

RESEARCH

Open Access



# Safety analysis of new medications in clinical trials: a simulation study to assess the differences between cause-specific and subdistribution frameworks in the presence of competing events

Astrid Genet<sup>1\*</sup>, Kathrin Bogner<sup>2</sup>, Ralf Goertz<sup>2</sup>, Sarah Böhme<sup>1</sup> and Friedhelm Leverkus<sup>1</sup>

## Abstract

Safety is an essential part of the evaluation of new medications and competing risks that occur in most clinical trials are a well identified challenge in the analysis of adverse events. Two statistical frameworks exist to consider competing risks: the cause-specific and the subdistribution framework. To date, the application of the cause-specific framework is the standard practice in safety analyses. Here we analyze how the safety analysis results of new medications would be affected if instead of the cause-specific the subdistribution framework was chosen. We conducted a simulation study with 600 participants, equally allocated to verum and control groups and a 30 months follow-up period. Simulated trials were analyzed for safety in a competing risk (death) setting using both the cause-specific and subdistribution frameworks. Results show that comparing safety profiles in a subdistribution setting is always more pessimistic than in a cause-specific setting. For the group with the longest survival and a safety advantage in a cause-specific setting, the advantage either disappeared or a disadvantage was found in the subdistribution analysis setting. These observations are not contradictory but show different perspectives. To evaluate the safety of a new medication over its comparator, one needs to understand the origin of both the risks and the benefits associated with each therapy. These requirements are best met with a cause-specific framework. The subdistribution framework seems better suited for clinical prediction, and therefore more relevant for providers or payers, for example.

**Keywords** Competing risks, Drug development, Cause-specific versus subdistribution hazard ratio, Safety analysis, Simulation study

## Background

Safety data are an essential part of the clinical evaluation of new medicinal products and regulatory submissions. However, their analysis might be challenged by the existence of competing risks. These are intercurrent events, defined as mutually exclusive events (death, other adverse events, change of treatment, noncompliance, end of study, etc.) whose occurrence precludes the event of interest from happening [1]. Competing risks are common. They are present in the vast majority of clinical

\*Correspondence:

Astrid Genet  
astrid.genet@Pfizer.com

<sup>1</sup> Pfizer Pharma GmbH, Linkstraße 10, Berlin 10785, Germany

<sup>2</sup> AMS Advanced Medical Services GmbH, Am Exerzierplatz 2, Mannheim 68167, Germany

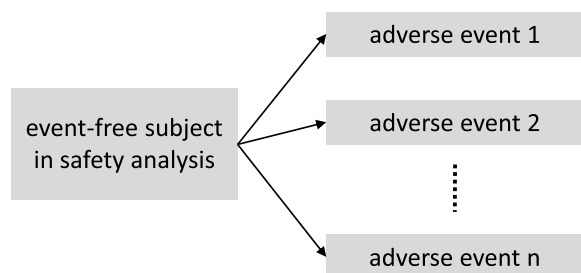


© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

trials [2, 3] and might bias the results [3, 4]. They represent a well-recognized problem in the analysis of adverse events [5, 6] and general recommendations urge the use of survival techniques that methodically account for the presence of competing risks [2, 5–7]. These techniques acknowledge that for a given adverse event there are other types of risks that occur at the same time.

The standard survival data situation corresponds to a Markov process with the two states: “event-free” and “event”. Splitting the “event” state into more states corresponding to different causes (“event 1”, “event 2”, “dead”, etc.) results in a Markov model for competing risks [8]. The analytical object in the presence of competing risks is the same as in standard marginal survival analysis: to estimate the probabilities, also named risks, and hazard rates of the event of interest over time and, if relevant, to assess whether there are differences between groups. However, a competing risks setting that extends the capabilities of analysis of two state survival models to deal with multi-state models (cf. Fig. 1) is required, when subjects can experience more than one event [9]. The risk of the event of interest over time is estimated among the risk of other competing events whose occurrence precludes it from happening. The concepts of risks and rates generalize easily to the competing risk situation: hazard rates become cause-specific hazard rates and risks become cumulative incidences [10].

Two statistical frameworks exist to perform survival analysis in the presence of competing risks: the cause-specific and the subdistribution settings. All standard methods for survival data apply to the cause-specific setting [11, 12] which focuses on the cause-specific hazard function. This function estimates the probability of each type of event separately, right-censoring individuals at the time of the competing event, as well as for loss of follow-up, withdrawal, or at the end of the observation time. For the subdistribution setting, specific approaches were developed that based on the cumulative incidence function [13, 14]. This function focuses on the cumulative incidence (or “subdistribution”) from a particular cause



**Fig. 1** Graphical presentation of the Markov multi-state model in a competing risks setting

and does not treat competing events as censored observations. Consequently, individuals remain at risk for the event of interest even after they experienced the competing risk. This contra intuitive representation is however necessary to predict the correct cumulative incidence functions [13].

These settings differ in their definitions. The aim of this study is to compare their properties and to recommend how to perform safety analyses in clinical research and regulatory submissions. We investigate whether systematic differences exist between the estimates obtained with each approach and define to what extent the interpretation of the results of survival analysis depends on the choice of one or the other setting. For both settings non-parametric approaches [4, 8, 14, 15] as well as regression models [13, 16] exist. Classical hazard-based methods for survival data apply when analyzing cause-specific hazards: Kaplan–Meier and Nelson–Aalen estimators as well as the Cox proportional hazards regression model. These methods, however, do not allow to draw inference for subdistribution functions of competing risks. Specific approaches were developed: the Aalen–Johansen estimator and the Fine and Grey model. This paper focuses on (semi-) parametric approaches: cause-specific (Cox regression) and subdistribution hazard regression (Fine & Gray model). Both offer two major advantages in comparison to the non-parametric approaches. First, they allow to adjust for covariates when assessing and comparing event probabilities over time and thus provide more insight into the mechanisms that lead to the occurrence of an event. Second, they allow to use a fitted model to make predictions (e.g., for certain attributes of the population under study).

In the **Methods** section, we provide a brief, nontechnical description of the cause-specific and subdistribution settings in survival analysis. Detailed technical descriptions can be found elsewhere [8, 13, 16, 17]. A short introduction to the non-parametric estimators can be found in the Additional file 1. To examine and to compare the properties of the cause-specific and subdistribution settings in survival analysis, a simulation study was conducted. It covers all possible practical outcomes: from superiority to inferiority of the medical intervention and from small to large effect sizes. In the **Results** section, we report the results of safety analyses performed on each simulated dataset with a cause-specific and a subdistribution setting. Finally, the practical interest of both approaches is discussed in terms of their implications and relevance for safety analyses of new medicinal products and regulatory submissions. We insist on the point that the appropriate choice of either cause-specific Cox models or Fine-Gray subdistribution hazard models depends on the precise question of interest to the

researcher or stakeholder: their differences make them complementary rather than interchangeable: they represent different points of views and therefore answer different research questions.

**Methods**

**Cause-specific hazard regression**

The Cox proportional hazards model [18] is expressed by the hazard function  $h(t)$  presented in Eq. 1. In this model,  $h(t)$  is determined by a set of covariates and expressed as:

$$h(t) = h_0(t) \times e^{\beta'X} \tag{1}$$

where  $t$  is the time,  $X$  is a vector of covariates,  $\beta'$  is the vector of regression coefficients that measures the effect size of each covariate on the hazard and  $h_0(t)$  is the baseline hazard, under the assumption that all explanatory variables are either set to zero ( $X=0$ ) or represent average values. The quantities of interest in the Cox model are the hazard ratios (HR)  $e^{\beta_j}$ , where  $j \in \{1, 2, \dots, c\}$  represents the  $c$  covariates considered in the analysis. HR are relative measures of an effect between different values taken by a covariate. They do not provide any information about the absolute risks. A very common covariate investigated in medical and pharmacological research is the group attribution. In this case, in a binary setting with two groups, the HR associated with this covariate is the ratio of the rates of occurrence of an event in both groups. A value equal to one indicates no differences between the groups, a value of less than one indicates a higher and a value of more than one a lower rate of occurrence in the reference group. The cause-specific Cox proportional hazard model is a natural extension of the standard Cox proportional hazard regression where a model is fitted separately to each cause-specific hazard by censoring all individuals who experienced one of the competing risks before the event of interest.

**Subdistribution hazard regression**

While in the Cox model the hazard for the event of interest only depends on its own (cause-specific) hazard, Fine and Gray [13] proposed a model that expresses an instantaneous hazard function  $h(t)$  by the cumulative (subdistribution) hazard function  $F(t)$  that is described in Eq. 2. The subdistribution model contains an additive component and the instantaneous risk of occurrence of the event of interest  $k$ ,  $F_k(t)$ , depends on all cause-specific risks. It can be expressed as:

$$F_k(t) = F_{0,k}(t) \times e^{\sum_i \gamma_i X_i} \tag{2}$$

where  $k$  is the event of interest while  $i \in \{1, 2, \dots, n\}$  represents all the competing events (including  $k$ ) considered in the analysis. Analogous to the expression of the Cox

model presented in Eq. 1,  $t$  is the time and,  $X$  is a vector of covariates. Similar to  $\beta$  in the Cox model,  $\gamma$  is the vector of regression coefficients measuring the effect size of each covariate on the cumulative hazard.  $F_{0,k}(t)$  is the baseline cumulative hazard, that is the cumulative hazard under  $X=0$ . The regression coefficients  $\gamma$  can be interpreted in a similar way as the  $\beta$  from a Cox model, except that they are relative measures of risk between the values of certain covariate, taking into account competing events occur that preclude the occurrence of the event of interest. This means that the size of the effect due to each competing event on the HR for the event of interest cannot be isolated. It should be noted that the model considers an extended risk set where individuals are still at risk for the event of interest even after they experienced the competing risk. Fine and Gray acknowledged that this is unnatural but necessary in order to get a model that correctly predicts cumulative incidence functions [13].

**Simulation study**

A simulation study was conducted to investigate the differences in the results of safety analysis performed in presence of competing risks when the subdistribution setting is chosen instead of the standard cause-specific setting. Three possible outcomes were considered, i. e. (1) superiority of the verum group compared to the control group, (2) inferiority of the verum group compared to control group, and (3) equivalence between both groups.

As commonly done in biometrics, HR was estimated to compare the risk of occurrence of the adverse event of interest between the verum and control groups [19, 20]. HR in the cause-specific setting ( $HR_{cs}$ ) was fitted by a Cox regression model and in the subdistribution setting ( $HR_{sd}$ ) by a Fine and Gray model.

The following assumptions were made:

- a. Each study comprised 600 patients allocated into two study groups (verum and control) in a 1:1 ratio.
- b. Two competing event types (the adverse event of interest and death as competing risk) were simulated with event times for both types following an exponential distribution. We selected a common and simple one-parameter event time distribution that implies a time constant hazard rate  $h(t)=h$  that makes it easy to control the characteristics of the simulated data [21]. The hazard rates of the exponential distributions were defined according to the targeted median time to event  $t_{0,5}$  for a given treatment group and a given event type:

$$h = \frac{\log(2)}{t_{0,5}}$$

- c. Administrative censoring occurred after 30 months if neither the primary event (adverse event of interest) nor the competing event (death) had occurred in a patient by then.
- d. The characteristics of the distribution of the competing event death were kept constant across all simulated scenarios. Median survival was set to 20 and 10 months for the verum and control groups, respectively, which corresponds to a HR of 0.5, in favor of the verum group.
- e. Median time to first adverse event was incremented in 1-month intervals between 1 and 20 months in the verum and control groups resulting in 400 patterns (i. e. 20×20), hereafter referred to as “conditions of interest”.

The following statistics were reported for the competing risk analysis of each condition of interest:

- a. Median time to adverse event (AE), median time to death and their corresponding standard deviations;
- b.  $HR_{cs}$  and  $HR_{sd}$ , their corresponding 95% confidence intervals and two-sided  $p$ -values to investigate for group differences.

The simulation of each of these conditions of interest was repeated 1 000 times. This value was chosen as a reasonable trade-off between the required accuracy of the results and the available computational resources. The statistics were separately assessed on each of the 1 000 datasets generated for each condition on interest and then pooled together according to Rubin’s rule (cf. Fig. 2) [22].

**Presentation of results**

For an initial assessment of whether changing the setting from cause-specific to subdistribution leads to a change in the three possible outcomes of the safety analysis

performed in a competing risk setting, results were classified into nine possible categories. These categories were defined in a two-step process:

First, starting from a cause-specific setting, the results of survival analysis of the 400 conditions of interest were classified into the possible outcome categories:

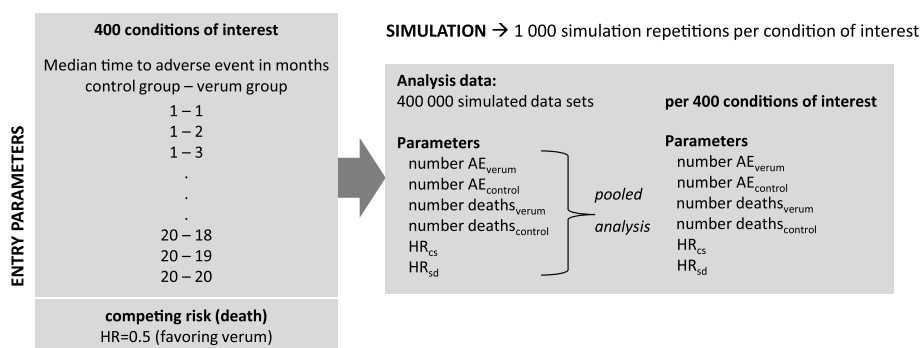
- (a) Superiority if  $HR_{cs} < 1$  and  $p\text{-value} \leq 0.05$
- (b) Inferiority if  $HR_{cs} > 1$  and  $p\text{-value} \leq 0.05$
- (c) Equivalence if  $HR_{cs} p\text{-value} > 0.05$

Second, for each condition of interest, we assessed whether  $HR_{sd}$  fell in the same outcome category as the  $HR_{cs}$  or in one of the two possible alternative categories. This resulted in nine possible outcome categories when switching from the cause-specific to the subdistribution setting. The proportion of the 400 conditions of interest falling in each of these nine possible categories was reported.

Heat maps provide a graphical overview of the results, from the classification of the 400 conditions of interest to the observed differences between the true HR and the outcomes of the survival analysis performed in both the cause-specific and subdistribution settings.

**Software**

All analyses were conducted using R version 3.6.1 [23]. The R package *survival* was used to fit the Cox proportional hazards model [17, 24]. The R package *cmprsk* was used to fit the Fine and Gray model [25]. The pooled-analysis of the parameters of the 1 000 simulated repetitions for each of the 400 condition of interest was done by the R package *mice* [26]. Heat maps for the graphical presentation of results were created with the R package *ggplot2* [27]. Detailed information on how to use these R packages can be found in the original publication for each package.



**Fig. 2** Overview of the simulation study procedure, the input parameters, and estimated parameters

### Results

The true value of the HR ( $HR_{true}$ ) for safety analysis, defined as the entry value given to simulate the occurrence of safety events, is known. The HR of the competing risk death was kept constant to 0.5 in favor of verum over all simulations. Each condition of interest was categorized according to  $HR_{true}$ : superiority ( $HR_{true} < 1$ ), inferiority ( $HR_{true} > 1$ ) or equivalence ( $HR_{true} = 1$ ) of the verum group compared to the control group. In 47.5% of the simulated conditions of interest, verum was safer than control (superiority). The amount of the conditions of interest where verum was less safe than control (inferiority) was the same. In the remaining 5% of simulated conditions of interest, verum and control were equally safe (equivalence) (cf. Fig. 3).

Estimated competing risks  $HR_{cs}$  of the 400 conditions of interest ranged from 0.05 to 20.16 and  $HR_{sd}$  from 0.15 to 11.36.

As is detailed in Fig. 3, 35.8% of the competing risks safety analysis performed in a cause-specific setting resulted in superiority, 35.8% in inferiority and 28.4% in equivalence of the verum group compared to the control group. The slight differences between  $HR_{cs}$  and  $HR_{true}$  can be easily explained. Unlike  $HR_{true}$ ,  $HR_{cs}$  is calculated in a competitive risks setting where patients who experienced the competing event (death) before the event of interest were censored. Censoring leads to a reduced number of patients, especially towards the end of the observation period which may as well reduce the

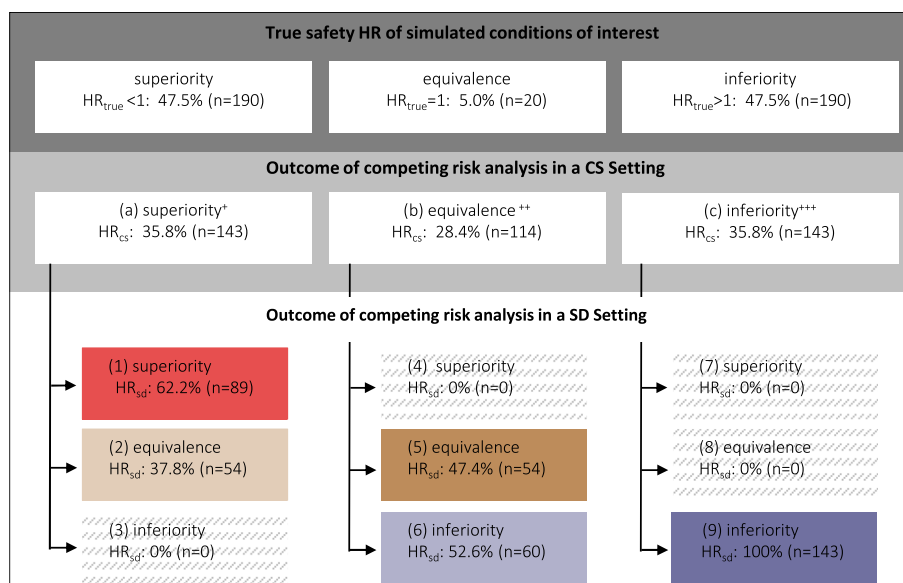
statistical power needed to detect existing differences between the treatment groups.

Among the conditions of interest that show superiority of verum in a cause-specific setting ( $n = 143$ ), 62.2% still showed superiority when analyzed in a subdistribution setting (category 1). For the remaining 37.8%, the superiority of the verum group disappeared in the subdistribution setting. Statistical tests were not significant and the outcome category changed from superiority of verum to equivalence (category 2). A change from superiority of verum in the cause-specific setting to inferiority of verum in the subdistribution setting was not observed (category 3; cf. Fig. 3).

Of the conditions that showed equivalence between verum and control in a cause-specific setting ( $n = 114$ ), none showed superiority when analyzed in a subdistribution setting (category 4). Equivalence between the two treatment groups remained in about half of the conditions of interest (47.4%; category 5) while the other half (52.6%, category 6) turned to inferiority of verum when analyzed in a subdistribution setting (cf. Fig. 3).

Among the conditions of interest that resulted in inferiority of verum in a cause-specific setting ( $n = 143$ ), all remained significantly disadvantageous for verum (category 9) when analyzed in a subdistribution setting, the other possible outcomes of category 7 and category 8 were not observed in the study (cf. Fig. 3).

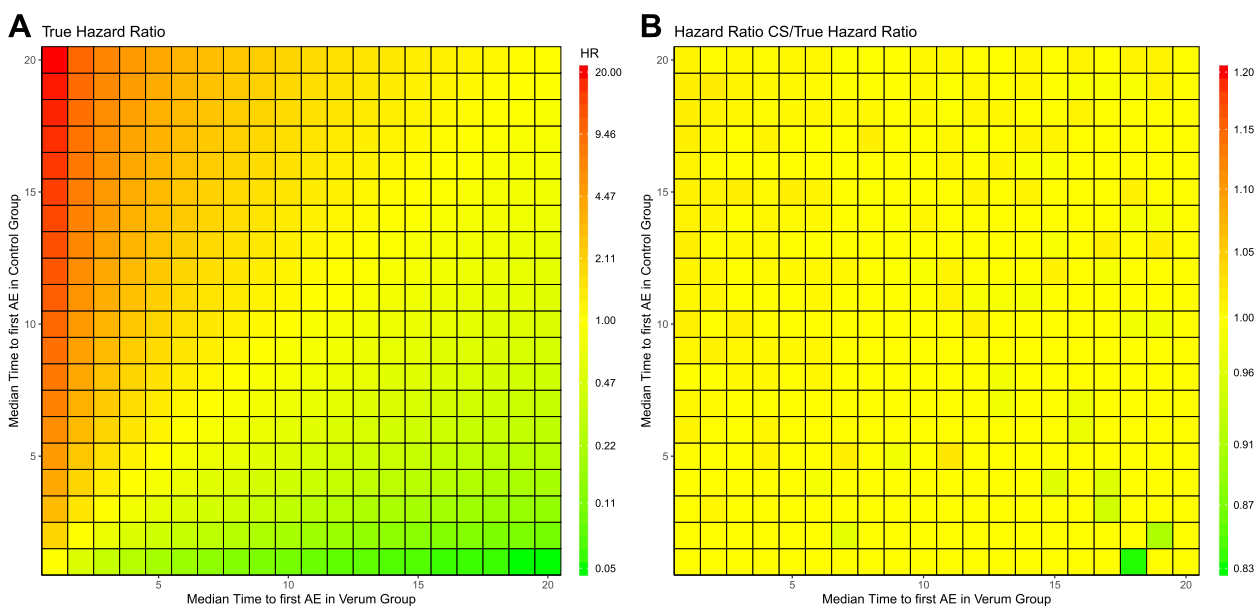
The heat map of panel (A) in Fig. 4 presents the  $HR_{true}$  of the entry values for the 400 conditions of interest in



Definitions of outcome categories in cause-specific (CS) & subdistribution (SD) settings:  
 \* superiority:  $HR < 1$  & p-values  $> 0.05$ ; \*\* equivalence: p-value  $> 0.05$ ; \*\*\* inferiority:  $HR > 1$  & p-values  $> 0.05$

Note: HR of the competing risk death was kept constant to 0.5 in favor of verum over all simulations in the cause-specific and subdistribution settings

**Fig. 3** Classification of results of competing risk analyses in cause-specific and subdistribution settings into outcome categories



**Fig. 4** Heat map of **A**  $HR_{true}$  and **B** the ratio  $(HR_{cs}/HR_{true})$ . Legend: Each heat map in figure 4 plots the hazard ratios of all 400 conditions of interest of simulated safety analysis with competing risks by median time to first adverse event in verum and control groups

the simulated safety study with a constant HR of the competing risk (death) of 0.5 in favor of the verum group. The verum group is superior to the control group if the median time to first adverse event is longer than in the control group ( $HR_{true} < 1$ , green shadings, lower right part). Conversely, the verum group is inferior to the control group if the median time to first adverse event is shorter in the verum group than in the control group ( $HR_{true} > 1$ , red shadings, upper left part). If the median time to first adverse event is the same for both groups, they are considered equivalent ( $HR_{true} = 1$ , yellow shadings, the diagonal separating lower right and upper left parts). Figure 4, panel (B) shows for each condition of interest in the simulation study in the cause-specific setting the ratios of  $HR_{cs}$  and  $HR_{true}$ . Accordingly, a ratio around the value of 1 (yellow shadings) indicates no difference between  $HR_{cs}$  and  $HR_{true}$ . This is observed for most ratios of the 400 conditions of interest in the simulated safety study in a cause-specific setting; independent of median time to first adverse event in verum and control groups. For ratios with values less than 1 (green shadings),  $HR_{cs}$  is lower than  $HR_{true}$ . This is observed for some conditions of interest, especially when the median time to first adverse event is much higher in the verum than in the control group. Only in these cases is a deviation in  $HR_{cs}$  observed in favor of the verum group. Ratios with values above 1 indicates higher  $HR_{cs}$  than  $HR_{true}$  values which is not observed in the simulated data.

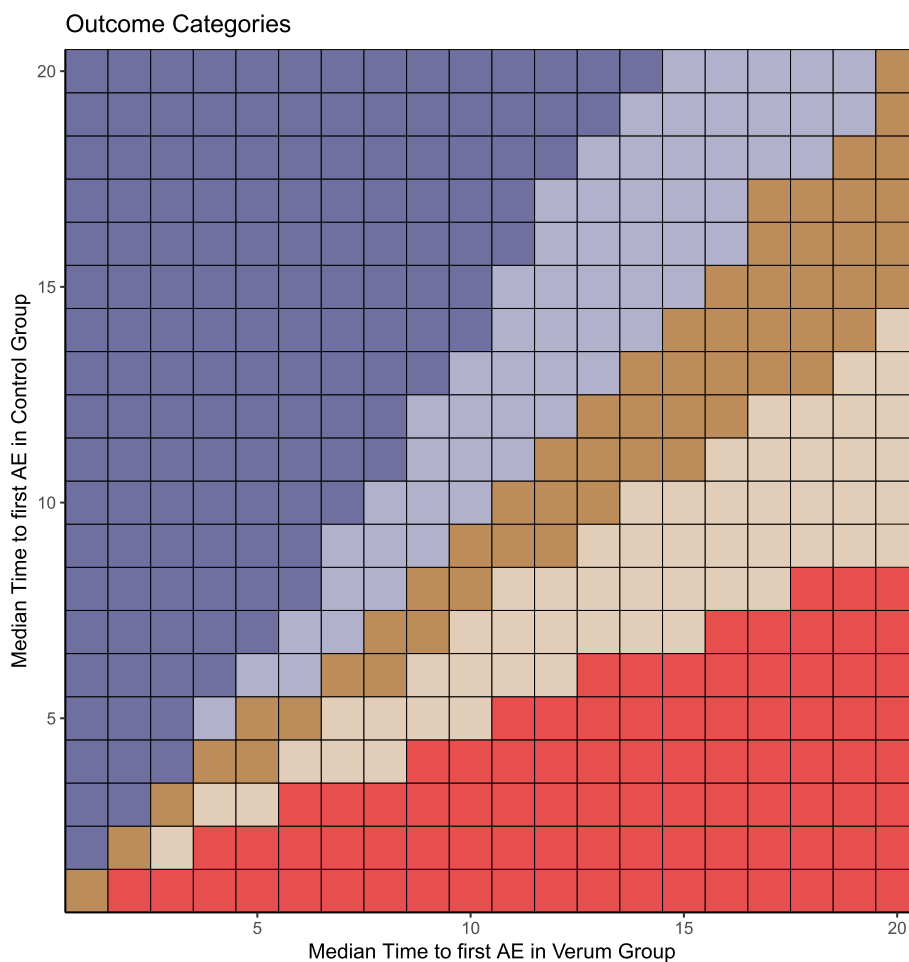
Finally, the heat map in Fig. 5 indicates the categories into which the conditions of interest are classified

according to the  $HR_{sd}$  from simulated safety analysis after switching from the cause-specific to the subdistribution setting (see also Fig. 3).

The safety analysis in the subdistribution setting, as in the cause-specific setting, resulted in superiority of the verum group over the control group if the median time to first adverse event in the control group is short and occurs earlier than in the verum group (red shading in Fig. 5; corresponds to category 1 in Fig. 3). However, when the median time to first adverse event in the control group increases, but is still shorter than in the verum group, results in the subdistribution setting no longer show superiority, but equivalence between both groups (light brown shading in Fig. 5; corresponds to category 2 in Fig. 3).

For the conditions of interest, for which in the cause-specific setting the safety in both groups was equal, the analysis in the subdistribution setting also shows equivalence, if the median time to first adverse event for both groups is close to each other (dark brown shading in Fig. 5; corresponds to category 5 in Fig. 3). If the median time to first adverse event however is earlier in the verum than in the control group, the outcome changes from equivalence in the cause-specific setting to inferiority of the verum group in the subdistribution setting (light blue shading, in Fig. 5; corresponds to category 6 in Fig. 3).

For all conditions of interest, for which in the cause-specific setting an inferiority of the verum group was the result of safety analysis, this is also confirmed in the



**Fig. 5** Map of outcome categories of safety analysis with competing risks in the subdistribution setting. Legend: Figure 5 plots all outcome categories of safety analyses for all 400 conditions of interest by median time to first adverse event in verum and control groups. The colors represent the outcome categories described in Fig. 3

subdistribution setting (dark blue shading in Fig. 5; corresponds to category 9 in Fig. 3).

All other possible outcome categories (3), (4), (7), and (8) of safety analysis when switching from cause-specific to subdistribution setting are not present in the simulation study data (see also Fig. 3).

In sum, the results of the simulated safety analysis with death as a competing event show that comparing safety profiles in a subdistribution setting is always more pessimistic than in a cause-specific setting. For the group with the longest survival and the safety advantage there is either no more advantage or a newly found disadvantage compared to its analysis in the cause-specific setting.

**Discussion**

**Understanding the etiology of risks for clinical evaluation**

Defining the benefit/risk balance of medications in comparison to that of the standard of care in a given indication, implies to understand the origin of both the

risks and the benefits associated with each therapy. The decision is based on acceptable trade-offs. Addressing epidemiological questions of etiology has long been recognized as the strength of the cause-specific setting [4, 12, 28–31], because of the censoring at the competing event. Censoring equals “disallowing” competing events so that censored patients could still experience the event of interest. Considering this hypothetical population, in which the event of interest would eventually happen for everyone, prevents competing events to get in the way when one is interested in comparing instantaneous rates of occurrence of the event of interest, between an intervention and its comparator. However, this hypothetical population may not be suitable for all research questions.

In the subdistribution setting, an extended risk set is considered where individuals remain even after they experienced the competing event. Fine and Gray themselves acknowledged that this approach is unnatural, since a patient who experienced the competing event

of death is no longer at risk of developing an adverse event. The consequence of this construction is that causal parameters are not accurately estimated. But this extended risk set is necessary to correctly predicts cumulative incidence functions, particularly useful for prediction. In this sense, the method is best suited to predict clinical outcomes on patients at risk (and eventually assess the impact of covariates on those outcomes) [4, 12, 28–31]. By pointing out the treatment with the lowest probability of all types of events within a given time frame, it brings a perspective on the data that may be of interest for providers, payers, or policymakers who need to predict the burden on human and financial resources of clinical events on patients enrolled in the care system [16, 32].

#### **Cause-specific and subdistribution framework when survival competes safety**

The results of our simulation study give a clear picture of the differences between both safety analysis settings. When analyzing safety data, prolonged survival in one group will mostly translate into a higher probability of adverse events in a subdistribution setting, where the risk is assessed by combining the hazards of all competing events within a single cumulative incidence function over the entire follow-up period. Our simulation shows that the results of the subdistribution analysis are always more pessimistic than the results of the cause-specific analysis. For the group with the longest survival and the safety advantage a change of the analysis setting translates in either a smaller advantage, no more advantage, a larger or even a newly found disadvantage.

However, the outcomes of neither the cause-specific nor the subdistribution settings are biased, they just answer different research questions. The subdistribution outcome reflects the effect of treatment on both safety and survival, with no possibility to differentiate between the two, while the cause-specific analysis reflects the effect of treatment on safety only.

When very serious adverse events are considered and longer living comes at the price of unbearable safety events, the outcome of the subdistribution analysis could be used to compare the safety profiles of both medicinal products. However, in most cases, prolonged survival is still very much desirable despite the occurrence of minor or manageable adverse events. In this case, the effect of survival present in the subdistribution outcome does not allow to interpret the safety profile of the intervention.

#### **Recommendations for clinical evaluations**

As a general rule, we recommend, to first describe the competing risks as well as their expected impact on the analysis. When competing risks have been identified,

competing risks analysis should be preferred to marginal analysis when the number of competing events in the study is at least equal to that of the event of interest [33], or when the absolute percentage of competing events is greater than 10% [13]. When competing risks analysis is indicated, we recommend a cause-specific setting, together with a justification of the choice of the competing events considered. This recommendation is in line with the suggestions made by the Committee for Medicinal Products for Human Use of the European Medicines Agency [7] in its Composite variable strategies.

The Cox proportional hazards model that is routinely presented in clinical study reports should remain the standard approach. The presentation of Kaplan-Meier estimates is also justified, although said to overestimate cumulative event probabilities [4, 6, 33, 34]. Kaplan-Meier in a cause-specific setting represents the absolute risk of having an event of interest, as if nothing else could happen before [29]. In comparison, Aalen-Johansen estimates the fraction of patients who will experience an event of interest within the given time frame, given the presence of other precluding events. The cause-specific setting therefore allows many more subjects to experience the event of interest. This explains the observation, also made in our simulation, that Kaplan-Meier estimates are systematically larger than those derived from the Aalen-Johansen method. Although this effect should be known and understood, we do not agree with the terminology commonly used in the literature that Kaplan-Meier “overestimates” the incidence of events. This wording implies that one setting delivers correct estimates and the other not, while it is in fact a matter of context.

As an alternative estimator to Kaplan-Meier for the same function, the Nelson-Aalen estimator could be considered [35]. Our simulation confirmed that it delivers the same information as the Kaplan-Meier estimator in comparative analysis, but its understanding is less straightforward. For this reason, the Nelson-Aalen estimator is less popular than the Kaplan-Meier estimator in time-to-event analysis since its first publication in the late 1950s [36]. As clinical study reports are also meant to be reviewed by non-statisticians such as medical experts and epidemiologists within the frame of regulatory activities and clinical evaluations, the well-known and commonly presented Kaplan-Meier curve should be favored. One might argue that there is no harm in presenting both, but we do not recommend it as a standard approach. Clinical reports usually contain large amounts of analyses, and the non-essential presentation of the Nelson-Aalen estimator for each endpoint, might cause most readers to feel overwhelmed.



## Limitations

In this study, we chose to keep the time to competing event constant in both study groups and across all simulated scenarios. It was therefore not possible to investigate further discrepancies between both settings on various times to occurrence of the competing events. It would be interesting to confirm that the conclusion of this work remains valid for a wide range of time to competing event. Also, the number of patients in both study groups was kept constant across the simulated scenarios. An interesting question to investigate would be how sample size influences the results. The size of the trial impacts the statistical significance that is the  $p$ -value and the breadth of the confidence interval of the estimates. Gaining deeper insights into the role of sample size is particularly interesting for the special case of rare disease and pediatric trials where only small numbers of eligible trial participants are available. Finally, event times were simulated with an exponential distribution. This simple, known, parametric distribution is widely used to simulate survival data to investigate the properties of the Cox Model [37]. It offers an easy control of the regression coefficients and has proportional hazards, which is advantageous for the implementation. However, it assumes that the baseline hazard function is constant, which is not always the case, especially in the setting of AEs, that tend to occur shortly after starting treatment and death, where some patients might be too sick to rescue when they enter the trial.

An exponential distribution was deemed sufficient for this application, where the focus was to compare methodological approaches rather than to perform a realistic description of various survival time data. However, more complex statistical approaches have been described [37–40] and it would be interesting to investigate how the simulation framework influences the results.

## Conclusions

When analyzing survival data in the presence of competing events, there is no absolute right or wrong when it comes to the choice between a cause-specific and a subdistribution setting. The decision rather depends on the research question at hand. We claim that the risk/benefit profile of a medication is better assessed in a cause-specific setting. The authorities in charge assess the effect of the intervention on the risk of experiencing adverse events. They need estimates of the instantaneous risk of adverse events while on treatment, as well as separate estimates of the effect of the intervention on the competing events. These requirements can be met in a cause-specific setting but not in a subdistribution setting where a single cumulative incidence function that includes all

the risks in presence is estimated. The subdistribution setting may be relevant, however, if economic questions should be answered or when both events are similar in the clinical harm (e. g. Death and extremely serious adverse events that tremendously impact patients' well-being and Quality of life). The Kaplan-Meier estimate of the survival function, or its complement, and the Cox proportional hazard model for comparative analysis should remain the standard approach in clinical study reports. In the presence of competing risks, they should be embedded in a cause-specific setting and the choice of the competing events in the analysis should be justified.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12874-023-01985-7>.

**Additional file 1.** Non-parametric inference of survival data.

## Acknowledgements

Not applicable.

## Authors' contributions

AG, SB, and FL were involved in the conception of this analysis. AG, SB, FL, KB and RG discussed and agreed on the methods used for the analysis. AG conducted the literature search and computed the simulation study. AG, KB und RG analyzed the data, prepared tables and figures, and wrote the initial draft of the manuscript. All authors revised the manuscript. All authors read and approved the final manuscript.

## Funding

Not applicable.

## Availability of data and materials

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

Received: 13 January 2023 Accepted: 25 June 2023

Published online: 13 July 2023

## References

- Allignol A, Beyersmann J, Schmoor C. Statistical issues in the analysis of adverse events in time-to-event data. *Pharm Stat.* 2016;15(4):297–305. <https://doi.org/10.1002/pst.1739>.
- Koller MT, Raatz H, Steyerberg EW, Wolbers M. Competing risks and the clinical community: irrelevance or ignorance? *Stat Med.* 2012;31(11–12):1089–97. <https://doi.org/10.1002/sim.4384>.
- van Walraven C, McAlister FA. Competing risk bias was common in Kaplan-Meier risk estimates published in prominent medical journals. *J*

- Clin Epidemiol. 2016;69:170–173.e178. <https://doi.org/10.1016/j.jclinepi.2015.07.006>.
4. Schuster NA, Hoogendijk EO, Kok AAL, Twisk JWR, Heymans MW. Ignoring competing events in the analysis of survival data may lead to biased results: a nonmathematical illustration of competing risk analysis. *J Clin Epidemiol*. 2020;122:42–8. <https://doi.org/10.1016/j.jclinepi.2020.03.004>.
  5. Stegherr R, Beyersmann J, Jehl V, Rufibach K, Leverkus F, Schmoor C, Friede T. Survival analysis for Adverse events with VarYing follow-up times (SAVVY): rationale and statistical concept of a meta-analytic study. *Biom J*. 2021;63(3):650–70. <https://doi.org/10.1002/bimj.201900347>.
  6. Stegherr R, Schmoor C, Beyersmann J, Rufibach K, Jehl V, Brückner A, et al. Survival analysis for Adverse events with VarYing follow-up times (SAVVY)-estimation of adverse event risks. *Trials*. 2021;22(1):420. <https://doi.org/10.1186/s13063-021-05354-x>.
  7. European Medicines Agency. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials EMA/CHMP/ICH/436221/2017. 2020. [https://www.ema.europa.eu/documents/scientific-guideline/ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical-principles\\_en.pdf](https://www.ema.europa.eu/documents/scientific-guideline/ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical-principles_en.pdf).
  8. Borgan Ø. Three contributions to the encyclopedia of biostatistics: the Nelson-Aalen, Kaplan-Meier, and Aalen-Johansen. 1997.
  9. Therneau TM, Crowson C, Atkinson E. Multi-state models and competing risks. 2020. p. 1–29.
  10. Andersen PK, Geskus RB, de Witte T, Putter H. Competing risks in epidemiology: possibilities and pitfalls. *Int J Epidemiol*. 2012;41(3):861–70. <https://doi.org/10.1093/ije/dyr213>.
  11. Geskus RB. Data analysis with competing risk and intermediate states. Chapman and Hall/CRC; 2016.
  12. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med*. 2007;26(11):2389–430. <https://doi.org/10.1002/sim.2712>.
  13. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94(446):496–509. <https://doi.org/10.2307/2670170>.
  14. Gray RJ. A class of K-sample tests for comparing the cumulative incidence of a competing risk. *Ann Stat*. 1988;16(3):1141–54.
  15. Edwards JK, Hester LL, Gokhale M, Lesko CR. Methodologic issues when estimating risks in pharmacoepidemiology. *Curr Epidemiol Rep*. 2016;3(4):285–96. <https://doi.org/10.1007/s40471-016-0089-1>.
  16. Klein JP, van Houwelingen HC, Ibrahim JG, Scheike TH. Handbook of survival analysis. 1st ed. 2013. (10.12.2013 ed.).
  17. Therneau TM, Grambsch PM. Modeling survival data: extending the Cox model. 1st ed. New York: Springer; 2000. <https://doi.org/10.1007/978-1-4757-3294-8>.
  18. Cox DR. Regression models and life-tables. *J R Stat Soc Ser B Methodol*. 1972;34(2):187–220.
  19. Collet D. Modelling survival data in medical research. 2015.
  20. Zwiener I, Blettner M, Hommel G. Survival analysis part 15 of a series on evaluation of scientific publications. *Dtsch Arztebl Int*. 2011. <https://doi.org/10.3238/arztebl.2011.0163>.
  21. Klein JP, Moeschberger ML. Survival analysis - techniques for censored and truncated data. 2nd ed. New York: Springer; 2003.
  22. Rubin DB. Multiple imputation for nonresponse in surveys. 1987.
  23. RCoreTeam. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2019.
  24. Therneau TM, Lumley T, Elizabeth A, Cynthia C. R package survival. 2021. <https://github.com/therneau/survival>.
  25. Gray B. R package cmprsk: competing risks regression. In: Subdistribution analysis of competing risks. 2022. <https://cran.r-project.org/web/packages/cmprsk/index.html>.
  26. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. *J Stat Softw*. 2011;45(3):1–67.
  27. Wickham H. ggplot2 elegant graphics for data analysis. 2nd ed. Cham: Springer; 2016.
  28. Austin PC, Allignol A, Fine JP. The number of primary events per variable affects estimation of the subdistribution hazard competing risks model. *J Clin Epidemiol*. 2017;83:75–84. <https://doi.org/10.1016/j.jclinepi.2016.11.017>.
  29. Austin PC, Lee DS, Fine JP. Introduction to the analysis of survival data in the presence of competing risks. *Circulation*. 2016;133(6):601–9. <https://doi.org/10.1161/circulationaha.115.017719>.
  30. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. *Am J Epidemiol*. 2009;170(2):244–56. <https://doi.org/10.1093/aje/kwp107>.
  31. Van Der Pas S, Nelissen R, Fiocco M. Different competing risks models for different questions may give similar results in arthroplasty registers in the presence of few events. *Acta Orthop*. 2018;89(2):145–51. <https://doi.org/10.1080/17453674.2018.1427314>.
  32. Pepe MS, Mori M. Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Stat Med*. 1993;12(8):737–51. <https://doi.org/10.1002/sim.4780120803>.
  33. Berry SD, Ngo L, Samelson EJ, Kiel DP. Competing risk of death: an important consideration in studies of older adults. *J Am Geriatr Soc*. 2010;58(4):783–7. <https://doi.org/10.1111/j.1532-5415.2010.02767.x>.
  34. Satagopan JM, Ben-Porat L, Berwick M, Robson M, Kutler D, Auerbach AD. A note on competing risks in survival data analysis. *Br J Cancer*. 2004;91(7):1229–35. <https://doi.org/10.1038/sj.bjc.6602102>.
  35. Colosimo E, Ferreira F, Oliveira M, Sousa C. Empirical comparisons between Kaplan-Meier and Nelson-Aalen survival function estimators. *J Stat Comput Simul*. 2002;72:299–308. <https://doi.org/10.1080/00949650212847>.
  36. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53:457–81.
  37. Bender R, Augustin T, Blettner M. Generating survival times to simulate Cox proportional hazards models. *Stat Med*. 2005;24(11):1713–23. <https://doi.org/10.1002/sim.2059>.
  38. Beyersmann J, Allignol A, Schumacher M. Competing risks and multistate models with R. New York: Springer; 2012.
  39. Beyersmann J, Latouche A, Buchholz A, Schumacher M. Simulating competing risks data in survival analysis. *Stat Med*. 2009;28(6):956–71. <https://doi.org/10.1002/sim.3516>.
  40. Wan F. Simulating survival data with predefined censoring rates for proportional hazards models. *Stat Med*. 2017;36(5):838–54. <https://doi.org/10.1002/sim.7178>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

