A nonparametric proportional risk model to assess a treatment effect in time-to-event data

Lucia Ameis, Oliver Kuss, Annika Hoyer, Kathrin Möllenhoff

Suggested Citation:
Ameis, L., Kuß, O., Hoyer, A., & Möllenhoff, K. (2024). A nonparametric proportional risk model to assess a treatment effect in time-to-event data. Biometrical Journal, 66(4), Article 2300147.
https://doi.org/10.1002/bimj.202300147

Received: 30 May 2023 | Revised: 26 March 2024 | Accepted: 28 March 2024

DOI: 10.1002/bimj.202300147

Biometrical Journal

RESEARCH ARTICLE

A nonparametric proportional risk model to assess a treatment effect in time-to-event data

Lucia Ameis[1,2] | **Oliver Kuss**[3] | **Annika Hoyer**[4] | **Kathrin Möllenhoff**[1,2]

[1]Institute of Medical Statistics and Computational Biology (IMSB), Faculty of Medicine, University of Cologne, Cologne, Germany

[2]Mathematical Institute, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

[3]Institute for Biometrics and Epidemiology, German Diabetes Center, Leibniz Institute for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany

[4]Biostatistics and Medical Biometry, Medical School OWL, Bielefeld University, Bielefeld, Germany

Correspondence
Kathrin Möllenhoff, Institute of Medical Statistics and Computational Biology (IMSB), University of Cologne, Cologne, Germany.
Email: kathrin.moellenhoff@uni-koeln.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Numbers: RTG 2624, 427806116

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

Abstract

Time-to-event analysis often relies on prior parametric assumptions, or, if a semiparametric approach is chosen, Cox's model. This is inherently tied to the assumption of proportional hazards, with the analysis potentially invalidated if this assumption is not fulfilled. In addition, most interpretations focus on the hazard ratio, that is often misinterpreted as the relative risk (RR), the ratio of the cumulative distribution functions. In this paper, we introduce an alternative to current methodology for assessing a treatment effect in a two-group situation, not relying on the proportional hazards assumption but assuming proportional risks. Precisely, we propose a new nonparametric model to directly estimate the RR of two groups to experience an event under the assumption that the risk ratio is constant over time. In addition to this relative measure, our model allows for calculating the number needed to treat as an absolute measure, providing the possibility of an easy and holistic interpretation of the data. We demonstrate the validity of the approach by means of a simulation study and present an application to data from a large randomized controlled trial investigating the effect of dapagliflozin on all-cause mortality.

KEYWORDS
hazard ratio, number needed to treat, risk, time-to-event analysis, treatment effect

1 | INTRODUCTION

In medical research, time-to-event data measuring the time until a specific event, for example, the time until death or the time until occurrence of a particular disease, are very common. In 2022, nearly 1000 articles including the phrase "time-to-event" were collected on PubMed (NCBI, 1996). Therefore, proper analysis of this type of data is of great interest. Well-known models include proportional hazard (PH) models like Cox's PH model (Cox, 1972) or the Weibull PH model (Kleinbaum & Klein, 2012) and proportional odds (PO) models like the log-logistic model (Kleinbaum & Klein, 2012), resulting in a focus on the hazard ratio (HR) and the odds ratio (OR) as estimates for the treatment effect. This contradicts the preference of reporting the relative risk (RR) when analyzing data given in form of 2×2 contingency tables (Sistrom & Garvan, 2004).

The RR is characterized by its easy interpretability: If two groups, say A and B, have an RR of r for an event, then group A is r times more likely to experience the event relative to group B. Consequently, an RR of 1 means an equal risk for both groups. If r larger than 1, group A has a larger risk than group B, and vice versa if it is smaller than 1 (Rosner, 2016). The RR should always be reported in combination with an absolute risk measure like the numbers needed to treat, to classify the relative measure. Once this has been done, even complex statistical results can be communicated easily (Sistrom & Garvan, 2004).

The misinterpretation of the HR as an RR has a long tradition, starting with even basic literature using the two terms interchangeably (Kalbfleisch & Prentice, 2002; Klein & Moeschberger, 2003). However, this is incorrect: both measures indicate the same direction in regards to the treatment effect and hence have similar interpretation, but are technically not the same (Sutradhar & Austin, 2018). The HR is a conditional measure, based on instantaneous rates (Hernán, 2010), while the RR is not. Therefore, the two values should be strictly distinguished and handled with care. In addition, an increase in studies reporting nonproportional hazard rates has been noted in recent years (Royston & Parmar, 2014), putting the PH assumption into question. More precisely, the Cox model and the underlying PH assumption have been criticized recently, as the PH assumption is only rarely assessed in practice, or, the model is even used regardless of the presence of nonproportional hazards (Jachno et al., 2019). Although Cox's model evaluates the mean HR over time in this case (Struthers & Kalbfleisch, 1986) and therefore still allows for meaningful interpretation, it nonetheless should not be confounded with the mean RR.

It is well-known that the OR approximates the RR if the event of interest has a low prevalence (Sistrom & Garvan, 2004). Otherwise, these measures do not coincide, which led to widespread misinformation in the past (Schwartz et al., 1999). In addition, just as the PH assumption, the PO assumption is sometimes not suitable. To fill in the gap of the estimation of the RR for time-to-event data, Kuss and Hoyer recently proposed a parametric proportional risk (PPR) model for a two-group situation as typically given in randomized controlled trials (RCTs) by using the exponentiated-uniform (EU) distribution (Kuss & Hoyer, 2021). This circumvents the problem of the incorrect interpretation of the HR and the OR, respectively.

Based on this idea, this paper proposes a nonparametric proportional risk (NPPR) estimator for the RR that is easy to implement, robust, and in particular independent of the underlying cumulative distribution function (CDF). For the construction the Kaplan–Meier estimator, Kaplan & Meier (1958) is used, allowing for the inclusion of right-censored data. The NPPR estimator also enables inference on a time interval, which is advantageous compared to a prior nonparametric method based on the ratio of two Kaplan–Meier estimators (Lachin, 2010) that focuses on a single time point. By means of a percentile bootstrap (Efron, 1981), we calculate confidence intervals (CIs) that allow to test for equal CDFs.

In addition to an RR measure, an absolute measure is needed to enable a holistic interpretation of the data (The Academy of Medical Sciences, 2017). Therefore, estimators for the risk difference and its inverse, the number needed to treat (NNT, Hildebrandt et al., 2009; Porta, 2016), are derived from the NPPR estimator. Especially, the NNT stands out due to its simple interpretability: If the NNT equals x_t at time point t, x_t patients have to be administered the treatment, so that one more event up to t is prevented compared to the control group (Hildebrandt et al., 2009; Porta, 2016).

The paper is structured as follows: First, we introduce the NPPR model and show how to estimate the treatment effect, given by the mean RR over time. Based on this, we present a formula for the risk difference and the NNT. Afterward, we report a small simulation study that compares the NPPR estimator to the RR estimated by applying the PPR model, the HR estimated by Cox's PH model, and the OR estimated using a log-logistic PO model. Finally, the practical usability of the model is illustrated by its application to data from the Dapagliflozin and Prevention of Adverse Outcomes in Heart Failure (DAPA-HF) trial (McMurray et al., 2019), a large RCT investigating the effect of dapagliflozin on all-cause mortality.

2 | METHODOLOGY

2.1 | The NPPR model

We consider a situation with two groups, in terms of RCTs given by a treatment (indexed by 1) and a control (indexed by 0) group, respectively. The corresponding (unknown) CDFs describing the probability of having experienced the outcome of interest up to a specific time point, are assumed to be proportional. This implies that their ratio, the RR, is constant over time, which we denote as proportional risk (PR) assumption.

To prevent mathematical and interpretative problems, we will restrict our analysis to the time interval $T = (t_{min}, t_{max}) \subset \mathbb{R}$, whereby t_{min} denotes the larger of the first hypothetical event time points in each group and t_{max} denotes the smaller of the latest hypothetical event time points. Mathematically, this is done to (1) avoid dividing by zero and (2) avoid artificially underestimating the RR. After the first of the two CDFs reaches (or asymptotically comes close to) 1, the RR will, by definition, converge to 1. Of note, a meaningful interpretation of an RR outside this time range is hardly possible. Now, let F_1 and F_0 denote the corresponding unknown CDFs in the treatment and control group, respectively. Given the PR assumption evaluating the ratio of probabilities at any time point $t \in T$, yields a treatment effect of

$$\frac{F_1(t)}{F_0(t)} = r$$

for a constant $r > 0$. From this, we define $\beta \in \mathbb{R}$ so that the equation

$$\exp(-\beta) = \frac{F_1(t)}{F_0(t)} = r$$

holds. Solving for β we obtain

$$\beta := -\log\left(\frac{F_1(t)}{F_0(t)}\right). \tag{1}$$

Moving to the log scale allows for a more symmetric interpretation of β. More precisely, a positive β corresponds to a positive treatment effect, a negative β to a negative one, and $\beta = 0$ to no effect, respectively. The main goal is to estimate β without any parametric assumptions on the CDFs. Therefore, we estimate F_1 and F_0 by means of the nonparametric Kaplan–Meier estimators $\hat{S}_1$, $\hat{S}_0$, using the relationship between CDFs and their corresponding survival functions. The resulting estimated CDFs will be denoted by $\hat{F}_i = 1 - \hat{S}_i$, $i = 0, 1$. As the underlying data are possibly right-censored, we need this slightly more complicated method of estimation instead of using the empirical CDFs.

Let $\tilde{T}_1$ and $\tilde{T}_0$ be the ordered sets of observed event time points for the two groups. We set $\tilde{t}_{min} = \max(\min(\tilde{T}_1), \min(\tilde{T}_0))$ and $\tilde{t}_{max} = \min(\max(\tilde{T}_1), \max(\tilde{T}_0))$. By defining

$$\tilde{T} = (t \mid t \in \tilde{T}_1 \text{ or } t \in \tilde{T}_0,\ \tilde{t}_{min} \le t \le \tilde{t}_{max}),$$

we restrict the time interval such that the fraction of $\hat{F}_1$ and $\hat{F}_0$ can be evaluated properly and to comply with the restrictions described above. Of note, tied observations are retained in $\tilde{T}_1$, $\tilde{T}_0$, and $\tilde{T}$ for later inclusion. Inserting the estimated CDFs instead of the true unknown CDFs in Equation (1) for any t, $\tilde{t}_{min} \le t \le \tilde{t}_{max}$, yields a (time-dependent) estimator $\hat{\beta}_t$ for β. Conceivably, the method might benefit from further restricting $\tilde{T}$, for example, if the data suggest a high variability at the beginning or at the end of the study.

To arrive at a single estimate for β, we compute a weighted mean of the $\hat{\beta}_t$ where the weights are derived from their variance. Precisely, as the Kaplan–Meier estimator is asymptotically normally distributed with a variance which can be approximated by Greenwood's formula (Breslow & Crowley, 1974), we estimate the variance of $\hat{\beta}_t = -\log\left(\frac{\hat{F}_1(t)}{\hat{F}_0(t)}\right) = -\log\left(\hat{F}_1(t)\right) + \log\left(\hat{F}_0(t)\right)$ for every $t \in \tilde{T}$ by applying the delta method (Oehlert, 1992) to $\log(\hat{F}_i(t))$, $i = 0, 1$. This yields

$$\hat{\sigma}_i^2(t) := \widehat{\mathrm{var}}\left(\log(\hat{F}_i(t))\right) = \frac{1}{\hat{F}_i(t)^2}\widehat{\mathrm{var}}\left(\hat{S}_i(t)\right), \quad i = 0, 1.$$

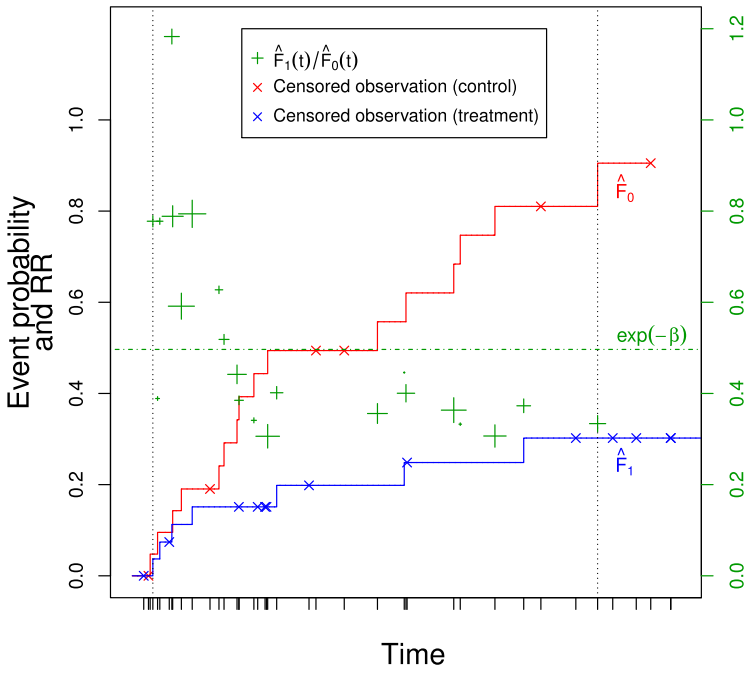

FIGURE 1 Visualization of the nonparametric proportional risk (NPPR) estimator. The size of the pluses symbolizes the weight given to the respective estimated relative risk (RR) at this time point. The larger the plus, the larger the weight. For a visualization on a logarithmic scale, we refer to the supplementary material (Figure A).

Since $\hat{F}_1(t)$ and $\hat{F}_0(t)$ are independent and consequently their covariance is equal to 0, we define the weight function by

$$\omega(t) := \widehat{\mathrm{var}}\left(\hat{\beta}_t\right) = \frac{1}{\hat{F}_1(t)^2}\widehat{\mathrm{var}}\left(\hat{S}_1(t)\right) + \frac{1}{\hat{F}_0(t)^2}\widehat{\mathrm{var}}\left(\hat{S}_0(t)\right)$$

for $t \in \tilde{T}$. This choice was inspired by inverse-variance weighting (Hartung et al., 2008) and the weighted least square method (Mandel, 1984). Consequently, with $W := \sum_{t\in\tilde{T}} \frac{1}{\omega(t)}$, a nonparametric estimator of β, the NPPR estimator, is defined by

$$\hat{\beta}_{NPPR} = \frac{1}{W}\sum_{t\in\tilde{T}} \frac{1}{\omega(t)}\hat{\beta}_t = -\frac{1}{W}\sum_{t\in\tilde{T}} \frac{1}{\omega(t)} \log\left(\frac{\hat{F}_1(t)}{\hat{F}_0(t)}\right). \tag{2}$$

As stated above, tied observations were retained in $\tilde{T}$. Therefore, if an event time point t' is included m times over both T_1 and T_0, the term $\frac{1}{\omega(t')}\hat{\beta}_{t'}$ is also included m times in the sum in (2). A visualization of the situation is given in Figure 1. Also, the estimator does, in general, estimate the mean RR over time. This holds even if the PR assumption is violated. In this case, the estimator still provides a summary of the RR.

The NPPR estimator $\hat{\beta}$ is a consistent estimator for β, which is a direct consequence of its construction, the continuous mapping theorem (van der Vaart, 1998), and the fact that the Kaplan–Meier estimators are consistent for every $t \in \tilde{T}$ (Andersen et al., 1993).

2.2 | Risk difference and NNT

In order to ensure a holistic interpretation of the data, an absolute measure is needed in addition to the relative measure provided by the RR. The risk difference for the unknown true CDFs is defined by

$$RD(t) = F_0(t) - F_1(t)$$

for $t \in T$. Using the PR assumption, we can write $F_1(t) = \exp(-\beta)F_0(t)$ and substitute $F_1(t)$ in the formula. This yields

$$RD(t) = (1 - \exp(-\beta))F_0(t).$$

TABLE 1 Overview of the three different assumptions and their abbreviations.

Assumption	Abbr.	Mathematical formulation
Proportional risks	PR	It exists $r > 0$ so that $\frac{F_1(t)}{F_0(t)} = r$, with $F_i(t)$, $i = 0, 1$, as CDFs
Proportional hazards	PH	It exists $r > 0$ so that $\frac{h_1(t)}{h_0(t)} = r$, with $h_i(t)$, $i = 0, 1$, as hazard functions
Proportional (failure) odds	PO	It exists $r > 0$ so that $\frac{\frac{F_1(t)}{1-F_1(t)}}{\frac{F_0(t)}{1-F_0(t)}} = r$, with $F_i(t)$, $i = 0, 1$, as CDFs

Abbreviation: CDF, cumulative distribution function.

Now we can insert the estimated CDF $\hat{F}_0$ corresponding to the Kaplan–Meier estimator $\hat{S}_0$ and the NPPR estimator $\hat{\beta}_{NPPR}$ obtained by (2). The resulting estimator of the risk difference is defined by

$$\widehat{RD}(t) = \left(1 - \exp(-\hat{\beta}_{NPPR})\right)\hat{F}_0(t)$$

for $\tilde{t}_{min} \leq t \leq \tilde{t}_{max}$. Of note, this still does depend on the time t and given the PR assumption monotonically increases if $\exp(-\hat{\beta}_{NPPR}) < 1$ and decreases if $\exp(-\hat{\beta}_{NPPR}) > 1$, respectively.

The NNT is defined as the reciprocal of the risk difference. Therefore, an estimator is given by

$$\widehat{NNT}(t) = \frac{1}{\widehat{RD}(t)} = \frac{1}{\left(1 - \exp(-\hat{\beta}_{NPPR})\right)\hat{F}_0(t)}.$$

It is also possible to estimate the NNT based on $\hat{F}_1(t)$:

$$\widehat{NNT}_{alt1}(t) = \frac{1}{\widehat{RD}_{alt1}(t)} = \frac{1}{\left(\exp(-\hat{\beta}_{NPPR}) - 1\right)\hat{F}_1(t)}.$$

This is beneficial, for example, if the treatment group size significantly exceeds the control group size since in this case $\hat{F}_1(t)$ would be a more reliable estimator than $\hat{F}_0(t)$. Another alternative is defined using both $\hat{F}_0(t)$ and $\hat{F}_1(t)$:

$$\widehat{NNT}_{alt2}(t) = \frac{1}{\widehat{RD}_{alt2}(t)} = \frac{1}{\left(\hat{F}_0(t) - \hat{F}_1(t)\right)}.$$

However, this estimator strongly depends on the precise time point used for evaluation since it is subject to the jumps of both $\hat{F}_1(t)$ and $\hat{F}_0(t)$, while $\widehat{NNT}(t)$ and $\widehat{NNT}_{alt1}(t)$ only depend on the jumps of one of the curves. In case of a small data set, one could look at all three estimators and infer on the range of the respective CIs (calculated using the percentile bootstrap (Efron, 1981) as described below).

3 | SIMULATION STUDY

3.1 | Setting and data generation

In order to evaluate the performance of the NPPR estimator compared to an alternative method, we conducted a simulation study. As competitors we chose the PPR model for estimating the RR, the Cox's PH model estimating the HR and a log-logistic PO model estimating the OR. An overview of the different assumptions and the different models are provided in Tables 1 and 2. The PPR model is based on the EU distribution (Kuss & Hoyer, 2021), and the corresponding CDFs are given by

$$F_{EU,i}(t) = (\theta_i t)^{\alpha}, \quad t \leq \frac{1}{\theta_i}, \text{ for } i = 0, 1,$$

TABLE 2 Overview of all models. Note, that the Weibull proportional hazard (PH) model is only used for data generation and not for parameter estimation in the simulation study. More information on the relative risk (RR), the hazard ratio (HR), and the odds ratio (OR) of the different models as well as the calculation of the formula is presented in the supplementary material Section 3.0.

Model name	Abbr.	Asm.	Type	Mathematical description	$\widehat{RR}/\widehat{HR}/\widehat{OR}$
Nonparametric proportional risk model	NPPR	PR	Nonparametric	$\hat{\beta}_{NPPR} = -\frac{1}{W}\sum_{t\in\tilde{T}}\frac{1}{\omega(t)}\log\left(\frac{\hat{F}_1(t)}{\hat{F}_0(t)}\right)$	$\widehat{RR}_{NPPR} = \exp(-\hat{\beta})$
Parametric proportional risk model	PPR	PR	Parametric	$F_{EU,i}(t) = (\theta_i t)^{\alpha}$	$\widehat{RR}_{PPR} = \left(\frac{\hat{\theta}_1}{\hat{\theta}_0}\right)^{\hat{\alpha}}$
Cox's proportional hazards model	Cox's PH	PH	Semiparametric	$h_{Cox}(t, X_i) = h_0(t)\exp(\beta X_i)\ h_0(t)$ baseline hazard $X_i = \begin{cases} 1 \text{ treatment} \\ 0 \text{ control} \end{cases}$	$\widehat{HR}_{Cox} = \exp(\hat{\beta})$
Log-logistic proportional odds model	LL PO	PO	Parametric	$F_{LL,i}(t) = 1 - \frac{1}{1+\left(\frac{t}{b_i}\right)^{a}}$	$\widehat{OR}_{LL} = \left(\frac{\hat{b}_0}{\hat{b}_1}\right)^{\hat{a}}$
Weibull proportional hazards model	Weibull PH	PH	Parametric	$F_{Weib,i}(t) = 1 - e^{-\left(\frac{t}{\lambda_i}\right)^{k}}$	$\widehat{HR}_{Weibull} = \left(\frac{\hat{\lambda}_0}{\hat{\lambda}_1}\right)^{\hat{k}}$

Abbreviations: PH, proportional hazard; PO, proportional odds; PR, proportional risk.

in which the shape parameter $\alpha > 0$ is assumed to be the same for both the treatment and the control group and the scale parameters $\theta_i > 0$, $i = 0, 1$, are assumed to be group-specific. By calculating maximum-likelihood estimates using the R function *optim* for θ_i, $i = 0, 1$, and α, the estimated RR is given by

$$\widehat{RR}_{PPR} = \left(\frac{\hat{\theta}_1}{\hat{\theta}_0}\right)^{\hat{\alpha}}.$$

Of note, the OR and HR are identical and vary over time, as can be shown by inserting the CDF into the respective definition (see supplementary material Section 3.0 for details). For the log-logistic PO model, we chose the parameterization as provided by the R package *flexsurv* (Jackson, 2016):

$$F_{LL,i}(t) = 1 - \frac{1}{1+\left(\frac{t}{b_i}\right)^{a}}, \text{ for } i = 0, 1,$$

in which the shape parameter $a > 0$ and the scale parameters $b_i > 0$, $i = 0, 1$, are similarly assumed to be the same and group-specific, respectively. Again, using `optim` to calculate maximum-likelihood estimates, the estimated (failure) OR is given by

$$\widehat{OR}_{LL} = \left(\frac{\hat{b}_0}{\hat{b}_1}\right)^{\hat{a}}.$$

Outcomes of the simulation study were bias, mean squared error (MSE), and empirical coverage of the estimated RR, respectively. In addition, we evaluated numerical robustness in terms of the number of converged simulation runs. Precisely, it might happen that all events of one group have already occurred before the first event of the other group occurs. Under these circumstances, or when all participants were censored, $\tilde{T}$ might be empty. Then, the NPPR estimator is not defined. Among the simulations, these cases are few in numbers and were excluded from further analysis. Similarly, data sets yielding estimates obtained by the PPR model, Cox's PH model, and the log-logistic PO model of an absolute effect ($|-\log(\widehat{RR}_{PPR})|$, $|-\log(\widehat{HR}_{Cox})|$, $|-\log(\widehat{OR}_{LL})|$) larger than 3 were removed. This choice was informed by the analysis of the simulated data with Cox's PH model. That is, limiting the analysis to estimated HRs smaller than 20 ($\approx \exp(3)$) was the most conservative way to exclude estimations suffering from obvious numerical issues. Of note, the NPPR estimator

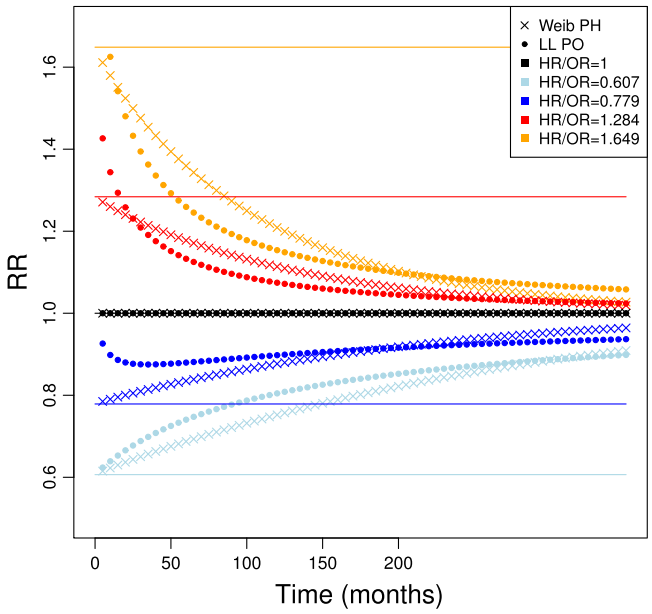

FIGURE 2 The relative risk (RR) of the Weibull proportional hazard (PH) model and the log-logistic (LL) proportional odds (PO) model over time in comparison with the constant value indicated by the straight lines. HR, hazard ratio; OR, odds ratio.

exceeded this limit in no case. The numerical robustness describes the number of estimates used for further analysis after the removal. All initial analyses have been done using R version 4.2.2. Analyses with Cox's PH model, the log-logistic PO model, and the simulation studies without censoring were added later and have been done using R version 4.3.0.

The simulation setting was inspired by the DAPA-HF trial discussed as a case study example in Section 4. Instead of focusing on the secondary outcome "death of all causes," we based the simulations on the data set regarding the studies primary outcome "worsening heart failure or cardiovascular death." Figure C (see supplementary material) contains an analysis of this data set analogous to Section 4. We simulated a PR, PH, and a PO situation with different true underlying effects for the RR, HR, or OR, respectively. For the generation of data satisfying the PR assumption, the PPR model was used and the log-logistic PO model for PO data. For a PH scenario, we used a Weibull PH model. The corresponding CDFs are given by

$$F_{Weib,i}(t) = \begin{cases} 1 - \mathrm{e}^{-\left(\frac{t}{\lambda_i}\right)^k} & \text{for } t \geq 0 \text{ for } i = 0, 1. \\ 0 & \text{for } t < 0 \end{cases}$$

Here, $k > 0$ is the shape parameter, which is assumed to be equal for both groups, and λ_i, $i = 0, 1$, are the group-specific scale parameters. In all scenarios, we used the parameters obtained from the placebo group of the data set as underlying truth to simulate data for the reference group, where we chose true underlying parameters $\beta = 0,\ 0.5,\ 0.25,\ -0.25,\ -0.5$ which correspond to an RR/HR/OR equal to $1,\ 0.607,\ 0.779,\ 1.284$, and 1.649 (see Figure 2 for a visualization of the RRs over time). All configurations used for the simulations are summarized in Table 3. For a visualization of the CDFs, we refer to Figure E in the supplementary material.

To achieve different amounts of censoring (30%, 50%, and 70%), we simulated censoring times from an appropriate uniform distribution. Also, we simulated studies without any censoring. For each of these combinations, we varied the number of study participants (50, 100, and 500). All simulation results were obtained from simulating 1000 studies. Combining the PR, PH, and PO cases resulted in 180 simulation scenarios representing a situation with no, a realistic, and a more extreme effect on both sides of the null effect. In all cases, we first drew the group assignment using inverse transformation sampling of a binomial distribution with probability $p = 0.5$. Therefore, the group sizes of the treatment and control group in the individual simulated study vary with an expected number of half the number of study participants in each group. However, the number of all patients always equals n. Afterward, corresponding to the declared underlying

TABLE 3 Parameters of the parametric proportional risk (PPR) model, the Weibull proportional hazard (PH) model, and the log-logistic proportional odds (PO) model corresponding to the true underlying effects for all different cases.

True underlying effect		Parameters								
		PPR			Weibull PH			log-logistic PO		
β	RR/HR/OR	α	θ_0	θ_1	k	λ_1	λ_0	a	b_0	b_1
0	1	0.859	0.009	0.009	0.916	88.296	88.296	0.973	70.817	70.817
0.5	0.607	0.859	0.009	0.005	0.916	152.428	88.296	0.973	70.817	118.362
0.25	0.779	0.859	0.009	0.007	0.916	116.012	88.296	0.973	70.817	91.553
−0.25	1.284	0.859	0.009	0.012	0.916	67.201	88.296	0.973	70.817	54.777
−0.5	1.649	0.859	0.009	0.016	0.916	51.146	88.296	0.973	70.817	42.370

Abbreviations: HR, hazard ratio; OR, odds ratio; RR, relative risk.

true effect, we drew survival times from the PPR model, the Weibull PH model, or the log-logistic PO model, again using inverse transformation sampling.

In case of censoring, given the censoring proportion we then drew censoring times from the uniform distribution—for the specific parameters, see the supplementary material (Table A) - for each participant. Finally, the observed time of each individual was defined as minimum of the survival and censoring time and the status adjusted accordingly. In case of no censoring, the observed time of each individual is equal to the survival time.

3.2 | Estimation

We compared the NPPR estimator with the RR estimated using the PPR model only if for the underlying data-generating model the PR assumption was satisfied. Due to numerical problems, we fitted the PPR model only if censoring was present. For the performance of the PPR model in case of PH data, we refer to Kuss & Hoyer (2021). Similarly, we fitted Cox's PH model and the log-logistic PO model only for PR data as well as for the data that fulfill the respective assumption. Of note, the Weibull PH and the log-logistic PO models violate the PR assumption. However, the NPPR estimator is in this case still applicable to estimate the mean RR over time. We used this measure as substitute for an estimator of the true underlying HR/OR and preceded with the analysis if these parameters were identical. This was done to examine the behavior of the NPPR model, if the PR assumption does not hold, which can be the case in practical applications. In correspondence to this, we interpreted the estimated HR/OR for the PR data using Cox's PH and the log-logistic PO models as mean HR/OR and used it as substitute to estimate the true RR in order to mimic the often seen practice to use the RR and the HR/OR interchangeably (Schwartz et al., 1999; Sutradhar & Austin, 2018).

The PPR and the log-logistic PO models were estimated using the R function *optim* as described above. Cox's PH model was fitted using the function *coxph* from the *survival* (Therneau, 2023; Therneau & Grambsch, 2000) package.

3.3 | Results

In the following, we will present the results of the simulation study. As the NPPR model directly estimates $\hat{\beta}_{NPPR} = -\log(\widehat{RR}_{NPPR})$, we transform the estimate $\widehat{RR}_{PPR}$ obtained by the PPR model to $-\log(\widehat{RR}_{PPR})$ for an easier comparability. We preceded accordingly with the estimated $\widehat{HR}_{Cox}$ and $\widehat{OR}_{LL}$. Tables 4–6 display the bias and the MSE for the different scenarios if censoring was present, respectively. The bias and MSE corresponding to the scenarios without censoring are displayed in Table B (see supplementary material). For details on the coverage and the numerical robustness, we refer to the supplementary material (Tables C–J).

For a quick introduction, Figure 3 displays a comparison between the estimated values for a censoring rate of 50% and 500 participants if the PR assumption is fulfilled. If the true underlying effect β equals 0, the NPPR estimator has a larger interquartile range (IQR) than the PPR model with equal median. In all remaining cases, the median is closer to the true value, while the IQR is still a bit larger. We observe that the NPPR estimator produces fewer outliers compared to the PPR model. The number of outliers and the IQR of Cox's PH model and the log-logistic PO model are similar to those of the NPPR model. As expected, the median is larger in case of $\beta \neq 0$. Similar findings are achieved for situations with a censoring rate of 30% and only in case of a high censoring rate, that is 70%, the number of outliers is comparable (see

TABLE 4 Bias and mean squared error (MSE) of $\hat{\beta}_{NPPR}$ (NPPR, nonparametric proportional risk), $-\log(\widehat{RR}_{PPR})$ (PPR, parametric proportional risk), $-\log(\widehat{HR}_{Cox})$ (Cox's PH model), and $-\log(\widehat{OR}_{LL})$ (log-logistic PO model) if the PPR model is the true underlying model.

			Bias				MSE			
Effect	**Censoring (%)**	**Participants**	**NPPR**	**PPR**	**Cox**	**LL**	**NPPR**	**PPR**	**Cox**	**LL**
0.00	30	500	0.000	−0.006	0.000	0.000	0.011	0.140	0.013	0.026
0.00	30	100	−0.002	−0.017	−0.001	−0.006	0.052	0.365	0.065	0.132
0.00	30	50	0.010	0.021	0.007	0.010	0.092	0.436	0.139	0.250
0.00	50	500	0.005	0.001	0.002	0.006	0.016	0.005	0.018	0.030
0.00	50	100	0.004	0.007	0.008	0.011	0.082	0.047	0.093	0.162
0.00	50	50	−0.011	0.000	−0.009	−0.015	0.153	0.111	0.192	0.316
0.00	70	500	−0.003	−0.002	−0.002	−0.003	0.027	0.016	0.028	0.040
0.00	70	100	−0.011	−0.005	−0.012	−0.012	0.145	0.084	0.146	0.211
0.00	70	50	0.016	0.017	0.025	0.027	0.280	0.205	0.324	0.463
0.50	30	500	0.002	0.196	0.495	0.426	0.013	0.144	0.259	0.210
0.50	30	100	0.009	0.314	0.490	0.444	0.066	0.530	0.314	0.337
0.50	30	50	0.007	0.349	0.494	0.476	0.129	0.583	0.397	0.521
0.50	50	500	−0.004	0.070	0.367	0.378	0.015	0.020	0.151	0.171
0.50	50	100	0.003	0.193	0.359	0.399	0.084	0.203	0.222	0.316
0.50	50	50	−0.012	0.242	0.347	0.414	0.168	0.291	0.328	0.520
0.50	70	500	−0.003	0.000	0.138	0.221	0.031	0.019	0.050	0.093
0.50	70	100	−0.014	0.011	0.141	0.234	0.169	0.101	0.186	0.293
0.50	70	50	−0.033	0.037	0.162	0.269	0.288	0.237	0.342	0.529
0.25	30	500	0.003	0.107	0.305	0.217	0.011	0.118	0.105	0.071
0.25	30	100	0.017	0.160	0.309	0.248	0.051	0.358	0.158	0.183
0.25	30	50	0.002	0.212	0.293	0.234	0.118	0.422	0.229	0.356
0.25	50	500	0.004	0.046	0.209	0.206	0.014	0.010	0.059	0.069
0.25	50	100	0.009	0.104	0.204	0.221	0.071	0.092	0.129	0.196
0.25	50	50	−0.014	0.127	0.175	0.197	0.157	0.193	0.232	0.361
0.25	70	500	−0.002	−0.004	0.069	0.113	0.022	0.014	0.028	0.045
0.25	70	100	0.004	0.005	0.079	0.126	0.130	0.079	0.142	0.211
0.25	70	50	0.000	0.033	0.104	0.166	0.277	0.195	0.322	0.479
−0.25	30	500	−0.001	−0.089	−0.304	−0.215	0.011	0.144	0.104	0.070
−0.25	30	100	0.000	−0.189	−0.290	−0.222	0.056	0.365	0.145	0.181
−0.25	30	50	−0.014	−0.175	−0.298	−0.252	0.116	0.446	0.227	0.362
−0.25	50	500	0.003	−0.027	−0.183	−0.188	0.015	0.008	0.050	0.064
−0.25	50	100	0.005	−0.071	−0.166	−0.192	0.070	0.079	0.115	0.179
−0.25	50	50	0.012	−0.117	−0.162	−0.194	0.148	0.157	0.207	0.348
−0.25	70	500	−0.007	−0.003	−0.075	−0.120	0.025	0.016	0.032	0.052
−0.25	70	100	−0.025	−0.027	−0.101	−0.156	0.154	0.096	0.167	0.250
−0.25	70	50	0.001	−0.028	−0.103	−0.154	0.310	0.215	0.348	0.493
−0.50	30	500	0.003	−0.219	−0.485	−0.419	0.012	0.193	0.248	0.200
−0.50	30	100	−0.014	−0.306	−0.487	−0.454	0.061	0.467	0.307	0.334
−0.50	30	50	0.006	−0.300	−0.463	−0.448	0.119	0.600	0.358	0.478
−0.50	50	500	−0.004	−0.102	−0.377	−0.390	0.016	0.052	0.159	0.180
−0.50	50	100	0.007	−0.200	−0.353	−0.388	0.080	0.171	0.208	0.302
−0.50	50	50	0.009	−0.232	−0.359	−0.419	0.160	0.244	0.326	0.501
−0.50	70	500	−0.010	−0.008	−0.140	−0.225	0.029	0.017	0.047	0.089
−0.50	70	100	−0.005	−0.027	−0.149	−0.244	0.157	0.111	0.182	0.280
−0.50	70	50	0.053	−0.031	−0.149	−0.250	0.311	0.261	0.373	0.546

Abbreviations: LL, log-logistic; PH, proportional hazard; PO, proportional odds.

TABLE 5 Bias and mean squared error (MSE) of $\hat{\beta}_{NPPR}$ (NPPR, nonparametric proportional risk) and $-\log(\widehat{HR}_{Cox})$ (Cox's proportional hazard [PH]) if the Weibull PH model is the true underlying model.

			Bias		MSE	
Effect	**Censoring (%)**	**Participants**	**NPPR**	**Cox**	**NPPR**	**Cox**
0.00	30	500	−0.004	−0.005	0.010	0.012
0.00	30	100	−0.007	−0.001	0.049	0.063
0.00	30	50	−0.002	−0.006	0.108	0.138
0.00	50	500	−0.001	−0.004	0.016	0.018
0.00	50	100	−0.005	−0.003	0.076	0.086
0.00	50	50	−0.024	−0.020	0.157	0.179
0.00	70	500	−0.001	0.004	0.028	0.027
0.00	70	100	−0.015	−0.020	0.141	0.151
0.00	70	50	0.009	0.007	0.302	0.344
0.50	30	500	−0.128	0.002	0.027	0.012
0.50	30	100	−0.128	0.015	0.071	0.063
0.50	30	50	−0.126	0.012	0.116	0.128
0.50	50	500	−0.091	−0.001	0.026	0.018
0.50	50	100	−0.082	0.012	0.098	0.106
0.50	50	50	−0.073	0.034	0.189	0.217
0.50	70	500	−0.061	−0.001	0.036	0.032
0.50	70	100	−0.051	0.015	0.167	0.162
0.50	70	50	−0.076	0.061	0.343	0.397
0.25	30	500	−0.066	0.003	0.014	0.011
0.25	30	100	−0.066	−0.005	0.051	0.058
0.25	30	50	−0.062	0.018	0.099	0.124
0.25	50	500	−0.051	0.000	0.017	0.016
0.25	50	100	−0.047	0.000	0.075	0.086
0.25	50	50	−0.041	0.006	0.174	0.201
0.25	70	500	−0.033	0.002	0.027	0.027
0.25	70	100	−0.028	0.004	0.146	0.149
0.25	70	50	−0.022	0.035	0.309	0.358
−0.25	30	500	0.056	−0.010	0.014	0.012
−0.25	30	100	0.073	−0.003	0.053	0.064
−0.25	30	50	0.069	−0.008	0.105	0.128
−0.25	50	500	0.058	0.009	0.018	0.016
−0.25	50	100	0.050	−0.002	0.087	0.095
−0.25	50	50	0.049	−0.010	0.149	0.184
−0.25	70	500	0.030	0.002	0.032	0.031
−0.25	70	100	0.008	−0.024	0.152	0.161
−0.25	70	50	0.056	−0.016	0.321	0.344
−0.50	30	500	0.120	−0.012	0.024	0.012
−0.50	30	100	0.138	0.000	0.071	0.065
−0.50	30	50	0.119	−0.023	0.120	0.140
−0.50	50	500	0.101	−0.001	0.024	0.016
−0.50	50	100	0.096	−0.015	0.085	0.090
−0.50	50	50	0.120	−0.006	0.165	0.183
−0.50	70	500	0.060	−0.002	0.033	0.028
−0.50	70	100	0.045	−0.036	0.148	0.151
−0.50	70	50	0.103	−0.015	0.350	0.348

TABLE 6 Bias and mean squared error (MSE) of $\hat{\beta}_{NPPR}$ (NPPR, nonparametric proportional risk) and $-\log(\widehat{OR}_{LL})$ (log-logistic PO, proportional odds) if the log-logistic PO model is the true underlying model.

			Bias		MSE	
Effect	**Censoring (%)**	**Participants**	**NPPR**	**LL**	**NPPR**	**LL**
0.00	30	500	−0.002	−0.003	0.009	0.026
0.00	30	100	0.008	0.009	0.043	0.124
0.00	30	50	0.010	0.020	0.096	0.294
0.00	50	500	−0.005	−0.005	0.014	0.028
0.00	50	100	0.000	0.001	0.077	0.160
0.00	50	50	0.007	0.016	0.145	0.324
0.00	70	500	0.008	0.011	0.028	0.039
0.00	70	100	−0.013	−0.014	0.160	0.236
0.00	70	50	−0.013	−0.024	0.331	0.517
0.50	30	500	−0.206	0.009	0.052	0.027
0.50	30	100	−0.211	0.011	0.089	0.125
0.50	30	50	−0.211	0.018	0.149	0.313
0.50	50	500	−0.163	0.008	0.041	0.030
0.50	50	100	−0.164	0.013	0.100	0.160
0.50	50	50	−0.191	−0.009	0.204	0.364
0.50	70	500	−0.107	−0.002	0.039	0.038
0.50	70	100	−0.119	0.004	0.180	0.229
0.50	70	50	−0.108	0.078	0.363	0.548
0.25	30	500	−0.105	0.001	0.020	0.025
0.25	30	100	−0.111	−0.003	0.062	0.139
0.25	30	50	−0.118	−0.007	0.102	0.262
0.25	50	500	−0.082	0.001	0.021	0.031
0.25	50	100	−0.069	0.022	0.074	0.149
0.25	50	50	−0.080	0.007	0.165	0.332
0.25	70	500	−0.059	0.000	0.030	0.040
0.25	70	100	−0.046	0.025	0.146	0.210
0.25	70	50	−0.073	−0.003	0.293	0.442
−0.25	30	500	0.099	−0.010	0.020	0.026
−0.25	30	100	0.089	−0.025	0.055	0.128
−0.25	30	50	0.115	0.012	0.110	0.296
−0.25	50	500	0.090	0.012	0.023	0.030
−0.25	50	100	0.091	0.003	0.075	0.136
−0.25	50	50	0.065	−0.036	0.163	0.327
−0.25	70	500	0.056	0.001	0.030	0.039
−0.25	70	100	0.061	0.010	0.153	0.207
−0.25	70	50	0.055	−0.035	0.313	0.515
−0.50	30	500	0.208	−0.007	0.053	0.026
−0.50	30	100	0.223	0.010	0.100	0.142
−0.50	30	50	0.219	−0.010	0.154	0.327
−0.50	50	500	0.164	−0.001	0.045	0.036
−0.50	50	100	0.158	−0.016	0.103	0.159
−0.50	50	50	0.166	−0.028	0.189	0.338
−0.50	70	500	0.108	−0.011	0.039	0.040
−0.50	70	100	0.105	−0.026	0.167	0.233
−0.50	70	50	0.133	−0.028	0.319	0.481

Abbreviation: LL, log-logistic.

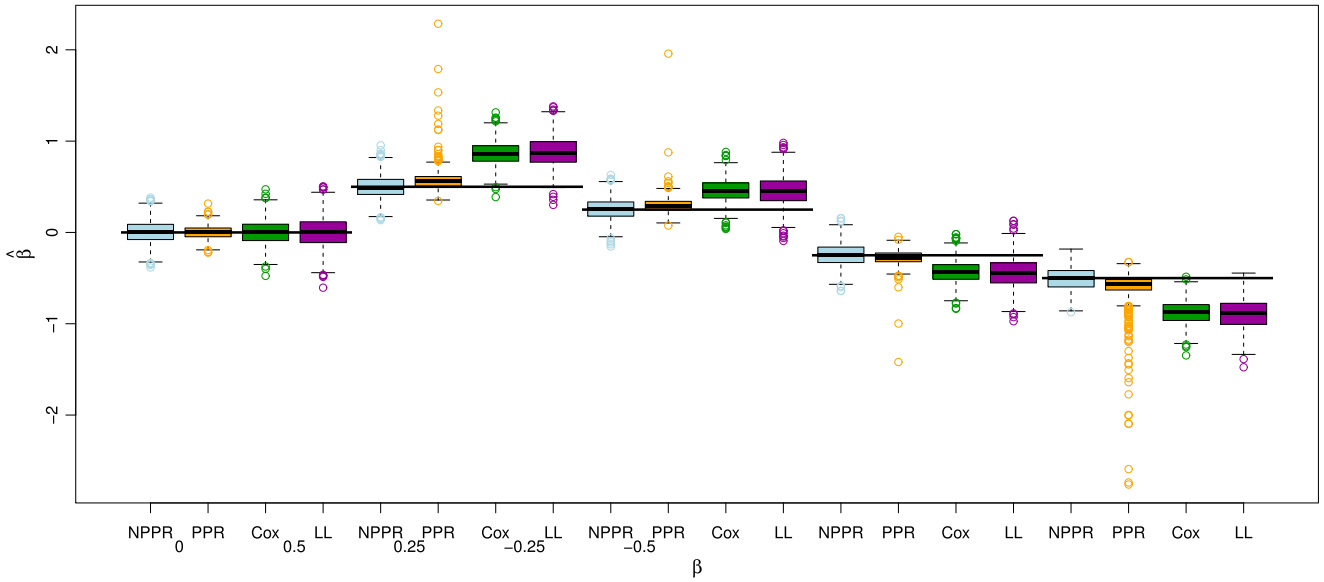

FIGURE 3 Comparison of boxplots for the estimated $\hat{\beta}_{NPPR}$ (NPPR, nonparametric proportional risk), $-\log(\widehat{RR}_{PPR})$ (PPR, parametric proportional risk), $-\log(\widehat{HR}_{Cox})$ (Cox), and $-\log(\widehat{OR}_{LL})$ (LL, log-logistic) for different choices of β if the PPR model is the underlying model, each in case of 500 participants and 50% censoring rate. Horizontal bars indicate the true underlying effect. Boxplots for 30% and 70% censoring rates and for the proportional hazard/proportional odds (PH/PO) cases are displayed in the supplementary material (Figures H–M).

supplementary material Figures F–G). For a comparison of different numbers of participants in case of a moderate effect of $\beta = 0.25$, under no censoring and assuming that all respective assumptions of the models are fulfilled, see Figure N of the supplementary material.

3.3.1 | Bias

Tables 4–6 and B (see supplementary material) depict the bias. In general, the bias is larger if the underlying model assumption is violated. In this case, a larger true effect results in a larger bias for all models.

It turns out that if the censoring rate is 50% or smaller the NPPR estimator consistently outperforms the PPR model for all underlying effects except of $\beta = 0$. This also holds for almost all other configurations, with only a few exceptions. For instance, considering a censoring rate of 70% and a true effect of $\beta = -0.25$ the bias of the NPPR model is for some configurations slightly larger than the one presented by the PPR model. If there is no treatment effect, that is, $\beta = 0$, the difference in biases mostly ranges in order of a magnitude of 0.004, which is very small. Only in case of a rather high amount of censoring, that is, 50%, and a small sample size of 50 participants this difference, given by 0.011, is noticeably larger. As expected, an application of Cox's PH model or the log-logistic PO model results in general in a larger bias, up to an absolute value of 0.495 (true value $\beta = 0.50$, 30% censoring, 500 participants) and 0.476 (true value $\beta = 0.50$, 30% censoring, 50 participants), respectively. In case of no censoring, the biases even exceed an absolute value of 0.500 multiple times for both models. If the true value of β is positive, both models tend to overestimate it, and underestimate it if the true value is negative. This is consistent with the trend observed for the HR and OR of the EU model over time (see Figure D in the supplementary material). Only in case of $\beta = 0$, the biases of all models are comparable in size, since with no effect present the PR, PH, and PO assumptions are fulfilled at the same time.

If the Weibull PH model is the true underlying model, the NPPR estimator has a slightly larger bias when trying to approximate the HR. It still never exceeds an absolute value of 0.138 if censoring is present, and an absolute value of 0.173 if there is no censoring. The negative values of β tend to be overestimated, while the positive ones are underestimated. Again, this behavior is consistent with the trend observed for the RR over time (see Figure 2). Similar results are observed if the log-logistic PO model is the true underlying model. The bias still never exceeds an absolute value of 0.223 and 0.259, respectively. Both Cox's PH and the log-logistic PO models performed adequate if the respective assumption is fulfilled. We note that the NPPR estimator is more precise for smaller censoring rates and a higher number of participants, as expected. Overall, it performs satisfactory and reliable in all cases under consideration.

Interestingly, all models tend to show a smaller bias for larger censoring rates if the respective assumption is violated. This most likely results from the censoring method, since later observations tend to be censored more often. Therefore, the events tend to concentrate on earlier time points for higher censoring rates. As seen in Figures 2 and D (see supplementary material), the RR, HR, and OR differ more at later time points. Consistent with this trend, the observed biases of all models are largest in case of no censoring if the respective assumption is violated.

3.3.2 | MSE

Tables 4–6 and B (see supplementary material) also depict the observed MSE. If the PPR model was assumed to be the true model, we note the same tendencies as seen with the bias. The NPPR estimator presents a smaller MSE for a low censoring rate of 30% in all cases. Considering 50% censoring it is overall smaller or equal to the one observed with the PPR model if $\beta \neq 0$ and for less than 500 participants if $\beta = 0.25$ or $\beta = -0.25$. The MSE of the NPPR estimator never exceeds 0.311 (true value $\beta = -0.50$, 70% censoring, 50 participants). On the other hand, the PPR model shows an MSE up to 0.600 (true value -0.50, 30% censoring, 50 participants). Again, the observed MSEs of Cox's PH model and the log-logistic PO model are in general rather large for $\beta \neq 0$ since the PH/PO assumption is violated. However, even in case of $\beta = 0$, the NPPR outperforms these two models. If the Weibull PH model is the true underlying model, the NPPR estimator presents a similar MSE as for the PR case. It never exceeds 0.350 (true value $\beta = -0.50$, 70% censoring, 50 participants). Cox's PH model performs comparable indicating a slightly larger variance as the bias is generally smaller. If the log-logistic PO model is the true underlying model, the observed MSEs behave similarly. In this case, the MSE of the NPPR model never exceeds 0.363. If no censoring is present, the MSE of the NPPR model only exceeds 0.100 twice (PO data, 50 participants, true value $\beta = -0.5,\ 0.5$). Overall, we conclude that the MSEs corresponding to the NPPR model show promising behavior, also in case of a violation of the PR assumption.

3.3.3 | Coverage

As in general, the covariances $\text{cov}(\hat{\beta}_t, \hat{\beta}_{t'})$ for $t \neq t'$, $t, t' \in \tilde{T}$, are unknown and it is therefore impossible to determine the variance of $\hat{\beta}$ in (2) directly, the CIs for the NPPR estimator were constructed according to the percentile bootstrap approach (Efron, 1981). Precisely, for each study with n participants, we drew an independent sample with replacement of size n from the simulated study. Consequently, the number of participants in each individual group is random and can differ from the one in the originally simulated study. The group assignment and status were not changed throughout. From this sample, $\hat{\beta}^*$ was estimated again using the NPPR estimator as described in Section 2. This procedure was repeated 500 times, yielding $\hat{\beta}^{*(1)}, \ldots, \hat{\beta}^{*(500)}$. From these values, the empirical 2.5%-quantile and 97.5%-quantile, respectively, were determined, defining the corresponding 95% CI. For the PPR and the log-logistic PO models, we used the multivariate delta method (van der Vaart, 1998) to construct the CIs. CIs for Cox's PH model were directly extracted from output of the *coxph* function.

Details on the coverage for each simulated case are presented in the supplementary material (Tables C–F). In general, if the PPR model is the true underlying model, the simulated coverage of the CI for the NPPR estimator is very close to the desired confidence level of 95%. It rarely falls below this value, the smallest value is given by 93.4% and the highest coverage is given by 98.2%. The latter is reached in case of a true effect of -0.25, 70% censoring, and 50 participants, which underlines the fact that CIs become rather conservative if the number of events is low.

Overall, the PPR model performs comparably. The smallest simulated coverage equals 89.4% and is observed in case of a true effect equal to 0.50 with a censoring rate of 30% and 100 participants. On the other hand, the smallest simulated coverage of Cox's PH model if censoring is present is equal to 80% ($\beta = -0.5$, 30% censoring, and 500 participants). The cases presenting a smaller coverage also present a larger bias. The coverage is lowest in case of no censoring (see Table F). Interestingly, it decreases with an increasing number of participants, most likely due to a more precise estimation of the HR, that is, close to the time-dependent RR but not equal to it. Therefore, the resulting bias and smaller variance combine to a lower coverage. However, in general, the coverage is very close to 95% especially for larger censoring rates. The coverage of the log-logistic PO model indicates a numerical problem since it often exceeds 99% and is never less than 94.4%. This is most likely caused by both the PO assumption and the parametric assumption being violated. If the Weibull PH model is the true underlying model, the coverage of the HR, approximated by the mean RR, is overall too low, which is a direct consequence of the violated PR assumption. However, we conclude that also in this case, for most of the

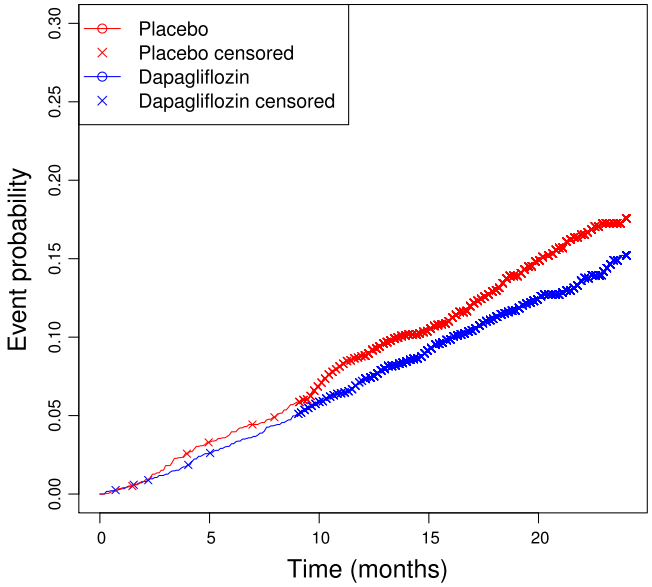

FIGURE 4 Event probability of the secondary outcome, all-cause mortality, and event time points during the DAPA-HF trial estimated with the Kaplan–Meier estimator.

configurations the approximation of the 95% level is still rather precise. If the log-logistic PO model is the true underlying model, the results are similar. The coverage of the OR is overall slightly smaller consistent with the bias being slightly larger. Both Cox's PH and the log-logistic PO models perform as expected if the respective assumption is fulfilled.

3.3.4 | Numerical robustness

The NPPR estimator proves to be overall robust, independent of the true underlying model. Assuming the PPR model, failures (less than 12 in 1000 simulation runs) were only observed in five of the 45 scenarios. This is also true for the Weibull PH model and the log-logistic PO model as underlying model, where the NPPR model only failed a very few times ($\leq$13), occurring only in case of a high censoring rate of 70% and a small sample size of 50 participants. Concerning the numerical robustness, the PPR model is clearly outperformed by the NPPR model, showing a higher robustness in almost every case. Cox's PH and the log-logistic PO models both mostly present an even greater robustness. For the sake of brevity, details are deferred to the supplementary material (Tables G–J).

4 | CASE STUDY: DAPA-HF TRIAL

To illustrate the NPPR estimator, we use data from the DAPA-HF trial (McMurray et al., 2019). This randomized, double-blind, placebo-controlled trial evaluated dapagliflozin, a sodium glucose cotransporter-2 (SGLT-2) inhibitor, for reducing severe cardiovascular outcomes in patients with heart failure, and reduced ejection fraction. In 410 sites in 20 countries, 4744 patients were treated for a median observation time of 18.2 months. Here, we report the results for the trial's secondary outcome, all-cause mortality, 274 (of 2373) patients experienced this outcome in the treatment group, but 324 (of 2371) in the control group. As we had no access to the original data, we digitized the Kaplan–Meier estimates from the original paper by the open software tool WebPlotDigitizer, version 3.8 (Rohatgi, 2015) and extracted the data by using the algorithms and R tools of Guyot et al. (2012). In order to validate this extraction process, we calculated an HR from the extracted data given by 0.837 (95% CI [0.712, 0.983]), which is essentially the same as the HR and its CI in the original paper, which was 0.83 (95% CI [0.71, 0.97]).

Figure 4 displays the estimated CDFs from the DAPA-HF trial. The assumption of PR seems to be reasonable. Using the method described in Section 2, we obtain $\hat{\beta}_{NPPR} = 0.178$ (95% CI [0.031, 0.346]), which corresponds to an RR of 0.837 (95% CI [0.708, 0.970]). The patients in the treatment group are therefore only 83.7% at risk of death for any cause relative to the patients in the placebo group. CIs were estimated using the bootstrap as described in Section 3.3.3. To visualize the

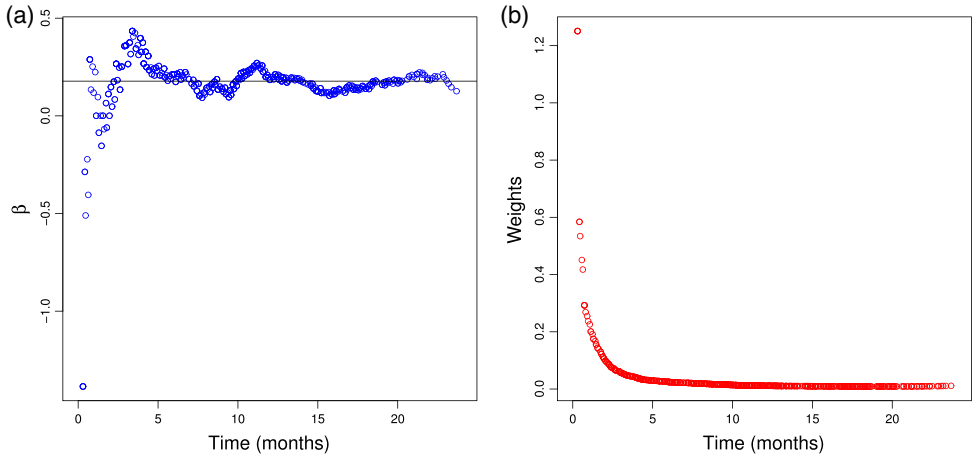

FIGURE 5 (a) Estimated $\hat{\beta}_t$ for the DAPA-HF trial at every event time point. The estimated $\hat{\beta}_{NPPR}$ overall is displayed by the horizontal line. (b) Weights at every event time point.

estimation process, Figure 5a shows the estimated $\hat{\beta}_t$ for each event time point. The weighted mean defining the NPPR estimator as defined in (2) is displayed by the solid line. The weights at each event time point are shown in Figure 5b.

The estimated NNT quickly decreases over time, likely due to the small number of reported events in the study. For instance, at the end of study ($t = 24$), only 29 (95% CI [19.256, 53.947]) patients treated with dapagliflozin would prevent one death relative to the placebo treatment. A detailed visualization of the NNT is shown in Figure O in the supplementary material.

5 | DISCUSSION AND CONCLUSION

In this paper, we proposed an NPPR model to assess a treatment effect in case of a two-group situation. Thereby, we solved the conceptional problems the HR and OR have shown in the past without losing the advantages of a nonparametric estimation method. By deducing the NNT from the model, we further provide an absolute measure in addition to the estimated RR. As no further specification beyond the PR assumption is necessary, the model is broadly applicable and provides a promising tool for the analysis of numerous applications, ranging from preclinical toxicology studies to late phase clinical trials, as, for instance RCTs. While not regarded in this paper, left truncated data could also be included (Tsai et al., 1987).

Using a bootstrap approach (Efron, 1981), we can construct CIs for the NPPR estimator and derive those of the estimated RR by multiplying with -1 and exponentiating. Due to the duality between CIs and statistical hypothesis tests, this further provides a test for the difference/ratio of the CDFs. To a certain degree, this holds true even if the PR assumption is violated since in this case the NPPR estimator evaluates the mean RR over time. Similar to the mean HR over time (Hernán, 2010) only regarding the mean RR from the beginning to the end of study might limit its interpretability. Therefore, again inspired by the mean HR (Hernán, 2010), we suggest to estimate the mean RR from the beginning up to multiple time points if the PR assumption is violated. However, if the estimated CDFs cross, the effect could be underestimated due to negative and positive $\hat{\beta}_t$ cancelling out and the method should be used with care. Hence, the NPPR estimator would highly benefit from the development of a goodness-of-fit test for the PR assumption.

Of course, the NPPR estimator comes along with some limitations. Up to this point, the model contains only one (binary) variable and it is not possible to include further variables in the model. A possible extension could be based on further stratification but might result in subgroup sample sizes that are too small for a proper evaluation of the Kaplan–Meier estimators. Moreover, the consideration of continuous variables is a challenging problem, which demands some future work. Another drawback, especially if compared to the parametric model, is the necessity of using bootstrap to calculate a CI as a formula for the variance of the estimator is not available yet. Alternatively, a different approach based on nonparametric maximum-likelihood estimation could circumvent these problems as well as supersede the calculation of weights.

Also, one important starting point of further research is the inclusion of competing risks, which is not possible as of now. A feasible way to address this issue without the need for a complicated construction might be to replace $1 - \hat{S}_i(t)$,

$i = 0, 1$, with the respective Aalen–Johansen estimator (Aalen & Johansen, 1978) since it is less susceptible to bias resulting from competing risks (Aalen & Johansen, 1978). However, in combination with an extension to multiple covariates, the incorporation of competing risks could offer an easier-to-interpret alternative (Andersen & Keiding, 2012) to the Fine–Gray subdistribution hazard model (Fine & Gray, 1999).

Another related topic would be a comparison of the PH and PR assumptions in terms of which is fulfilled more often in practice. This is a limitation of the given data example since the data do not seem to violate either.

We demonstrated that the NPPR estimator shows a very good performance if the PR assumption is fulfilled. For small to moderate censoring rates, it mostly outperforms the PPR model and is numerically more robust. The comparison with Cox's PH and the log-logistic models also shows that the direct estimation of the RR is advantageous over the approximation with the mean HR or the OR. If the PR assumption is not fulfilled it still shows satisfying behavior. However, if there are only few events, either due to high censoring rates or generally small sample sizes, problems can arise or even make an application impossible. In those cases, while still less robust, the PPR model can provide a better solution. The NPPR estimator, in general, estimates the mean RR over time. Therefore, applications outside of a PR scenario might also be of interest. Furthermore, similarly to the discussion about the PH assumption required from Cox's model, possible problems resulting from a violated PR assumption should be taken into consideration.

In summary, we strongly believe that the NPPR estimator is a useful addition to the existing tools for the analysis of time-to-event data, which not only circumvents the technical problems of the HR and the OR, respectively, but is also easy to interpret and comes along without an assumption on the underlying survival distribution.

ACKNOWLEDGMENTS

This work has been supported by the Research Training Group "Biostatistical Methods for High-Dimensional Data in Toxicology" (RTG 2624, P7) funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation - Project Number 427806116).

Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The R code, which can be used to reproduce the analysis of the case study and the simulation results, is available at https://github.com/LuciaAmeis/NPPR-model-for-time-to-event-data.

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the Supporting Information section.

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

ORCID

Lucia Ameis https://orcid.org/0009-0002-9420-6984
Oliver Kuss https://orcid.org/0000-0003-3301-5869
Annika Hoyer https://orcid.org/0000-0002-0241-9951
Kathrin Möllenhoff https://orcid.org/0000-0001-7861-3892

REFERENCES

Aalen, O. O., & Johansen, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics*, *5*(3), 141–150. http://www.jstor.org/stable/4615704

Andersen, P. K., Borgan, O., Gill, R. D., & Keiding, N. (1993). *Statistical models based on counting processes*. Springer Series in Statistics. Springer. Corrected second printing, 1997. https://doi.org/10.1007/978-1-4612-4348-9

Andersen, P. K., & Keiding, N. (2012). Interpretability and importance of functionals in competing risks and multistate models. *Statistics in Medicine*, *31*(11-12), 1074–1088. https://doi.org/10.1002/sim.4385

Breslow, N., & Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Annals of Statistics*, *2*(3), 437–453. https://doi.org/10.1214/aos/1176342705

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 187–202. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

Efron, B. (1981). Censored data and the bootstrap. *Journal of the American Statistical Association*, *76*(374), 312–319. https://doi.org/10.1080/01621459.1981.10477650

Fine, J. P., & Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, *94*(446), 496–509. https://doi.org/10.1080/01621459.1999.10474144

Guyot, P., Ades, A. E., Ouwens, M. J. N. M., & Nicky, J. W. (2012). Enhanced secondary analysis of survival data: Reconstructing the data from published Kaplan-Meier survival curves. *BMC Medical Research Methodology*, *12*, 9. https://doi.org/10.1186/1471-2288-12-9

Hartung, J., Knapp, G., & Sinha, B. (2008). *Statistical meta-analysis with applications*. Wiley Series in Probability and Statistics. Wiley. https://doi.org/10.1002/9780470386347

Hernán, M. A. (2010). The hazards of hazard ratios. *Epidemiology*, *21*(1), 13–15. Erratum in: *Epidemiology*, *22*(1), 134, (2011). https://doi.org/10.1097/EDE.0b013e3181c1ea43

Hildebrandt, M., Vervölgyi, E., & Bender, R. (2009). Calculation of NNTs in RCTs with time-to-event outcomes: A literature review. *BMC Medical Research Methodology*, *9*(21). https://doi.org/10.1186/1471-2288-9-21

Jackson, C. (2016). flexsurv: A platform for parametric survival modeling in R. *Journal of Statistical Software*, *70*(8), 1–33. https://doi.org/10.18637/jss.v070.i08

Jachno, K., Heritier, S., & Wolfe, R. (2019). Are non-constant rates and non-proportional treatment effects accounted for in the design and analysis of randomised controlled trials? A review of current practice. *BMC Medical Research Methodology*, *19*(103). https://doi.org/10.1186/s12874-019-0749-1

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, *53*(282), 457–481. https://doi.org/10.2307/2281868

Kalbfleisch, J. D., & Prentice, R. L. (2002). *The statistical analysis of failure time data* (2nd ed.). Wiley Series in Probability and Statistics. Wiley. https://doi.org/10.1002/9781118032985

Klein, J. P., & Moeschberger, M. L. (2003). *Survival analysis: Techniques for censored and truncated data* (2nd ed.). Statistics for Biology and Health. Springer. https://doi.org/10.1007/b97377

Kleinbaum, D. G., & Klein, M. (2012). *Survival analysis: A self-learning text* (3rd ed.). Statistics for Biology and Health. Springer. https://doi.org/10.1007/978-1-4419-6646-9

Kuss, O., & Hoyer, A. (2021). A proportional risk model for time-to-event analysis in randomized controlled trials. *Statistical Methods in Medical Research*, *30*(2), 411–424. https://doi.org/10.1177/0962280220953599

Lachin, J. M. (2010). *Biostatistical methods: The assessment of relative risks* (2nd ed.). Wiley Series in Probability and Statistics. Wiley. https://doi.org/10.1002/9780470907412

Mandel, J. (1984). *The statistical analysis of experimental data*. Dover Publications.

McMurray, J. J. V., Solomon, S. D., Inzucchi, S. E., Køber, L., Kosiborod, M. N., Martinez, F. A., Ponikowski, P., Sabatine, M. S., Anand, I. S., Bělohlávek, J., Böhm, M., Chiang, C.-E., Chopra, V. K., de Boer, R. A., Desai, A. S., Diez, M., Drozdz, J., Dukát, A., Ge, J., … Langkilde, A.-M. (2019). Dapagliflozin in patients with heart failure and reduced ejection fraction. *New England Journal of Medicine*, *381*(21), 1995–2008. https://doi.org/10.1056/NEJMoa1911303

NCBI. (1996). *PubMed*. https://pubmed.ncbi.nlm.nih.gov/?term=time-to-event

Oehlert, G. W. (1992). A note on the delta method. *American Statistician*, *46*(1), 27–29. https://doi.org/10.2307/2684406

Porta, M. (2016). *A dictionary of epidemiology*. Oxford University Press. Current online version. https://doi.org/10.1093/acref/9780199976720.001.0001

Rohatgi, A. (2015). *WebPlotDigitizer*. https://automeris.io/WebPlotDigitizer/

Rosner, B. (2016). *Fundamentals of biostatistics* (8th ed.). Cengage Learning.

Royston, P., & Parmar, M. K. (2014). An approach to trial design and analysis in the era of non-proportional hazards of the treatment effect. *Trials*, *15*(314). https://doi.org/10.1186/1745-6215-15-314

Schwartz, L. M., Woloshin, S., & Welch, H. G. (1999). Misunderstandings about the effects of race and sex on physicians' referrals for cardiac catheterization. *New England Journal of Medicine*, *341*(4), 279–283. https://doi.org/10.1056/NEJM199907223410411

Sistrom, C. L., & Garvan, C. W. (2004). Proportions, odds, and risk. *Radiology*, *230*(1), 12–19. https://doi.org/10.1148/radiol.2301031028

Struthers, C. A., & Kalbfleisch, J. D. (1986). Misspecified proportional hazard models. *Biometrika*, *73*(2), 363–369. https://doi.org/10.2307/2336212

Sutradhar, R., & Austin, P. C. (2018). Relative rates not relative risks: Addressing a widespread misinterpretation of hazard ratios. *Annals of Epidemiology*, *28*(1), 54–57. https://doi.org/10.1016/j.annepidem.2017.10.014

The Academy of Medical Sciences. (2017). *Sources of evidence for assessing the safety, efficacy and effectiveness of medicines*. https://acmedsci.ac.uk/file-download/86466482

Therneau, T. M., & Grambsch, P. M. (2000). *Modeling survival data: Extending the Cox model*. Springer. https://doi.org/10.1007/978-1-4757-3294-8

Therneau, T. M. (2023). *A package for survival analysis in R*. R Package version 3.5-7. https://CRAN.R-project.org/package=survival

Tsai, W., Nicholas, P. J., & Wang, M. (1987). A note on the product-limit estimator under right censoring and left truncation. *Biometrika*, *74*(4), 883–886. https://doi.org/10.2307/2336484

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. https://doi.org/10.1017/CBO9780511802256

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ameis, L., Kuss, O., Hoyer, A., & Möllenhoff, K. (2024). A nonparametric proportional risk model to assess a treatment effect in time-to-event data. *Biometrical Journal*, *66*, 2300147. https://doi.org/10.1002/bimj.202300147