Competing Risks and Multistate Models

Claudia Schmoor¹, Martin Schumacher², Jürgen Finke³, and Jan Beyersmann²

Abstract

Complex clinical endpoints are present in studies in cancer. Especially in studies on hematopoietic stem-cell transplantation (HSCT), various risks exist after HSCT. Patients can experience acute and chronic graft versus host disease (GVHD) or need to undergo immunosuppressive therapy (IST), a relapse can occur, or patients can die after relapse or without former relapse (nonrelapse mortality, NRM). Sometimes, endpoints can be reasonably combined in a composite endpoint, as, for example, relapse and NRM are combined into disease-free survival (DFS). In this case, standard survival techniques, as Kaplan–Meier estimation of the DFS probability, can be applied.

Often, interest focuses on endpoints for which competing risks are present, as, for example, GVHD, with death without prior GVHD as competing risk. This results in a competing risks model, a special case of a multistate model. A more complex multistate model is required when the effects of events occurring in the course of the study on further disease process shall be investigated, as, for example, the effect of GVHD on relapse and NRM. Another endpoint of interest is time under IST. As patients usually experience multiple episodes of IST, thus switching back and forth between “IST” and “no IST” during follow-up, the multistate model used for analysis must be adapted for this event structure.

The aim of this nontechnical report is to explain use and interpretation of Cox-type regression models suitable for the different situations in a randomized trial on the effects of anti-T-cell globulin as GVHD prophylaxis. Clin Cancer Res; 19(1); 12–21. ©2012 AACR.

Introduction

The analysis of time-to-event endpoints (“survival analysis”) is a major application of advanced statistical methodology in medical research (1). In cancer studies, a variety of time-to-event endpoints are studied (2–5). Standard survival analyses consider composite endpoints, which every patient must experience at some point in time, although potentially after study closure. An example is progression-free survival, which is the time until progression or death without prior progression, whatever occurs first. The analysis of the single components of such a composite time-to-event endpoint, called competing risks or, synonymously, competing events, produces more specific results (6–8). This is the analysis of time to first event and type of first event. Subsequent events, e.g. death after progression, may be analyzed in a multistate framework (9); this approach also covers the analysis of the intermediate events as time-dependent covariates.

The multistate framework models events as transitions between states and includes competing risks as a special case. The occurrence of a competing risk is modelled as a transition out of an initial state, e.g. no progression, into a competing risk state, e.g. progression. The transition takes place at the time of the first event.

The aim of the present article is to explain use and interpretation of Cox-type regression models for competing risks and multistate models for more complex event structures. Our presentation focuses on concepts rather than mathematical technicalities. A recently published study on the use of an anti-T-cell globulin (ATG-F) to lower the incidence of graft-versus-host disease (GVHD; refs. 10, 11) after hematopoietic stem-cell transplantation (HSCT) will be used to illustrate the analysis step-by-step.

Clinical Cancer Research has recently published excellent research articles on competing risks regression models (12, 13). We complement this work in 2 ways. Firstly, we offer a more basic and nontechnical step-by-step introduction to competing risks as an integral part of a worked data example. Secondly, we move beyond competing risks and, hence, a first-event analysis by explaining how more complex event patterns can be addressed within a multistate framework. The transition to multistate models is achieved by first introducing competing risks from a multistate perspective. Multistate models can then be viewed as arising in a sequence of competing risks situations.
Data example

Between May 2003 and February 2007, 201 leukemia patients in early \( n = 107 \) or advanced disease status \( n = 94 \) entered a randomized, multicenter open-label, phase III trial to compare standard GVHD prophylaxis plus pretransplant ATG-Fresenius (ATG-F, \( n = 103 \)) versus standard GVHD prophylaxis alone (control, \( n = 98 \)). All patients received myeloablative conditioning before HSCT from matched unrelated donors. Stem-cell source was bone marrow \( n = 37 \) or peripheral blood \( n = 164 \). Procedures and results have been published in detail previously (10, 11).

The primary study aim was to show a reduction of severe acute GVHD (aGVHD) of grade III-IV, which is defined to occur only within the first 100 days post-HSCT, by adding ATG-F to standard GVHD prophylaxis. When analyzing the event aGVHD III-IV the competing event “death without prior aGVHD III-IV” has to be considered. It has to be ensured that ATG-F does not decrease the number of aGVHD III-IV events by increasing the number of competing events. A harmful effect of ATG-F on the competing event would not have been acceptable. Therefore, the primary endpoint was chosen as the composite endpoint “aGVHD grade III-IV or death without prior aGVHD III-IV” (Fig. 1, scenario I). As data are complete within the first 100 days, this endpoint is amenable both to logistic regression as in the primary efficacy analysis (10) and to a time-to-event analysis as in the present article. The time-to-event analysis censors event-free individuals on day 100. Secondary time-to-event endpoints were chronic GVHD (cGVHD), relapse, nonrelapse mortality (NRM), disease free survival, overall survival, and time being under immunosuppressive therapy (IST).

Figure 1. Multistate models in different scenarios. A, scenario I: effect of treatment on aGVHD III-IV. B, scenario II: effect of treatment on cGVHD. C, scenario III: effect of cGVHD on relapse and NRM. D, scenario IV: effect of treatment on IST.
As prespecified for the primary efficacy analysis, the covariates disease status (early vs. advanced) and stem-cell source (bone marrow vs. peripheral blood) were included in all regression models for adjustment of the treatment effect. For simplicity and as is common in the literature, we illustrate these results with plots of unadjusted event proportions. A more involved alternative would have been plots based on the regression model (10).

Scenario I: Effect of treatment on aGVHD III–IV

Analysis of composite endpoint "aGVHD III–IV or death" with the all-events hazard. Time-to-event data are typically only incompletely observed. Patients who have not experienced an event until study closure are (right) censored. This incomplete information compromises use of proportions:

\[
\frac{\text{number of observed events}}{\text{number of patients}}
\]

because the number of events has only partially been observed. The symbol # will be used in the following to indicate "number of". Survival analysis is therefore based on hazards.

The empirical all-events hazard for the composite endpoint "aGVHD III–IV or death" at time \(s\) is

\[
\frac{\text{#observed aGVHD III–IV or death events at time } s}{\text{#patients without an aGVHD III–IV or death event and not censored before time } s}
\]

This fraction can be computed in the presence of censoring, because it only relies on observed data. It describes the event probability in a small time interval under the condition that the event has not occurred before. The sum of these fractions over time gives the Nelson–Aalen estimator (14) of the cumulative hazard, which is not bounded by 1 and fractions over time gives the Nelson–Aalen estimator (14) of the cumulative hazard, which is not bounded by 1 and

\[
\text{transformation of } \frac{\text{#observed events}}{\text{#patients}}
\]

over all observed event times \(s \leq t\).

At day 100, the Kaplan–Meier estimate of the incidence of the composite endpoint "aGVHD III–IV or death" is 0.214 in the ATG-F group and 0.347 in the control group (Supplementary Fig. S1A). These values are identical to the empirical proportions 22/103 in the ATG-F group and 34/98 in the control group, because the data are complete within the first 100 days. Comparing the treatment groups with a standard Cox regression model resulted in a HR of ATG-F versus control of 0.66, 95% confidence interval (CI: 0.38–1.13, Table 1). The HR is less than 1, indicating a beneficial effect of ATG-F; however, the CI includes the value 1. Thus, the difference between ATG-F and control is not significant at the 5% level.

Analysis of competing events "aGVHD III–IV" and "death" with event-specific hazards. The hazard-based approach allows for more specific analyses of the competing events. Instead of the empirical all-events hazard at time \(s\) shown above, we now consider the empirical event-specific hazard for aGVHD III–IV,

\[
\frac{\text{#observed aGVHD III–IV events at time } s}{\text{#patients without an aGVHD III–IV or death event and not censored before time } s}
\]

and the empirical event-specific hazard for "death without prior aGVHD III–IV",

\[
\frac{\text{#observed deaths without prior aGVHD III–IV at time } s}{\text{#patients without an aGVHD III–IV or death event and not censored before time } s}
\]

The additive split of the all-events hazard into the event-specific hazards neither makes an assumption about dependence or correlation of the competing risks nor does it speculate about preventing 1 of the competing risks. The Cox regression model can again be used to relate an event-specific hazard to individual covariate information. The difference to the analysis of the all-events hazard is that there are as many event-specific hazards and analyses as there are competing risks. All of the event-specific hazards should be analyzed, because the cumulative event probability of a competing risk, the so-called cumulative incidence function, depends on all event-specific hazards.

Because the Kaplan–Meier curve \(S(t)\) estimates the all-events-free proportion at time \(t\), 1 minus the Kaplan–Meier curve estimates the all-events proportion up to time \(t\). Checking the increments of the Kaplan–Meier curve shows that the estimated all-events proportion up to time \(t\) equals the sum of

\[
S(t) = 
\frac{\text{#observed aGVHD III–IV or death events at time } s}{\text{#patients without an aGVHD III–IV or death event and not censored before time } s}
\]

over all observed event times \(s \leq t\), where \(S(s–)\) is the value of the Kaplan–Meier curve (which is a step function) just before time \(s\). Thus, the increments are the product of the probability of being all-events-free and the all-events hazard at time \(s\).

Replacing in this presentation the empirical all-events hazard with the empirical event-specific aGVHD III–IV hazard yields the Aalen–Johansen estimator (16) of the cumulative aGVHD III–IV hazard as sum of

\[
S(t) = 
\frac{\text{#observed aGVHD III–IV events at time } s}{\text{#patients without an aGVHD III–IV or death event and not censored before time } s}
\]

over all observed event times \(s \leq t\). The Aalen–Johansen estimator therefore generalizes the Kaplan–Meier estimator to multiple event types. However, 1 minus the Aalen–Johansen estimator does not have a Kaplan–Meier representation in terms of the event-specific aGVHD III–IV hazard, because \(S(s–)\) depends on the all-events hazard.
In case of no censoring, this formula for calculation of the cumulative incidence of aGVHD III-IV equals the simple proportion

\[
\frac{\text{#observed aGVHD III–IV events}}{\text{#patients}}.
\]

The Aalen–Johansen estimator of the cumulative incidence function for "death without prior aGVHD" is obtained in the same way as above by replacing the empirical all-events hazard with the empirical event-specific hazard for "death without prior aGVHD". The sum of both Aalen–Johansen estimators equals the all-events proportion (aGVHD III-IV or death).

Comparing treatment groups with the Cox regression model for the event-specific hazards resulted in a HR of ATG-F versus control of 0.48, 95% CI (0.24–0.96), with respect to the event aGVHD III-IV, and in a HR of ATG-F versus control of 1.17, 95% CI (0.47–2.92), with respect to the competing event death (Table 1).

This shows a significant beneficial effect of ATG-F versus control on the reduction of aGVHD III-IV, whereas there is essentially no difference between treatment groups with respect to "death without prior aGVHD III-IV". This result was aimed for when planning the study, but for showing that ATG-F decreases the risk of aGVHD III-IV without increasing the competing event "death without prior aGVHD III–IV", the composite endpoint was chosen as primary endpoint. No difference was seen with regard to the competing event, but by adding this component to the

### Table 1. Overview of results of different scenarios

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Endpoint</th>
<th>HR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I: Effect of treatment on aGVHD III–IV (ATG-F vs. control)</strong></td>
<td>Composite endpoint</td>
<td>0.66 (0.38–1.13)</td>
</tr>
<tr>
<td></td>
<td>aGVHD III–IV or death</td>
<td>0.48 (0.24–0.96)</td>
</tr>
<tr>
<td></td>
<td>Event-specific hazards</td>
<td>1.17 (0.47–2.92)</td>
</tr>
<tr>
<td></td>
<td>aGVHD III–IV</td>
<td>0.47 (0.23–0.94)</td>
</tr>
<tr>
<td></td>
<td>Death without prior aGVHD III–IV</td>
<td>1.35 (0.54–3.38)</td>
</tr>
<tr>
<td><strong>II: Effect of treatment on cGVHD (ATG-F vs. control)</strong></td>
<td>Event-specific hazards</td>
<td>0.33 (0.20–0.54)</td>
</tr>
<tr>
<td></td>
<td>cGVHD</td>
<td>0.59 (0.33–1.06)</td>
</tr>
<tr>
<td></td>
<td>Death without prior cGVHD</td>
<td>0.38 (0.23–0.61)</td>
</tr>
<tr>
<td></td>
<td>Subdistribution hazards</td>
<td>1.17 (0.66–2.07)</td>
</tr>
<tr>
<td></td>
<td>cGVHD</td>
<td>0.38 (0.23–0.61)</td>
</tr>
<tr>
<td></td>
<td>Death without prior cGVHD</td>
<td>1.17 (0.66–2.07)</td>
</tr>
<tr>
<td><strong>III: Effect of cGVHD on relapse and NRM (cGVHD vs. no cGVHD)</strong></td>
<td>Event-specific hazards</td>
<td>0.50 (0.26–0.96)</td>
</tr>
<tr>
<td></td>
<td>Relapse</td>
<td>1.18 (0.52–2.65)</td>
</tr>
<tr>
<td><strong>IV: Effect of treatment on IST (ATG-F vs. control)</strong></td>
<td>Event-specific transition hazards</td>
<td>0.31 (0.18–0.55)</td>
</tr>
<tr>
<td></td>
<td>Alive, no IST → alive under IST</td>
<td>2.02 (1.41–2.91)</td>
</tr>
<tr>
<td></td>
<td>Alive, under IST → alive, no IST</td>
<td>0.84 (0.56–1.25)</td>
</tr>
<tr>
<td></td>
<td>Composite endpoint</td>
<td>0.84 (0.56–1.25)</td>
</tr>
<tr>
<td></td>
<td>Overall survival</td>
<td>0.84 (0.56–1.25)</td>
</tr>
</tbody>
</table>
primary endpoint, the difference between treatment groups was attenuated.

**Summary of the event-specific hazards analyses of “aGVHD III-IV” and “death” using subdistribution hazards.** As an advantage, the previous competing risks analysis provides more specific results than the analysis of the composite endpoint. As a disadvantage, there are now 2 results instead of 1. Often, interest focuses on a specific competing event such as aGVHD III-IV, and researchers may wish to summarize the analyses of the event-specific hazards in terms of the resulting effect on the cumulative incidence function.

To this end, Fine and Gray (17) suggested a Cox-type model for the so-called subdistribution hazard. This method has become a popular technique (18, 19), and was, for example, requested in the review process of the primary efficacy analysis of the study (10).

The empirical subdistribution hazard for the competing risk of interest aGVHD III–IV is

\[
\text{HR}_{\text{aGVHD III–IV}} = \frac{S(t|\text{aGVHD III–IV})}{S(t|\text{not aGVHD III–IV})} \cdot \frac{S(t|\text{death})}{S(t|\text{death, no aGVHD III–IV})}
\]

The computational subtlety here is that the denominator includes patients with an observed death event as “patients without an aGVHD III-IV event and not censored” until their censoring time. The mechanism is illustrated in the Supplementary Table S1. The inclusion of observed deaths in the denominator compromises interpreting the subdistribution hazard as a momentary event rate or hazard (20). It merely allows for directly translating results onto the cumulative incidence scale: one can show that 1 minus the Kaplan–Meier formula where as hazard the subdistribution hazard is inserted equals the Aalen–Johansen estimator of the cumulative incidence function. This also shows that the cumulative incidence function is a function of 2 event-specific hazards or of 1 subdistribution hazard.

regarding computational matters of the Fine and Gray model, one can note that, whenever censoring is due to administrative reasons only, after rearrangement of the data in a way that patients who did not experience the event of interest enter the calculation with their real censoring time (in case the competing event did not occur) or with their expected censoring time (in case the competing event did occur), standard Cox regression analysis software can be used for fitting the Fine and Gray model. Comparing the treatment groups with respect to aGVHD III-IV with the Cox regression model for the subdistribution hazards resulted in a HR of ATG-F versus control of 0.47, 95% CI (0.23–0.61). The results for the event-specific aGVHD III-IV hazard and for the subdistribution hazard of the cumulative aGVHD III-IV incidence are virtually identical in this study. This is because HRs obtained from an analysis of an event-specific hazard and from an analysis of the corresponding subdistribution hazard have been found to be similar in general, if censoring is heavy, or if treatment shows no effect on the competing hazard (22). Both criteria are met here, the first one because the present time-to-event analysis imposes censoring on day 100 on all individuals, who were event-free up to that time. The analyses do not necessarily yield similar results, which will now be illustrated for the endpoint cGVHD.

**Scenario II: Effect of treatment on cGVHD**

We now consider the event cGVHD that by definition occurs after day 100 post HSCT. So, the analysis has to be restricted to patients at risk for cGVHD (90 and 80 patients of whom 27 and 49 patients experienced cGVHD in the ATG-F and control group, respectively). Again, we have to consider the competing event “death without prior cGVHD” (Fig. 1, scenario II) which occurred in 29 patients in the ATG-F group and in 21 patients in the control group. ATG-F has a strong effect on the incidence of cGVHD and no effect on the incidence of the event “death without prior cGVHD”, as can be seen from the Aalen–Johansen estimates of the cumulative incidence functions in Fig. 2A and B.

Comparing the cumulative incidence functions of the treatment groups via the subdistribution hazards as described in the last chapter resembles this result. The Cox regression model for the subdistribution hazards yields a HR of 0.38, 95% CI (0.23–0.61), with respect to the event cGVHD, and of 1.17, 95% CI (0.66–2.07), with respect to the event “death without prior cGVHD”.

We now look at the comparison of the treatment groups with respect to the event-specific hazards. With respect to the event cGVHD, the result is similar to the result of the subdistribution hazards analysis: the HR of ATG-F versus control is 0.33, 95% CI (0.20–0.54). With respect to the competing event “death without prior cGVHD”, the result is different from the result of the subdistribution hazards analysis: The HR of ATG-F versus control is 0.59, 95% CI (0.33–1.06), meaning a nonsignificant but reduced hazard of ATG-F with respect to “death without prior cGVHD” as compared with control (Table 1).

The different results of both analysis approaches with respect to the event “death without prior cGVHD” may be explained as follows. The main difference between both approaches is that the number of observed events is related to different risk sets. In the subdistribution hazards analysis, the number of “death without prior cGVHD” events is related to the number of all patients over time (similar in ATG-F and control). In the event-specific hazards analysis, the number of “death without prior cGVHD” events is related to the number of patients without cGVHD over time. Because ATG-F reduces the number of cGVHD events, the risk set in the ATG-F group is larger than the risk set in the control group. As a consequence, the event-specific hazard for “death without prior cGVHD” is lower in the ATG-F group as compared with control.

**Scenario III: Effect of cGVHD on relapse and NRM**

So far, the endpoint GVHD was analyzed as 1 of 2 competing events without investigating what happens to patients after they have experienced GVHD. If interest
focuses on the effect of GVHD on further course of disease, more complex multistate models are required. Multistate models can be viewed as a sequence of competing risks situations. It is known that cGVHD is associated with antileukemic effects, and, thus, its occurrence may have a protective effect on the development of relapse. On the other hand, it may also lead to an increased risk of NRM (23).

For an analysis of these effects, a multistate model as depicted in Fig. 1, scenario III is necessary. Relapse and NRM are again competing events that have to be considered in the analysis. In the analysis of the effect of cGVHD, it additionally has to be considered that the patients develop cGVHD during the course of the study, and that it is not known at the start of the study if a patient will develop cGVHD. It is quite common in the literature about the analysis of such time-dependent events as cGVHD that they are erroneously treated as if known in the beginning, that is, a simple 2 group comparison is conducted comparing patients who will develop the event over time with patients who will not develop the event over time (see literature search by van Walraven and colleagues, ref. 24). This violates principle (1) ‘Do not condition on the future’ established by Andersen and Keiding (20) for the analysis of survival data, and is known to produce a bias in the estimation of the effect (25).

Thus, the analysis of the effect of cGVHD on relapse and NRM has to be done using appropriate statistical methods considering the time-dependent nature of cGVHD as the Cox regression model for the event-specific hazards for relapse and for NRM including cGVHD as time-dependent covariate (see ref. 9, chapter 11). The central point in the modeling of cGVHD as time-dependent covariate is that patients start in the group of "no cGVHD" patients and enter the group of "cGVHD" patients not until they experience the cGVHD event during course of the study. For example, for the
endpoint relapse, this amounts to comparing the transition hazard

\[
\frac{\text{patients without relapse or NRM and not censored but without cGVHD before time} \times \text{observed relapses after prior cGVHD at time}}{	ext{patients without relapse or NRM and not censored but with cGVHD before time}}
\]

with the transition hazard

\[
\frac{\text{patients without relapse or NRM and not censored but without cGVHD before time} \times \text{observed relapses without prior cGVHD at time}}{	ext{patients without relapse or NRM and not censored but with cGVHD before time}}
\]

In other words, in Fig. 1, scenario III, the transition hazard working along the arrow from "cGVHD" to "relapse" is compared with the transition hazard working along the arrow from "HSCT" to "relapse". The same principle applies for the endpoint NRM.

In the ATG-F trial, 63 patients experienced a relapse and 52 patients died without former relapse (NRM). The analysis resulted in a HR of cGVHD versus no cGVHD of 0.50, 95% CI (0.26–0.96), with respect to the event relapse, and in a HR of cGVHD versus no cGVHD of 1.18, 95% CI (0.52–2.63), with respect to the event NRM (Table 1). So, also in our trial, the occurrence of cGVHD shows a protective effect on the development of relapse. A graphical illustration of the effects of time-dependent covariates can best be done by showing Nelson–Aalen estimates of the cumulative hazard functions, as can be found in the original publication of the results of our trial (11).

**Scenario IV: Effect of treatment on time under immunosuppressive therapy**

Another endpoint of interest in the ATG-F trial was the impact of ATG-F on time under IST (11). This is regarded as an indicator for GVHD burden. As patients usually experience multiple episodes of IST during follow-up, multistate models for this event structure are required. The statistical challenge is that patients can switch back and forth between "IST" and "no IST"; in addition, they can also die or be censored (Fig. 1, scenario IV).

We can now define the (empirical) event-specific transition hazard for, for example, switching from "no IST" to "IST,"

\[
\frac{\text{patients alive without IST and not censored before time} \times \text{observed changes to IST at time}}{	ext{patients alive without IST and not censored before time}}
\]

The Cox regression model can again be used to investigate the impact of ATG-F. An individual patient contributes as many data lines to such an analysis as there are observed changes from the status "no IST". A single data line covers 1 status episode, containing the start date and the end date of the current episode of "no IST". This implies the assumption that the transition times are observed exactly. Only if the individual changes to "IST" will the change be counted as an event. The 2 other possible changes are censoring and death.

In a practical data analysis, there are additional issues, which do not come up in a standard competing risks analysis. One consideration pertains the possible inclusion of current duration of IST status or number of previous switches in the Cox model. In the simpler scenario III, one might consider including "time of cGVHD" in the Cox model. Another issue is robust variance estimation (26). This is sometimes recommended because one individual may contribute several data lines to the analysis. These issues are beyond the scope and technical level of the present article; readers are referred to Andersen and Keiding (14). In particular, their data example illustrates the situation where the time spent in an intermediate state is prognostic. These authors also explain probability estimation for such more complex event structures. The idea is to extend the Kaplan–Meier estimate to a matrix-valued Aalen–Johansen estimator, where the empirical hazards for all event types enter.

In our study, the HR for receiving IST was 0.31, 95% CI (0.18–0.55), and that for being able to stop IST was 2.02, 95% CI (1.41–2.91), ATG-F versus control, respectively (Table 1). The interpretation is that, in terms of the event-specific transition hazards, ATG-F strongly promotes stopping IST, and similarly it strongly prevents receiving IST again. Figure 3A and B shows the probability of survival free of IST, and that of survival and still under IST. All patients start in status "IST", as, according to the study protocol, they have to receive cyclosporine A until 100 days posttransplantation. Sixty-one patients switched directly from "IST" to "death" or were censored. One hundred and forty patients switched to "no IST" of whom 57 patients switched back to "IST" at least once. The latter fact implies that the curves in Fig. 3A and B are not monotone. Note that the fact that the probability of survival under IST in Fig. 3B is lower in the ATG-F group does not imply worse prognosis of patients requiring IST despite ATG-F. This reflects the fact that patients in the ATG-F group predominantly live free of IST whereas patients in the control group predominantly live under IST. In fact, this curve would drop down to 0 in the ATG-F group, if every patient in this group would move back to "IST" at least once. The same principle applies with the transition hazard working along the arrow from "HSCT" to "relapse". The same principle applies with the transition hazard working along the arrow from "IST" to "death" or were censored.

The sum of the probability of survival free of IST and of the probability of survival under IST equals the probability of overall survival. The overall survival probabilities in Fig. 3C were estimated by the Kaplan–Meier estimator, as overall survival is a composite endpoint without competing risks. Comparing treatment groups with a standard Cox model for the all-events hazard leads to a HR of 0.84, 95% CI (0.56–1.25; Table 1).

**Discussion**

A variety of time-to-event endpoints are studied in cancer research. In our data example, further composite endpoints were disease-free survival and overall survival. Also overall survival can be regarded as composite, because one can distinguish between cancer-related mortality and noncancer mortality (27, 28). If the composite time-to-event endpoint is all-encompassing in the sense that every patient will at some point in time experience it (although potentially after study closure), standard survival analyses using Kaplan–Meier curves and Cox regression analyses for the all-events hazard are adequate.
If a more specific analysis is desired, competing risks have to be investigated. The converse is also true: If a time-to-event endpoint is analyzed that is not all-encompassing, the competing risks have to be considered, too. An example is the endpoint “progression” with competing risk “death without prior progression”. It has recently been shown that the analysis of “progression” does frequently not account for competing risks, which compromises interpretation of the results (2).

In a competing risks analysis, the Kaplan–Meier curve should not be used to estimate the cumulative incidence functions, because it inevitably overestimates the true event probability.
probabilities. The Aalen–Johansen estimator should be used instead. Using the Cox model (or the log-rank test) for 1 event-specific hazard alone is correct, but yields an incomplete analysis. The other competing risk must be considered, too. In practice, the other competing risk is sometimes informally analyzed, reporting crude proportions (# observed competing events)/(# patients). This is only of limited use as a consequence of censored data. One should analyze all event-specific hazards, using the Cox model, and should also look at the Aalen–Johansen estimator of all cumulative incidence functions. All these analyses should be reported for a comprehensive understanding of the results.

The analysis of the event-specific hazards investigates the etiology of an event, meaning the direct association of treatment with the instantaneous hazards of the event of interest and the competing event, respectively. If, in addition a formal summary analysis in terms of the event incidence is desired, a Cox-type analysis of the corresponding subdistribution hazard may be reported. This investigates the association of treatment with the probability of experiencing the interesting event, which is directly influenced by the association of treatment with the instantaneous hazard of the competing event.

The analysis of the event-specific hazards as presented in this article only relies on the fact that their sum equals the usual all-events hazard. In the competing risks literature (29), risk-specific latent times are sometimes associated with the single event-specific hazards, or it is speculated what will happen, if 1 competing risk is eliminated. These considerations require an "independent" or "dependent" competing risks concept. But this does not pertain to the analyses of the present article.

When interest focuses on the course of disease after an event has occurred or on endpoints occurring in multiple episodes over time, as IST in our study, multistate models for more complex event structures are required. The probabilities of survival under IST and of survival free of IST are estimated with the Aalen–Johansen estimator. Patient groups are compared by modeling the transition hazards between the states alive and free of and alive under IST using Cox regression. Methods still under development include direct regression models and the calculation of simultaneous confidence bands for the transition probabilities.

Study planning remains a challenge in competing risks. Latouche and Porcher (30) give a useful account of the available approaches; Schulgen and colleagues (31) give a practical illustration in terms of a trial in diabetes patients. Typically, sample size calculations rely on focusing on 1 competing risk. As argued earlier, this was not acceptable for the primary efficacy analysis in our study. As a consequence and to ensure that a decreased number of aGVHD III-IV events would not be due to an increased number of untimely deaths, the primary efficacy analysis had been planned for a composite endpoint. Simulation techniques applied to empirical data could help to better understand the impact and consequences of competing risks and other, more complex endpoint structures (32). Future research is needed to further develop planning of and to investigate feasibility of studies that aim for a competing risks analysis of the competing risk of interest while ensuring noninferiority for the other competing risk.

Disclosure of Potential Conflicts of Interest

J. Finke has an honorarium from speakers bureau of Fresenius. No potential conflicts of interest were disclosed by the other authors.

Authors’ Contributions

Conception and design: C. Schmoor, M. Schumacher
Development of methodology: C. Schmoor, M. Schumacher, J. Beyersmann
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): J. Finke
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): C. Schmoor, M. Schumacher, J. Finke, J. Beyersmann
Writing, review, and/or revision of the manuscript: C. Schmoor, M. Schumacher, J. Finke, J. Beyersmann
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): C. Schmoor

Acknowledgments

The authors thank the referees for valuable suggestions that improved this article.

Grant Support

The ATG-F trial was funded by Fresenius Biotech GmbH, Germany. J. Beyersmann was supported by Grant BI 4500/1-1 of the German Research Foundation (DFG).

Received May 25, 2012; revised October 4, 2012; accepted October 25, 2012; published OnlineFirst November 21, 2012.

References

24. van Walraven C, Davis D, Forster A, Wells G. Time-dependent hazard was common in survival analyses published in leading clinical journals. J Clin Epidemiol 2004;57:672–82.