Statistics
in Medicine

Special Issue Paper

Received 25 November 2010, Accepted 4 August 2011 Published online 14 November 2011 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.4385

Interpretability and importance of functionals in competing risks and multistate models

Per Kragh Andersen*† and Niels Keiding

The basic parameters in both survival analysis and more general multistate models, including the competing risks model and the illness–death model, are the transition hazards. It is often necessary to supplement the analysis of such models with other model parameters, which are all functionals of the transition hazards. Unfortunately, not all such functionals are equally meaningful in practical contexts, even though they may be mathematically well defined. We have found it useful to check whether the functionals satisfy three simple principles, which may be used as criteria for practical interpretability. Copyright © 2011 John Wiley & Sons, Ltd.

Keywords: competing risks; illness–death model; survival analysis

1. Introduction

The theory of competing risks has a long history with somewhat independent developments in mathematical theory as motivated by applications to actuarial science, biostatistics and epidemiology, and reliability theory, and this diversity has not always benefited each of the applications. In particular, mechanistic interpretations from reliability are not always useful for epidemiological applications, and the important distinction between competing risks on one side and disability models (actuarial)/illness–death models (epidemiological) on the other is sometimes forgotten. In this paper, our aim is to give priority to the biostatistical and epidemiological applications and to critically evaluate a number of old and new ideas in the theory and practice of competing risks and illness–death models in our area from the three following principles that we have found helpful:

(1) Do not condition on the future;
(2) Do not regard individuals at risk after they have died; and
(3) Stick to this world.

The general exposition will be based on elementary multistate Markov and semi-Markov models, which will be recapitulated in Section 3 after a brief survey of basic survival analysis in Section 2. The basic competing risks model is a $(k+1)$ state model with transitions only possible from state 0 to each of the k-absorbing states; we only need $k = 2$ to make our principal points. This model is to be contrasted to the three-state illness–death model with transitions possible $0 \to 1 \to 2$ and $0 \to 2$. For simple survival analysis, it is by now general knowledge that a distribution can be equivalently characterized through its hazard rate, probability density, or survival function, and the concept and interpretation of independent right censoring are also well understood: among those who are still alive, additional information of being uncensored should provide no further insight into the future failure risk. Our focus, which we will present in Section 4, is on how meaningful and how identifiable various functionals and parameterizations remain when generalized from the simple survival situation to the aforementioned two more general models. Topics include latent survival times and their independence properties, interpretation of

Department of Biostatistics, University of Copenhagen, Copenhagen, Denmark
**Correspondence to: Per Kragh Andersen, Department of Biostatistics, University of Copenhagen, Copenhagen, Denmark.*
†*E-mail: p.k.andersen@biostat.ku.dk*

subdistribution hazards, conditional probability of being ill given alive, pattern–mixture representations, and age at disease occurrence in a world where one cannot die.

2. Survival analysis

Survival analysis is about drawing inference about a positive random variable T denoted *survival time*. Much focus in survival analysis is on the various obstacles to observing T when data are collected in real time so that some values are incompletely observed. However, this should not detract attention from the target of the effort: the distribution of a well-defined, positive, and finite survival time T if there were no obstacles to observation. The generic example is time T to death from any cause, either considered alone or in combination with some relevant event in the life history of the individual. An example of such an event is disease relapse in which case T is the relapse-free survival time.

The most common form of incomplete observation is *right censoring*, where the value of T is only known if $T \leqslant C$, the time of right censoring. If $T > C$, then this is the only available information. In addition, T is sometimes observed with *delayed entry*, that is, conditionally on $T > L$, the time of left truncation. If $T \leqslant L$, then nothing is observed.

For *valid* inference on the distribution of T, (L, C) must, in some suitable sense, be *independent* of T. Our preferred definition [1, chapter III] amounts to stating that given that we know that a person is alive at time t (i.e., $T > t$), the further information that the person is uncensored may not change this person's instantaneous probability of dying. It is important to notice, however, that even though incomplete observation of T is most often inevitable, this assumption of independent censoring may rarely be checked on the basis of the data that are typically available. We return to that point in Section 4.1.

The distribution of T may be characterized by its *survival function* $S(t)$ given by

$$S(t) = P(T > t) = 1 - P(T \leqslant t) = 1 - F(t) \tag{1}$$

where $F(t)$ is the *distribution function*. When T is a proper random variable,

$$\lim_{t\to\infty} F(t) = 1.$$

For an absolutely continuous distribution, the *hazard rate function*, $h(t)$ is defined as

$$h(t) = -\frac{d\log(1 - F(t))}{\mathrm{d}t}. \tag{2}$$

It follows that a useful interpretation is that, for a small $\mathrm{d}t > 0$, the hazard is approximately the conditional probability per time unit of failure before time $t + \mathrm{d}t$ given survival till t:

$$h(t) \approx P(T \leqslant t + \mathrm{d}t \mid T > t)/\mathrm{d}t. \tag{3}$$

From Equation (2), it follows that

$$F(t) = 1 - \exp(-\int_0^t h(u)\mathrm{d}u), \tag{4}$$

that is, there is a one-to-one correspondence between the *rate* $h(\cdot)$ and the *risk* $F(\cdot)$. This has the consequence that a statistical model (e.g., a regression model) specified for the rate immediately implies a model for the risk. This is useful because hazard-based models are frequent in survival analysis: $\log(h(t))$ is unbounded and well suited for a description via a linear predictor. Thus, the Cox regression model [2] is

$$\log(h(t \mid Z)) = \log(h_0(t)) + \beta^{\mathsf{T}} Z \tag{5}$$

where Z is a vector of explanatory variables and $h_0(t)$, the *baseline hazard*, is the hazard rate function when $Z = 0$. In Equation (5), the interpretation of the vector, β of regression coefficients, is that for the kth component, $\exp(\beta_k)$ is the *hazard ratio* associated with a one-unit change in Z_k. From Equations (4) and (5), it then follows that

$$\log(-\log(1 - F(t \mid Z))) = \log(H_0(t)) + \beta^{\mathsf{T}} Z$$

with $H_0(t) = \int_0^t h_0(u)\mathrm{d}u$ the cumulative baseline hazard.

2.1. Inference

On the basis of independently right-censored data $(\widetilde{T}_i, D_i; i = 1, \ldots, n)$ with $\widetilde{T}_i = T_i \wedge C_i$ and $D_i = I(T_i \leqslant C_i)$ and where the complete but only partially observed $T_1, \ldots, T_n$ are assumed to be independent and identically distributed, the common survival function may be estimated using the well-known Kaplan–Meier estimator

$$\widehat{S}(t) = \prod_{\widetilde{T}_i \leqslant t} \left(1 - \frac{D_i}{Y_0(\widetilde{T}_i)}\right) \tag{6}$$

with $Y_0(t) = \sum_i I(\widetilde{T}_i \geqslant t)$ being the number *at risk* at time t.

In general, the likelihood function based on observations $(\widetilde{T}_i, D_i, Z_i; i = 1, \ldots, n)$ can be expressed in terms of the hazard function

$$L = \prod_i (h(\widetilde{T}_i \mid Z_i))^{D_i} S(\widetilde{T}_i \mid Z_i).$$

This fact provides another motivation for modeling survival data using the hazard function. However, although nonparametric estimation of the hazard function requires some sort of smoothing, the *cumulative hazard*

$$H(t) = \int_0^t h(u)\mathrm{d}u$$

may, in an independent and identically distributed situation, be estimated by the *Nelson–Aalen* estimator

$$\widehat{H}(t) = \sum_{\widetilde{T}_i \leqslant t} \frac{D_i}{Y_0(\widetilde{T}_i)}. \tag{7}$$

In contrast to the hazard function, the cumulative hazard does not have a simple probabilistic interpretation. The following interpretation does, however, apply: Put a single subject on test at time 0 and follow it over time; when this first subject fails, say at time $T^{(1)}$, replace it by a similar subject that has survived until that time and follow the second subject over time; when the second subject fails, say at time $T^{(2)} > T^{(1)}$, replace it by a third one that has survived until time $T^{(2)}$ and so on. The cumulative hazard $H(t)$ is then the expected number of 'renewals' in that experiment by time t. Despite this indirect interpretation, estimation of $H(t)$ may still be useful because, on a plot of $\widehat{H}(t)$ versus t, the local slope will approximate the hazard function, which in particular implies that the cumulative hazard is often useful for model checking purposes.

3. Multistate models

It is seen from Equation (3) that the interpretation of the hazard function involves *conditioning on the past*, which is a fundamental concept in the study of survival data and invites basing the theory on stochastic processes [1]. Another example of such a conditioning is the conditional survival function $P(T > t \mid T > s)$ given survival till some time s with $t > s > 0$. This function is particulary useful for left-truncated data, where the unconditional survival function $S(t)$ may be ill determined because of little or no information for small values of t.

3.1. The two-state model for survival data

In the stochastic process approach, survival data are described via the two-state model depicted in Figure 1. In this model, $S(t)$ and $F(t)$ are the *state occupation probabilities* for states 'Alive' $X(t) = 0$ and 'Dead' $X(t) = 1$, respectively, that is,

$$S(t) = P(X(t) = 0), \quad F(t) = P(X(t) = 1),$$

whereas the hazard function is the $0 \to 1$ *transition intensity* with the interpretation (cf. Equation (3))

$$h(t)\mathrm{d}t \approx P(X(t + \mathrm{d}t) = 1 \mid X(t) = 0)$$

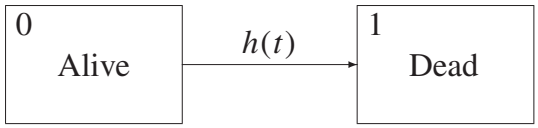

Figure 1. The two-state stochastic process $X(t) = 0$ or 1 for survival data.

for small $dt > 0$. State 0 is *transient*

$$\lim_{t\to\infty} P(X(t) = 0) = 0$$

whereas state 1 is *absorbing*. This is a consequence of the random variable

$$T = \inf_{t>0}(X(t) \neq 0),$$

the sojourn time spent in state 0, being proper. Because every individual begins in state 0 ($S(0) = 1$), $S(t)$ and $F(t)$ are also the *transition probabilities*

$$S(t) = P_{00}(0,t) = P(X(t) = 0 \mid X(0) = 0),$$

$$F(t) = P_{01}(0,t) = P(X(t) = 1 \mid X(0) = 0)$$

whereas the conditional survival function is the transition probability

$$P_{00}(s,t) = P(X(t) = 0 \mid X(s) = 0).$$

A description of survival data as a stochastic process may, at a first glance, seem more complicated than necessary. However, the benefits of this approach become apparent in studies where more than a single event (death) is of interest. Such studies could deal with death from a number of different causes, or they could deal with both disease occurrence and death with or without the disease. In such studies, the occurrence of an event may be modeled as a transition into a state in a multistate model.

3.2. *Competing risks*

The situation with death from a number of separate causes (e.g., 2) can be depicted as the *competing risks model*, Figure 2, where compared with Figure 1, the final state is split into two.

In Figure 2, every one is in the transient state 0 'Alive' at time 0 ($P(X(0) = 0) = 1$) and makes a transition out of state 0 at the time T of death. The random variable,

$$T = \inf_{t>0}(X(t) \neq 0)$$

is proper and represents the sojourn time spent in state 0. The transition is either to the absorbing state 1 'Dead from cause 1' or to the absorbing state 2 'Dead from cause 2', and the cause of death is the random variable $D = X(\infty) = X(T)$. As for survival data, right censoring (at C) may prevent observation of

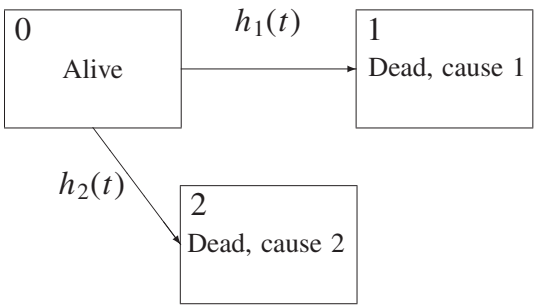

Figure 2. The three-state stochastic process $X(t) = 0, 1,$ or 2 for competing risks (with two causes of death).

T and of the cause of death, and the right-censored competing risks data are then $(\widetilde{T}_i, D_i; i = 1, \ldots, n)$ with $\widetilde{T}_i = T_i \wedge C_i$ and $D_i = X(\widetilde{T}_i)$ (i.e, $D_i = 0$ when $T_i > C_i$). Also, left truncation may be present though we shall skip the details.

Analysis of independently right-censored competing risks data amounts to studying the joint distribution of (T, D), that is, the distribution of what would have been observed in a population without censoring. The basic parameters for this distribution are the *cause-specific hazards* $h_1(t), h_2(t)$ with

$$h_j(t) = \lim_{\mathrm{d}t \to 0} P(X(t + \mathrm{d}t) = j \mid X(t) = 0)/\mathrm{d}t. \tag{8}$$

These are the transition intensities for $X(t)$. Alternatively, the joint distribution may be described by the so-called *cumulative incidence functions*

$$F_j(t) = P(T \leqslant t, D = j), \quad j = 1, 2, \tag{9}$$

from which the *marginal distribution* of T can be found. Thus, the marginal survival function is

$$S(t) = 1 - F_1(t) - F_2(t).$$

Note that although

$$\lim_{t \to \infty} S(t) = 0$$

(T is *proper*), we have for $j = 1, 2$ that

$$\lim_{t \to \infty} F_j(t) = P(D = j) < 1.$$

This has the consequence that the random variables

$$T_{0j} = \inf_{t>0}(X(t) = j), \quad j = 1, 2, \tag{10}$$

are *improper* because, for example, $T_{01} = \infty$ for an individual i with $D_i = 2$.

The survival function $S(t)$ is the state occupation probability for state 0 and also the '$0 \to 0$ transition probability' (because $P(X(0) = 0) = 1$):

$$S(t) = P(X(t) = 0) = P_{00}(0, t) = P(X(t) = 0 \mid X(0) = 0).$$

Similarly,

$$F_j(t) = P(X(t) = j) = P_{0j}(0, t) = P(X(t) = j \mid X(0) = 0), \quad j = 1, 2.$$

These probabilities are simple functionals of the cause-specific hazards

$$S(t) = \exp\left(-\int_0^t (h_1(u) + h_2(u))\mathrm{d}u\right)$$

and

$$F_j(t) = \int_0^t S(u)h_j(u)\mathrm{d}u, \, j = 1, 2. \tag{11}$$

Note also that transition probabilities given $X(s) = 0$ may be derived:

$$P_{00}(s, t) = \exp\left(-\int_s^t (h_1(u) + h_2(u))\mathrm{d}u\right)$$

and

$$P_{0j}(s, t) = \int_s^t S(u)/S(s)h_j(u)\mathrm{d}u = \int_s^t \exp\left(-\int_s^u (h_1(x) + h_2(x))\mathrm{d}x\right) h_j(u)\mathrm{d}u, \, j = 1, 2$$

(while, obviously, $P_{jk}(s, t) = 0$ for $j \neq 0$ and $k \neq j$).

3.2.1. Inference. These probabilities may be estimated nonparametrically using the *Aalen–Johansen* estimators [1, 3] that are obtained by plugging-in the *Nelson–Aalen* estimators for the cumulative cause-specific hazards. Thus,

$$H_j(t) = \int_0^t h_j(u)\mathrm{d}u, \quad j = 1, 2$$

is estimated by

$$\widehat{H}_j(t) = \sum_{\widetilde{T}_i \leq t} \frac{I(D_i = j)}{Y_0(\widetilde{T}_i)},$$

leading to

$$\widehat{F}_j(t) = \sum_{\widetilde{T}_i \leq t} \widehat{S}(\widetilde{T}_i-)\frac{I(D_i = j)}{Y_0(\widetilde{T}_i)}$$

where $\widehat{S}$ is the Kaplan–Meier estimator for the distribution of T, that is, on the basis of failures from any cause.

The likelihood based on independently right-censored competing risks data

$$L = \prod_{i=1}^{n}(h_1(\widetilde{T}_i))^{I(D_i=1)}(h_2(\widetilde{T}_i))^{I(D_i=2)}S(\widetilde{T}_i) \tag{12}$$

can be expressed in terms of the cause-specific hazards. Models may, therefore, be formulated as a model for the cause-specific hazards. A common choice is a Cox model (5) yielding parameter estimates, which are ratios between cause-specific hazards.

At this stage, it is important to notice a crucial difference between the simple two-state model for survival data and the competing risks model. Although in survival analysis, see Equations (2) and (4), there is a one-to-one correspondence between the rate h and the risk F, this is no longer the case for competing risks. In evaluating a single 'risk', that is, a cumulative incidence $F_j(t)$, *both rates* (cause-specific hazards) $h_1(t), h_2(t)$ are required. This means, first of all, that even though interest may focus on a single cause of death (e.g., cause 1) to estimate the corresponding risk $F_1(t)$, it is necessary to estimate both $h_1(t)$ and $h_2(t)$. Second, following Equation (11), the relationship between the risk and the covariates will not be simple even if simple models are fitted to the cause-specific hazards. As an example, one may study a single covariate Z affecting $h_1(t)$ via a Cox model, $h_1(t \mid Z) = h_{10}(t)\exp(\beta Z)$, whereas $h_2(t)$ does not depend on Z. In that case, the cumulative cause 1 incidence is

$$F_1(t \mid Z) = \int_0^t \exp(-\int_0^u (h_{10}(x)\exp(\beta Z) + h_2(x))\mathrm{d}x)h_{10}(u)\exp(\beta Z)\mathrm{d}u$$

which is not a simple function of Z.

This fact has led to the development of models that directly link the cumulative incidence to covariates. The most frequently used such model, which Fine and Gray [4] introduced, specifies

$$\log(-\log(1 - F_j(t \mid Z))) = \log(\widetilde{H}_j(t)) + \beta_j^{\mathsf{T}} Z \tag{13}$$

where $\widetilde{H}_j(t)$ is an increasing but otherwise unspecified function. The Fine–Gray model thus provides parameters, which describe the relationship between the covariates and the cause j risk. For example, for a binary covariate Z_1 with an estimated regression coefficient $\widehat{\beta}_1 > 0$, it follows that for all values, Z_2^0, for the other covariates in the model, we have that

$$\widehat{F}_j(t \mid Z_1 = 1, Z_2^0) > \widehat{F}_j(t \mid Z_1 = 0, Z_2^0).$$

Thus, the positive regression coefficient has the *qualitative* meaning that individuals with $Z_1 = 1$ have a uniformly increased cause 1 cumulative incidence compared with those with $Z_1 = 0$. However, as we shall see in later sections, the interpretation of the value of $\exp(\beta_1)$ is not at all obvious, so the *quantitative* meaning of the regression coefficient is not simple, though Equation (13) may be useful for prediction.

3.2.2. Examples. Practical examples of studies involving competing risks questions are numerous. Kalbfleisch and Prentice [5, Section 8.2.3] exemplified the classical case with different causes of death using the animal carcinogenesis data of [6]. In that study, the authors followed-up radiated mice until failure in either a germ-free or in a conventional laboratory environment, and three different causes of death were all of potential interest. Analysis of these data thus typically involves comparison of both cause-specific hazards and cumulative incidences between the two environments for all causes of death.

In other studies, one failure cause may be of primary interest, and focus is then on the analysis of that cause in the presence of competing risks. As such an example, Thomsen *et al.* [7] have studied the occurrence of affective disorders in patients with obesity (compared with patients with osteoarthritis). The authors presented an analysis not only focused on explanatory variables for the cause-specific hazard of affective disorders but also predicted cumulative incidence curves for that event for various covariate patterns taking the competing risk of death into account.

Another classical application of competing risks methodology is taken from studies of bone marrow transplantation as a treatment in leukaemia [8, 9]. Here, the competing endpoints are relapse of the disease and nonrelapse-related deaths (death in remission). Thus, relapse is taken to be a final (absorbing) state, though obviously, patients experiencing a relapse are still at risk of dying. The argument is that relapse as well as death in remission signal a failure of the treatment with bone marrow, and patients who relapse are in need of alternative treatment. Sometimes, the two endpoints are combined into a single one, thus analyzing 'relapse-free survival'. However, because risk factors for relapse and death in remission often differ (e.g., the effect of graft-versus-host disease tends to differ between the two endpoints) such a combined end-point may not always be useful. Therefore, analysis of such data typically includes both models for the rates (cause-specific hazards) of both relapse and death in remission and for the corresponding risks (cumulative incidences), for example, [10].

It is a general feature of all these examples that the central parameters are the cause-specific hazards and also that suitably chosen functionals are necessary to complete the interpretation of the analyses.

3.3. The illness–death model

Taking the simple two-state model for survival data as the starting point, we obtain the illness–death model (Figure 3) by splitting the initial 'Alive' state into two.

This model is useful when, in the same individuals, both the occurrence of a (chronic) disease and death are studied. Everyone begins in the initial state 'Alive and disease free' ($X(0) = 0$). From this initial state, an individual who develops the disease makes a transition to the state 'Alive and diseased' ($X(t) = 1$), whereas an individual who dies without getting the disease makes a transition to the state 'Dead' ($X(t) = 2$). A diseased individual ($X(t) = 1$) will eventually die and make a transition to the state 'Dead' ($X(t) = 2$). Thus, states 0 and 1 are transient, whereas state 2 is absorbing. The more complicated situation, where a diseased individual may get 'cured' and move back to the disease-free state 0, is also of interest, and that model is the basis for the study of recurrent events [11], but we will not study that situation further here.

The illness–death model may be defined via two proper random variables. These are the sojourn time, T_0, spent in state 0

$$T_0 = \inf_{t>0}(X(t) \neq 0),$$

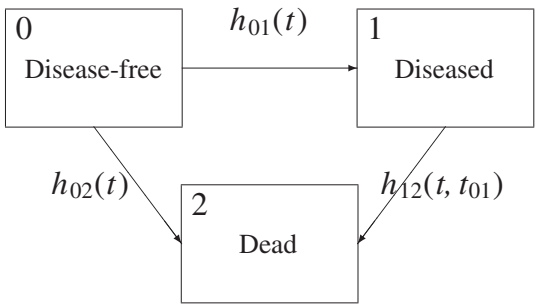

Figure 3. The three-state illness–death stochastic process $X(t) = 0, 1,$ or 2 for occurrence of a (chronic) disease and death.

and the time, T, to death

$$T = \inf_{t>0}(X(t) = 2),$$

where, obviously, $T_0 \leqslant T$. Thus, $T_0 = T$ corresponds to a $0 \to 2$ transition at T and $T_0 < T$ to a $0 \to 1$ transition at T_0 and a $1 \to 2$ transition at T. Right censoring (at C) may, as in the previous sections, prevent complete observation of T_0 and T. Thus, an individual with $C < T_0$ is censored while in state 0 and an individual with $T_0 < C < T$ is censored while in state 1. If $C > T$, then (T_0, T) is completely observed. Left truncation (at L) may also be present. We will skip the details, which turn out to be somewhat more complicated than in the previous sections, because a closer discussion of what is observed at L is needed. Thus, for $T_0 < L < T$, separate situations occur whether or not the value of T_0 is observed at L.

The right-censored observations may be summarized as

$$(\widetilde{T}_0, \widetilde{T}, X(\widetilde{T}_0), X(\widetilde{T}))$$

where $\widetilde{T}_0 = T_0 \wedge C$ and $\widetilde{T} = T \wedge C$. The target of the analysis is the distribution of (T, T_0) in a population without censoring. This distribution may be derived from the three transition intensities. These are the disease rate

$$h_{01}(t) = \lim_{\mathrm{d}t \to 0} P(X(t + \mathrm{d}t) = 1 \mid X(t) = 0)/\mathrm{d}t, \tag{14}$$

the mortality rate without disease

$$h_{02}(t) = \lim_{\mathrm{d}t \to 0} P(X(t + \mathrm{d}t) = 2 \mid X(t) = 0)/\mathrm{d}t, \tag{15}$$

and the mortality rate among the diseased, that is, the $1 \to 2$ transition rate. Here, the situation is more complicated if this also depends on the time,

$$T_{01} = \inf_{t>0}(X(t) = 1) \tag{16}$$

of disease occurrence. Note that T_{01} may be infinite (namely, for individuals moving directly from state 0 to state 2, i.e., for those in whom $T_0 = T$). However, when conditioning on the past (at t) in the definition of

$$h_{12}(t, t_{01}) = \lim_{\mathrm{d}t \to 0} P(X(t + \mathrm{d}t) = 2 \mid X(t) = 1, T_{01} = t_{01})/\mathrm{d}t, \tag{17}$$

attention is restricted to those in state 1 at time t (for whom $T_{01} < t < \infty$). Note that if $h_{12}(\cdot) = 0$ then, formally, the illness–death model is the competing risks model. We return to a discussion of this in the later sections.

When $h_{12}(t, t_{01})$ is independent of the time t_{01} of disease occurrence, the illness–death process $X(\cdot)$ is *Markov*, if not, then it is *semi-Markov*. The simplest semi-Markov model arises when $h_{12}(t, t_{01})$ is only a function of the *duration* $t - t_{01}$ in state 1 at time t.

The state occupation probabilities, $P(X(t) = j), j = 0, 1, 2$ and the transition probabilities $P(X(t) = j \mid X(s) = k), s < t, j, k = 0, 1, 2$ are explicit functions of the three transition rates (both in the Markovian and the non-Markovian situation). Thus, $P_{00}(s, t)$ is given by the same expression as in the competing risks model, whereas $P_{11}(s, t \mid t_{01}) = P(X(t) = 1 \mid X(s) = 1, T_{01} = t_{01}), t_{01} < s < t$ is given by

$$P_{11}(s, t \mid t_{01}) = \exp\left(-\int_s^t h_{12}(u, t_{01})\mathrm{d}u\right)$$

and $P(X(t) = 1 \mid X(s) = 0)$ by

$$P_{01}(s, t) = \int_s^t P_{00}(s, u)h_{01}(u)P_{11}(u, t \mid u)\mathrm{d}u.$$

These expressions simplify in the case of a *Markov* illness–death model.

A functional of these probabilities (and thereby of the transition rates) with important epidemiological applications is the disease *prevalence*

$$P(X(t) = 1 \mid X(t) = 0 \text{ or } X(t) = 1), \tag{18}$$

that is, the fraction of the individuals alive with the disease.

3.3.1. Inference. Inference for the illness–death model will depend on whether the Markov assumption is made. For the transition rates out of state 0, standard hazard models may be applied. However, to apply, for example, a Cox model for the $1 \to 2$ transition rate, a choice of 'baseline' time variable must be made (usually t or duration $t - t_{01}$). Furthermore, if both time variables affect the rate, then the other one must be suitably accounted for, for example, using time-dependent covariates [12]. Alternatively, if t is the baseline time variable, then functions of t_{01} may be taken into account as time-fixed covariates.

State occupation probabilities may be estimated nonparametrically using the Aalen–Johansen estimator [3, 13], or $P(X(t) = 1)$ may be estimated by the difference between the Kaplan–Meier estimators for the distributions of T and T_0 [14]. For the Markov illness–death model, cumulative transition intensities may be estimated nonparametrically by the Nelson–Aalen estimator and transition probabilities by the Aalen–Johansen estimator. Meira-Machado *et al.* [15] introduced estimators for $P(X(t) = k \mid X(s) = j), s < t, j, k = 0, 1, 2$, which do not rely on the Markov assumption. The latter estimator also builds on observed survival distributions: $\widehat{S}$ for T and $\widehat{S}_0$ for T_0. Thus, $\widehat{P}_{00}(s,t)$ is simply $\widehat{S}_0(t)/\widehat{S}_0(s)$ as in the Markovian case, whereas

$$\widehat{P}_{01}(s,t) = \frac{\widehat{E}(\phi_{st}(T_0, T))}{\widehat{S}_0(s)}$$

and

$$\widehat{P}_{11}(s,t) = \frac{\widehat{E}(\widetilde{\phi}_{st}(T_0, T))}{\widehat{E}(\widetilde{\phi}_{ss}(T_0, T))}.$$

Here, $\phi_{st}(u,v) = I(s < u \leqslant t, v > t)$ and $\widetilde{\phi}_{st}(u,v) = I(u \leqslant s, v > t)$ and $\widehat{E}(\phi_{st}(T_0, T))$ is the 'Kaplan–Meier integral'

$$\widehat{E}(\phi_{st}(T_0, T)) = \sum_i w_i \phi_{st}(\widetilde{T}_{0i}, \widetilde{T}_i)$$

with w_i equal to minus the jump at $\widetilde{T}_i$ for $\widehat{S}$. Without censoring, the estimator for any $P_{hj}(s,t)$ reduces to the relative frequency of processes in state j at time t among those in state h at time $s < t$.

3.3.2. Examples. A classical example of an illness–death model (in fact, allowing for recovery) is the disability insurance example [16], where focus is on estimation on the (assumed constant) transition intensities.

In other examples, state 1 is an intermediate event in the disease process and focus may be on the way in which occurrence of this event affects the mortality and whether accounting for the intermediate event improves the precision of mortality predictions. Thus, Andersen *et al.* [17] have studied data from a clinical trial in liver cirrhosis patients with esophageal varices where the intermediate event is variceal bleeding. The authors studied treatment effects on the three transition rates and compared survival predictions with or without accounting for bleeding episodes.

The *prevalence* is a classical epidemiological measure of disease frequency in a population, and the concept has been discussed within a stochastic process framework [18]. Estimation of the prevalence has also been discussed as 'the probability of being in response function' [19] and later, as mentioned earlier, on the basis of the difference between two Kaplan–Meier estimators [14].

As a final example, we can mention the disease course after bone marrow transplantation discussed in Section 3.2 as a competing risks situation. If the target also includes studying the mortality of patients after relapse, the relevant model is now the illness–death model. It has been demonstrated for this example how various functionals derived from the basic transition intensities were necessary to fully appreciate the *net effect* of the various transitions on the final course of events [20, 21]. This process was tentatively termed 'survival synthesis'.

3.4. Summary

In Sections 3.2 and 3.3, we have studied extensions of the simple survival data situation using multistate processes. We exemplified by the competing risks model and the illness–death model and refrained from looking at more complicated models, though the considerations in what follows would be equally relevant for such situations. In both models, the basic parameters were the transition intensities, and other

probabilistic aspects of the models were functionals of these. The practical examples from the literature showed the usefulness of these functionals and illustrated that even in cases where a single event is of primary scientific interest (e.g., a specific cause of death), analysis of the entire process is often justified.

The substantive context made conditioning on the past (or the present, e.g., when defining the prevalence) natural, whereas we never conditioned on the future because we do not find it justified to explain what happens tomorrow based on knowledge to be collected in a more distant future. We saw that a number of random variables, both proper and improper, could be defined, though focus was everywhere on the proper random variables and we never speculated what might happen in a hypothetical population where the variables, which may be infinite in this world, exist and are finite for everyone. Finally, we only conditioned on being in a transient state and not on having reached one of the absorbing states, for example, having died from a particular cause. This is because we do not find further follow-up of deceased individuals justified.

We believe that these three principles

(1) Do not condition on the future;
(2) Do not condition on having reached an absorbing state; and
(3) Stick to this world

are universal and make the functionals under consideration interpretable. This then means that we do believe that functionals violating one or more of these principles have questionable interpretations and thereby have a limited usefulness.

In the next section, we will discuss various functionals, which do not obey our three principles and for which we therefore question their usefulness.

4. Functionals violating our three principles

4.1. Competing risks

4.1.1. The subdistribution hazard. For survival data, we noted in Section 2 that there is a one-to-one correspondence between the rate (2) and the risk (4). We further noted that the hazard rate has a convenient probabilistic interpretation (Equation (3)). Applying the mapping (2) to a cumulative incidence for the competing risks model leads to the so-called *subdistribution hazard* [22]. The subdistribution hazard for cause j,

$$\widetilde{h}_j(t) = -\frac{d\log(1-F_j(t))}{\mathrm{d}t}$$

is the hazard for the improper random variable $T_{0j} = \inf_t(X(t) = j)$ and has the probabilistic 'interpretation'

$$\widetilde{h}_j(t) \approx P(X(t+\mathrm{d}t) = j \mid X(t) \neq j)/\mathrm{d}t,$$

that is, the instantaneous rate of failure per time unit from cause j among those *who are either alive or have died from causes other than* j at time t. This quantity violates our principle 2, and we therefore question the usefulness of the subdistribution hazard. We find the interpretation of regression coefficients in the Fine–Gray model difficult, as these are log(subdistribution hazard ratios), see also Section 3. This difficulty was also stated directly in [4], but nonetheless, users of the model often seem to be unaware of this fact and quote subdistribution hazard ratios more or less as if they were equally interpretable as ordinary hazard ratios [23].

However, as also discussed in Section 3, models that directly link the cumulative incidence to covariates are, indeed, useful as long as it is kept in mind that for the Fine–Gray model, the scale on which cumulative incidence curves are parallel does not have a simple probabilistic interpretation. Therefore, models with other link functions than the 'cloglog' link have been investigated [10, 24, 25].

4.1.2. Latent failure time distributions. We noted in Section 3.2 that the random variables $T_{0j} = \inf_t(X(t) = j), j = 1, 2$ are improper with a distribution given by the cumulative incidence. In the

absence of competing risks, the distribution function would have been obtained by applying the mapping (4) to the cause-specific hazard, for example, for cause j

$$\widetilde{F}_j(t) = 1 - \exp(-\int_0^t h_j(u)\mathrm{d}u). \tag{19}$$

This function is the distribution function for a latent but proper random variable, say $\widetilde{L}_j$, the time to failure from cause j under *independence* of the competing risks, that is, equal to what would have been the distribution function for T_{0j} in a population where this is the only cause operating [5]. Here, independence of competing risks is often taken to mean independence of $\widetilde{L}_1, \widetilde{L}_2$, though weaker assumptions exist. The assumption is unidentifiable on the basis of the available data as summarized in Section 3.2 [5, 26, 27], and therefore, a study of latent failure times relates to a hypothetical world where some causes are not operating, and it violates our principle 3. In the latent failure time approach to competing risks, one imagines the existence of failure times $\widetilde{L}_1, \widetilde{L}_2$ with joint survival function $Q(t_1, t_2) = P(\widetilde{L}_1 > t_1, \widetilde{L}_2 > t_2)$, where $\widetilde{L}_j$ is the time to failure from cause j. Observations then include $T = \min(\widetilde{L}_1, \widetilde{L}_2)$ and the corresponding cause D (i.e., $D = j$ if $T = \widetilde{L}_j$). However, as indicated earlier, $Q(\cdot, \cdot)$ is unidentifiable on the basis of the available data. What may be identified are $S(t) = Q(t, t)$ and other functionals of the cause-specific hazards. This is because the likelihood (12) is defined via the cause-specific hazards, and although the cause-specific hazards are uniquely given from $Q(t_1, t_2)$ as

$$h_j(t) = -\frac{\partial \log Q(t_1, t_2)}{\partial t_j}\rfloor_{t_1=t_2=t},$$

the converse is not true. In fact, an example of two different joint survival functions with the same cause-specific hazards has been given, where one corresponds to independent latent failure times and the other to dependent latent failure times [5].

Various attempts to circumvent this problem have been put forward, for example, showing that by assuming the joint distribution to belong to a class generated by a given copula, the marginal distributions are identifiable [28]. However, such an assumption cannot be evaluated against the data.

The literature on elimination of causes of death has a long history and is still very much alive. It is worth briefly recapitulating the famous discussion on how to model the possible effect of elimination of smallpox.

Daniel Bernoulli presented a model for the infection dynamics of smallpox at a meeting in the Académie des Sciences in Paris in 1760 [29], illustrated with numerical calculations based on Halley's life table from 1693 and assumptions on (age-independent) incidence of smallpox and (age-independent) mortality for smallpox-infected individuals. (Bernoulli's paper has now been put into a contemporary infectious disease epidemiology perspective [30].)

A few months later d'Alembert presented an alternative approach, very close to our present-day non-parametric calculations of mortality 'in a world with no smallpox' [31]. Karn [32] gave a comprehensive account of both approaches, as well as a sketch of the prehistory of smallpox inoculation. She carefully added the following:

> It is necessary to warn the reader that the theory of the life-table with a given disease eliminated as developed by Bernoulli, D'Alembert, Tremblay and Duvillard supposes that the mortality from the given disease is non-selective, i.e. that the population after removal of disease A is as susceptible to diseases B, C, D, etc. as it was before the elimination of that disease. This may possibly be true of certain diseases, but if a disease like phthisis or small-pox were eliminated the surviving population might be more subject to death from other diseases.

We find Karn's wise and cautious advice as important today and hope that 'what if' questions are explicitly put into the context of *sensitivity analysis*. Thus, one may speculate how much the distribution of causes of death may change if, by some intervention, one were able to modify the rate of one cause of death by a certain amount without changing the remaining cause-specific hazards. In the spirit of that, multistate models for bone marrow transplantation have been studied calculating the consequences on, for example, the risk of relapse under a number of hypothetical scenarios concerning changes in the rate of developing graft-versus-host disease [21, 33]. Obviously, one may see calculations based on an assumption of 'independent' competing risks as an extreme case of such an activity, and although this may provide some bounds for what might happen under various interventions, we would like to emphasize that such calculations remain hypothetical and without support in the data at hand.

Greenland [34] took a slightly different approach, recommending to handle the complexity of competing risks in biological contexts by always respecting the multivariate nature of the problem, that is, avoid singling out one isolated feature.

The likelihood (12) factorizes as $L = L_1 L_2$, where

$$L_j = \prod_{i=1}^{n} (h_j(\widetilde{T}_i))^{I(D_i=j)} \exp(-H_j(\widetilde{T}_i)),\ j = 1, 2.$$

Note that, for example, L_1 is the likelihood that would be obtained for $h_1(t)$ if cause 1 were the only one operating and where individuals failing from cause 2 are censored at their time of failure. That is, if no parameters are common for the two cause-specific hazards, inference for $h_1(t)$ may be performed by, formally, censoring individuals at the time of a cause 2 failure and vice versa. Also note that this has nothing to do with an assumption of 'independent competing risks'; it is solely a consequence of the likelihood factorization. This fact may at a first glance seem surprising because why is a Cox model for the rate $h_1(t)$ censoring for cause 2 failures valid, whereas 'one minus the Kaplan–Meier estimator based on cause 1 failures' is not a valid estimator for the risk $F_1(t)$? The answer has to do with the lack of a one-to-one correspondence between the rate and the risk as we emphasized in Section 3.2. As explained in Section 2, when we, in survival analysis, estimate the risk as $1-\widehat{S}(t)$ in the presence of censored observations, we make inference for the risk parameter in a population without censoring under the assumption of independent censoring. That is, we assume the population without censoring to be potentially observable. If censoring is due to failures from competing risks, then the population without censoring is quite hypothetical, and in the presence of competing risks, the risk $F_j(t)$ of failure from cause j is given by the cumulative incidence (11) and should be estimated accordingly by the Aalen–Johansen estimator. However, inference for the 'local' parameter $h_j(t)$ may be carried out by, technically, censoring for failures from competing causes as a consequence of the likelihood factorization. This includes not only the Cox model but also techniques like the logrank test and the Nelson–Aalen estimator.

Only when the rate of competing causes is *low*, that is, $h_2(t) \approx 0$, do we have

$$1 - \exp(-\int_0^t h_1(u)\mathrm{d}u) \approx \int_0^t \exp\left(-\int_0^u (h_1(x) + h_2(x))\mathrm{d}x\right) h_2(u)\mathrm{d}u. \tag{20}$$

However, this merely tells us that if competing risks are negligible, then they may be neglected!

We conclude this subsection by emphasizing that the Kaplan–Meier estimator should be used only for estimation of the distribution function for a *proper* random variable T. This is because if T is not proper, then there will be competing risks 'censoring the observation of T', and in that the case, the population without censoring is not a useful one to make inference for. An exception would be when Equation (20) holds, that is, when the competing risks are negligible. We are then faced with the problem of ascertaining whether censoring is *independent*, and as mentioned in Section 2, this typically cannot be carried out on the basis of the available data. The question of independent censoring is, therefore, usually a topic for discussion. *Administrative* censoring, that is, the situation when individuals are alive but still in the study at the date of study closure may often be taken as 'independent'. However, reasons for individuals dropping out of the study before the date of study closure should always be scrutinized carefully to ascertain whether the mechanisms leading to dropout can safely be taken to be independent of the failure process.

4.1.3. The conditional probability curve. As discussed in Section 3.2, a classical application of competing risks methodology is taken from studies of bone marrow transplantation as a treatment in leukaemia where the competing endpoints are relapse of the disease and death in remission. Motivated by this example, a *conditional probability curve* has been suggested as a summary measure [9, 35]. In this context, the suggested summary curve is the conditional probability of relapse by t given no nonrelapse-related death by t, that is,

$$P(X(t) = 1 \mid X(t) \neq 2) = P(X(t) = 1 \mid X(t) = 0 \text{ or } X(t) = 1).$$

If computed as a prediction at time 0, it conditions on the future and violates our principle 1, and if computed at time t, it conditions on possibly having reached the absorbing state 1 and violates our principle 2. We will argue that this is a perfectly reasonable functional in the *illness–death model* (where it is simply the prevalence (18)). In this latter model, referring still to the bone marrow transplantation

example, the prevalence is the fraction of the patients alive who have experienced a relapse. Regarded in the framework of a competing risks model, it is the fraction of the patients who are either alive with a relapse or who have died after a relapse out of those patients who are either alive or have died after a relapse. The problem is that the classical example from bone marrow transplantation may not always be regarded as a bona fide competing risks situation because the state 'relapse', as mentioned previously, is not really absorbing. When applying competing risks methodology to this situation, one should therefore consider carefully whether this is, in fact, justified or whether inference using the illness–death model would be more appropriate.

4.1.4. Pattern–mixture parametrization. As discussed in Section 3.2, the object of competing risks is the joint distribution of time to failure T and cause of failure D. Technically, this joint distribution can of course be factorized as the conditional distribution of T given D and the marginal distribution of D. Such a pattern–mixture parametrization has been studied [36, 37]. Considering the conditional distribution of failure time given failure cause, it conditions on the future and violates our principle 1, and we therefore question the usefulness of this parametrization. Clearly, for purposes of individual patient predictions, it is completely useless. Thus, knowing the results from an analysis of that model, the doctor may say to the patient that 'if you die from lung cancer then you should stop smoking' or 'among those who will die from heart disease, men have a hazard which is twice that for women'! Note that the 'selection' parametrization of the distribution of (T, D), that is, the marginal distribution of T times the conditional distribution of D given T [38] does not violate our principles as neither conditioning on the future nor follow-up of deceased subjects is involved.

4.2. The illness–death model

4.2.1. Pattern–mixture parametrization. Just as in the competing risks model, pattern–mixture parameterizations have been studied in the illness–death model for the distribution of T_0, the sojourn time spent in state 0. Thus, the so-called 'flow graph models' [39] use this formulation throughout. In fact, this is a standard parametrization for semi-Markov models [40], and although it is mathematically feasible, it does condition on the future and violates our principle 1.

4.2.2. Semicompeting risks. The situation where a terminal event censors a nonterminal event but not vice versa has been studied [41–43] under the name of 'semicompeting risks'. This is seen to be exactly the situation for the illness–death model where death without the disease ($0 \to 2$ transition) prevents the disease occurrence ($0 \to 1$ transition), whereas subjects experiencing the disease ($0 \to 1$ transition) are still at risk of death ($1 \to 2$ transition). The authors argue that although the primary concern for clinicians is patient survival, 'there is often scientific interest in the distribution of a non-terminal event, unconditionally on censoring' and though 'the quantity is controversial it is meaningful to many researchers.' However, because it relates to a hypothetical world where patients cannot die, it violates our principle 3 and we question its usefulness.

5. Discussion

We have looked at analysis of multistate models as 'generalized survival analysis'. For simple survival analysis, only a single transition hazard is present, a fact that limits the amount of functionals to be studied. On the other hand, multistate models allow several possible transitions and, therefore, include several transition hazards, and we have seen that the added generality also allows more functionals of these transition hazards to be studied. Although all the functionals discussed are mathematically well defined, we have argued that they do not all possess a useful interpretation from an applied point of view. What is interpretable in a practical context, and what is not, is of course, to some extent, a matter of taste and a matter of which applications one has in mind. We have put forward three principles (Section 3.4), which we believe enhance both the interpretation and the practical applicability of a given functional:

(1) Do not condition on the future;
(2) Do not condition on having reached an absorbing state; and
(3) Stick to this world.

We emphasize that we do not regard these principles as inherently obvious axioms. On the contrary, we have come to them inductively, as a convenient summary from a broad experience of what makes sense in practice.

References

1. Andersen PK, Borgan Ø, Gill RD, Keiding N. *Statistical Models Based on Counting Processes*. Springer: New York, 1993.
2. Cox DR. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society Series B* 1972; **34**:187–220.
3. Aalen OO, Johansen S. An empirical transition matrix for nonhomogenous Markov chains based on censored observations. *Scandinavian Journal of Statistics* 1978; **5**:141–150.
4. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association* 1999; **94**:496–509.
5. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*, 2nd ed. Wiley: New York, 2002.
6. Hoel DG, Walburg HE. Statistical analysis of survival experiments. *Journal of the National Cancer Instiute* 1972; **49**:361–372.
7. Thomsen AF, Kvist TK, Andersen PK, Kessing LV. Increased relative risk of subsequent affective disorder in patients with a hospital diagnosis of obesity. *International Journal of Obesity* 2006; **30**:1415–1421.
8. Prentice RL, Kalbfleisch JD, Peterson AV, Flournoy N, Farewell V, Breslow NE. The analysis of failure times in the presence of competing risks. *Biometrics* 1978; **34**:541–554.
9. Pepe MS, Mori M. Kaplan-Meier, marginal, or conditional probability curves in summarizing competing risks failure time data? *Statistics in Medicine* 1993; **12**:737–751.
10. Klein JP, Andersen PK. Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics* 2005; **61**:223–229.
11. Cook RD, Lawless JF. *The Statistical Analysis of Recurrent Events*. Springer: New York, 2007.
12. Andersen PK, Keiding N. Multi-state models for event history analysis. *Statistical Methods in Medical Research* 2002; **11**:91–115.
13. Datta S, Satten GA. Validity of the Aalen-Johansen estimators of stage occupation probabilities and Nelson-Aalen estimators of integrated transition hazards for non-Markov models. *Statistics & Probability Letters* 2001; **55**:403–411.
14. Pepe MS. Inference for events with dependent risks in multiple endpoint studies. *Journal of the American Statistical Association* 1991; **86**:770–778.
15. Meira-Machado L, Una-Alvarez J, Cadarso-Suarez C. Nonparametric estimation of transition probabilities in a non-Markov illness-death model. *Lifetime Data Analysis* 2006; **12**:325–344.
16. Sverdrup E. Estimates and test procedures in connection with stochastic models for deaths, recoveries and and transfers between different states of health. *Skandinavisk Aktuarietidsskrift* 1965; **48**:184–211.
17. Andersen PK, Esbjerg S, Sørensen TIA. Multi-state models for bleeding episodes and mortality in liver cirrhosis. *Statistics in Medicine* 2000; **19**:587–599.
18. Keiding N. Age-specific incidence and prevalence: a statistical perspective (with discussion). *Journal of the Royal Statistical Society, Series A* 1991; **154**:371–412.
19. Temkin NR. An analysis for transient states with application to tumor shrinkage. *Biometrics* 1978; **34**:571–580.
20. Keiding N. Event history analysis and inference from observational epidemiology. *Statistics in Medicine* 1999; **18**: 2353–2363.
21. Keiding N, Klein JP, Horowitz MM. Multistate models and outcome prediction in bone marrow transplantation. *Statistics in Medicine* 2001; **20**:1871–1885.
22. Gray RJ. A class of K-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics* 1988; **16**:1141–1154.
23. Kim HT. Cumulative incidence in competing risks data and competing risks regression analysis. *Clinical Cancer Research* 2007; **13**:559–565.
24. Fine JP. Regression modeling of competing crude failure probabilities. *Biostatistics* 2001; **2**:85–97.
25. Scheike TH, Zhang M-J, Gerds TA. Predicting cumulative incidence probability by direct binomial regression. *Biometrika* 2008; **95**:205–220.
26. Cox DR. The analysis of exponentially distributed life-times with two types of failure. *Journal of the Royal Statistical Society series B* 1959; **21**:411–421.
27. Tsiatis AA. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences, USA* 1975; **72**:20–22.
28. Zheng M, Klein JP. Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika* 1995; **82**:127–138.
29. Bernoulli D. Essai d'une nouvelle analyse de la mortalitè, causeè par la petite vèrole, et des avantages de l'inoculation pour le prèvenir. *Histoire avec le Mèmoires, Acadèmie Royal des Sciences Paris* 1766; **1760**:1–45.
30. Dietz K, Heesterbeek JAP. Daniel Bernoulli's epidemiological model revisited. *Mathematical Biosciences* 2002; **180**:1–21.
31. d'Alembert JLR. Sur l'application du calcul des probabilités à l'inoculation de la petite vérole. In *Opuscules Mathematiques*, Vol. 2, JLR d'Alembert (ed.), 1761; 26.
32. Karn MN. An inquiry into various death-rates and the comparative influence of certain diseases on the duration of life. *Annals of Eugenics* 1931; **4**:279–326.
33. Klein JP, Keiding N, Copelan EA. Plotting summary predictions in multistate survival models: probabilities of relapse and death in remission for bone marrow transplantation patients. *Statistics in Medicine* 1993; **12**:2315–2332.

34. Greenland S. Epidemiological measures and policy formulation: lessons from potential outcomes. *Emerging Themes in Epidemiology* 2005; **2**:5.
35. Allignol A, Latouche A, Yan J, Fine JP. A regression model for the conditional probability of a competing event: application to monochlonal gammopathy of unknown significance. *Applied Statistics* 2011; **60**:135–142.
36. Larson MG, Dinse GE. A mixture model for the regression analysis of competing risks data. *Applied Statistics* 1985; **34**:201–211.
37. Elashoff RM, Li G, Li N. An approach to joint analysis of longitudinal measurements and competing risks failure time data. *Statistics in Medicine* 2007; **26**:2813–2835.
38. Nicolaie MA, van Houwelingen HC, Putter H. Vertical modeling: a pattern mixture approach for competing risks modeling. *Statistics in Medicine* 2010; **29**:1190–1205.
39. Huzurbazar AV. *Flowgraph Models for Multistate Time-to-Event Analysis*. Wiley: New York, 2005.
40. Phelan MJ. Estimating the transition probabilities from censored Markov renewal processes. *Statistics & Probability Letters* 1990; **10**:43–47.
41. Fine JP, Jiang H, Chappell R. On semi-competing risks. *Biometrika* 2001; **88**:907–919.
42. Peng L, Fin JP. Regression modeling of semicompeting risks data. *Biometrics* 2006; **63**:96–108.
43. Hsieh J-J, Wang W, Ding AA. Regression analysis based on semicompeting risks data. *Journal of the Royal Statistical Society series B* 2008; **70**:3–20.