SPECIAL ISSUE PAPER

Pharmaceutical Statistics

(wileyonlinelibrary.com) DOI: 10.1002/pst.1739

Published online 1 March 2016 in Wiley Online Library

Statistical issues in the analysis of adverse events in time-to-event data

Arthur Allignol,[a]* Jan Beyersmann,[a] and Claudia Schmoor[b]

The aim of this work is to shed some light on common issues in the statistical analysis of adverse events (AEs) in clinical trials, when the main outcome is a time-to-event endpoint. To begin, we show that AEs are always subject to competing risks. That is, the occurrence of a certain AE may be precluded by occurrence of the main time-to-event outcome or by occurrence of another (fatal) AE. This has raised concerns on 'informative' censoring. We show that, in general, neither simple proportions nor Kaplan–Meier estimates of AE occurrence should be used, but common survival techniques for hazards that censor the competing event are still valid, but incomplete analyses. They must be complemented by an analogous analysis of the competing event for inference on the cumulative AE probability. The commonly used incidence rate (or incidence density) is a valid estimator of the AE hazard assuming it to be time constant. An estimator of the cumulative AE probability can be derived if the incidence rate of AE is combined with an estimator of the competing hazard. We discuss less restrictive analyses using non-parametric and semi-parametric approaches. We first consider time-to-first-AE analyses and then briefly discuss how they can be extended to the analysis of recurrent AEs. We will give a practical presentation with illustration of the methods by a simple example.

Keywords: Aalen-Johansen; competing risks; safety; survival

1. INTRODUCTION

The analysis of safety in terms of adverse events (AEs) is relevant in almost all clinical trials [1]. Generally, these are reported as incidence proportions calculated by crude rates, that is, the number of patients with AEs divided by the number of patients in the sample [2–4]. AEs however can occur at any point in time during the patient's time under observation, which is not taken into account by such crude estimates. This is particularly relevant if follow-up durations differ between treatment groups, for example, if an experimental treatment successfully prolongs survival. Also, note that individual follow-up durations will almost always differ between patients because of staggered study entry. Moreover, in a time-to-event setting, a crude rate estimator that ignores censoring can be highly biased, in particular when the proportion of dropouts differs between treatment groups [2–4].

For these reasons, survival analysis techniques are advocated for analysing AEs when the main outcome of a clinical trial is a time-to-event endpoint. We further argue that AEs are always subject to competing risks. That is, the occurrence of the AE of interest might be precluded by the occurrence of the main time-to-event outcome, or another, fatal, AE (e.g. [5–7]). This has raised concerns on informative censoring. For instance, Nishikawa *et al.* [5] also advocate the use of competing risks analysis to deal with 'informative' censoring. The implication is that the occurrence of competing events should be correctly taken care of when analysing AEs. For instance, the simple incidence proportion is a correct estimator only if vital and AE statuses are known for all patients. Kaplan–Meier estimates of AE occurrence must not be used, but common survival techniques for hazards that censor the competing event are still valid, but incomplete analyses. They must be complemented by an analogous analysis of the competing event for inference on the cumulative AE probability. For simplification, we restrict the presentation mainly to methods that consider just the time-to-first AE. We will, however, briefly indicate how the hazard-based techniques that are crucial for analysing time-to-first AE can also be used for recurrent AEs.

As depicted in Figure 1, we consider the simple situation in which patients may either experience at least one AE of a certain type or die without prior AE. In the following, the short-term death always means death without prior AE. Patients enter the study free of AE and alive and may experience at least one AE or die without prior AE, whichever comes first. The number of patients who experience at least one AE is denoted by #AE, and the number of study patients is denoted by n, such that #AE $\leq n$. So our main target quantity is the probability to experience at least one AE.

The aim of this paper is to explain and illustrate basic concepts of the correct analysis of AEs in patients with different observation times due to right-censoring. We illustrate the methods by using data from a clinical trial in oncology. We have chosen not to explain the medical context, because of two reasons. First, we have to make the data anonymous because of confidentiality at this stage. Second, we regard this as an advantage because we want to avoid subject matter considerations and discussions in order to concentrate just on the methodological aspects of the different analyses. We start with an illustration of the methods in

[a] *Institute of Statistics, Ulm University, Ulm, Germany*

[b] *Clinical Trials Unit, University Medical Center Freiburg, Freiburg, Germany*

**Correspondence to: Arthur Allignol, Institute of Statistics, Ulm University, Ulm, Germany.*
E-mail: arthur.allignol@uni-ulm.de

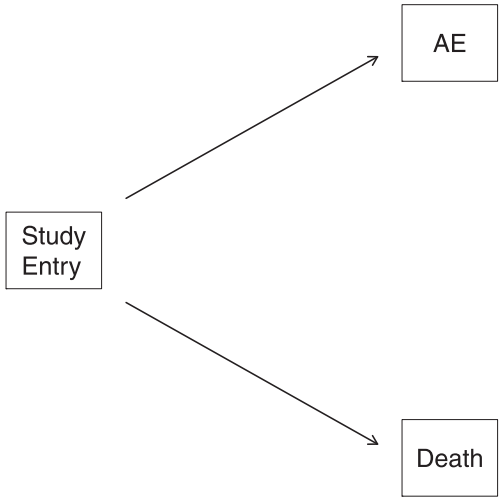

Figure 1. Graphical display of the data situation. After study entry, a patient may either experience at least one adverse event (AE) (=AE), or die without prior AE (=Death).

the simple situation where the data are complete in the following sense: we consider a data set of 200 patients with complete follow-up information about AEs and vital status up to 2 years after study entry for all patients. After 2 years, 54 patients experienced at least one AE, while 48 died without prior AE. The remaining 100 patients experienced neither of the competing events up to 2 years. In the sequel, the illustration of the methods in the situation of incomplete data is performed by modifying our data set by adding artificial right-censoring such that patients will have different follow-up times.

The paper is organised as follows: Section 2 deals with the non-parametric estimation of the probability of AE. Section 3 presents the competing risks setting more formally and discusses the use of the incidence rate (IR) for summarising the risk of AE. In Section 4, semi-parametric models for the comparison of treatment groups are presented. Section 5 gives some more details on competing risks and censoring, which is an omnipresent topic in the analysis of competing risks data, and also briefly outlines how the present methodology may be used when AEs are recurrent. The paper concludes with a discussion in Section 6.

2. NON-PARAMETRIC ESTIMATION OF THE PROBABILITY OF ADVERSE EVENTS

2.1. Complete data

We begin by considering complete data without censoring, that is, up to a certain time in follow-up, say τ, it is known for every patient if an AE or death has occurred. This simplified situation allows for an accessible and natural introduction towards the competing risks situation. For instance, the simple fact that, with an infinite follow-up time, $\mathrm{P(AE)} + \mathrm{P(Death)} = 1$ gets sometimes lost in the situation of censoring. For instance, if follow-up ends at one common calendar time for all patients, but patient entry was staggered, the individual follow-up durations will differ, and the aforementioned information will, at best, only be known up to the minimum follow-up time. The complete data situation also permits to introduce the quantities we want to estimate, that is, the quantities that are approximated when considering right-censored data.

In the complete data situation without censoring, the incidence proportion is calculated by the crude rate

$$\frac{\#\mathrm{AE}}{n}, \tag{1}$$

where #AE is the number of patients with at least one AE in $[0, \tau]$. The quotient $\#\mathrm{AE}/n$ is the correct estimator of the probability to experience at least one AE, $\mathrm{P(AE)} = \mathrm{P(AE\ in\ } [0, \tau])$, in $[0, \tau]$.

Patients may actually die before experiencing an AE; thus, death is a competing event (competing risk) for AE. In other words, death may preclude the observation of an AE – after death, the AE cannot occur anymore. Thus, with an infinite follow-up time,

$$\frac{\#\mathrm{AE}}{n} + \frac{\#\mathrm{Death}}{n} = 1. \tag{2}$$

Still considering complete data without censoring, the estimated probability of experiencing at least one AE within some time-interval $[0, t]$ is given by

$$\hat{\mathrm{P}}(\mathrm{AE\ in\ } [0, t]) = \frac{\#\mathrm{AE\ in\ } [0, t]}{n}. \tag{3}$$

Then on $[0, t]$, the estimated probability to experience the composite event, that is, AE or death without prior AE, is

$$\begin{aligned} &\hat{\mathrm{P}}(\mathrm{AE\ in\ } [0, t]) + \hat{\mathrm{P}}(\mathrm{Death\ w/o\ prior\ AE\ in\ } [0, t]) \\ &= \frac{\#\mathrm{AE\ in\ } [0, t]}{n} + \frac{\#\ \mathrm{Death\ in\ } [0, t]}{n} \\ &= \frac{\#\ \mathrm{AE\ or\ death\ (whatever\ comes\ first)\ in\ } [0, t]}{n} \\ &= 1 - \hat{\mathrm{P}}(T > t), \end{aligned} \tag{4}$$

with T the time to first AE or death without prior AE, whatever comes first.

In our simple example of 200 patients with complete follow-up up to 2 years (54 with AE and 48 deaths prior to AE), the probability to experience an AE in time interval $[0, 2]$, $\mathrm{P(AE\ in\ } [0, 2])$, is correctly estimated by the incidence proportion calculated by the crude rate $54/200 = 0.27$. The estimated probability to experience an AE plus the estimated probability to die without prior AE, $\frac{\#\ \mathrm{Death\ in\ } [0, 2]}{n} = 48/200 = 0.24$, sum up to the estimated probability to experience the composite event, that is, AE or death, $1 - \hat{\mathrm{P}}(T > t) = 1 - 98/200 = 0.51$.

Sometimes, the Kaplan–Meier estimator is used to estimate $P(\mathrm{AE\ in\ } [0, t])$ by treating death without prior AE as censored observation. Numerous technical arguments explaining why the Kaplan–Meier estimator is a biased estimator of $\mathrm{P(AE\ in\ } [0, t])$ have been given [8]. An intuitive one is that 1 minus the Kaplan–Meier estimator aims at approximating a distribution function, that is, it tends towards one as t gets larger. However, as noted before, $\mathrm{P(AE\ in\ } [0, t]) + \mathrm{P(Death\ in\ } [0, t])$ tends towards 1 as t gets larger. Thus, in a true competing risks situation, the probability of AE is strictly smaller than 1, and consequently, the Kaplan–Meier estimator to estimate $P(\mathrm{AE\ in\ } [0, t])$ will be biased upwards.

2.2. Right-censored data

We now consider the situation, where the event time T defined previously is not observed for all patients until all times t, $t \leq \tau$. We only observe the minimum of a right-censoring time and

the event time T. In this situation, the quantities presented in (3) and (4) cannot be computed anymore as the numerator is now unknown. It is remarkable that this well-known difficulty has led to a widespread use of appropriate statistical methods for censored time-to-event data in efficacy analyses, but not in the analysis of safety data.

The probability $P(T > t)$ of not experiencing the composite event can be estimated by the Kaplan–Meier estimator censoring patients who experience their first AE or death after their observed–censored event time,

$$\hat{P}_{KM}(T > t) = \prod_{u \le t} \left(1 - \frac{\#\text{ AE or death at } u}{\#\text{ under observation before } u}\right), \quad (5)$$

where # under observation before u denotes the number of patients under observation with neither AE nor death before u, that is, the so-called risk set, and where the product is over all observed unique (AE or death) event times u in $(0, t]$. The Kaplan–Meier estimator aims at approximating

$$1 - \frac{\#\text{ AE or death in } [0, t]}{n}. \quad (6)$$

Note that without censoring, (5) and (6) are equal.

The incidence proportion should not be used to estimate $P(\text{AE in } [0, t])$ because it is biased. For example,

$$\frac{\#\text{ AEs in } [0, t]}{n}$$

estimates $P(\text{AE in } [0, t] \text{ and } T \le \text{censoring time})$. This is not a relevant quantity because a patient's safety concern is directed towards $P(\text{AE in } [0, t]) \ge P(\text{AE in } [0, t] \text{ and } T \le \text{censoring time})$. In words, the incidence proportion of AE underestimates the probability of AE in the presence of censoring.

As already stated in Section 2.1, the Kaplan–Meier estimator by treating death without prior AE as censored observation should also not be used to estimate $P(\text{AE in } [0, t])$, because it overestimates the probability of AE.

The Aalen–Johansen estimator of the cumulative incidence function (CIF) is the correct method for estimating the probability of AE in the presence of competing risks. The CIF of AE, denoted by $P(T \le t, \text{AE})$, is the expected proportion of patients experiencing an AE over the course of time.

To derive an estimator for the CIF of AE, we use the fact that 1 minus the Kaplan–Meier estimator of the probability of not experiencing the composite event AE or death up to time t, $1 - \hat{P}(T > t)$, can be written as follows:

$$1 - \hat{P}(T > t) = \sum_u \hat{P}(T > u-) \cdot \frac{\#\text{ AE or death at } u}{\#\text{ under observation before } u}, \quad (7)$$

with $\hat{P}(T > u-)$ denoting the Kaplan–Meier estimator of the probability of not experiencing the composite event AE or death just before time u. Equation 7 is easily shown by checking the increments $\hat{P}(T > u-) - \hat{P}(T > u)$ at jump times u. The intuition behind the right-hand side of (7) is that the probability of experiencing one of the events AE or death $1 - \hat{P}(T > t)$ is obtained by 'summing' the probability of not experiencing one of the events up to time u times the conditional probability to experience one of the events exactly at time u given no prior event. In other words, the right-hand side of (7) can be interpreted as the sum over empirical probabilities to have an event – either AE or death – at an observed event time u, $u \le t$. This sum then is the empirical probability to have an AE or death event in $[0, t]$.

The same reasoning leads to an estimator of the CIF of AE. Instead of summing over the empirical probability of experiencing the composite event of AE or death, we sum over the empirical probability of experiencing an AE, that is,

$$\hat{P}(T \le t, \text{AE}) = \sum_u \hat{P}(T > u-) \cdot \frac{\#\text{ AE at } u}{\#\text{ under observation before } u}. \quad (8)$$

In the absence of censoring, (8) equals

$$\frac{\#\text{ AE in } [0, t]}{n}, \quad (9)$$

which again highlights that the incidence proportion is a correct estimator for complete data, but not for censored data. In contrast, the Kaplan–Meier-type estimator for AE occurrence is always biased, and it does not equal (9) in the absence of censoring.

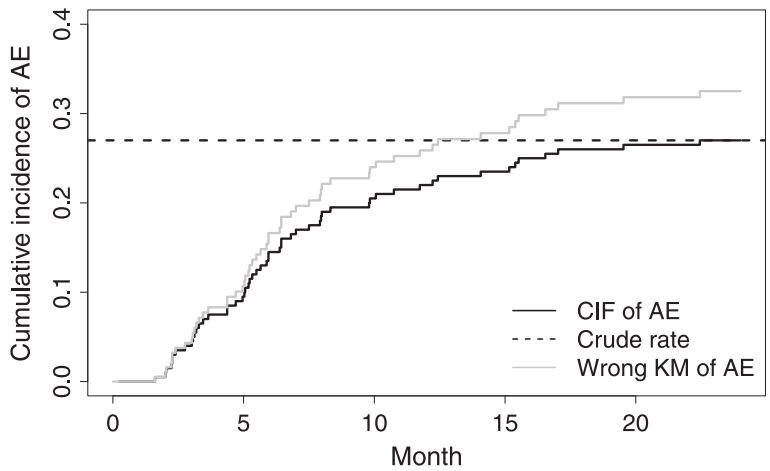

Figure 2. Probability of adverse event (AE) estimated in the complete data set by the Aalen–Johansen estimator of the cumulative incidence function (CIF) (black), the incidence proportion calculated as the crude rate (dashed line), and the wrong Kaplan–Meier estimator (grey).

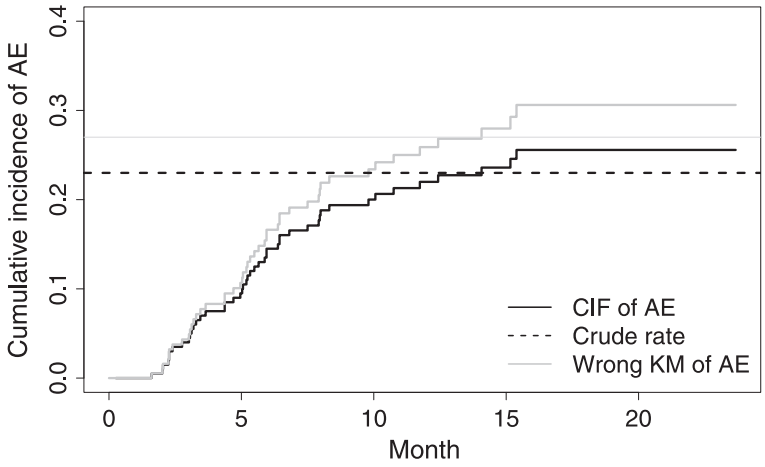

Figure 3. Probability of adverse event (AE) estimated in the right-censored data set by the Aalen–Johansen estimator of the cumulative incidence function (CIF) (black), the incidence proportion wrongly calculated as the crude rate (dashed line), and the wrong Kaplan–Meier estimator (grey). The horizontal grey line gives the incidence proportion computed from the complete data.

We finally note that

$$1 - \hat{\mathrm{P}}(T > t) = \hat{\mathrm{P}}(T \leq t, \mathrm{AE}) + \hat{\mathrm{P}}(T \leq t, \mathrm{Death}), \qquad (10)$$

that is, 1 minus the Kaplan–Meier estimator of the probability of not experiencing the composite event AE or death up to time t, is equal to the sum of the Aalen–Johansen estimators of the probabilities of experiencing the competing events AE and death up to time t. The balance Equation 10 is violated, if the terms on the right-hand side are estimated via Kaplan–Meier.

2.3. Illustration

The different methods for estimation of the probability of an AE are now illustrated in our simple example. We start with the data set being complete up to 2 years. Figure 2 shows the incidence proportion calculated as the crude rate (1), the Aalen–Johansen estimator of the CIF (8), and the wrong Kaplan–Meier estimator treating death prior to AE as censored observations. The Aalen–Johansen estimator of the CIF is equal to the incidence proportion 0.27 at the plateau – as expected – because the incidence proportion is a correct estimator for P(AE in [0, 2]) in the situation of complete data. It can be seen that the wrong Kaplan–Meier estimator clearly overestimates the probability of AE.

For illustration of the methods in the situation of right-censored data, we add artificial right-censoring such that patients will have different follow-up times. Right-censoring times were generated following a uniform distribution on [6, 24] and independently of the data. At the end of the follow-up after 2 years, now, 46 patients with AE and 42 deaths were observed. The incidence proportion calculated as the crude rate (1) is now 0.23, while the Aalen–Johansen estimator of the CIF (8) is equal to 0.26 at the plateau (Figure 3). This illustrates that the use of simple proportions to estimate P(AE in [0, t]) leads to underestimation in the presence of right-censoring. Figure 3 also displays the wrong Kaplan–Meier of AE, which again overestimates the probability of AE.

3. THE COMPETING RISKS MODEL

More formally, the competing risks model considers the time until some first event T, for example, AE or death without prior AE, as well as the type of event at time T denoted by E. E equals 1 if an AE is observed at time T, 2 if a death without prior AE occurred at T. Thus, competing risks data consist of the tuple (T, E), whose observation might be subject to right-censoring C. Thus, observed are $(T \wedge C, \mathbf{I}(T \leq C) \cdot E)$, where $\wedge$ denotes minimum and $\mathbf{I}(T \leq C)$ denotes the indicator function being 1 if $T \leq C$ and 0 otherwise.

3.1. Survival analysis is based on hazards

The modelling of competing risks data (and survival data in general) is built on hazards. For instance, the Kaplan–Meier estimator of the composite event AE or death, whatever comes first, is based on estimates of the all-event hazard $\alpha(t)$ (times the length of infinitesimally small time steps dt)

$$\widehat{\alpha(t)\mathrm{d}t} = \frac{\#\text{ AE or death at } t}{\#\text{ under observation before } t}, \qquad (11)$$

which decomposes into two so-called event-specific hazards $\alpha_{\mathrm{AE}}(t)\mathrm{d}t + \alpha_{\mathrm{Death}}(t)\mathrm{d}t$, estimated by

$$\frac{\#\text{ AE at } t}{\#\text{ under observation before } t} + \frac{\#\text{ Death at } t}{\#\text{ under observation before } t}. \qquad (12)$$

The decomposition in (12) motivates the Nelson–Aalen estimator of the cumulative hazard to experience an AE,

$$\int_0^t \widehat{\alpha_{\mathrm{AE}}(u)du} = \sum_u \frac{\#\text{ AE at } u}{\#\text{ under observation before } u}. \qquad (13)$$

It is evident from the numerator of (13) that only AE events are counted for computing the Nelson–Aalen estimator of the cumulative hazard to experience an AE. One way to do that in practice is to censor the competing event, that is, death without prior AE. That is, inference for the *hazard* of AE can be performed by, formally, censoring patients who died without prior AE, and vice versa. Note that this has nothing to do with an 'independent competing risks' assumption, which postulates the existence of hypothetical, independent latent times until AE and until death, respectively [9]. It is just a practical trick to compute the Nelson–Aalen estimator. However, we cannot 'censor away'

competing events when computing probabilities as (8) depend on all event-specific hazards through $P(T > u-)$. Thus, censoring by a competing risk is 'informative' in the sense that the competing event influences the probability estimate.

In the competing risks setting, all event-specific hazards should be analysed in order to obtain a full picture of the data [10], as the CIF depends on all event-specific hazards via $P(T > u-)$; see (8). Analogously, it is recommended to look at the CIFs of all competing risks (Figure 4).

3.2. Incidence rate

Often, the IR (also called incidence density) of AE

$$\frac{\#\text{AE}}{\text{Population time at risk}} \tag{14}$$

is used to summarise the risk of a patient to experience at least one AE. The IR of AE is a valid estimator of the hazard of AE assuming it to be time constant. If the event type E is observed for all individuals, an estimator of the cumulative AE probability can be derived if the hazard of AE estimated by the IR is combined with an estimator of the competing hazard [11]

$$\frac{\text{IR AE}}{\text{All-event IR}} = \frac{\#\text{AE}/\text{Population time at risk}}{(\#\text{AE} + \#\text{Deaths})/\text{Population time at risk}} = \frac{\#\text{AE}}{n}. \tag{15}$$

The last Equation 15 is only valid for infinite follow-up as only then $\#\text{AE} + \#\text{Deaths} = n$, but it illustrates that both the IR of AE and of death without prior AE should be looked at, because both enter the estimation of the AE probability. Thus, the same considerations as outlined previously for hazards in general apply to the IRs. An estimator of the cumulative AE probability that is valid with right-censored data is

$$\frac{\text{IR AE}}{\text{All-event IR}} \cdot (1 - \exp(t \cdot \text{All-event IR})), \tag{16}$$

which is the parametric analogue of formula (8) under the assumption of constant event-specific hazards. We note that although the IR is a valid estimator of the hazard of AE – assuming the latter to be constant – the constant hazard assumption has been criticised in medical applications [12,13].

For illustration in our simple example, Figure 5 displays the Nelson–Aalen estimates of the cumulative hazard of AE and death

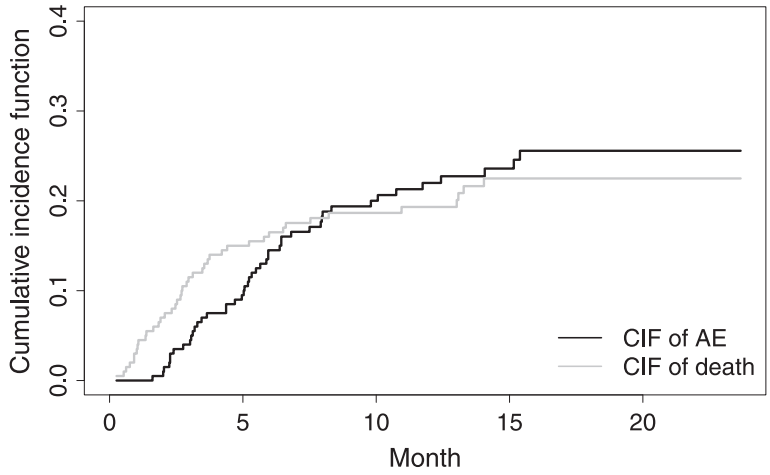

Figure 4. Probability of adverse event (AE) estimated in the right-censored data set by the Aalen–Johansen estimator of the cumulative incidence function (CIF) (black) along with the Aalen–Johansen estimates of the CIF of death (grey).

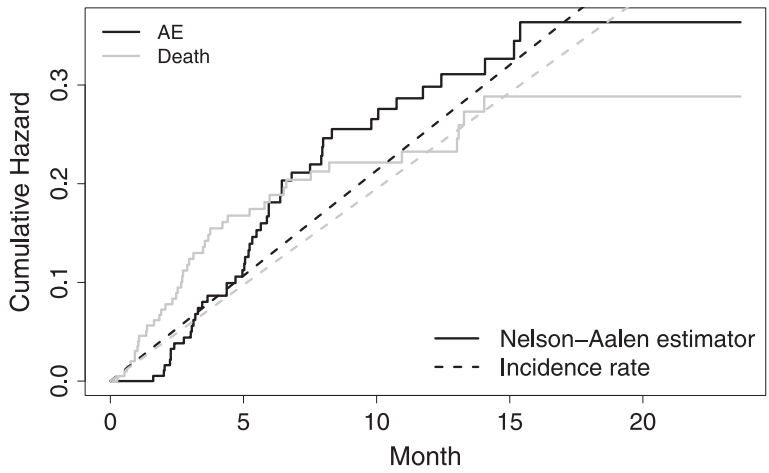

Figure 5. Cumulative hazard of adverse event (AE) (black) and death (grey) estimated by the Nelson–Aalen estimator (solid lines) and the corresponding incidence rates (dashed lines).

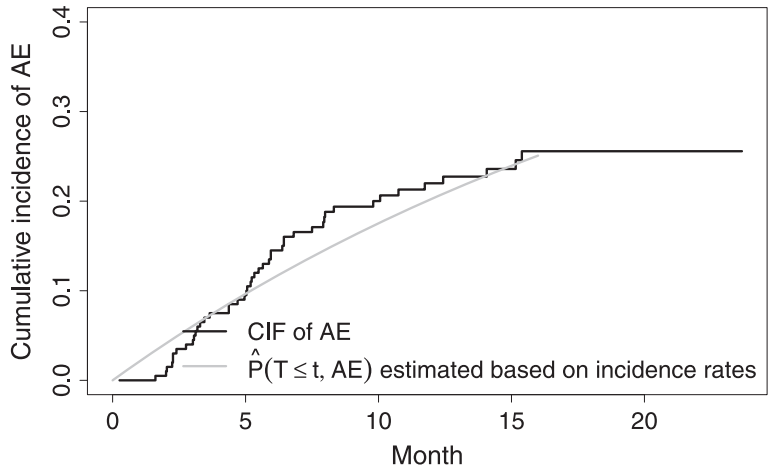

Figure 6. Probability of adverse event (AE) estimated by the Aalen–Johansen estimator in black and estimated based on the incidence rates (grey). CIF, cumulative incidence function.

along with the cumulative IRs of AE and death without prior AE computed from the right-censored data set. Recall that the IRs estimate the event-specific hazards if we assume them to be constant. Figure 5 suggests that this assumption might be violated for both the hazard of AE and of death without prior AE. This is further suggested by Figure 6 that shows the CIF of AE estimated by the Aalen–Johansen estimator (8) and via the IRs of AE and death without prior AE (16). Interestingly, both approaches are quite close to each other at the last event time (around month 15). The interpretation in the present data example is that although a model of constant event-specific hazards is not supported by the data, estimating the AE probability in this simple competing risks model leads to a meaningful estimate of the plateau of the CIF of AE. This is in contrast to both the crude rate (which does not account for censoring) and the inappropriate Kaplan–Meier estimator (which does not account for competing risks) from Figure 3. However, care should be taken of not extrapolating beyond the last event time.

One reviewer also pointed out that the incidence proportion, although in general inappropriate as a probability estimator, may be given an interpretation as the cumulative IR at the average follow-up time,

$$\frac{\#\text{AE}}{n} = \frac{\text{Population time at risk}}{n} \cdot \text{IR AE}.$$

In the presence of right-censoring, this relation allows for an interpretation of the incidence proportion with respect to average follow-up, under the assumption of a constant event-specific hazard of AE.

4. COMPARISON OF TREATMENT GROUPS

For comparison of treatment groups in a clinical trial with survival data, the most widely used regression model is the Cox proportional hazards model. The model assumes that the hazard for the composite event AE or death dependent on treatment groups $Z = 0$ and $Z = 1$ is of the form

$$\alpha(t|Z) = \alpha_0(t)\exp(\beta Z), \tag{17}$$

where $\alpha_0(t)$ is an unspecified, positive, baseline hazard and $\exp(\beta)$ is the hazard ratio of treatment group $Z = 1$ as compared with treatment group $Z = 0$.

A similar model can be fitted to the event-specific hazard of AE,

$$\alpha_{\text{AE}}(t|Z) = \alpha_{\text{AE};0}(t)\exp(\beta_{\text{AE}}Z), \tag{18}$$

where $\alpha_{\text{AE};0}(t)$ is an unspecified baseline event-specific hazard and $\exp(\beta_{\text{AE}})$ is an event-specific hazard ratio. In perfect analogy to the Nelson–Aalen estimator (13), fitting a Cox proportional hazards model for the hazard of AE can be performed *in practice* by censoring the death events. However, a complete picture is to be had only if an analogous model is also fitted to the hazard of death without prior AE

$$\alpha_{\text{Death}}(t|Z) = \alpha_{\text{Death};0}(t)\exp(\beta_{\text{Death}}Z). \tag{19}$$

Remember that the CIF of AE (8) depends on a highly nonlinear way on both event-specific hazards; thus, both event-specific hazards should be analysed to understand the shape of the event probabilities. This often raises some interpretational issues and has led to the development of regression models that are directly interpretable in terms of the CIF, the most prominent being the Fine and Gray model [14].

The idea of Fine and Gray is to consider an alternative hazard notion that reestablishes a one-to-one relationship between hazards and probabilities. The subdistribution hazard $\lambda(t)dt$ of AE is estimated by the following:

$$\frac{\#\text{ AE at } u}{\#\text{ patients not censored and w/o AE before } u}. \tag{20}$$

The Fine and Gray model then assumes proportionality of the subdistribution hazards

$$\lambda(t) = \lambda_0(t)\exp(\gamma Z), \tag{21}$$

with $\lambda_0(t)$ the baseline subdistribution hazard and $\exp(\gamma)$ the subdistribution hazard ratio. Note that in (20), patients who have experienced a death without prior AE are still considered to be at risk for the subdistribution hazard to experience an AE. The technical challenge in (20) is that dead patients should only be kept in the modified risk set until their potential censoring time. Fine and Gray [14] solved this using inverse probability of censoring

weighting and empirical process arguments. The supplementary material of [15] illustrates a simpler solution that applies to clinical trials with only administrative censoring. One advantage of the Fine and Gray approach is that there is a one-to-one relationship between the subdistribution hazard and the CIF

$$1 - \exp\left(-\int_0^t \lambda(u)du\right) = \mathrm{P}(T \leq t, \mathrm{AE}),$$

such that the subdistribution hazard ratios are directly interpretable in terms of the CIF.

Both modelling event-specific hazards and subdistribution hazard have their merits. On the one hand, the subdistribution hazard analysis allows for a direct probability interpretation, but the subdistribution hazard in itself and the associated subdistribution hazard ratio do not have a clear biological interpretation [16]. On the other hand, the interpretation of the event-specific hazards requires greater care in the sense that both event-specific hazards models should be interpreted side by side in order to understand the treatment effect on the CIF of interest. But that is only through the event-specific hazards that we understand why we see a certain effect on the event probabilities [9].

5. FURTHER METHODOLOGICAL ASPECTS

5.1. Competing risks and censoring

We have demonstrated that censoring an observed competing event yields valid inference for the event-specific hazard but such analyses are required for all event-specific hazards in order to arrive at probability statements: a Kaplan–Meier-type estimator for AE that censors observed death events is biased, but a Nelson–Aalen estimator that follows the same censoring rules is the correct non-parametric estimator of $\int_0^t \alpha_{\mathrm{AE}}(u)du$. In our experience, this has caused some confusion in practice.

One issue is that a typical assumption on censoring in a survival analysis is that those censored have the same momentary risk of an event as those still in the risk set (e.g. [17], p. 38). This assumption holds for the so-called *random censoring* that assumes a censoring time that is independent of the event time. In our data example, administrative censoring fulfils the random censoring assumption, but censoring by a competing risk does not: a patient who has been observed to die is *not* at risk of an AE anymore.

In fact, random censoring is an unnecessarily restrictive assumption, and the counting process literature [18,19] makes a more subtle *independent censoring assumption*. (Note that the terminology varies in the literature; we follow the counting process approach.) The aforementioned assumption on censoring has compared the momentary risk of those still under observation with those censored. The independent censoring assumption starts with the completely observed case as in Section 2.1 and assumes that the momentary risk of an event is not changed by *additional* knowledge on the censoring *process*. By censoring process, we mean individual on/off mechanisms that equal 1 as long as the individual is under observation and that equal 0 otherwise. Clearly, the additional knowledge of administrative censoring times fulfils the independent censoring assumption.

The point is that censoring by a competing risk also fulfils the independent censoring assumption. Consider again the completely observed case and additionally assume knowledge on the occurrence of competing events. This knowledge will not change momentary risks, because it was already available in the completely observed case. The bottom line is simple: observed competing events should be removed from the risk set, because, for example, a person who has died will not experience an AE anymore. This is what censoring a competing event achieves. Because probabilities depend on all hazards involved, it is required to perform such analyses for all event-specific hazards.

5.2. Target quantities in the presence of competing risks

We have argued that occurrence of AE is always subject to competing risks, which entails that all the event-specific hazards of the competing risks at hand should be considered. The latter is in particular true for probability statements which are, for example, relevant for *prediction* of patient outcomes. To this end, we have mainly focused on the probability to experience at least one AE (within $[0, \tau]$).

However, competing risks do not only present a technical challenge in that multiple hazards are present. There are also multiple ways to formalise group comparisons. For instance, consider the situation where a treatment is beneficial in that it reduces $\alpha_{\mathrm{Death}}(t)$ but does not affect the momentary AE risk by leaving $\alpha_{\mathrm{AE}}(t)$ unchanged. It is then fairly straightforward to show [9, Sec. 4.5] that treatment reduces the probability of death without prior AE for all times but, as a consequence of this, also increases the probability of AE for all times.

The practical implication for AE analyses is, in our current experience, limited: comparable event-specific hazards for AE will often lead to comparable CIFs for AE, although this is not guaranteed. We also believe that the question of how to formalise group comparisons in the presence of competing risks has not been fully resolved yet, although there are some efforts in this direction; see, for example, [20], and [21] using ideas for composite endpoints.

5.3. Beyond time-to-first adverse event

We have focused on time to first AE of a certain type. Key points were that they are always subject to competing risks and that survival methodology is required, although not the Kaplan–Meier estimator. We will now briefly outline that the hazard-based techniques generalise to recurrent AE events. Probability estimation, however, will be more involved. Let us consider one type of AE that may occur more than once in a single patient and as long as the patient is still alive. Formally, this is immediately accounted for in the simple IR formula

$$\frac{\#\mathrm{AE}}{\text{Population time at risk}},$$

if we re-interpret #AE as the number of observed AEs (in $[0, \tau]$), possibly recurrent, such that an individual may now contribute more than one AE event. Some authors have viewed this as a 'dubious concept' [22], but as [23] points out, it is an advantage that does, of course, rely on the constant hazard assumption. This assumption is now more restrictive than in the time-to-first event setting, because it is now also assumed that the AE intensity remains constant irrespective of the different number of prior AEs of the patients.

Computationally equally simple is generalising the Nelson–Aalen estimator,

$$\sum_u \frac{\#\text{ of observed AEs at } u}{\#\text{ under observation before } u},$$

where an individual now contributes to the numerator in the aforementioned display as many summands as it experiences AEs. While this estimator avoids the constant hazard assumption, it still assumes that the momentary AE intensity is the same for all patients irrespective of the individual AE history. Alternatively, we may view this Nelson–Aalen estimator as an estimator of the *partially conditional AE rate*. That is, the summand

$$\frac{\#\text{ of observed AEs at } u}{\#\text{ under observation before } u}$$

estimates an average over all individuals currently at risk, that is, alive and under observation. The estimate is partially conditional on the present at-risk status and averages in that it allows for the individual momentary risks to differ according to the individual courses of disease. Chapter 8 of [18] gives an accessible introduction to the details involved.

A popular regression model for recurrent events is the so-called Andersen–Gill model [19], an extension of the Cox model using the counting process paradigm. The extension is that a proportional hazards assumption is made for the intensity of a counting process that counts recurrent AEs; past AE occurrences may be included as time-dependent covariates.

Further summary functionals such as probability estimates may be derived but are typically complex. One possibility would be to model recurrent AEs in a multi-state model extension of Figure 1 and to use a general matrix-valued version of the Aalen–Johansen estimate [9,19]. See also [7,24] for further summary functionals in the context of recurrent events in the presence of a terminal event, as well [6] for an estimation of the cumulative duration of AEs in this same context.

5.4. Death after adverse event

The focus of this paper has been to highlight statistical issues in the analysis of time-to-first-AE events. Any analysis of subsequent events will rely on an adequate analysis of the time-to-first-event situation. In the previous subsection 5.3, we have discussed how hazard-based analyses extend to recurrent AEs. We now briefly consider the related issue that the occurrence of a first AE event may change the momentary risk of subsequent events, including both AEs and primary study outcomes such as overall survival.

To begin, note that the situation at hand was used as a *definition* of competing risks by [8] who 'define[d] a competing risk as an event whose occurrence either precludes the occurrence of another event under examination or fundamentally alters the probability of occurrence of this other event'. We have used the more agnostic definition of time-to-first-event T and type-of-first-event E, but there is no practical discrepancy. On the other hand, the definition in [8] immediately connects to the topic of the present subsection.

In subsection 5.2, we have discussed that the occurrence of a competing risk fulfils the independent censoring assumption. The consequence is that such an observed competing event may be coded as a censoring for the analyses of the other event-specific hazards. Note that these properties must be interpreted within the time-to-first-event setting. For instance, consider the situation of coding a first AE as a censoring for the analysis of the hazard of death *without prior AE*. This must not be interpreted as an analysis of the overall survival hazard or even for the hazard of death *after first AE*. Such analyses may be achieved by including AE occurrence as time-dependent covariates in a Cox regression model. Beyersmann *et al.* [9, Chapter 11] explain such analyses and their connection to multistate models. Multistate models may be used in this context to derive probability estimates [25].

This aspect is closely related to the momentary risk of recurrent AEs discussed in subsection 5.3. Recall that, for example, using the IR for recurrent AEs by simply counting the number of all observed AEs in the numerator does not only rely on a constant hazard assumption but also assumes that the momentary risk is not changed by previous AE events. On the other hand, the Andersen–Gill extension of the Cox model allows to model the impact of such previous events.

6. DISCUSSION

In general, simple incidence proportions should not be used for estimating the probability of AE in the time-to-event setting, unless vital and AE statuses are known for all patients. Indeed, AEs can occur at any point in time during the patient's time under observation, which is not taken into account by the calculation of crude rates. Moreover, in a time-to-event setting, a crude estimator that ignores censoring will be biased. This is precisely the reason why survival techniques like the Kaplan–Meier estimator or Cox regression are used for efficacy outcomes. The same rationale applies to the analysis of AEs. Furthermore, AEs are always subject to competing risks, that is, the occurrence of a certain AE may be precluded by the main outcome of interest or another fatal AE.

Guidelines on the statistical evaluation and the reporting of AEs give mixed recommendations. ICH E9 on the statistical analysis of AEs [1], ICH E3 on the statistical reporting of AEs [26], and also the Council for International Organizations of Medical Sciences working group on management of safety information from clinical trials [27] state that incidences should be presented relating the number of subjects experiencing an AE to the number of subjects at risk and that, depending on the situation, survival analysis or 'life table' methods should be considered for calculation. But none of these guidelines mentions that competing risks have to be taken into account and it remains unclear what kind of survival analyses techniques are recommended exactly.

In the competing risks setting, the CIF – estimated non-parametrically by the Aalen–Johansen estimator – is the correct quantity for expressing the probability of AE. Moreover, standard survival techniques *for hazards* that censor the competing event are still valid, but incomplete analyses, still valid because inference on the hazard of AE only counts AE events, but incomplete as the probability to experience an AE depends on all event-specific hazards. Thus, the competing event should also be subject to such an analysis. However, the Kaplan–Meier estimator that censors the competing event is not a valid estimator for the probability of AE because it implicitly assumes that all patients will eventually experience the AE. The IR of AE can be used to estimate the hazard of AE assuming it to be constant. As for the non-parametric or semi-parametric estimators, the IR of the competing event should also be looked at. In our data example, the constant event-specific hazards assumption was violated, but the CIF of AE estimated within this model and evaluated at the last observed AE time outperformed both the crude rate and the inappropriate Kaplan–Meier estimator.

Comparison of (treatment) groups with respect to AE can be performed via a proportional event-specific hazards model. It reflects the direct effect of treatment on the instantaneous risk of AE. For a complete understanding of the competing risks data, the competing event should also be subject to such an analysis.

In contrast, the proportional subdistribution hazards model for AE results in a direct comparison of (treatment) groups in terms of the CIF of AE. In a sense, the latter model provides a summary analysis, but this is only through the study of all event-specific hazards that one can understand the shape of the CIF for the event of interest [9]. For instance, analysing both event-specific hazards as well as the subdistribution hazard is what is advocated by guidelines on competing risks analysis [28].

We reiterate that censoring by a competing event is independent in the sense that it retains the form of the competing process intensity [19] but it is informative as it impacts probabilities. Thus, analysing the hazard of AE (with the Nelson–Aalen estimator or Cox models) by censoring the competing event is still valid, but this analysis must be performed in turn for the hazard of death without prior AE.

We finally note that the Nelson–Aalen estimator, IRs, and Cox proportional hazards models generalise to recurrent AEs; the basic idea being that the counting processes underpinning survival analysis techniques can count more than one event per patient [18, Ch. 8].

REFERENCES

[1] ICH Expert Working Group. ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials E9, 1995. Available at: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf (accessed 27.8.2015).

[2] O'Neill RT. Statistical analyses of adverse event data from clinical trials special emphasis on serious events. *Drug Information Journal* 1987; **21**(1):9–20.

[3] Gait JE, Smith S, Brown SL. Evaluation of safety data from controlled clinical trials: the clinical principles explained. *Drug Information Journal* 2000; **34**(1):273–287.

[4] Moore K, van der Laan M. Application of time-to-event methods in the assessment of safety in clinical trials. *Design and Analysis of Clinical Trials with Time-to-Event Endpoints*. Taylor & Francis, Chapman & Hall/CRC, Boca Raton, 2009; pp. 455–482.

[5] Nishikawa M, Tango T, Ogawa M. Non-parametric inference of adverse events under informative censoring. *Statistics in Medicine* 2006; **25**(23):3981–4003.

[6] Wang J, Quartey G. Nonparametric estimation for cumulative duration of adverse events. *Biometrical Journal* 2012; **54**(1):61–74.

[7] Gong Q, Tong B, Strasak A, Fang L. Analysis of safety data in clinical trials using a recurrent event approach. *Pharmaceutical Statistics* 2014; **13**(2):136–144.

[8] Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Statistics in Medicine* 1999; **18**(6):695–706.

[9] Beyersmann J, Allignol A, Schumacher M. *Competing Risks and Multistate Models with r*. Springer: New York, 2011.

[10] Beyersmann J, Scheike TH. *Classical Regression Models for Competing Risks*, Handbook of Survival Analysis. CRC Press: Boca Raton, 2013.

[11] Beyersmann J, Gastmeier P, Schumacher M. Incidence in ICU populations: how to measure and report it? *Intensive Care Medicine* 2014; **40**:871–876.

[12] Grambauer N, Schumacher M, Dettenkofer M, Beyersmann J. Incidence densities in a competing events analysis. *American Journal of Epidemiology* 2010; **172**(9):1077–1084.

[13] Kraemer HC. Events per person-time (incidence rate): a misleading statistic? *Statistics in Medicine* 2009; **28**(6):1028–1039.

[14] Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association* 1999; **94**(446):496–509.

[15] Schmoor C, Schumacher M, Finke J, Beyersmann J. Competing risks and multistate models. *Clinical Cancer Research* 2013; **12**:12–21.

[16] Andersen PK, Keiding N. Interpretability and importance of functionals in competing risks and multistate models. *Statistics in Medicine* 2012; **31**(11-12):1074–1088.

[17] Kleinbaum DG, Klein M. *Survival Analysis* (3rd edn). Springer: New York, 2012.

[18] Aalen OO, Borgan Ø, Gjessing HK. *Survival and Event History Analysis*, Statistics for Biology and Health. Springer-Verlag: New York, 2008.

[19] Andersen PK, Borgan Ø, Gill RD, Keiding N. *Statistical Models Based on Counting Processes*. Springer-Verlag: New-York, 1993.

[20] Tai BC, Wee J, Machin D. Analysis and design of randomised clinical trials involving competing risks endpoints. *Trials* 2011; **12**(1):127–137.

[21] Rauch G, Beyersmann J. Planning and evaluating clinical trials with composite time-to-first-event endpoints in a competing risk framework. *Statistics in Medicine* 2013; **32**(21):3595–3608.

[22] Windeler J, Lange S. Events per person year – a dubious concept. *British Medical Journal* 1995; **310**(6977):454–456.

[23] Vandenbroucke J. Continuing controversies over 'risks and rates' – more than a century after William Farr's 'on prognosis'. *Soz.—Präventivmed* 2003; **48**:216–218.

[24] Cook RJ, Lawless JF. *The Statistical Analysis of Recurrent Events*. Springer: New York, 2007.

[25] Zinßer T, Schumacher M. Modeling time-dependency in analysis of adverse events by the example of the CLEOPATRA study, 2015. submitted to *Pharmaceutical Statistics*.

[26] ICH Expert Working Group. ICH Harmonised Tripartite Guideline: Structure and Content of Clinical Study Reports E3, 1995. Available at: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E3/E3_Guideline.pdf (accessed 27.8.2015).

[27] Council for International Organizations of Medical Sciences – CIOMS Working Group VI. *Management of Safety Information from Clinical Trials*. CIOMS: Geneva, 2005.

[28] Latouche A, Allignol A, Beyersmann J, Labopin M, Fine JP. A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. *Journal of Clinical Epidemiology* 2013; **66**(6):648–653.