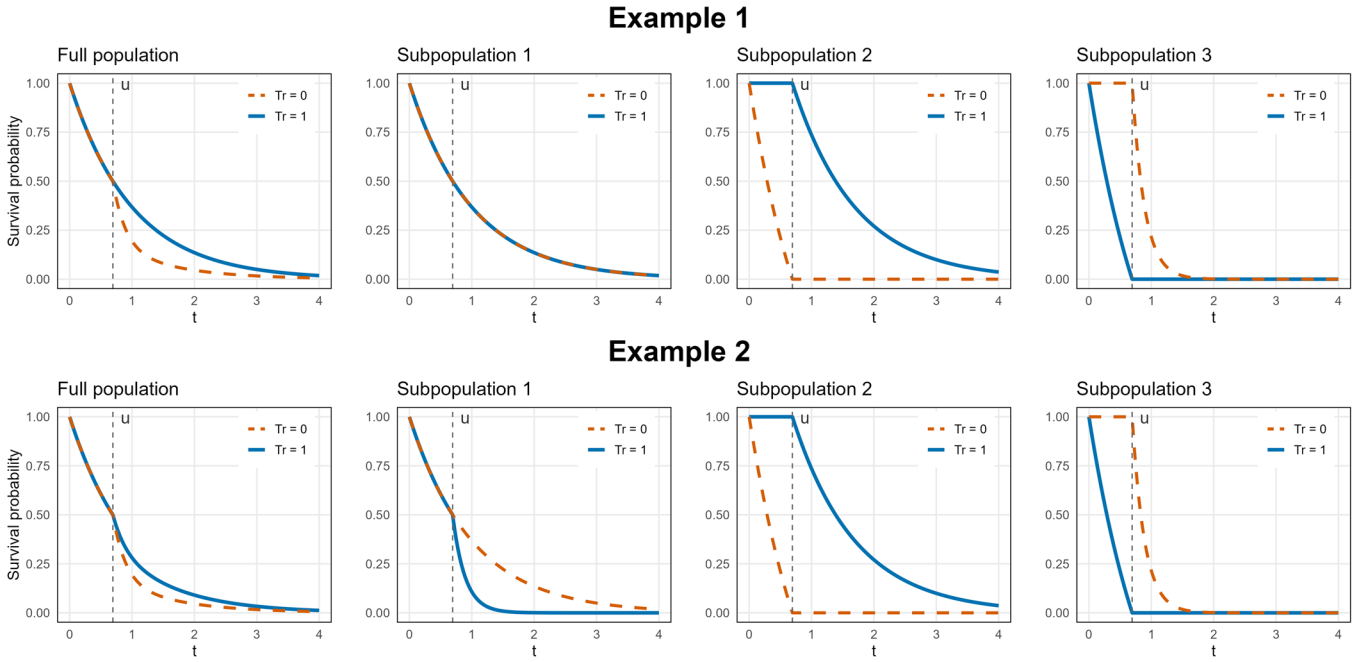


FIGURE 1 | The first two examples described in Section 3.2. The left plot represents the survival functions with (solid blue line) and without treatment (dashed red line) for the full population. The three following plots illustrate the survival functions in the three subpopulations as described in the text.

$t \geq u$. Finally, we show that there are situations in which $\mathrm{HR}_C(t)$ is not even defined for all t in the joint support of T_0 and T_1.

These findings reveal that the "equal-until-u" case, far from being trivial, exposes fundamental limitations of both the conventional and causal hazard ratios. These shortcomings indicate that any attempt to infer time-dependent treatment effects must be grounded in additional structural assumptions.

3.2 | Construction of Counterexamples

For the first two examples, we assume that there are three subpopulations of equal size. This may be modeled by a random variable S, with $P(S=1)=P(S=2)=P(S=3)=1/3$. The survival functions in the three subdistributions will be denoted by $S_{i|S=s}(t)=P(T_i>t|S=s)$, $i=0,1$, $s=1,2,3$. Moreover for ease of notation, we define $u=\log(2)$.

In all examples, it is assumed that T_0 and T_1 are conditionally independent given by S, that is,

$$P(T_0 \leq t_0, T_1 \leq t_1|S=s) = P(T_0 \leq t_0|S=s)\,P(T_1 \leq t_1|S=s),$$

for all $s=1,2,3$, $t_0,t_1>0$. An illustration of Example 1 and Example 2 is given in Figure 1.

Example 1. For the first example, the distributions of T_0 and T_1 in the subpopulations $s=1,2,3$ are given by

$$S=1: \quad S_{0|S=1}(t) = S_{1|S=1}(t) = \exp(-t).$$

$$S=2: \quad S_{0|S=2}(t) = (2\,\exp(-t)-1)\,1_{\{t\leq u\}},$$
$$S_{1|S=2}(t) = 1_{\{t\leq u\}} + (2\exp(-t))\,1_{\{t>u\}}.$$

$$S=3: \quad S_{0|S=3}(t) = 1_{\{t\leq u\}} + (2^5\exp(-5t))\,1_{\{t>u\}},$$
$$S_{1|S=3}(t) = (2\,\exp(-t)-1)\,1_{\{t\leq u\}}.$$

A straightforward calculation yields that the survival function of T_0 is

$$S_0(t) = \exp(-t)1_{\{t\leq u\}} + \tfrac{1}{3}(\exp(-t)+2^5\exp(-5t))1_{\{t>u\}},$$

and the survival function of T_1 is

$$S_1(t) = \exp(-t).$$

It is not hard to see, that, for all $t>u$,

$$S_0(t) < S_1(t).$$

On the other hand, for all $t \geq u$, it holds

$$P(T_0 \geq t, T_1 \geq t|S=2) \leq P(T_0 \geq t|S=2) = 0,$$

and

$$P(T_0 \geq t, T_1 \geq t|S=3) \leq P(T_1 \geq t|S=3) = 0.$$

Hence,

$$P(T_0 \geq t, T_1 \geq t) = P(T_0 \geq t, T_1 \geq t, S=1);$$

consequently the causal hazard ratio $\mathrm{HR}_C(t)$ reduces to the causal hazard ratio in Subpopulation 1. Due to the conditional independence of T_0 and T_1 given S, this is just the regular hazard ratio in this subpopulation, and it follows that $\mathrm{HR}_C(t)=1$ for all

15214036, 2025, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bimj.70100 by Martha Krebsforschungszentrum - Deutsches Krebsforschungszentrum (DKFZ) , Wiley Online Library on [02/03/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$t \geq u$. Moreover, symmetry considerations yield that $\mathrm{HR}_C(t) = 1$ for $t < u$.

Example 2. For the second example, the distributions of T_0 and T_1 in the subpopulations $s = 1, 2, 3$ are given by

$$
\begin{aligned}
S = 1: \quad & S_{0|S=1}(t) = \exp(-t), \\
& S_{1|S=1}(t) = \exp(-t)1_{\{t \leq u\}} + 2^4 \exp(-5t)1_{\{t>u\}} \\
S = 2: \quad & S_{0|S=2}(t) = (2\ \exp(-t) - 1)\, 1_{\{t \leq u\}}, \\
& S_{1|S=2}(t) = 1_{\{t \leq u\}} + (2\exp(-t))\, 1_{\{t>u\}}. \\
S = 3: \quad & S_{0|S=3}(t) = 1_{\{t \leq u\}} + (2^5 \exp(-5t))\, 1_{\{t>u\}}, \\
& S_{1|S=3}(t) = (2\ \exp(-t) - 1)\, 1_{\{t \leq u\}}.
\end{aligned}
$$

The survival function of T_0 is again

$$S_0(t) = \exp(-t)1_{\{t \leq u\}} + \tfrac{1}{3}(\exp(-t) + 2^5 \exp(-5t))1_{\{t>u\}},$$

and the survival function of T_1 is

$$S_1(t) = \exp(-t)1_{\{t \leq u\}} + \tfrac{1}{3}(2\exp(-t) + 2^4 \exp(-5t))1_{\{t>u\}}.$$

By elementary considerations,

$$S_0(t) = S_1(t) \text{ for } t \leq u, \quad S_0(t) < S_1(t) \text{ for } t > u,$$

demonstrating a survival benefit for treatment. On the other hand, using analogous arguments as for the first example,

$$\mathrm{HR}_C(t) = 1 \text{ for } t < u, \quad \mathrm{HR}_C(t) > 1 \text{ for } t \geq u.$$

Example 3. Finally, a third example can be constructed by removing Subpopulation 1 from either of the two examples above (that is, considering the distribution $P(\cdot|S \neq 1)$). Then, for $t \geq u$

$$P(T_0 \geq t, T_1 \geq t) = 0,$$

that is neither $\mathrm{HR}_C(t)$ nor any other causal effect measure at time t are meaningfully defined.

The reader may be skeptical about the validity of these examples, since values $t > u$ are not part of the joint support of T_0 and T_1 in *each subgroup*. Yet specifying the support of T_0 and T_1 on a subgroup (or individual) level would inevitably imply assumptions about their dependence structure, which—as common in causal inference—is left unspecified in Assumptions 1.

3.3 | Interpretation

The first finding is that the conventional hazard ratio does not necessarily represent a causal effect for a sensible subpopulation at risk at time u, even when the survival curves of the two survival times are equal up to time u. This implies, in particular, that the conventional hazard ratio HR and the causal hazard ratio HR_C do not necessarily coincide in this setting. Consequently, the interpretation of $\mathrm{HR}(u)$ as a treatment effect at time u is at least questionable.

While this is surprising and diametral to at least part of the literature, one may still argue that $\mathrm{HR}(u)$ has a straightforward causal effect interpretation in this setting—however on the population at baseline and not on any population at risk at time u. Taking Example 1 and assuming no knowledge about the subgroup affiliation, the survival probability with and without treatment is the same before u and better for treatment after u. Without any further information, a reasonable individual would prefer the treatment option for the reason that they expect a better survival after time u.

The second finding is that there are situations where the causal hazard ratio $\mathrm{HR}_C(t)$ does not represent a satisfying stand-alone measure for the treatment effect at time t. The most striking may be Example 2, in which the survival with treatment is better for any point $t > u$, while the causal hazard ratio $\mathrm{HR}_C(t)$ is greater than 1 indicating that treatment is harmful in the group $\{T_0 \geq t, T_1 \geq t\}$. While there is no fundamental problem with the interpretation of the causal hazard ratio per se (treatment is indeed harmful in this subgroup), summarizing the treatment effect over time with the causal hazard ratio alone is clearly not satisfying. Without any knowledge about the subgroup affiliation, any reasonable individual would again prefer treatment over no treatment, while interpreting the causal hazard ratio as the (and not one particular) treatment effect at time t would suggest otherwise. In fact, it appears that in this situation the causal interpretation of the conventional hazard ratio (for the population at baseline) is more straightforward than the interpretation of the causal hazard ratio.

Consequently, the use of the causal hazard ratio $\mathrm{HR}_C(t)$ as an estimand reflecting the treatment effect at time t is questionable; even more so since—by Example 3—it may not even exist in a meaningful way. It is evident that these arguments are not restricted to the causal hazard ratio and extend to any causal effect on subsets of $\{T_0 \geq t, T_1 \geq t\}$. Altogether the findings in Sections 2 and 3 suggest that, without making further assumptions on the dependence structure of T_0 and T_1, no satisfying causal estimand for a population treatment effect at time t exists!

4 | Toward Time-Dependent Treatment Effects

There are essentially two possible responses to the shortcomings discussed above. The first is to conclude that a useful and coherent time-dependent treatment effect cannot be defined and advocate against any attempt to estimate or interpret it. The second approach, that will be discussed for the remainder of this paper, is to make additional structural assumptions that deliver a well-defined estimand at time t and enable some "degree of inference". Such assumptions unavoidably constrain the joint distribution of T_0 and T_1 and hence remain untestable—yet context and domain knowledge may render them plausible; for instance, as Martinussen et al. (2020) argued, T_0 and T_1 are usually expected to be positively correlated.

A future line of methodological research may hence work on deriving various assumptions and corresponding fully coherent causal estimands enabling inference for time-dependent treatment effects. The choice of estimand should reflect the scientific question at hand and need not coincide with the causal hazard

15214036, 2025, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bimj.70100 by Martha Krebsforschungszentrum - Deutsches Krebsforschungszentrum (DKFZ) , Wiley Online Library on [02/03/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ratio $\mathrm{HR}_C(t)$. Under certain assumptions it might even be possible to define an estimand for all individuals still at risk at time t—for example, the hazard ratio that would emerge if everyone were re-randomized at that moment. Stronger, more explicit assumptions could allow point identification and consistent estimation, whereas weaker conditions may yield only partial identification such as upper and lower bounds.

Illustrative results already exist. Martinussen (2022) showed that if $\mathrm{HR}(t) \equiv \mathrm{HR} < 1$ is constant and the joint distribution of T_0 and T_1 is induced by a frailty, then

$$\mathrm{HR}_C(t) < \mathrm{HR}(t) < 1.$$

Conversely, Section 3.4 of the corresponding arXiv preprint (Martinussen et al. 2018) specifies conditions that create a one-to-one relationship between $\mathrm{HR}_C(t)$ and $\mathrm{HR}(t)$, thereby allowing consistent estimation of $\mathrm{HR}_C(t)$.

The use of inference methods for time-dependent treatment effects in practice should always be guided by considerable caution. They must be accompanied by sensitivity studies spanning multiple assumption sets, and any results should be regarded as exploratory until they are confirmed, for instance by a properly designed trial with change of treatment at time t.

5 | Concluding Remarks

As pointed out by Beyersmann et al. (2025), time-to-event data is special. While censoring—discussed at length in that paper—is one such peculiarity, it is by no means the only one. Because of their temporal nature, the analysis of time-to-event endpoints differs fundamentally from that of many other continuous endpoints—such as clinical measurements taken at fixed time points. For example, *conditional on survival up to time* t, it is of substantial interest to know if the event happens in the near future (say in the interval $(t, t + \Delta t]$), a question that has no meaningful analogue for outcomes like blood-pressure readings. In modeling event or transition times, it has proven useful—both for theoretical development and practical interpretation—to frame the process in terms of infinitesimal event probabilities (hazards or intensities), even in settings where censoring is of little or no concern, such as in queuing theory (Kleinrock 1975, p. 48) or Hawkes processes (Hawkes 1971). Aalen (1989, p. 908) provided an insightful discussion of this issue connecting the concept of intensities to the so-called "French School of probability"; in particular, he states that "by giving the intensity concept a primary place, one recognizes that survival [...] data" are "not just any measurements which happen to be censored occasionally, but are measurements of times and should be analyzed as such."

Because hazard contrasts compare the instantaneous event rate given survival to time t, they are natural candidates for the treatment effect at time t—an inherently conditional quantity that can hence only be assessed using conditional tools. Although we agree with Beyersmann et al. (2025) that unconditional measures such as survival probabilities and their contrasts are extremely useful—and ought to be reported "much more routinely", they are not suited for this type of question. Unfortunately, it is precisely the conditional nature that impedes the definition and identification of time-dependent treatment effects. Conditioning on survival up to time t breaks the original comparability of randomized groups, making it impossible to define or identify a time-dependent treatment effect without further assumptions.

From a mathematical perspective, imposing usual assumptions for continuous or time-to-event endpoints in causal inference, these problems appear very discouraging; a meaningful and coherent treatment effect at time t may not even exist and identification of candidates such as the causal hazard ratio is virtually impossible. A practitioner might judge this verdict to be too pessimistic—and may argue that an experienced bistatistician being aware about the pitfalls described in the literature would be able to give at least an educated guess for a treatment effect, even if only in a naive or intuitive sense.

By introducing additional, medically plausible assumptions to constrain dependence structures, future inference procedures for time-dependent effects may be built on a rigorous mathematical foundation. As discussed before, obtained results may still be rather crude and should be interpreted with caution. However, it may happen that one must make a decision about a time-dependent treatment effect— for example, when determining the timepoint for a change of the treatment regime in the planning phase of a clinical trial. In these situations, we firmly believe that an educated guess should always be chosen over a wild one.

Acknowledgments

Open access funding enabled and organized by Projekt DEAL.

Conflicts of Interest

The author declares no conflicts of interest.

Data Availability Statement

Data sharing not applicable to this paper as no datasets were generated or analyzed during the current study.

References

Aalen, O. O. 1989. "A Linear Regression Model for the Analysis of Life Times." *Statistics in Medicine* 8, no. 8: 907–925.

Aalen, O. O., R. J. Cook, and K. Røysland. 2015. "Does Cox Analysis of a Randomized Survival Study Yield a Causal Treatment Effect?" *Lifetime Data Analysis* 21, no. 4: 579–593.

Beyersmann, J., C. Schmoor, and M. Schumacher. 2025. "Hazards Constitute Key Quantities for Analyzing, Interpreting and Understanding Time-to-Event Data." *Biometrical Journal* 67, no. 3: e70057.

Cox, D. R. 1972. "Regression Models and Life-Tables." *Journal of the Royal Statistical Society: Series B (Methodological)* 34, no. 2: 187–202.

Fay, M. P., and F. Li. 2024. "Causal Interpretation of the Hazard Ratio in Randomized Clinical Trials." *Clinical Trials* 21, no. 5: 623–635.

Hawkes, A. G. 1971. "Spectra of Some Self-Exciting and Mutually Exciting Point Processes." *Biometrika* 58, no. 1: 83–90.

Hernán, M. A. 2010. "The Hazards of Hazard Ratios." *Epidemiology* 21, no. 1: 13–15.

Hernán, M. A., and J. M. Robins. 2010. *Causal Inference: What if*. Chapman and Hall.

15214036, 2025, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bimj.70100 by Martha Krebsforschungszentrum - Deutsches Krebsforschungszentrum (DKFZ) , Wiley Online Library on [02/03/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Jachno, K. M., S. Heritier, R. L. Woods. et al. 2022. "Examining Evidence of Time-Dependent Treatment Effects: An Illustration Using Regression Methods." *Trials* 23, no. 1: 857.

Keiding, N. 1999. "Event History Analysis and Inference From Observational Epidemiology." *Statistics in Medicine* 18, no. 17-18: 2353–2363.

Kleinrock, L. 1975. *Queuing Systems, Volume I: Theory*. Wiley.

Martinussen, T. 2022. "Causality and the Cox Regression Model." *Annual Review of Statistics and Its Application* 9, no. 1: 249–259.

Martinussen, T., S. Vansteelandt, and P. K. Andersen. 2018. "Subtleties in the Interpretation of Hazard Ratios." arXiv:1810.09192.

Martinussen, T., S. Vansteelandt, and P. K. Andersen. 2020. "Subtleties in the Interpretation of Hazard Contrasts." *Lifetime Data Analysis* 26: 833–855.

Appendix A

In Section 2.2, it is mentioned that the given definition for a causal effect is somewhat restrictive since both the functional and the contrast function are assumed to be real-valued. In the following, an attempt for a more general definition is made that avoids the use of two different functionals and defines the causal effect directly as a contrast on the induced probability measures of the two counterfactuals. For the following definition, let G be a group with neutral element e. Further denote the inverse element of $g \in G$ as g^-.

Definition A.1. Let $\mathcal{P}$ denote a set of probability measures on a measurable space $(\Omega, \mathcal{F})$. Let Y_0 and Y_1 denote two measurable functions denoting counterfactual outcomes on $(\Omega, \mathcal{F})$ with $P_{Y_0}, P_{Y_1} \in \mathcal{P}$. Let $\psi : \mathcal{P} \times \mathcal{P} \to G$ be a functional satisfying

i. For all $Q \in \mathcal{P}$, $\psi(Q, Q) = e$.

ii. For all $Q_1, Q_2 \in \mathcal{P}$, $\psi(Q_1, Q_2) = \psi(Q_2, Q_1)^-$.

Then, $\psi(P_{Y_1}, P_{Y_0})$ is called a **population causal effect** for comparing Y_1 and Y_0 under the probability distribution P.

There may be settings, in which the direction of the effect is of little interest—it may then make sense to allow for a definition, say of an *unsigned causal effect*, in which G is replaced by a semigroup with neutral element and the antisymmetry condition in (ii) by a symmetry condition.

15214036, 2025, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bimj.70100 by Martha Krebsforschungszentrum - Deutsches Krebsforschungszentrum (DKFZ) , Wiley Online Library on [02/03/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License