Scuola di Dottorato
Università degli Studi di Milano-Bicocca

Department of Medicine and Surgery
PhD program in Public Health
Cycle XXXIV
Curriculum in Biostatistics and Clinical Research

Adverse events in survival data: from clinical questions to methods for statistical analysis

Candidate: TASSISTRO ELENA
Registration number: 786970

Tutor: Prof.ssa Maria Grazia Valsecchi
Co-tutor: Prof.ssa Laura Antolini

Coordinator: Prof. Guido Grassi

ACADEMIC YEAR 2020-2021

Contents

Introduction

In clinical studies on novel treatments the evaluation of both efficacy and safety outcomes should play an important role in the assessment of the benefit-risk profile of the therapies. Object of interest of safety evaluations are the *adverse events* (AEs), that are defined by the European Medicines Agency as "*any untoward medical occurrence in a patient or clinical investigation subject administered a pharmaceutical product and which does not necessarily have to have a causal relationship with this treatment*" [1]. From this definition, an AE can therefore be any unfavourable and unintended sign, symptom or disease which is temporally associated with the use of a treatment, whether or not considered related to the treatment.

The majority of the studies, however, is designed only with respect to detecting treatment effects in efficacy outcomes, since the evaluation of safety outcomes requires larger sample sizes and long trial durations [2]. This difference between efficacy and safety influences also the statistical analyses of the data. Indeed, whereas the statistical methodology to evaluate the efficacy of a treatment is continuously advancing, the analyses of safety outcomes are limited to descriptive methods reported among the initial descriptive results in clinical papers [2–7]. This limitation is due to a variety of factors, such as treatment discontinuation (caused by toxicity, death, disease progression, etc.), low power due to rare AEs or, when the outcome of interest is the time to first occurrence of an AE, multiplicity of the possible AEs or recurrent episodes of the same AE-type.

The primary focus of safety evaluations is to quantify the incidence of AEs, so the commonly used measures are the crude proportion of AE, that is the number of subjects developing an AE divided by the total number of subjects, and the epidemiological AE rate, that is the number of subjects developing an AE divided by the total time spent free from failure [6–9]. These quantities are not functions of time and do not account properly for the occurrence of one or more *competing risks*, which are events (death, relapse, etc.) that preclude or alter the probability of experiencing the event of interest (AE) [10]. A classical clinical example of competing events is death due to cardiovascular diseases versus death due to any other cause. In order to account for the presence of competing risks, one should resort to survival methodologies to analyse this type of data, such as the Kaplan-Meier, the Aalen-Nelson or the Aalen-Johansen estimators.

In studies where novel therapies need to be evaluated both in terms of survival time

outcomes (for example time to relapse in onco-hematology) and in terms of AEs (or severe AEs) the treatment failure can be defined as the first event occurring among relapse and serious AEs. In the presence of these competing risks, the occurrence of relapse as first event and the subsequent treatment change exclude the possibility of observing AEs related to the treatment under analysis. In principle, the analysis of AEs could be tackled from two different points of view. Approach 1 focuses on the description of the observed occurrence of AE as first event among the competing risks AE and relapse and on the comparison between the two treatments. In this case, treatment ability to protect from relapse has an impact on the chance of observing AEs due to the competing risks action. The more the treatment causes relapse, the less is the chance to observe an AE as first event (indirect protection) [6]. Approach 2 focuses on the description of the potential occurrence of AE in relapse free patients. To address this issue, one should consider the occurrence of AEs as if relapse would not exclude the possibility of observing AEs related to the treatment under analysis, thus in the absence of competing risks. In this way, the comparison between the treatment impact on AEs does not depend on the effect of treatment on relapse. This approach is particularly important if one needs to compare alternative treatments with differential indirect protections.

In both approaches, the theoretical quantities and the relative estimators used in clinical papers to describe AEs data should have these features:

(a) the estimator addresses for the presence of right censoring;

(b) the theoretical quantity and estimator are functions of time, i.e. can be calculated at different time points.

This thesis has two aims: the first is the review of the two approaches starting from the type of clinical question answered by the two types of analysis, and the identification of the suitable quantities and the commonly used estimators according to the aforementioned features. The second aim is to define a strategy to relax the assumption of independence between the potential times to the competing events, such as that to AE and the potential time to relapse, of the commonly used estimators in the second approach.

The thesis is organized as follows: in Chapter 1 a review of the basics of survival analysis is presented. In addition the notation used throughout the work is introduced. In Chapter 2 the standard methods commonly used to analyse AEs data represented by the crude proportion of AEs and the epidemiological AEs rate are described and we prove that they fail in at least one of the two features. We then show the use of crude incidence and cause-specific standard methods to be consistent with the two features for both approaches. In Chapter 3 we introduce the use of regression models, stratified Kaplan-Meier curves and inverse probability of censoring weighting to relax the assumption of independence by achieving conditional independence given covariates. In Chapter 4 we introduce, as motivating example, a study aiming at evaluating the occurrence of

osteonecrosis as possible AE of the front line chemotherapy treatment of children diagnosed with acute lymphoblastic leukaemia. A simulated data example is also presented in order to show the interpretation of the standard methods and the regression models where the assumption of independence is relaxed. In addition, we aim at making clear to the reader the impact of the assumption of independence between potential times to AEs and to relapse of the standard estimators used for the second approach through simulated datasets of increasing sample sizes. In Chapter 5 an extensive simulation protocol that shows the performance of the latter methods and the impact of not accounting for an unmeasured covariate is presented. The protocol is also extended changing the imbalance of the covariates and the hazard ratio of relapse. In Chapter 6 results of the simulations are presented. In Chapter 7, starting from on ongoing work on the analysis of AEs in patients affected by leukaemia, the recurrent events are briefly introduced and how to interpret the Aalen-Nelson estimator in this context is explained. In addition, an insight on the illness-death model to analyse the impact of the occurrence of the AE on the subsequent hazard of relapse is presented. The thesis ends with a discussion of all the findings presented in the previous chapters.

Chapter 1

Basics of survival analysis and notation

The aim of survival analysis is to study the *survival time*, which is the time elapsed from a certain starting point (e.g. first diagnosis of a given disease, surgical intervention, beginning of a treatment, birth) to the occurrence of an event (e.g. death, relapse, recovery or in general a prespecified event of interest) [11].

A typical problem of the survival analysis is given by the presence of *censoring*, due to the fact that the event of interest may not be observed on all subjects. As a consequence, the survival time is completely available only for some individuals. A survival time may be censored because the person does not experience the event of interest before the study ends (i.e. administrative censoring), is lost to follow-up during the study period or withdraws from the study because of reasons other than the event of interest. All these reasons are typical examples of *right censoring* that in general happens when the true survival time is equal to or greater than the observed survival time. However, two other situations can occur: the true survival time is less than or equal to the observed survival time (*left censoring*) or the true survival time is within a known time interval (*interval censoring*) [10].

Survival data on a sample of N observations are represented by a pair of variables (T_O, δ), where T_O is the observed survival (or failure) time and δ is the failure indicator ($\delta = 1$ if the subject develops the event of interest and $\delta = 0$ if his/her survival time is censored). Defining with T the true survival time one wishes to analyse and with C the censoring time, the observed survival time is $T_O = min(T, C)$. The assumption on which methods for censored survival analysis are based on is that the random variables T and C are independent (*independent censoring*). This means that the survival experience of subjects censored prior a given time t can be estimated by using data on the remaining subjects [11]. This assumption is also called *non-informative censoring* (e.g. administrative censoring). In situations where there is dependence between T and C (*informative censoring*) one should consider non standard methods that account for this problem.

1.1 Theoretical functions

The distribution function of the survival time random variable (r.v.) T known as *cumulative incidence function* (CIF) is

$$F(t) = \mathrm{P}(T \leq t)$$

and gives the probability that the event of interest occurs at a failure time less than or equal to t. Its derivative $\frac{d}{dt}F(t) = f(t)$ is the *probability density function* at the time point t. The complement to 1 of the CIF is the *survival function*

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

which represents the probability of surviving longer than time t. Since t ranges from 0 up to infinity, $S(t)$ is a non-increasing function of t and

$$S(t) = \begin{cases} 1, & \text{if } t = 0 \\ 0, & \text{if } t \to \infty \end{cases}$$

Another function at the basis of the survival analysis is the *hazard function*, defined as

$$h(t) = \lim_{\Delta t \to 0^+} \frac{\mathrm{P}(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}$$

It gives the instantaneous rate, which corresponds to the risk at time t of an individual to develop the event of interest in the next time unit, given that he/she is event-free up to time t. Since the hazard expresses a rate, rather than a probability, $h(t)$ ranges between 0 and infinity. The hazard function can also be rewritten as

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt}\log(S(t)) \tag{1.1}$$

which depends only on $S(t)$.

We define with

$$H(t) = \int_0^t h(u)du$$

the *cumulative hazard function*, which is the cumulative risk of an event occurring by time t. With some analytical calculation, from equation (1.1) one can derive a formula for the survival function

$$S(t) = \exp\left(-\int_0^t h(u)du\right) = \exp\left(-H(t)\right) \tag{1.2}$$

Another basic quantity of the survival analysis that can be used in the presence of

competing risks is the *crude incidence* (CI) probability

$$CI(t) = P(T \leq t; \delta = 1) \tag{1.3}$$

which corresponds to the absolute risk of failure due to the event of interest, in this case denoted by 1, up to time t. In (1.3) there is the additional random variable δ used to denote the type of event occurred by time t, in a scenario where failure may be originated by two, or even more, events.

1.2 Common estimators in survival analysis

To analyse survival data and then to obtain estimates of the survival, hazard and incidence functions one can use non-parametric, semi-parametric or parametric methods. In the following, examples of the first two categories are briefly described.

1.2.1 The Kaplan-Meier estimator

The *Kaplan-Meier* (KM) estimator [12], also known as the product limit estimator, is a non-parametric maximum likelihood (ML) method that can be used to estimate the survival function in the presence of survival data with right censored observations, assuming independence between the censoring time (i.e. time of relapse) and the true survival time (i.e. time of AE). Let $t_1 < t_2 < ... < t_j < ... < t_J$, $j = 1, ..., J$, be the $J \leq N$ distinct ordered failure times observed among the N subjects, δ_j the failure indicator and $d_j = \sum_{i=1}^{N} I(t_i = t_j; \delta_j = 1)$ the corresponding number of failures, $d_j \geq 1$. Let n_j be the number of subjects *at risk* of failure, i.e. the number of subjects who have either AE or relapse times greater than or equal to t_j. An estimate of $S(t)$ derived by the KM estimator is

$$\hat{S}(t) = \prod_{t_j \leq t} \frac{n_j - d_j}{n_j} \tag{1.4}$$

If the largest observed survival time is a censored survival time t^*, $\hat{S}(t)$ is undefined for times $t > t^*$, whereas if the largest observed survival time t_J is an uncensored observation, $n_J = d_J$ and so $\hat{S}(t) = 0$ for $t \geq t_J$. Of note, formula (1.4) is equal to the simple proportion of subjects survived at time t in the absence of censoring.

The plot of the KM estimate of the survival function is a step-function in which the estimated survival probabilities are constant between adjacent failure times and decrease at each failure time. The KM estimator could be multiplied by 100 in order to interpret it as the expected number of subjects surviving free from the event among 100 hypothetical subjects.

1.2.2 The Aalen-Nelson estimator

The *Aalen-Nelson* (AN) estimator [13–15] is a non-parametric ML method that can be used to estimate the cumulative hazard function in the presence of survival data with right censored observations, assuming independence between the censoring time (i.e. time of relapse) and the true survival time (i.e. time of AE). An estimate of $H(t)$ derived by the AN estimator is

$$\hat{H}(t) = \sum_{t_j \leq t} \frac{d_j}{n_j} \tag{1.5}$$

which represents an increasing step function with increments equal to $\frac{d_j}{n_j}$ at each distinct observed failure time. In case of a single event of interest, the AN estimate of the cumulative hazard can be interpreted as the expected number of events in the time interval a patient may experience as if he/she would develop subsequent events in time [16]. This interpretation is interesting in particular when recurrent events in time are collected. In this case, the AN estimator could be multiplied by 100 in order to interpret it as the expected number of events in time in 100 hypothetical subjects.

Of note, from equation (1.2), one can derive a formula to estimate the survival function through the AN estimator:

$$\hat{S}(t) = \prod_{t_j \leq t} \exp\left(-\frac{d_j}{n_j}\right) \tag{1.6}$$

1.2.3 Alternative KM formula and relationship between KM and AN estimators

An alternative (recursive) formula of the KM estimator is

$$\hat{S}(t) = 1 - \sum_{t_j \leq t} \frac{d_j}{n_j} \hat{S}(t_j-)$$

where $\hat{S}(0) = 1$. Of note, this formula is algebraically equivalent to equation (1.4).

The KM and AN estimators of the survival function presented in formulas (1.4) and (1.6) respectively are related each other. Considering the mathematical property that, for large values of n_j, $\exp(-1/n_j)$ is approximately equal to $1-1/n_j$, the survival probability estimated with the AN formula (1.6) approximates the survival probability estimated with the KM formula (1.4) at least in large samples [11].

The AN and KM estimators can also be used to obtain "smoothed" estimates of the hazard function. In particular, an estimate of the hazard can be obtained by the cumulative sum derived from the AN estimator represented in formula (1.5), which is by definition a non-decreasing function. As an alternative, one may consider the cumulative product of $(1 - \hat{h}(t))$ terms, with the complement to 1 of the KM estimate of formula

(1.4) used to obtain again a non-decreasing function

$$1 - \prod_{t_j \leq t} (1 - \hat{h}(t_j)) = 1 - \prod_{t_j \leq t} \left(1 - \frac{d_j}{n_j}\right)$$

1.2.4 The Cox Regression Model

The Cox regression model [17] is the most widely used method in the analysis of survival data with right censored observations for quantifying the impact of multiple covariates on a certain survival outcome, assuming independence between the true survival time and the censoring time (i.e. non-informative censoring) conditional on covariates [11]. Of note, this assumption of conditional independence is less restrictive with respect to the assumption of independence required by the KM estimator.

Let $\boldsymbol{X} = (X_1, X_2, ..., X_k, ..., X_K)$, $k = 1, ..., K$, be the vector of K explanatory variables or covariates. For sake of simplicity $\boldsymbol{X}$ will be used to indicate a vector of both time-fixed and time-dependent covariates. The Cox model assumes that the hazard function is

$$h(t, \boldsymbol{X}) = h_0(t) \exp(\boldsymbol{\beta}' \boldsymbol{X}) \tag{1.7}$$

where $h_0(t)$ is an unspecified non-negative function of time called *baseline hazard*, assumed to be the same for all subjects, and $\boldsymbol{\beta}$ is a vector of coefficients. Of note, if all covariates are equal to zero, the Cox model reduces to the baseline hazard function. That is, the exponential part of the formula (1.7) becomes equal to 1. This property is the reason why $h_0(t)$ is called *baseline function* [10].

Another assumption of the Cox model is that independent covariates affect the hazard in a multiplicative way: this is implied by the use of the exponential function for linking the covariates to the hazard [11].

Since $h(t, \boldsymbol{X})$ in formula (1.7) does not specify the form of $h_0(t)$ but it specifies only the hazard ratio for any two individuals with covariate vectors $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, the Cox model is a semi-parametric model. Because the hazard ratio

$$HR = \frac{h(t, \boldsymbol{X}_1)}{h(t, \boldsymbol{X}_2)} = \frac{h_0(t) \exp(\boldsymbol{\beta}' \boldsymbol{X}_1)}{h_0(t) \exp(\boldsymbol{\beta}' \boldsymbol{X}_2)} = \exp(\boldsymbol{\beta}' (\boldsymbol{X}_1 - \boldsymbol{X}_2))$$

does not depend on $h_0(t)$ (and so is constant over time) the Cox model is a *proportional hazards* (PH) regression model. Of note, if the two individuals are taken to have covariates vectors $\boldsymbol{X}$ and $\mathbf{0}$ the hazard ratio becomes

$$HR = \frac{h(t, \boldsymbol{X})}{h(t, \mathbf{0})} = \frac{h_0(t) \exp(\boldsymbol{\beta}' \boldsymbol{X})}{h_0(t)} = \exp(\boldsymbol{\beta}' \boldsymbol{X})$$

To estimate the vector of parameters $\boldsymbol{\beta}$ it is sufficient to maximize the partial likelihood [17], which correspond to the product, on all event times, of the ratio between the

hazard of a subject developing an event at time t_j and the hazard of all subjects at risk at the same time t_j:

$$L(\boldsymbol{\beta}) = \prod_{j=1}^{J} \frac{\exp(\boldsymbol{\beta}' \boldsymbol{X_j})}{\sum_{l \in R_j} \exp(\boldsymbol{\beta}' \boldsymbol{X_l})} \tag{1.8}$$

where J is the number of events and R_j is the number of subjects at risk of developing an event at time t_j.

1.2.5 The Aalen-Johansen estimator for competing risks

The non-parametric ML estimator of the crude incidence presented in formula (1.3) is given by the *Aalen-Johansen* (AJ) formula introduced in [18]

$$\hat{CI}_{AE}(t) = \sum_{t_j \leq t} \hat{S}(t_j-) \cdot \hat{h}_{AE}(t_j)$$

where $\hat{S}(t_j-) = \hat{P}(T > t_j-)$ is the KM estimator of the proportion of patients free from failure up to time t_j- and $\hat{h}_{AE}(t)$ is the instantaneous rate of AE, which corresponds to the proportion of patients experiencing an AE at time t over the total number of patients at risk, i.e. free from AE and relapse, at that time.

Of note, if only one event is involved (i.e. if competing risks are absent) the CI estimate is equal to 1-KM estimate.

1.3 Notation

In the following of the thesis we will consider the occurrence of an AE as the event of interest, whereas the occurrence of relapse acts as competing event. We will denote with T_{AE} the survival time from origin defined by the occurrence of AE in time. Similarly, the occurrence of a relapse in time defines the survival time T_{RL}. These survival times are called "potential" since only the minimum is observable as first event. The observable failure time is $T = min(T_{AE}, T_{RL})$ and the observable cause of failure is E (equal to 1 if AE, equal to 2 if relapse). In the presence of a right censoring time C, the observed time is $min(T, C)$ and $\Delta = I(T \leq C)$ is the failure indicator. Finally, $(t_i, \delta_i, \delta_i \cdot e_i)$, $i = 1, ..., N$, is used to denote the observed failure time, the failure indicator and the cause of failure, on a sized N sample, and t_j indicates the distinct observed failure times. Our focus will be on

- the estimation of the CI of AE, thus on the distribution of (T, Δ)
- the estimation of the survival (and related quantities) in an hypothetical world where relapse is absent, thus on the distribution of the potential T_{AE}.

When two binary covariates X_1 and X_2 are supposed to have an impact on the potential time T_{AE}, the survival function can be written (under exponential assumption) as

$$\begin{aligned} S_{AE}(t) = & P(X_1 = 0)P(X_2 = 0)exp(-\lambda_{00AE} \cdot t) + \\ & + P(X_1 = 0)P(X_2 = 1)exp(-\lambda_{01AE} \cdot t) + \\ & + P(X_1 = 1)P(X_2 = 0)exp(-\lambda_{10AE} \cdot t) + \\ & + P(X_1 = 1)P(X_2 = 1)exp(-\lambda_{11AE} \cdot t) \end{aligned} \tag{1.9}$$

This is an average of exponential distributions weighted for the proportion of patients with covariates values $X_1 = k, k = 0, 1$ and $X_2 = l, l = 0, 1$.

Chapter 2

Standard methods to analyse AE data

The analysis of AEs could be tackled from two different points of view. Approach 1 focuses on the description of the observed occurrence of an AE as first event among the competing risks AE and relapse. In this case, treatment ability to protect from relapse has an impact on the chance of observing an AE due to the competing risks action. The more the treatment protects from relapse, the greater is the chance to experience AEs [6].

Approach 2 focuses on the treatment action in the development of AEs in patients relapse free in time. To address this issue, one should consider the occurrence of AEs as if relapse would not exclude the possibility of observing AEs related to the treatment under analysis, thus in the absence of competing risks. In this way, the comparison between the treatment impact on AEs does not depend on the effect of treatment on relapse.

In both approaches, the theoretical quantities and the relative estimators used in clinical papers to describe AEs data should have these features:

(a) the estimator addresses for the presence of right censoring;

(b) the theoretical quantity and estimator are functions of time.

In this chapter, the standard methods represented by the crude proportion of AE and the epidemiological AE rate are described and we will prove that they fail in at least one of the two features. We will then show the use of the crude incidence and of the cause-specific standard methods to be consistent with the two features for both approaches.

2.1 The crude proportion of AE

The empirical crude proportion (CP) is defined as

$$CP = \sum_{i=1}^{N} \frac{I(t = t_i; \delta_i \cdot e_i = 1)}{N}$$

where $I(\cdot)$ identifies an indicator function, and it originates from the count of patients who fail due to AEs during the entire follow-up over a total of N patients, regardless of the individual follow-up length.

The CP estimator is consistent with the first approach of analysis, since it can be thought as a naïve estimate of the probability of observing AEs over the entire follow-up. Of note AEs are counted only if observed as first events, whereas relapse acts as competing risk. However, CP, although it is calculated at the last available event time, is not a function of time in the sense that is not calculated at different time points. In addition, it does not address properly for the presence of right censoring because N is fixed and the problem of censoring is disregarded. Thus, CP fails with respect to features (a) and (b).

2.2 The crude incidence of AE

The theoretical CP can be generalized in time by the crude incidence (CI) probability, which corresponds to the absolute risk of treatment failure due to AEs up to time t. The non-parametric ML estimator of the $CI_{AE}(t)$ is given by the AJ formula reviewed in paragraph 1.2.5 and here rewritten

$$\hat{CI}_{AE}(t) = \sum_{t_j \leq t} \hat{S}(t_j-) \cdot \hat{h}_{AE}(t_j) \tag{2.1}$$

where $\hat{S}(t_j-) = \hat{P}(T > t_j-)$ is the KM non-parametric ML estimator of the proportion of patients free from treatment failure up to time t_j- and

$$\hat{h}_{AE}(t_j) = \frac{\sum_{i=1}^{N} I(t_i = t_j; \delta_{ij} \cdot e_{ij} = 1)}{n_j} \tag{2.2}$$

is the instantaneous rate of AEs, which corresponds to the proportion of patients experiencing AEs at time t over the total number of patients at risk, i.e. free from AEs and relapse (and censoring), at that time. Of note, the denominator corresponds to the person-time spent at risk in the time window $[t_j, t_j + 1)$. Formula (2.2) can be written also as $\hat{h}_{AE}(t_j) = \frac{d_{jAE}}{n_j}$ where d_{jAE} is the number of patients developing AE at time t_j.

The AJ estimator of $CI_{AE}(t)$ is consistent with the first approach of analysis, since it can be thought as an estimate of the probability of treatment failure due to AEs over the course of time, where, since AEs are counted only if observed as first events, relapse acts as competing risk. One may note in formula (2.1) the indirect protection of relapse, that lowers down $\hat{S}(t_j-)$ when relapse occurs. The $CI_{AE}(t)$ satisfies both features: it addresses for the presence of right censoring, due to the non-parametric ML estimator property, and it is a function of time.

2.3 The epidemiological AE rate

The epidemiological AE rate is defined as

$$Rate = \frac{\sum_{i=1}^{N} I(t_i = t; \delta_i \cdot e_i = 1)}{\sum_{i=1}^{N} t_i}$$

and it originates from the count of patients observed to fail due to AEs during the entire follow-up divided by the total time spent free from treatment failure, i.e. spent free from both AEs and relapse. The AE rate represents the number of observed AEs per 1 unit of person-time spent at risk.

The AE rate can be thought as an estimate of the probability of observing AEs in the next time unit for a patient that is now free from AEs and relapse (and censoring), assuming this probability as constant. If this probability cannot be reasonably assumed constant, the AE rate can be interpreted as an "average" rate over the follow-up. The AE rate is consistent with the second approach of analysis, where the focus is on treatment action in the development of AEs in patients relapse free in time. Of note, the occurrence of relapse (or of right censoring) would imply a contribution to the denominator equal to time of relapse (or of right censoring) and a null contribution to the numerator. Thus, the AE rate does not fail with respect to feature (a). The AE rate can be proved to be the parametric ML estimator of the probability of observing AEs in the next time unit for a patient free from treatment failure

$$CSH_{AE}(t) = \lim_{\Delta t \to 0^+} \frac{P(t < T \leq t + \Delta t; E = 1 | T > t)}{\Delta t} \tag{2.3}$$

assuming this probability constant in time. The quantity presented in (2.3) corresponds to the cause-specific hazard (CSH) of AEs and can be estimated through the non-parametric ML estimator in (2.2). The AE rate, however, fails with respect to feature (b), being constant in time.

2.4 The cause-specific hazard of AE

The theoretical AE rate can be generalized in time easily by relaxing the assumption of constancy in time of the $CSH_{AE}(t)$ introduced in (2.3) and thus resorting to the non-parametric estimator in (2.2). The estimator of the $CSH_{AE}(t)$ is commonly obtained by the cumulative sum derived from the Aalen-Nelson (AN) formula

$$\hat{AN}_{AE}(t) = \sum_{t_j \leq t} 1 \cdot \hat{h}_{AE}(t_j) \tag{2.4}$$

which is by definition a non-decreasing function. The $AN_{AE}(t)$ estimator is consistent with the second approach of analysis, where, unlikely in (2.1), there is no indirect protection from relapse. One may observe by comparing (2.1) and (2.4) that $\hat{S}(t_j-)$, which is lowered down in (2.1) when relapse occurs, is replaced in (2.4) by the fixed value 1.

As an alternative to the $AN_{AE}(t)$ estimator where "smoothing" is obtained by cumulative sum, one may consider the cumulative product of $(1 - \hat{h}_{AE}(t))$ terms, with the complement to 1 used to obtain again a non-decreasing function, leading to the KM formula

$$\hat{KM}_{AE}(t) = 1 - \prod_{t_j \leq t} (1 - \hat{h}_{AE}(t_j)) \tag{2.5}$$

Both estimators (2.4) and (2.5) address for the presence of right censoring (feature (a)) since it is addressed in $\hat{h}_{AE}(t_j)$, and they are functions of time (feature (b)).

2.5 Issues on the meaning of AN and KM estimators

At first glance, the $AN_{AE}(t)$ and $KM_{AE}(t)$ curves could be interpreted only in terms of treatment action in determining AE occurrence as first event, regardless of the impact of relapse. This interpretation comes natural since $h_{AE}(t)$ is related only to the velocity of development of AEs in time unlikely $CI_{AE}(t)$, where $S(t_j-)$ in (2.1) is influenced by the development of relapse.

One may note, however, that the occurrence of relapse may "select" not at random patients excluded from the sub-sample of patients on which the instantaneous rate of AEs is calculated in (2.2). This issue on patients selection may influence the interpretation of $CSH_{AE}(t)$ and subsequently of the chosen smoother. The $CSH_{AE}(t)$ may not capture entirely the mechanism of treatment action in terms of AEs on AEs free patients represented by

$$H_{AE}(t) = P(t < T_{AE} \leq t + 1 | T_{AE} > t) \tag{2.6}$$

unless there is independence between T_{AE} and T_{RL}. In this case, the sub-sample of patients with $T > t$ in (2.3) is a random sample of patients with $T_{AE} > t$ in (2.6) and the interpretations of $AN_{AE}(t)$ and $KM_{AE}(t)$ hold. This independence, however, can be rarely assumed and cannot be tested.

Chapter 3

Hazard of AE: estimation and smoothing

To relax the assumption of independence between the two potential times T_{AE} and T_{RL} one possibility is to estimate the hazard of AE in strata defined by observed covariates assuming only within each stratum independence between T_{AE} and the additional selection due to T_{RL}. This approach can be carried out by averaging stratum estimates obtained either non parametrically or by the use of the Cox model leading to a weighted average survival probability. An alternative method is addressing the presence of selection due to relapse through inverse probability of censoring weighting (IPCW) [19, 20]. This method aims at creating a pseudo-population that is similar to the one observable in the absence of relapse by adding a weight to patients who do not develop relapse. On this pseudo-population a survival probability is then calculated.

3.1 Weighted average survival probability

The survival probability for each level of the observed covariates can be estimated through the KM estimator within each stratum or by a Cox PH model including these covariates among regressors. The overall average survival is determined by weighting the survival probabilities in each level of the covariates by the proportion of subjects having that covariate level. If both covariates X_1 and X_2 are observed, the weighted average

survival is calculated as

$$
\begin{aligned}
S_{AE}(t) = & \frac{\sum_{i=1}^{N} I(X_{1i}=0, X_{2i}=0)}{n} S_{AE}^{00}(t) + \\
& + \frac{\sum_{i=1}^{n} I(X_{1i}=0, X_{2i}=1)}{n} S_{AE}^{01}(t) + \\
& + \frac{\sum_{i=1}^{n} I(X_{1i}=1, X_{2i}=0)}{n} S_{AE}^{10}(t) + \\
& + \frac{\sum_{i=1}^{n} I(X_{1i}=1, X_{2i}=1)}{n} S_{AE}^{11}(t)
\end{aligned}
\tag{3.1}
$$

where $S_{AE}^{kl}(t)$ indicates the survival probability at time t obtained through the KM estimator or the Cox model for a patient having $X_1 = k$, $k = 0, 1$, and $X_2 = l$, $l = 0, 1$.

3.2 IPCW estimator

The unitary contribution of a subject i in the count of subjects at risk of experiencing an AE at time t is replaced by the unstabilized weight $\hat{w}_i^u(t) = \frac{1}{\hat{S}_{RL}^X(t|X_1,X_2)}$ that is inversely proportional to the estimate of the conditional probability of being relapse free (i.e. of remaining uncensored) $\hat{S}_{RL}^X(t|X_1, X_2)$ until time t. The lower is the probability of being relapse free, the greater are the weights. The estimate $\hat{S}_{RL}^X(t|X_1, X_2)$ can be based on the KM estimator within each level of the observed covariates or on the fit of a Cox PH model for relapse in which all prognostic factors for AE and for relapse are entered as covariates [21]. Once the weights are calculated, one can estimate the survival probability for time to AE in the absence of relapse using the KM estimator [22].

In detail, the use of the IPCW estimator can be summarized in the steps below:

1. Fit a model for the censoring mechanism (i.e. relapse) including all the covariates that could have an impact on the times of AE and of relapse. This type of model is required in order to assess how long a subject stays in the study without experiencing relapse. This means using the KM estimator or implementing a Cox PH model considering the development of relapse as the event of interest (and so subjects who experience an AE are now treated as censored observations).

2. Estimate, parametrically or semi-parametrically, the probability of remaining relapse free at each observed time point t, denoted by $\hat{S}_{RL}^X(t|X_1, X_2)$, for all subjects at risk at the same time t. In addition, also the probability of being relapse free at time t independently from covariates $\hat{S}_{RL}^X(t)$ is estimated.

3. Calculate the IPCW unstabilized weights $\hat{w}_i^u(t) = \frac{1}{\hat{S}_{RL}^X(t|X_1,X_2)}$ for each subject. In case of heavy censoring, these weights become very large, so one can calculate the stabilized weights substituting the numerator with the probability of being relapse free at time t independently from covariates, $\hat{w}_i^s(t) = \frac{\hat{S}_{RL}^X(t)}{\hat{S}_{RL}^X(t|X_1,X_2)}$.

4. Estimate the survival probabilities for time to AE in the absence of relapse weighting subjects according to the IPCW methodology at each observed time point of interest.

Chapter 4

Data examples

In this chapter we present the standard methods commonly used to analyse AEs applied to a motivating example on osteonecrosis in childhood acute lymphoblastic leukaemia and to a simulated dataset. In addition, we applied also the weighted average survival probability and the IPTW methods in order to account for the possible dependence between the potential times of the competing events.

4.1 Motivating data

The motivating example of this thesis was the evaluation of the occurrence of osteonecrosis (ON), a relatively rare disabling complication related to the administration of a front line intensive chemotherapy treatment in children newly diagnosed with acute lymphoblastic leukaemia (ALL). The aim was to assess this complication as related to the front line treatment, thus before death or the development of a relapse during or after the end of this treatment. We show the application of the standard methods commonly used to analyse AEs data to children with ALL enrolled in two subsequent multicenter clinical trials conducted in Italy with the Italian Association of Pediatrical Hematology and Oncology (AIEOP) [23]. In this study, ON is an AE of the treatment protocol administered to children with ALL and acts as competing risk for death (not due to relapse) or the first relapse (which is followed by another type of treatment).

We analysed data on 3668 children aged 1-17 years at diagnosis of ALL and of those, 87 experienced an ON during or after the end of the front line treatment (ON after transplant or relapse were excluded) while 715 children relapsed or died.

The crude proportion of ON is $\frac{87}{3668} = 0.023$, meaning that 2.3% of the study population developed an AE before relapse or death. In Figure 4.1 panel a) the crude incidence of ON calculated through the AJ estimator is displayed. The CI probability of failure due to an ON is lower that 2.5% after 4 years from diagnosis of ALL.

The epidemiological AE rate, calculated as the number of subjects experiencing an ON

over the total time at risk of developing an ON, a relapse or death, is $Rate = \frac{87}{20517.03} = 0.004$, meaning that 4 ONs per 1000 person-years occur. The estimates of the $CSH_{ON}(t)$ obtained through the AN and KM smoothed estimators are displayed in Figure 4.1 panel b). Multiplying the AN estimator by 100, at 4 years from diagnosis, the expected number of ONs in 100 hypothetical children is 2.5.

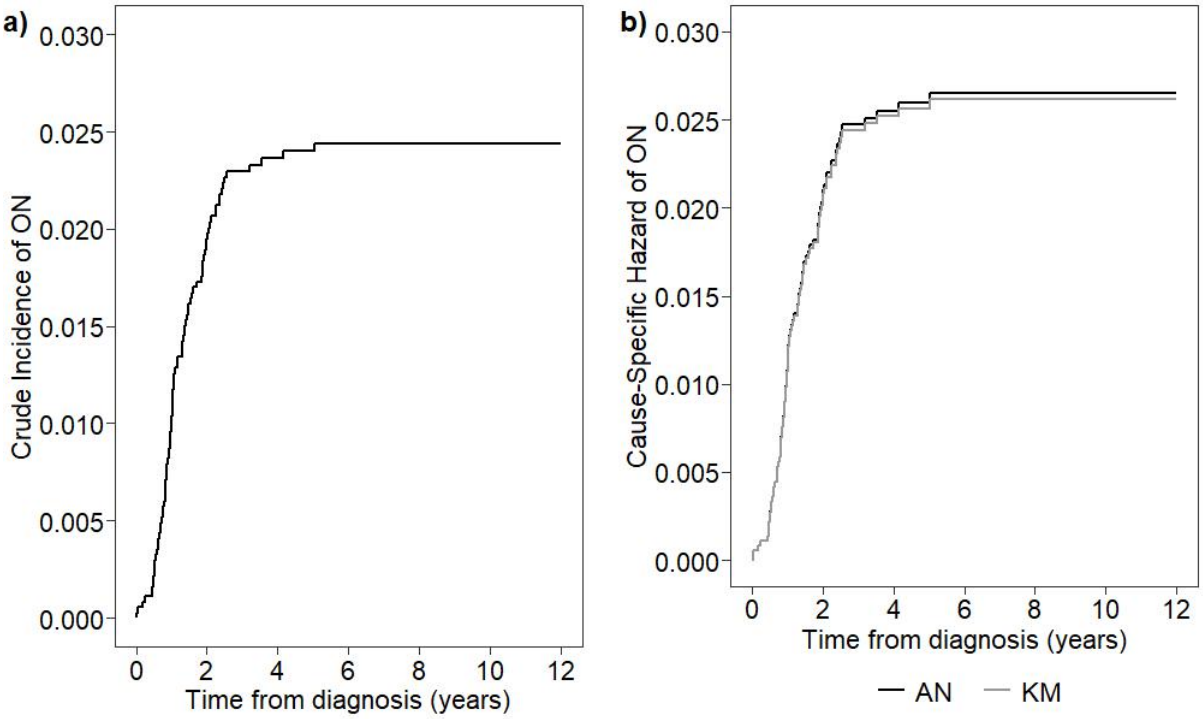

Figure 4.1: a) $CI_{ON}(t)$ estimated through the Aalen-Johansen formula; b) $AN_{ON}(t)$ and $KM_{ON}(t)$ smoothed curves of the $CSH_{ON}(t)$.

In the dataset two covariates are of relevance: age at diagnosis of ALL, since incidence of ON is higher with higher age, and risk group, which is the stratification of the children, based on genetic features and cytological/molecular early response to treatment and defines the intensity of treatment administered (the higher the risk, the higher the intensity). In order to correctly account fo the presence of a dependence between the potential time of ON and the potential time to relapse or death we implemented the methods proposed in Chapter 3, including first only risk group as covariate and then adding also age at diagnosis (> 10 versus ≤ 10 years).

In Figure 4.2 the estimates of the survival probability obtained with different methods are displayed. One can see that the naïve KM estimator and the weighted average method considering only risk group as covariate give the same estimates of the survival probability. Including also age at diagnosis as covariate, a similar but lower survival probability is obtained. The distance between the two groups of curves (naïve and curves accounting for 1 covariate and curves accounting for 2 covariates) suggests that the inclusion of the second covariate removes part of the dependence. However, overall the inclusion of the two

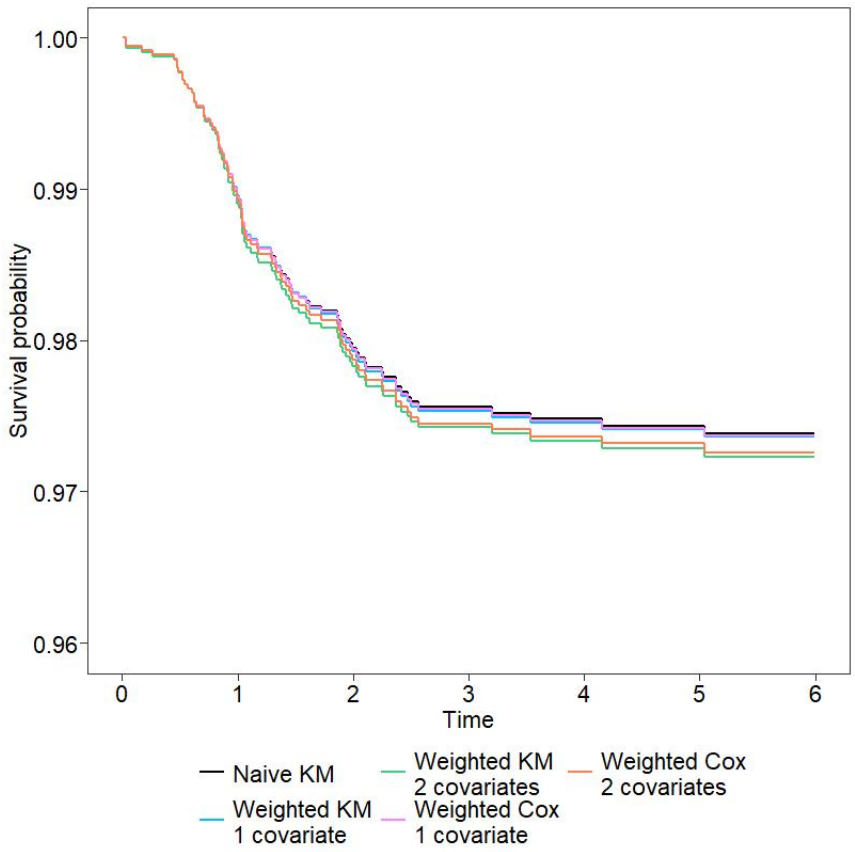

Figure 4.2: Survival probability estimates obtained with the naïve KM estimator, the weighted KM estimator stratifying only for 1 covariate (risk group) or for 2 covariates (risk group and age at diagnosis) and the weighted Cox model including only risk group or both risk group and age at diagnosis as covariates.

covariates does not change remarkably the survival probability from the naïve estimator. This could be due to the low ability of the covariates risk group and age at diagnosis to remove the dependence between the potential time of ON and the potential time of relapse or death. Yet another explanation could be that the two potential times do not have a strong dependence as indeed the pathways to relapse or death is not necessarily related to the determinants of ON.

4.2 Simulated data

The standard methods commonly used to analyse AEs are also calculated on a simulated example dataset of $N = 300$ subjects. The potential times T_{AE} and T_{RL} are simulated from exponential distributions with parameters depending on two binary covariates X_1 and X_2, with $P(X_1 = 1) = 0.3$ and $P(X_2 = 1) = 0.4$. The combination of X_1 and X_2 identifies a different hazard profile in patients experiencing an AE or a relapse:

- if $X_1 = 0$ and $X_2 = 0$, $T_{AE} \sim Exp(1)$ and $T_{RL} \sim Exp(2)$

- if $X_1 = 0$ and $X_2 = 1$, $T_{AE} \sim Exp(3)$ and $T_{RL} \sim Exp(6)$
- if $X_1 = 1$ and $X_2 = 0$, $T_{AE} \sim Exp(3)$ and $T_{RL} \sim Exp(5)$
- if $X_1 = 1$ and $X_2 = 1$, $T_{AE} \sim Exp(9)$ and $T_{RL} \sim Exp(15)$

One may note that, fixed $X_1 = 0$ (or $X_1 = 1$), if X_2 changes, both the hazard of AE and the hazard of relapse triple. For sake of simplicity, we did not consider the presence of censoring despite we already explained the influence of right censoring on each standard method. The distribution of the potential times T_{AE} and T_{RL} is presented in Figure 4.3 panel a). At first glance, the dependence between the two times is not evident. One can notice, however, that the times of patients with both covariates equal to 0 are systematically greater (median time T_{AE} equal to 0.7 and median time T_{RL} equal 0.4) than the remaining points. Similarly, for the other points, median values for T_{AE} and T_{RL} are: 0.2 and 0.1 ($X_1 = 0$, $X_2 = 1$), 0.2 and 0.2 ($X_1 = 1$, $X_2 = 0$), 0.1 and 0.0 (both covariates equal to 1). The correlation between times T_{AE} and T_{RL} is equal to 0.23. This moderate value is due to the absence of correlation within the four groups of patients identified by the covariates. However, the correlation between median times T_{AE} and T_{RL} within the four groups becomes equal to 0.94. The distribution of the failure time T, calculated as the minimum value between T_{AE} and T_{RL}, is displayed in Figure 4.3 panel b). Of note, the mean failure time is equal to 0.19.

In Table 4.1 quantities needed to analyse AE data with the standard methods are displayed. In particular, the distinct failure times t_j, the number of patients at risk n_j at each time point, the number of subjects experiencing an AE or a relapse as first event (d_{jAE} and d_{jRL} respectively), the survival probability $\hat{S}(t_j)$ estimated through the KM estimator and the instantaneous rate of AE $\hat{h}_{AE}(t_j)$ in the simulated sample of 300 patients are summarised.

In this dataset, 80 subjects develop an AE and 220 fail due to relapse. Then, the crude proportion of AE is $CP = \frac{80}{300} = 0.27$, meaning that 27% of the study population develop an AE. In Table 4.1, at each time t_j, the quantities needed to estimate the crude incidence $\hat{CI}_{AE}(t)$ through the AJ estimator are summarised and in Figure 4.4 panel a) the graph is shown. For example, the highest CI probability of failure due to AE 0.263 is reached at the last observed failure time 2.1. This means that the absolute risk of developing an AE up to time 2.1 is 26.3% when allowing for the presence of the competing risk relapse.

The epidemiological AE rate, calculated as the number of subjects developing an AE over the total time at risk of developing both an AE or a relapse, is $Rate = \frac{80}{56.4} = 1.42$ with the total time at the denominator derived from the sum of $t_j * d_{jAE}$ and $t_j * d_{jRL}$ in Table 4.1. In the table also the values of the estimates obtained with both the AN e KM smoothed estimators of the $CSH_{AE}(t)$ are reported and their curves are displayed in Figure 4.4 panel b). One can see that in the time interval $[0, 1.3]$, considering the

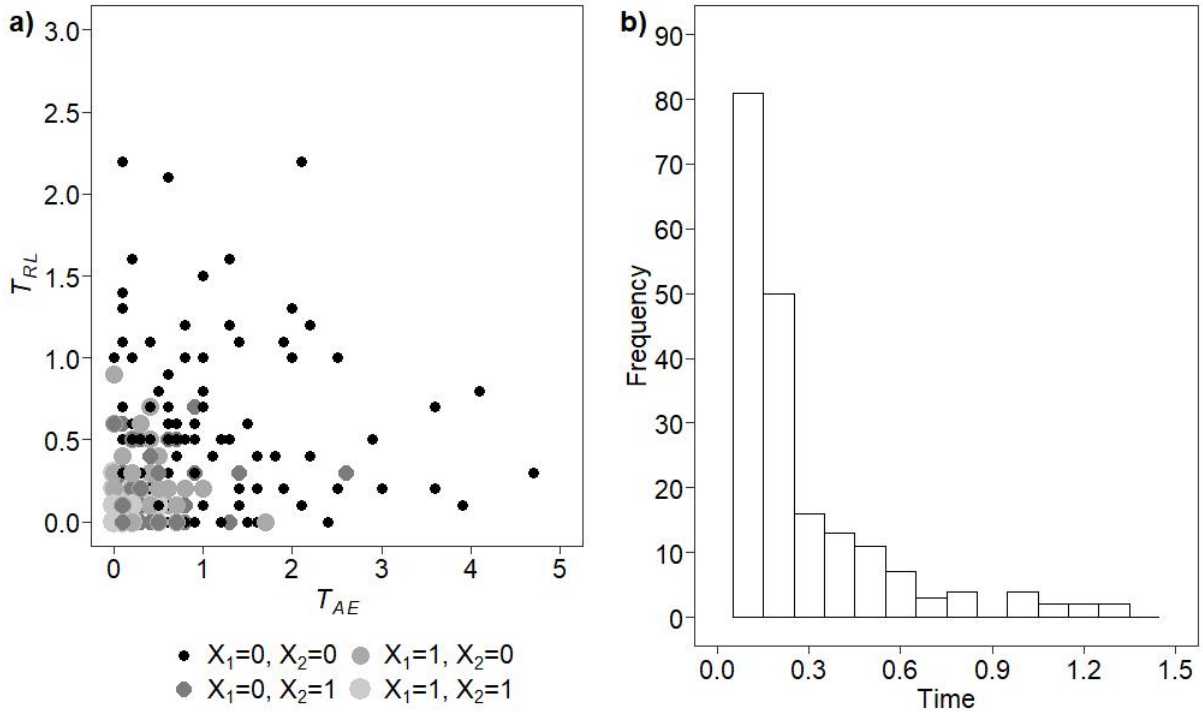

Figure 4.3: a) Scatterplot of T_{AE} and T_{RL} times according to the four groups identified by the binary covariates X_1 and X_2 and b) histogram of the distribution of the failure times calculated as the minimum value between T_{AE} and T_{RL}, in the simulated dataset of $N = 300$ subjects.

AN estimator, a subject at risk is expected to develop 1.2 AEs. Considering the KM estimator, at time 1.3 in an hypothetical sample of 100 subjects 73 AEs are expected to happen.

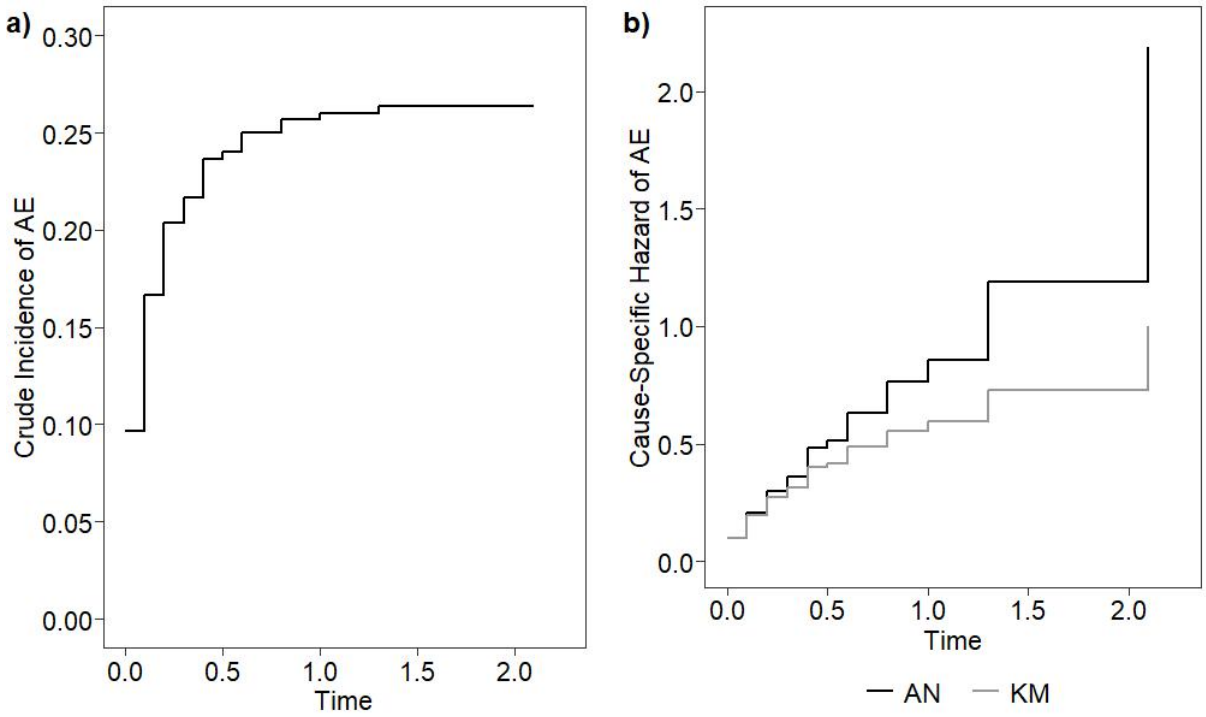

Figure 4.4: a) $CI_{AE}(t)$ estimated through the Aalen-Johansen formula; b) $AN_{AE}(t)$ and $KM_{AE}(t)$ smoothed curves of the $CSH_{AE}(t)$.

Table 4.1: Example

j	t_j	n_j	d_{jAE}	d_{jRL}	$\hat{S}(t_j)$	$\hat{h}_{AE}(t_j)$	$\hat{h}_{AE}(t_j)\hat{S}(t_j-)$	$\hat{CI}_{AE}(t_j)$	$\hat{AN}_{AE}(t_j)$	$\hat{KM}_{AE}(t_j)$
1	0.0	300	29	75	0.653	0.097	0.097	0.097	0.097	0.097
2	0.1	196	21	60	0.383	0.107	0.070	0.167	0.204	0.194
3	0.2	115	11	39	0.217	0.096	0.037	0.204	0.300	0.271
4	0.3	65	4	12	0.163	0.062	0.013	0.217	0.362	0.316
5	0.4	49	6	7	0.120	0.122	0.020	0.237	0.484	0.400
6	0.5	36	1	10	0.083	0.028	0.003	0.240	0.512	0.416
7	0.6	25	3	4	0.060	0.120	0.010	0.250	0.632	0.486
8	0.7	18	0	3	0.050	0.000	0.000	0.250	0.632	0.486
9	0.8	15	2	2	0.037	0.133	0.007	0.257	0.765	0.555
10	1.0	11	1	3	0.023	0.091	0.003	0.260	0.856	0.595
11	1.1	7	0	2	0.017	0.000	0.000	0.260	0.856	0.595
12	1.2	5	0	2	0.010	0.000	0.000	0.260	0.856	0.595
13	1.3	3	1	1	0.003	0.333	0.003	0.263	1.189	0.730
14	2.1	1	1	0	0.000	1.000	0.000	0.263	2.189	1.000

t_j are the distinct failure times; n_j is the number of patients at risk at time t_j; d_{jAE} and d_{jRL} are the number of patients developing AE or RL at time t_j, respectively; $\hat{S}(t_j)$ is the survival function at time t_j, estimated through the Kaplan-Meier estimator; $\hat{h}_{AE}(t_j)$ corresponds to the instantaneous rate of AE and it is the non-parametric ML estimator of the $CSH_{AE}(t)$; $\hat{h}_{AE}(t_j)\hat{S}(t_j-)$ is the product of the KM estimator of the proportion of patients free from treatment failure up to time t_j- and the instantaneous rate of AE; $\hat{CI}_{AE}(t_j)$ is the AJ estimator of $CI_{AE}(t)$ of AE; $\hat{AN}_{AE}(t_j)$ and $\hat{KM}_{AE}(t_j)$ are the AN and KM smoothed estimates of the $CSH_{AE}(t)$.

To enlighten the problems related to the interpretation of the $AN_{AE}(t)$ and $KM_{AE}(t)$ smoothed curves, we simulated different datasets according to the parameters specified at the beginning of this chapter with increasing sample size from $N = 100$ to $N = 1000$ and we compared these quantities with the theoretical cumulative hazard function of T_{AE} and with one minus the survival function respectively. The theoretical cumulative hazard function can be calculated as $-\log(S_{AE}(t))$, where the theoretical survival function $S_{AE}(t)$ is calculated following formula (1.9). In Figure 4.5 panels a) and b) the $AN_{AE}(t)$ and $KM_{AE}(t)$ smoothed curves calculated in the simulated datasets are displayed. One can notice that, independently from the sample size of the dataset and from the low correlation between T_{AE} and T_{RL}, the $AN_{AE}(t)$ and $KM_{AE}(t)$ smoothed curves do not fit the theoretical quantities, i.e. the cumulative hazard and one minus the survival function respectively.

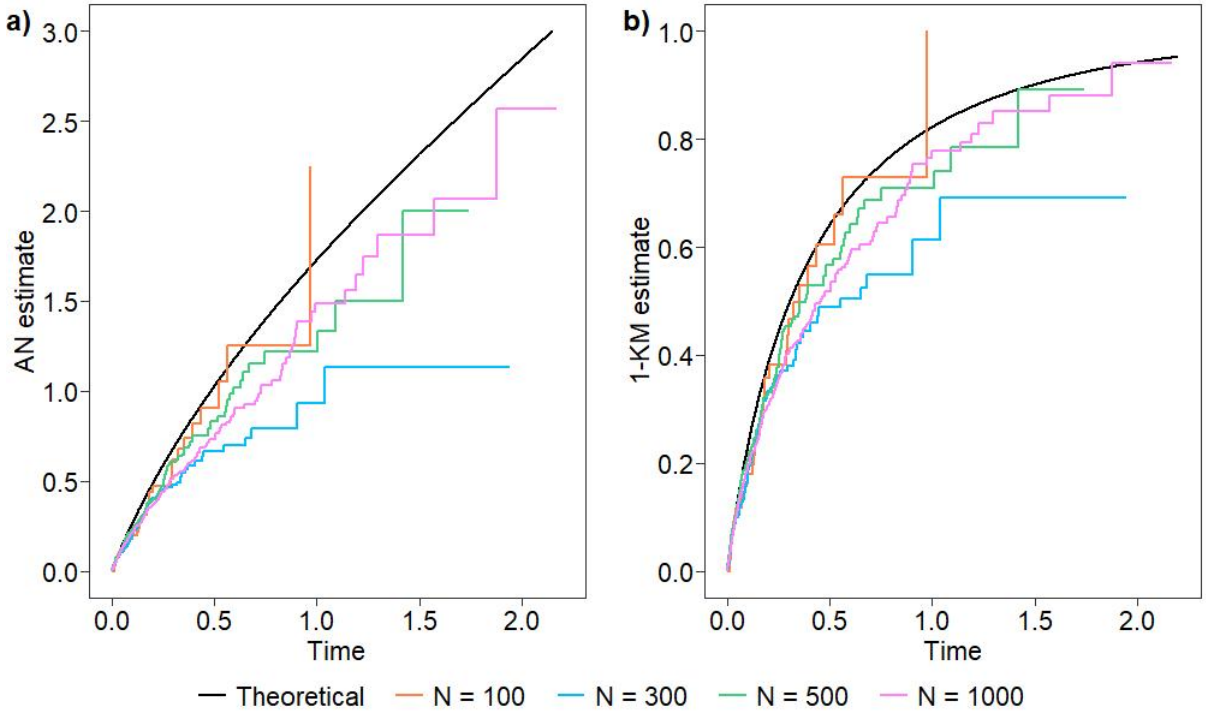

Figure 4.5: a) $AN_{AE}(t)$ and b) 1-$KM_{AE}(t)$ calculated for different sample sizes.

The weighted average survival obtained with the KM estimator and from the Cox model and the IPCW method proposed in Chapter 3 were calculated on the simulated dataset of 300 subjects. The survival curves obtained adjusting for both binary covariates X_1 and X_2 are displayed in Figure 4.6. One can observe that, comparing the estimates with the theoretical survival function, for small times the three methods perform well, whereas as the time increases the weighted average obtained with the Cox model (green line) gives the less biased estimate.

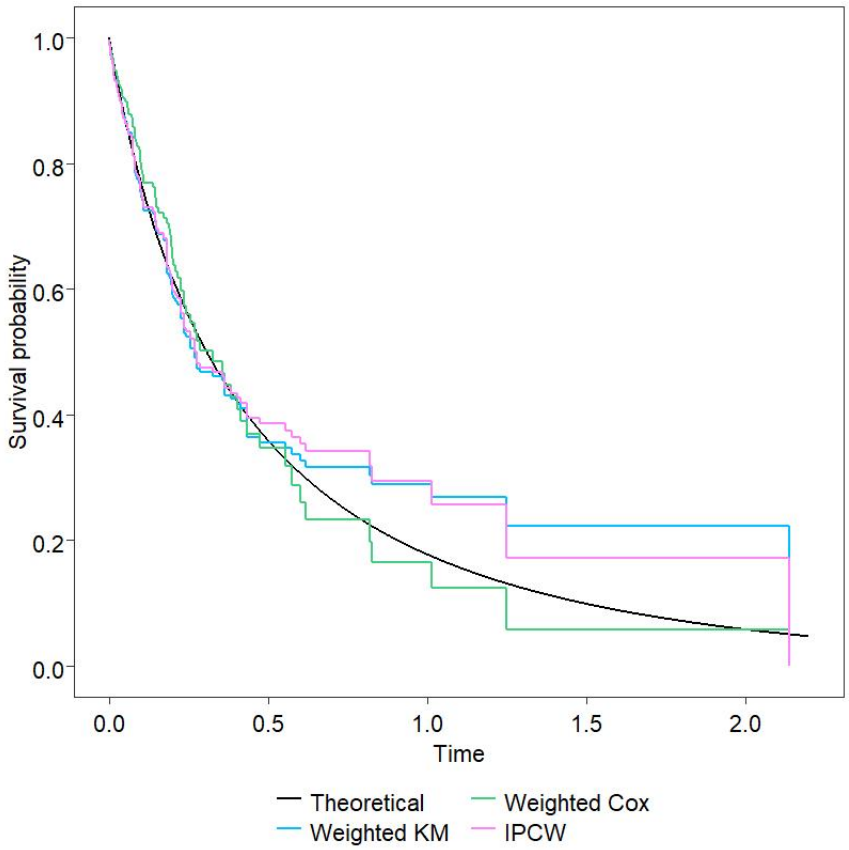

Figure 4.6: Survival probability estimates obtained with the weighted KM estimator, the weighted Cox model and the IPCW method. The black line is the theoretical survival function.

Chapter 5

Simulation protocol

5.1 Data generation

We generated 1000 datasets of $N = 300$ observations, using the inversion method proposed by Bender et al. [24]. First, we simulated two independent binary covariates X_1 and X_2 from Bernoulli distributions with probabilities of success $P(X_1 = 1) = 0.3$ and $P(X_2 = 1) = 0.4$. Then we generated the potential times T_{AE} and T_{RL} from exponential distributions with parameters depending on the two binary covariates. Finally, the observable failure time T was determined as the minimum of the potential times T_{RL} and T_{AE}. Of note, since in this thesis we are interested in the basic data structure and statistical techniques accounting for it rather then to clinical questions, we decided not to consider the presence of censoring. For the same reason, we simulated our data following constant event-specific hazards models [7].

In detail, we simulated 4 different scenarios varying the parameters of the exponential distribution from which the potential time T_{RL} is generated (summarised in Table 5.1):

- scenario 1 - the hazard of relapse is independent from the covariates values (it is always equal to 2)
- scenario 2 - only X_1 has an impact on the hazard of relapse (it changes from 2 when $X_1 = 0$ to 5 when $X_1 = 1$)
- scenario 3 - the hazard of relapse depends on X_1 and, when $X_1 = 0$, on X_2 (in this case the hazard of relapse for subjects with $X_2 = 1$ is tripled). In this scenario also an interaction between the two covariates has an impact on the hazard of relapse
- scenario 4 - fixed $X_1 = 0$ (or $X_1 = 1$), if X_2 changes, the hazard of relapse triples. In this case, X_2 generates a dependence between the potential times T_{RL} and T_{AE}.

In Figure 5.1 the theoretical survival functions of AE and of relapse, calculated according to formula (1.9) in the four scenarios are displayed. Of note, in all scenarios fixed

Table 5.1: Parameters of the exponential distributions from which the times of AE and of relapse are generated

Scenario	AE				Relapse			
	λ_{00AE}	λ_{01AE}	λ_{10AE}	λ_{11AE}	λ_{00RL}	λ_{01RL}	λ_{10RL}	λ_{11RL}
1	1	3	3	9	2	2	2	2
2	1	3	3	9	2	2	5	5
3	1	3	3	9	2	6	5	5
4	1	3	3	9	2	6	5	15

λ_{ijAE} and λ_{ijRL} are the parameters of the exponential distributions of T_{AE} and T_{RL}, respectively, when $X_1 = i$, $i = 0, 1$, and $X_2 = j$, $j = 0, 1$.

$X_1 = 0$ (or $X_1 = 1$), if X_2 changes, the hazard of AE triples. As a consequence, the survival function of AE is the same in all scenarios.

5.2 Estimated quantities

For each scenario, we calculated the true value of the survival probability in an hypothetical world where relapse is absent following formula (1.9) at two fixed time-points ($t = 0.2$ and $t = 0.3$), chosen in order to have a sufficient number of events (at least 60 events). In addition, we estimated the expected number of subjects at risk of developing an event (AE or relapse) at these time-points for an hypothetical sample of $N = 300$ patients as

$$\begin{aligned} & N \cdot P(T_{RL} > t, T_{AE} > t) = \\ = N \cdot [& P(X_1 = 0)P(X_2 = 0)exp(-\lambda_{00RL} \cdot t)exp(-\lambda_{00AE} \cdot t) + \\ & + P(X_1 = 0)P(X_2 = 1)exp(-\lambda_{01RL} \cdot t)exp(-\lambda_{01AE} \cdot t) + \\ & + P(X_1 = 1)P(X_2 = 0)exp(-\lambda_{10RL} \cdot t)exp(-\lambda_{10AE} \cdot t) + \\ & + P(X_1 = 1)P(X_2 = 1)exp(-\lambda_{11RL} \cdot t)exp(-\lambda_{11AE} \cdot t)] \end{aligned}$$

where the calculation of $P(T_{RL} > t, T_{AE} > t)$ in each stratum is obtained by the product of $P(T_{RL} > t)$ and $P(T_{AE} > t)$ given the conditional independence given the covariates.

In order to compare the methods that can be used to analyse AE data, for each dataset we estimated different quantities:

1. the naïve $KM_{AE}(t)$ survival estimate of the incidence of treatment failure only due to AE, as if relapse was removed without accounting for covariates (following formula (2.5))

2.1. the weighted average survival probability, obtained through the KM estimator in strata defined according to the observed covariate X_1, thus considering unobserved

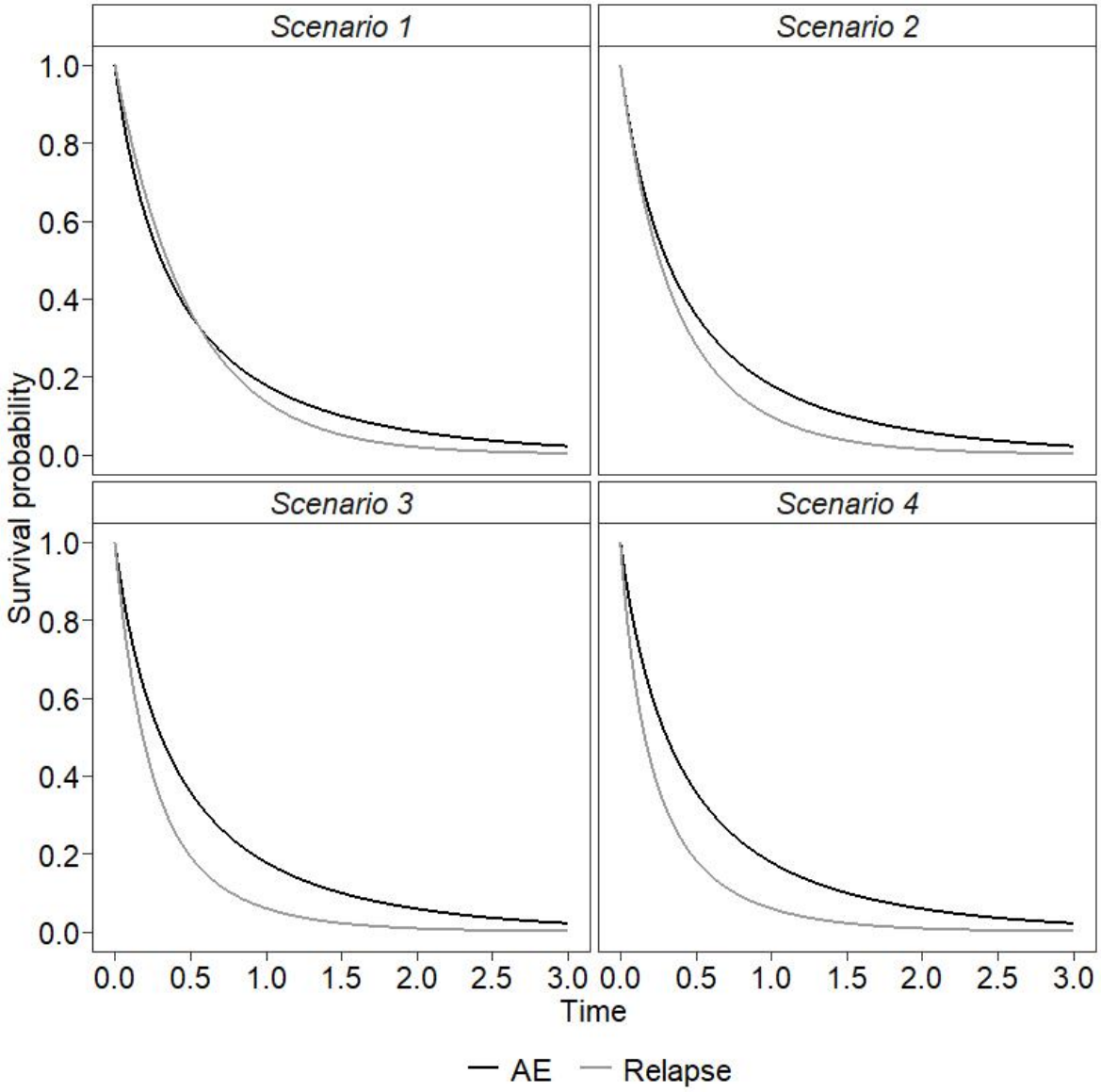

Figure 5.1: Theoretical survival functions of AE and of relapse in the four scenarios.

covariate X_2

2.2. the weighted average survival probability, obtained through the KM estimator in strata defined according to the observed covariates X_1 and X_2

2.3. the weighted average survival probability, obtained through the Cox model with only the observed covariate X_1, thus considering unobserved covariate X_2

2.4. the weighted average survival probability, obtained through the Cox model with the observed covariates X_1 and X_2

3.1. the KM survival probability on the pseudo-population obtained by IPCW estimator (Cox based) where weights are estimated according to the observed covariate X_1,

thus considering unobserved covariate X_2

3.2. the KM survival probability on the pseudo-population obtained by IPCW estimator (Cox based) where weights are estimated according to the observed covariates X_1 and X_2.

5.3 Extensions of the simulation protocol

5.3.1 Change in the imbalance of the covariates

We defined a new simulation scenario varying the imbalance of the covariates. First, we set the parameters of the Bernoulli distributions from which the binary covariates X_1 and X_2 are generated in order to construct the "worst" situation that may happen when analysing real data (named case A), that is both covariates are very frequent in the population under study. To do so, we changed the Bernoulli parameters from the original $P(X_1 = 1) = 0.3$ and $P(X_2 = 1) = 0.4$ to $P(X_1 = 1) = P(X_2 = 1) = 0.5$, meaning that 50% of the population have both covariates. Then, we set the Bernoulli parameters in order to obtain the "best" scenario (named case B), that is at least one covariate (here X_2) is rare in the study population. In this case, we kept fixed the prevalence of the first covariate $P(X_1 = 1) = 0.3$ and we changed that of the second one to $P(X_2 = 1) = 0.1$. In Figure 5.2 the theoretical survival probabilities of AE and relapse in the two scenarios are displayed.

5.3.2 Change in the hazard ratio of relapse

We defined another new simulation scenario changing the hazard ratio of relapse. We changed the parameters of the exponential distributions from which the hazard of relapse was generated in order to reduce the impact of the competing event (Table 5.2). First we set the parameters in order to have, fixed $X_1 = 0$ (or $X_1 = 1$), when X_2 changes, an increase of 2 times of the hazard of relapse (case A) and then an increase of 1.5 times (case B). In Figure 5.3 the theoretical survival probabilities of AE and relapse in the two scenarios are displayed.

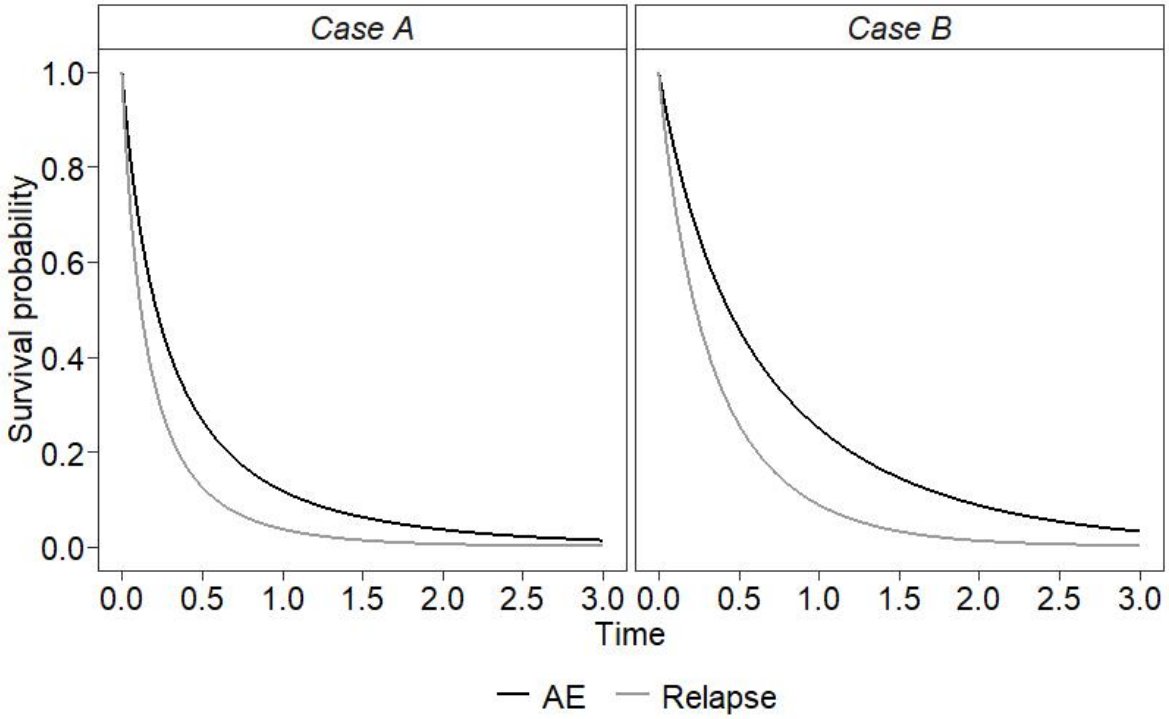

Figure 5.2: Theoretical survival functions of AE and of relapse in the scenarios obtained changing the imbalance of the covariates. In case A $P(X_1 = 1) = P(X_2 = 1) = 0.5$, in case B $P(X_1 = 1) = 0.3$ and $P(X_2 = 1) = 0.1$.

Table 5.2: Parameters of the exponential distributions from which the times of AE and of relapse are generated changing the hazard ratio of relapse

	AE				Relapse			
	λ_{00AE}	λ_{01AE}	λ_{10AE}	λ_{11AE}	λ_{00RL}	λ_{01RL}	λ_{10RL}	λ_{11RL}
Case A	1	3	3	9	2	4	5	10
Case B	1	3	3	9	2	3	5	7.5

λ_{ijAE} and λ_{ijRL} are the parameters of the exponential distributions of T_{AE} and T_{RL}, respectively, when $X_1 = i$, $i = 0, 1$, and $X_2 = j$, $j = 0, 1$.

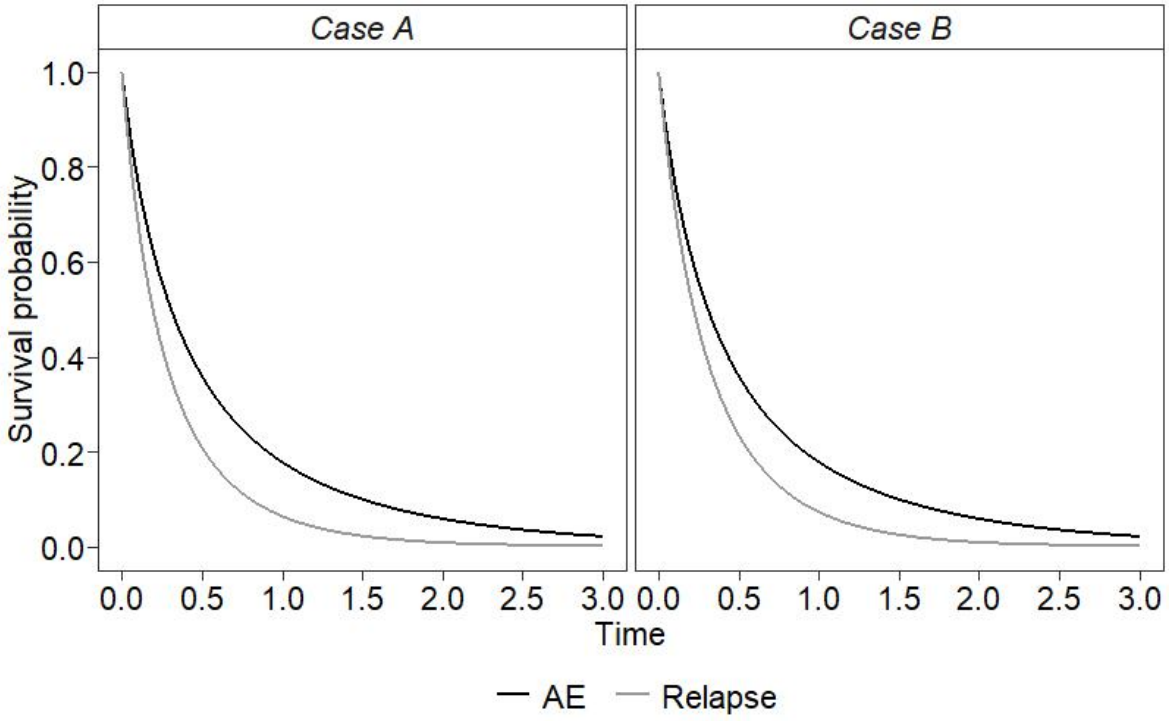

Figure 5.3: Theoretical survival functions of AE and of relapse in the scenarios obtained changing the imbalance of the covariates. In case A $\lambda_{00RL} = 2$, $\lambda_{01RL} = 4$, $\lambda_{10RL} = 5$, $\lambda_{11RL} = 10$, in case B $\lambda_{00RL} = 2$, $\lambda_{01RL} = 3$, $\lambda_{10RL} = 5$, $\lambda_{11RL} = 7.5$.

Chapter 6

Simulation results

Simulation results in the four scenarios are displayed in Figure 6.1. At each time-point the expected number of subjects at risk of developing an event is displayed and the distance between the estimate and the theoretical survival probability (bias) is represented in a boxplot for each of the 7 estimators. In addition, the theoretical survival probability, the mean of the estimates obtained with the different methods, the bias from the theoretical survival probability and the variance of the estimates are summarised in Tables 6.1 (scenarios 1 and 2) and 6.2 (scenario 3 and 4).

In scenario 1, where the hazard of relapse is independent from the covariates values, the results of all methods are similar: the median value of the difference of the survival estimated through each method and the theoretical survival is equal to 0. Of note, even the estimate of the survival obtained with the naïve KM estimator censoring relapsed subjects is also unbiased.

In scenario 2, where only X_1 has an impact on the hazard of relapse, the naïve KM estimator gives a biased survival probability. With the other methods unbiased estimates are obtained. Of note, the methods in which only the X_1 covariate is considered perform better with respect to those in which both covariates are included. This is due to the fact that X_2 does not have an impact on relapse.

In scenario 3 the hazard of relapse depends on X_1 and, when $X_1 = 0$, also on X_2. Methods in which only X_1 is included give biased estimates. Methods with X_1 and X_2 covariates give unbiased estimates with the exception of the IPCW estimator, where there is an underestimation of the survival probability. This is due to the fact that in this scenario also an interaction between the two covariates is present: X_2 has an impact on the hazard of relapse only when $X_1 = 0$. However this interaction is not accounted for in the model to estimate weights. To corroborate this result, Figure 6.2 shows the results of IPCW and weighted Cox approaches when X_1, X_2 and their interaction are considered. The reader may observe that the inclusion of the interaction overcomes the bias in the IPCW method, while it is not needed in the Cox model since, conditional on

X_1 and X_2, this model requires an assumption of independence between T_{AE} and T_{RL} that is present even if there is an interaction between X_1 and X_2. In this regard, the Cox model is robust also in the presence of an interaction between the covariates on the hazard of relapse.

In the last scenario displayed in Figure 6.1, where both X_1 and X_2 have an impact on the hazard of relapse, all methods including one covariate only give similar biased results. However, the distance between the estimated and the theoretical survival probabilities is lower than that obtained from the naïve KM estimator. The estimates from the weighted KM or the Cox model and from the IPCW are unbiased when the methods account for the presence of all covariates that have an impact on the hazard of relapse. Of note, the estimates from the IPCW have a greater variability with respect to the others (Table 6.2).

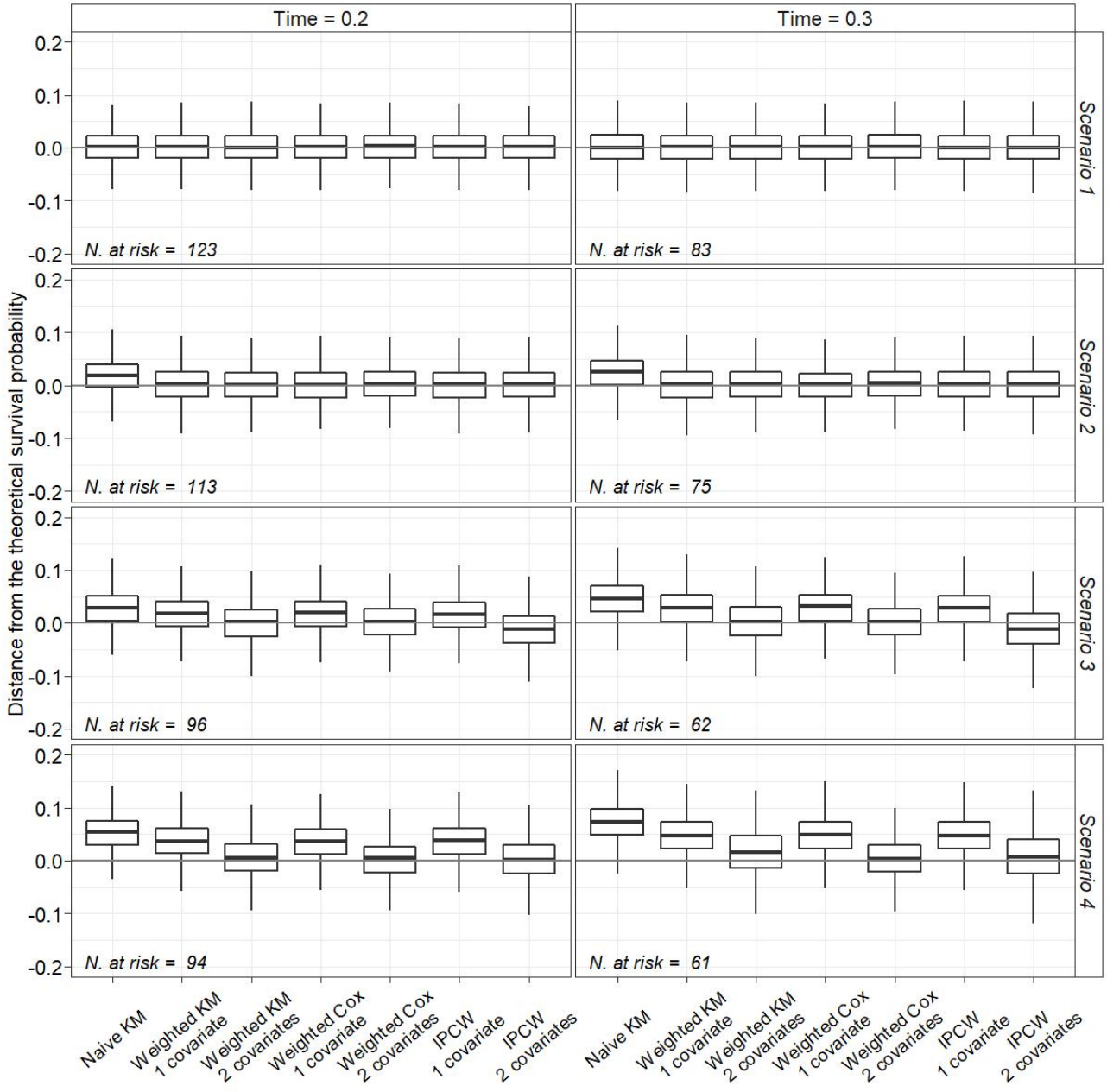

Figure 6.1: Simulation results in all scenarios at times $t = 0.2$ and $t = 0.3$. The grey horizontal line is the reference null bias.

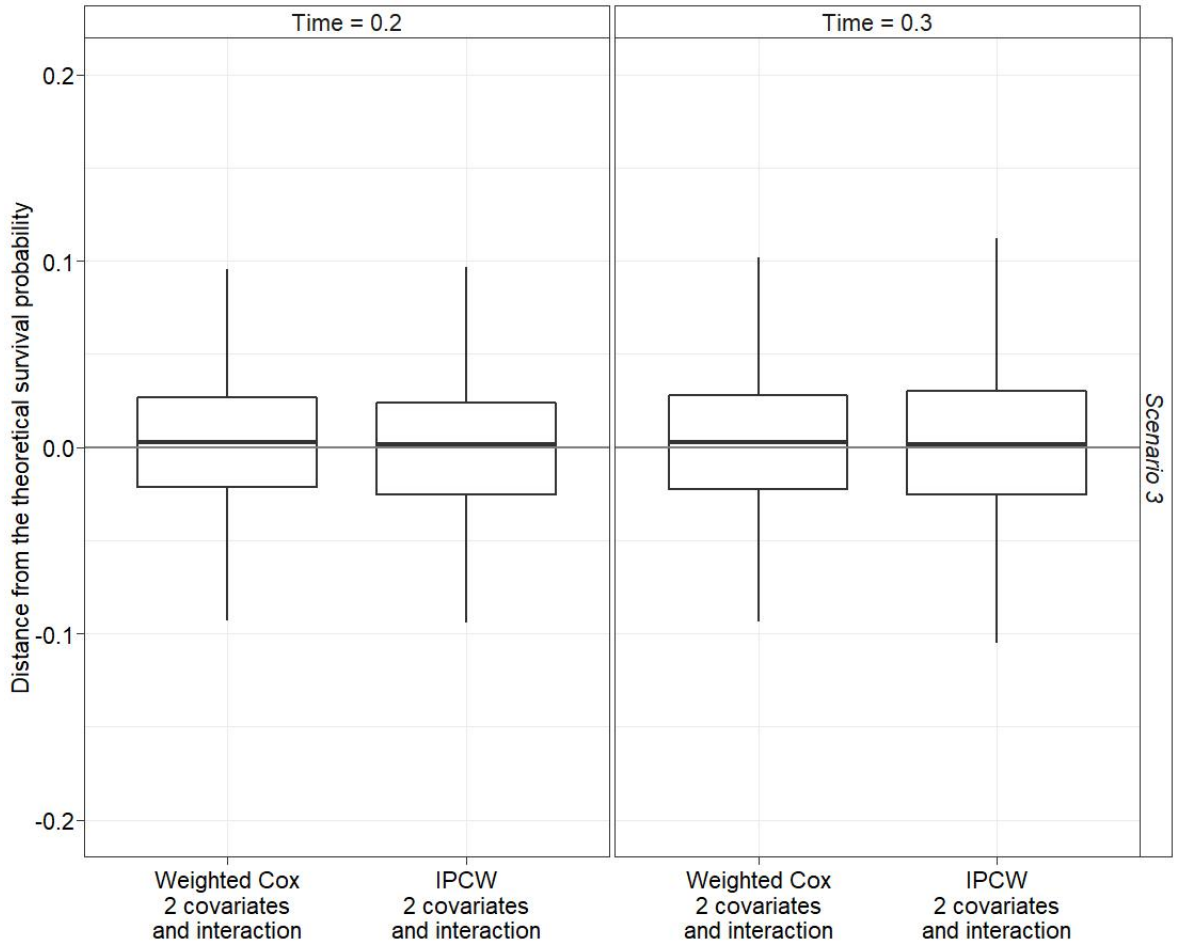

Figure 6.2: Simulation results for scenario 3 of the IPCW estimator accounting for the presence of an interaction between X_1 and X_2 in the estimate of the weights. The grey horizontal line is the reference null bias.

Table 6.1: Simulation results in scenarios 1 and 2 at times $t = 0.2$ and $t = 0.3$

Method	$Time = 0.2$				$Time = 0.3$			
	β	$Mean(\hat{\beta})$	$Bias \cdot 10^2$	$Var(\hat{\beta}) \cdot 10^3$	β	$Mean(\hat{\beta})$	$Bias \cdot 10^2$	$Var(\hat{\beta}) \cdot 10^3$
Scenario 1								
Naïve KM	0.6162	0.6173	0.1107	0.9832	0.5062	0.5071	0.0833	1.0989
Weighted KM 1 cov.	0.6162	0.6170	0.0871	0.9778	0.5062	0.5068	0.0545	1.0758
Weighted KM 2 covs.	0.6162	0.6169	0.0751	0.9705	0.5062	0.5067	0.0477	1.0708
Weighted Cox 1 cov.	0.6162	0.6169	0.0762	0.9725	0.5062	0.5070	0.0727	1.0626
Weighted Cox 2 covs.	0.6162	0.6182	0.2026	0.9551	0.5062	0.5081	0.1824	1.0366
IPCW 1 cov.	0.6162	0.6171	0.0911	0.9769	0.5062	0.5068	0.0573	1.0753
IPCW 2 covs.	0.6162	0.6169	0.0787	0.9685	0.5062	0.5066	0.0395	1.0612
Scenario 2								
Naïve KM	0.6162	0.6329	1.6793	1.0754	0.5062	0.5303	2.4066	1.2262
Weighted KM 1 cov.	0.6162	0.6170	0.0866	1.1490	0.5062	0.5077	0.1440	1.3257
Weighted KM 2 covs.	0.6162	0.6171	0.0948	1.1096	0.5062	0.5088	0.2573	1.2691
Weighted Cox 1 cov.	0.6162	0.6169	0.0786	1.1205	0.5062	0.5067	0.0507	1.2340
Weighted Cox 2 covs.	0.6162	0.6181	0.1955	1.0742	0.5062	0.5090	0.2794	1.1712
IPCW 1 cov.	0.6162	0.6171	0.0914	1.1539	0.5062	0.5079	0.1689	1.3061
IPCW 2 covs.	0.6162	0.6169	0.0787	1.1341	0.5062	0.5078	0.1538	1.2742

β is the estimated theoretical survival probability, $\hat{\beta}$ is the estimated survival probability in each of the 1000 samples, *bias* is calculated as difference between the average of the estimated survival probability on the 1000 samples and the theoretical survival probability.

Table 6.2: Simulation results in scenarios 3 and 4 at times $t = 0.2$ and $t = 0.3$

Method	$Time = 0.2$				$Time = 0.3$			
	β	$Mean(\hat{\beta})$	$Bias \cdot 10^2$	$Var(\hat{\beta}) \cdot 10^3$	β	$Mean(\hat{\beta})$	$Bias \cdot 10^2$	$Var(\hat{\beta}) \cdot 10^3$
Scenario 3								
Naïve KM	0.6162	0.6436	2.7411	1.1564	0.5062	0.5519	4.5641	1.3960
Weighted KM 1 cov.	0.6162	0.6330	1.6879	1.1900	0.5062	0.5350	2.8723	1.4414
Weighted KM 2 covs.	0.6162	0.6165	0.0387	1.3020	0.5062	0.5086	0.2381	1.7080
Weighted Cox 1 cov.	0.6162	0.6342	1.8033	1.1754	0.5062	0.5365	3.0288	1.3707
Weighted Cox 2 covs.	0.6162	0.6180	0.1839	1.2280	0.5062	0.5087	0.2433	1.4382
IPCW 1 cov.	0.6162	0.6312	1.5092	1.1895	0.5062	0.5341	2.7907	1.4023
IPCW 2 covs.	0.6162	0.6042	-1.1910	1.3751	0.5062	0.4951	-1.1153	1.7330
Scenario 4								
Naïve KM	0.6162	0.6689	5.2730	1.1702	0.5062	0.5792	7.2987	1.4296
Weighted KM 1 cov.	0.6162	0.6519	3.5720	1.2818	0.5062	0.5544	4.8186	1.6089
Weighted KM 2 covs.	0.6162	0.6219	0.5784	1.6244	0.5062	0.5214	1.5129	2.0540
Weighted Cox 1 cov.	0.6162	0.6512	3.5049	1.2536	0.5062	0.5543	4.8105	1.4761
Weighted Cox 2 covs.	0.6162	0.6186	0.2439	1.3757	0.5062	0.5095	0.3250	1.5812
IPCW 1 cov.	0.6162	0.6522	3.6039	1.2910	0.5062	0.5541	4.7882	1.6137
IPCW 2 covs.	0.6162	0.6176	0.1418	1.8397	0.5062	0.5116	0.5322	2.5230

β is the estimated theoretical survival probability, $\hat{\beta}$ is the estimated survival probability in each of the 1000 samples, *bias* is calculated as difference between the average of the estimated survival probability on the 1000 samples and the theoretical survival probability.

All the methods with the exception of the IPCW were compared in the extensions of the simulation protocol. We decided not to consider the IPCW method since we saw that already in scenario 4 of the main simulation it performs in the same way of the weighted Cox model.

Figure 6.3 shows the results obtained changing the imbalance of the covariates. As expected, in both cases only estimates obtained from the weighted KM and the weighted Cox model methods with the inclusion of both covariates are unbiased. Comparing the results of case A with the corresponding results of scenario 4 in Figure 6.1 one may observe the greater variability due to the lower number of patients at risk of developing an event (AE or relapse). In Table 6.3 one can observe that the variances of the estimates obtained from the weighted Cox model are 0.0017 and 0.0018 at times 0.2 and 0.3 respectively; these values are higher compared to the same values obtained in the main simulation (0.0014 and 0.0016 respectively, Table 6.2). The low number of subjects at risk is also the reason why the weighted KM does not perform very well. Comparing the results of case B with the corresponding results of scenario 4 in Figure 6.1 the estimates of all methods are less biased. This is particularly evident for the estimates derived from the methods including only X_1: for example, at time 0.2, in scenario 4 of Table 6.2 the bias is near to 0.035, whereas in Table 6.3 it is near to 0.011.

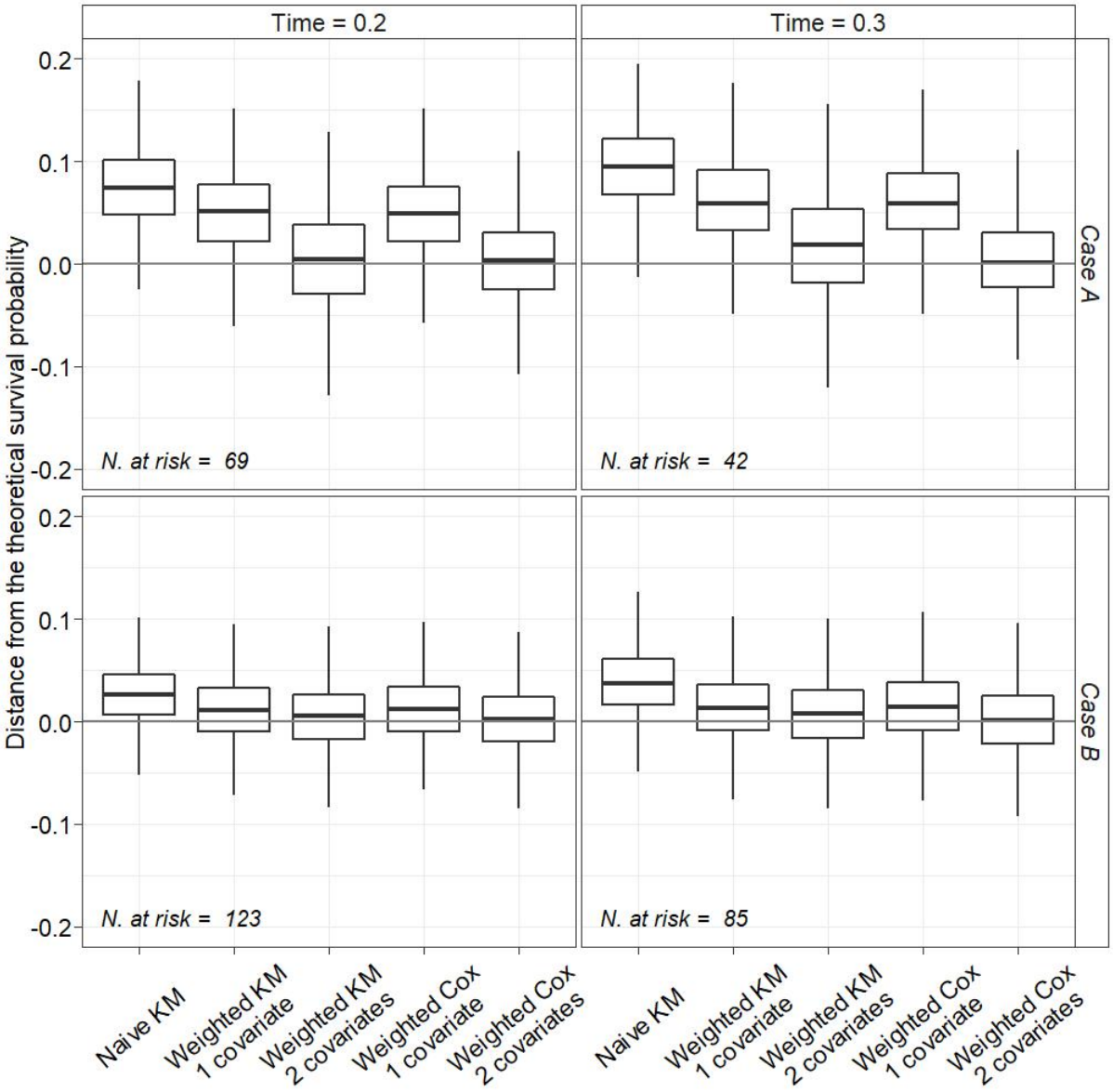

Figure 6.3: Simulation results for the variation of scenario 4 when $P(X_1 = 1) = P(X_2 = 1) = 0.5$ (Case A) or $P(X_1 = 1) = 0.3$ and $P(X_2 = 1) = 0.1$ (Case B). The grey horizontal line is the reference null bias.

Table 6.3: Simulation results for the variation of scenario 4 at times $t = 0.2$ and $t = 0.3$ when $P(X_1 = 1) = P(X_2 = 1) = 0.5$ (Case A) or $P(X_1 = 1) = 0.3$ and $P(X_2 = 1) = 0.1$ (Case B)

Method	$Time = 0.2$				$Time = 0.3$			
	β	$Mean(\hat{\beta})$	$Bias \cdot 10^2$	$Var(\hat{\beta}) \cdot 10^3$	β	$Mean(\hat{\beta})$	$Bias \cdot 10^2$	$Var(\hat{\beta}) \cdot 10^3$
Case A								
Naïve KM	0.5204	0.5944	7.3967	1.4747	0.4053	0.5008	9.5503	1.7725
Weighted KM 1 cov.	0.5204	0.5701	4.9693	1.6424	0.4053	0.4667	6.1181	2.0681
Weighted KM 2 covs.	0.5204	0.5248	0.4351	2.3194	0.4053	0.4244	1.9148	2.7203
Weighted Cox 1 cov.	0.5204	0.5687	4.8242	1.6026	0.4053	0.4653	6.0010	1.8410
Weighted Cox 2 covs.	0.5204	0.5237	0.3266	1.6985	0.4053	0.4092	0.3897	1.8263
Case B								
Naïve KM	0.7074	0.7334	2.6025	0.9304	0.6070	0.6461	3.9089	1.1861
Weighted KM 1 cov.	0.7074	0.7181	1.0722	1.0266	0.5062	0.6208	1.3815	1.3496
Weighted KM 2 covs.	0.7074	0.7119	0.4577	1.1028	0.5062	0.6148	0.7803	1.4059
Weighted Cox 1 cov.	0.7074	0.7188	1.1409	1.0143	0.5062	0.6219	1.4979	1.2622
Weighted Cox 2 covs.	0.7074	0.7089	0.1555	1.0641	0.5062	0.6087	0.1709	1.2976

β is the estimated theoretical survival probability, $\hat{\beta}$ is the estimated survival probability in each of the 1000 samples, *bias* is calculated as difference between the average of the estimated survival probability on the 1000 samples and the theoretical survival probability.

Figure 6.4 shows results when the parameters of the exponential distributions from which the hazard of relapse was simulated change. Comparing the results with the corresponding results of scenario 4 in Figure 6.1 one can observe that the lower is the hazard ratio of relapse, the lower is the bias of the estimated survival. For example, considering the estimate obtained from the weighted Cox model adjusted for X_1 and X_2 at time 0.2, when the hazard ratio of relapse is equal to 3 the bias from the theoretical survival probability is 0.0024 (Table 6.2), whereas when the hazard ratio of relapse is equal to 2 or to 1.5 the bias is 0.0020 or 0.0015 (case A and case B in Table 6.4 respectively). Of note, in this simulation setting the bias obtained from the naïve KM reduces (at time 0.2, from 0.053 when the hazard ratio is equal to 3 to 0.037 and 0.026 when it is equal to 2 or to 1.5 respectively) but it gives always the worst survival estimates.

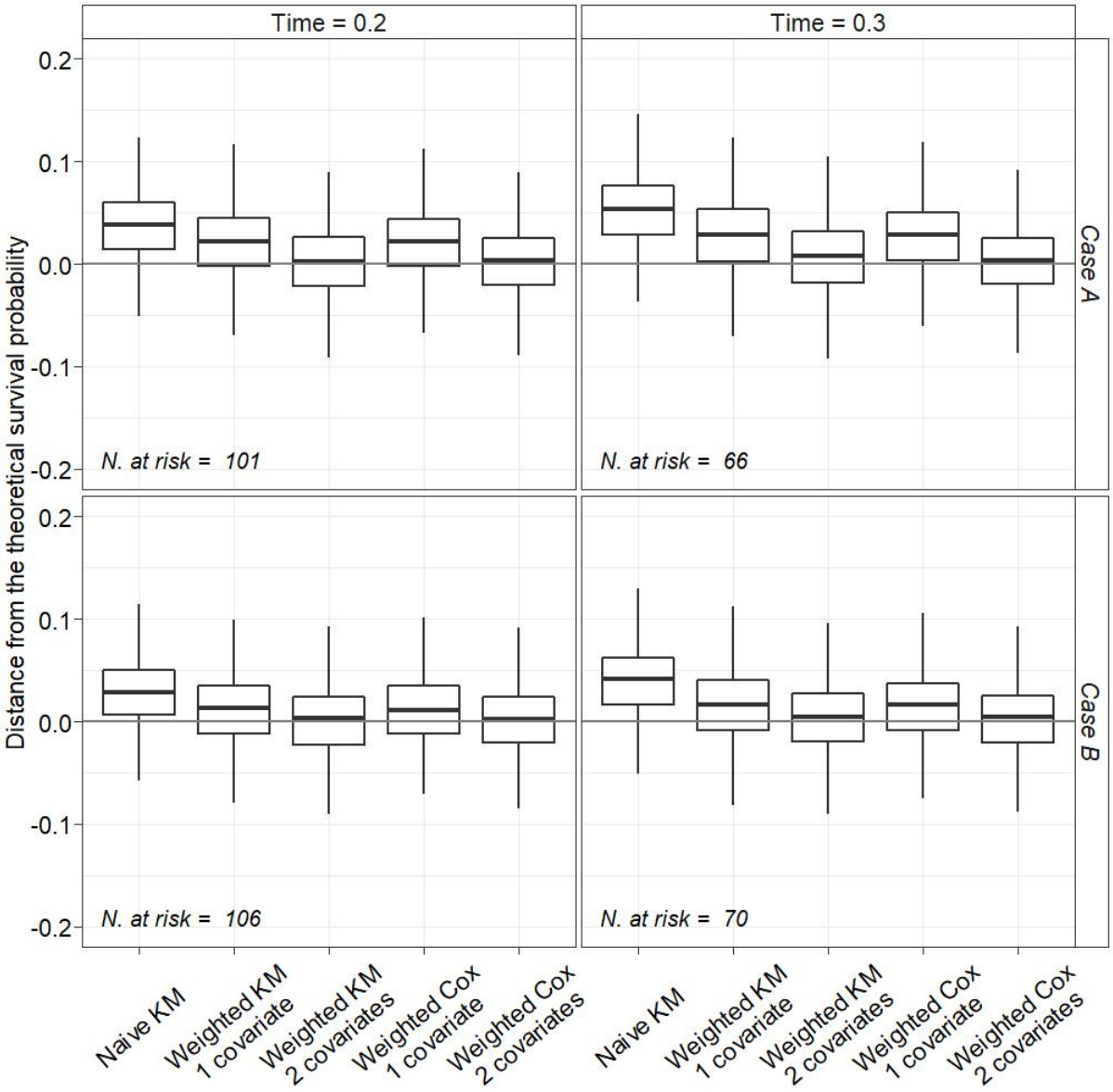

Figure 6.4: Simulation results for the variation of scenario 4 when fixed $X_1 = 0$ (or $X_1 = 1$), if X_2 changes, the hazard of relapse increases of 2 times (Case A) or of 1.5 times (Case B). The grey horizontal line is the reference null bias.

Table 6.4: Simulation results for the variation of scenario 4 at times $t = 0.2$ and $t = 0.3$ when fixed $X_1 = 0$ (or $X_1 = 1$), if X_2 changes, the hazard of relapse increases of 2 times (Case A) or of 1.5 times (Case B)

Method	$Time = 0.2$				$Time = 0.3$			
	β	$Mean(\hat{\beta})$	$Bias \cdot 10^2$	$Var(\hat{\beta}) \cdot 10^3$	β	$Mean(\hat{\beta})$	$Bias \cdot 10^2$	$Var(\hat{\beta}) \cdot 10^3$
Case A								
Naïve KM	0.6162	0.6532	3.7054	1.1321	0.4053	0.5008	9.5503	1.7725
Weighted KM 1 cov.	0.5204	0.6363	2.0144	1.2209	0.4053	0.4667	6.1181	2.0681
Weighted KM 2 covs.	0.5204	0.6178	0.1601	1.2982	0.4053	0.4244	1.9148	2.7203
Weighted Cox 1 cov.	0.5204	0.6357	1.9545	1.2015	0.4053	0.4653	6.0010	1.8410
Weighted Cox 2 covs.	0.5204	0.6182	0.2017	1.2173	0.4053	0.4092	0.3897	1.8263
Case B								
Naïve KM	0.7074	0.7334	2.6025	0.9304	0.5062	0.5582	5.2002	1.3261
Weighted KM 1 cov.	0.7074	0.7181	1.0722	1.0266	0.5062	0.5341	2.7882	1.4648
Weighted KM 2 covs.	0.7074	0.7119	0.4577	1.1028	0.5062	0.5127	0.6475	1.5152
Weighted Cox 1 cov.	0.7074	0.7188	1.1409	1.0143	0.5062	0.5331	2.6968	1.3512
Weighted Cox 2 covs.	0.7074	0.7089	0.1555	1.0641	0.5062	0.5086	0.2403	1.3359

β is the estimated theoretical survival probability, $\hat{\beta}$ is the estimated survival probability in each of the 1000 samples, *bias* is calculated as difference between the average of the estimated survival probability on the 1000 samples and the theoretical survival probability.

Chapter 7

Insights into event history following adverse events: recurrent adverse events and failure event following adverse/intermediate event

This chapter discusses insights into two additional topics: the analysis of recurrent AEs and the analysis of the impact of the occurrence of the AE on the subsequent hazard of relapse. The first topic is the generalization of the theoretical work carried out so far in this thesis to the broader scenario of repeated AEs, that may arise in practice. The second topic is on the assessment of the impact of the occurrence of the AE (and the subsequent decisions on the patient management) on the development of failure (relapse). This answers the subsequent question arising after the numerical quantification of the occurrence of the AE.

7.1 Analysis of recurrent AEs

This section is related to an ongoing work on the analysis of inflammatory/immune-related AEs in adult patients affected by leukaemia and treated with bosutinib. This type of AEs are "recurrent events" since they may occur more than once over the follow-up time for a given subject [10].

In this setting, the CP of AEs should not be used since it does consider that a patient could develop more than one AE. The epidemiological AEs rate could be used but it is important to underline that it is assumed to be constant in time, as in the case of only one AE. To analyse correctly data on recurrent AEs, one could use an extension of the AN estimator, which can be interpreted as the expected number of AEs in a sample of 100 subjects.

The motivating application is not affected by the problems due to the dependence between the potential time to AEs and the potential time to any other competing event, since in this particular case relapse was not considered. However, it is important to underline that in other settings one or more competing events could happen. In this case, one should account for the presence of a dependence among potential times, as in the case of only one AE.

Since results have not been yet published, we created a simulated dataset to show how can be interpreted the AN estimator in case of recurrent events. In Figure 7.1 two different ways of presenting the results of the AN estimator calculated on a dataset including recurrent events. In panel a) the AN estimate can be interpreted as the expected number of AEs a subject experiences in time. For example, at time 1.5 he/she develops 2 AEs. In panel b) the AN estimate can be viewed as the expected number of AEs in time in 100 hypothetical subjects. In this case, for example, at time 1.5 quite 200 AEs are expected in a population of 100 subjects.

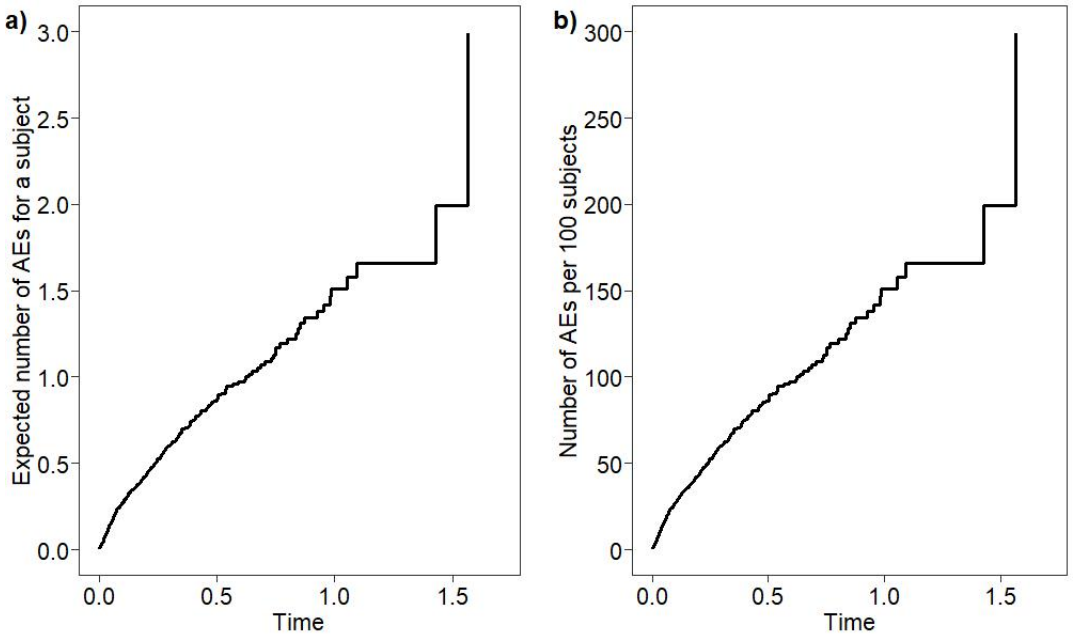

Figure 7.1: Two different ways of interpreting the AN estimator: a) expected number of AEs for a subject; b) number of AEs per 100 subjects.

7.2 Analysis of the impact of the occurrence of the AE on the subsequent hazard of relapse

When the focus moves to the impact that the occurrence of AEs have on the development of failure (relapse), the illness-death model could be a suitable theoretical

representation where the AE is thought of as a particular status of illness, that may have an influence on the development of failure. The illness-death model is the simplest multi-state model where the transition from the initial state 0 to the absorbing state 2 may involve an intermediate state 1, in this case the AE.

The standard approach of analysis in this setting is modelling the transition hazards from 0 to 2 and from 1 to 2, including time to state 1 as a time-varying covariate and measuring time from origin even after transition into state 1. The hazard from 1 to 2 can be also modelled separately using only patients in state 1, measuring time from illness and including time to state 1 as a fixed covariate. In the literature [26, 27] an approach was proposed through a model where time after the transition into state 1 is measured in both scales and time to state 1 is included as a time-varying covariate. Another possibility is a model where time after transition into state 1 is measured only from transition into state 1 and time to state 1 is included as a fixed covariate.

During my doctoral research activity I discussed through theoretical reasoning and simulation protocols, the use of these models and I developed a practical strategy aiming at:

(a) validate the properties of the illness-death process (Markov, semi-Markov, extended semi-Markov properties)

(b) estimate the impact of time to state 1 on the hazard from state 1 to 2

(c) quantify the impact that the transition into state 1 has on the hazard of the absorbing state.

This theoretical reasoning hesitated in a scientific paper that was published on Biometrical Journal in 2020 [28] and is attached to this thesis.

Discussion

In the majority of clinical studies on novel therapies safety outcomes (such as incidence of AEs) are analysed only through descriptive methods because of a variety of factors, such as treatment discontinuation, low power due to rare AEs, multiplicity of the possible AEs or recurrent episodes of the same AE-type. In this work we emphasized the critical aspects of the standard methods (i.e. crude proportion, AE rate, smoothed estimators of the cause-specific hazard) commonly used to analyse AEs data, proposing alternative solutions that permit to relax the assumption of independence between the potential time to AE and the potential time to relapse.

In the first part of this thesis we reviewed two different approaches starting from the type of clinical question when analysing AEs data: approach 1 consists in the description of the observed occurrence of an AE as first event. In this case, treatment ability to protect from relapse has an impact on the chance of observing AEs due to the competing risks action. When the aim is to describe the proportion of AEs, since the frequently presented crude proportion is not a function of time and does not properly account for censoring [6, 7], one can use the Aalen-Johansen estimator of the crude incidence of AEs, commonly used for competing risks analysis [8]: this quantity gives an estimate of the probability of treatment failure due to AEs over the course of time where relapse acts as competing event since AEs are counted only if observed as first events.

Approach 2 consists in the description of the potential occurrence of AE in relapse free patients. In this case, the commonly presented AE epidemiological rate correctly addresses for the presence of right censoring but it is not a function of time [5, 6]. At first glance, the estimators of the cause-specific hazard derived from the Aalen-Nelson or Kaplan-Meier formulas could be interpreted in terms of treatment action in determining the occurrence of an AE as first event regardless of the impact of relapse. However, the occurrence of relapse may "select" not at random patients excluded from the sub-sample of those on which the instantaneous rate of AEs is calculated. This selection, which is due to the dependence between the two potential times to AE and to relapse, leads to biased estimates. In particular, patients selection is stronger (and thus the bias is greater) when the imbalance of the covariates is strong and when the hazard ratio of relapse is high [25].

We proposed alternative methods, such as weighted average survival probability (estimated either by the Kaplan-Meier estimator or by the use of the Cox model) and inverse

probability of censoring weighting, and we proved through simulations that they overcome the problem due to the dependence between the potential times to AE and to relapse. In particular, we showed that one can handle patients selection, and thus obtain conditional independence between the two potential times, adjusting for all the observed covariates. Of note, even adjusting only for few observed covariates as in the reality due to unmeasured covariates, gives less biased estimates with respect to the estimate obtained from the naïve Kaplan-Meier. In fact we proved that the estimate obtained from the naïve Kaplan-Meier is always biased unless the hazard of relapse is independent from the covariates values (scenario 1). In an hypothetical scenario where all the covariates are observed, the weighted average survival estimate obtained either non parametrically or by the Cox model and the survival estimate from the inverse probability of censoring weighting would be unbiased (methods applied adjusting for both X_1 and X_2).

In addition, we point out that with the inverse probability of censoring weighting method one could obtained biased estimates when all the possible interactions between the observed covariates are not included in the model to estimate the weights (scenario 3). However, the inclusion of the interaction is not needed when the weighted Cox model is used, since conditional on the observed covariates, this model is robust in estimating the average survival. Nevertheless, a limitation in the use of the weighted average survival method is given by the fact that it may be applied only in the presence of binary (or categorical covariates), since if the covariate is continuous it is impossible to identify the subgroups in which the survival function is estimated.

We extended also the main simulation protocol changing the imbalance of the covariates or the hazard ratio of relapse. In both cases, results are similar to those obtained in the primary simulation: methods adjusted for the presence of both covariates give unbiased estimates, whereas those adjusted only for X_1 give less biased estimates with respect to those obtained with the naïve Kaplan-Meier estimator.

Of note, we did not include the presence of right censoring in the simulation protocol since the estimators we investigated already account for it. Thus, the inclusion of right censoring would represent a feature of the data that would not help to quantify the extent to bias of the estimators. This characteristic rather depends on intrinsic properties of the estimators and not on the possible presence of censoring as additional complexity in the data. A sensibility analysis to the presence of censoring could be however of interest and would be matter of future work.

In the final part of the thesis, starting from on ongoing work, we showed that from the point of view of the interpretation a complexity in the data given by the presence of recurrent events as in Chapter 7 improves the interpretation of the Aalen-Nelson estimator. This could be viewed as the expected number of AEs a subjects experiences in time or as the expected number of AEs in time in 100 hypothetical patients.

Regarding the possible developments of this work, a starting point could be the ap-

plication of the proposed models that relax the assumption of independence between the potential times in the context of recurrent events.

Bibliography

[1] European Medicines Agency (1995). ICH Topic E2A. Clinical safety data management: definitions and standards for expedited reporting. Available at: https://www.ema.europa.eu/en/ich-e2a-clinical-safety-data-management-definitions-standards-expedited-reporting.

[2] Hengelbrock J, Gillhaus J, Kloss S, Leverkus F (2016). Safety data from randomized controlled trials: applying models for recurrent events. *Pharmaceutical Statistics*, **15**(4):315-323.

[3] Siddiqui O (2009). Statistical methods to analyze adverse events data of randomized clinical trials. *Journal of Biopharmaceutical Statistics*, **19**(5):889-899.

[4] Bender R, Beckmann L, Lange S (2016). Biometrical issues in the analysis of adverse events within the benefit assessment of drugs. *Pharmaceutical Statistics*, **15**(4):292-296.

[5] Proctor T, Schumacher M (2016). Analysing adverse events by time-to-event models: the CLEOPATRA study. *Pharmaceutical Statistics*, **15**(4):306-314.

[6] Unkel S, Amiri M, Benda N, et al (2019). On estimands and the analysis of adverse events in the presence of varying follow-up times within the benefit assessment of therapies. *Pharmaceutical Statistics*, **18**(2):166-183.

[7] Beyersmann J, Schmoor C. The analysis of adverse events in randomized clinical trials. In S. Halabi & S. Michiels (Eds.), *Textbook of clinical trials in oncology: a statistical perspective*. Boca Raton, FL: CRC Press;2019:537-558.

[8] Allignol A, Beyersmann J, Schmoor C (2016). Statistical issues in the analysis of adverse events in time-to-event data. *Pharmaceutical Statistics*, **15**(4):297-305.

[9] Stegherr R, Beyersmann J, Jehl V, et al (2021). Survival analysis for AdVerse events with VarYing follow-up times (SAVVY): Rationale and statistical concept of a meta-analytic study. *Biometrical Journal*, **63**(3):650-670.

[10] Kleinbaum DG, Klein M (2005). Survival Analysis: A Self-Learning Text. *New: Springer Science and Business Media, LLC.*

[11] Marubini E, Valsecchi MG (1995). Analysing survival data from clinical trials and observational studies. *Chichester: J. Wiley.*

[12] Kaplan EL, Meier P (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**(282):457-481.

[13] Nelson W (1969). Hazard plotting for incomplete failure data. *Journal of Quality Technology*, **1**:27-52.

[14] Nelson W (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics*, **14**:945-965.

[15] Aalen O (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics*, **6**(4):701-726.

[16] Therneau TM, Grambsch PM (2000). Modeling Survival Data: Extending the Cox Model. *Springer, New York.*

[17] Cox DR (1972). Regression models and life-tables. *Journal of the Royal Statistical Society*, **34**(2):187-220.

[18] Aalen O, Johansen S (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics*, **5**(3):141-150.

[19] Robins JM, Rotnitzky A (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In: Jewell M, Dietz K, Farewell V, eds. *AIDS Epidemiology Methodological Issues.* Boston, MA: Birkhuser, 297-331.

[20] Cole SR, Hernan MA (2004). Adjusted survival curves with inverse probability weights. *Computer Methods and Programs in Biomedicine*, **75**(1):45-49.

[21] Robins JM, Finkelstein DM (2000). Correcting for noncompliance and dependent censoring in an aids clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics*, **56**(3):779-788.

[22] Willems SJW, Schat A, van Noorden MS, Fiocco M (2018). Correcting for dependent censoring in routine outcome monitoring data by applying the inverse probability censoring weighted estimator *Statistical Methods in Medical Research*, **27**(2):323-335.

[23] Parasole R, Valsecchi MG, Silvestri D, Locatelli F, et al (2018). Correspondence: Osteonecrosis in childhood acute lymphoblastic leukemia: a retrospective cohort study of the Italian Association of Pediatric Haemato-Oncology (AIEOP). *Blood Cancer J*, **8**(12):115.

[24] Bender R, Augustin T, Blettner M (2005). Generating survival times to simulate Cox proportional hazards models. *Statistics in Medicine*, **24**(11):1713-1723.

[25] Putter H, Fiocco M, Geskus RB (2007). Tutorial in biostatistics: Competing risks and multi-state models *Statistics in Medicine*, **26**(11):2389-2430.

[26] Iacobelli S, Carstensen B (2013). Multiple time scales in multi-state models *Statistics in Medicine*, **32**(30):5315-5327.

[27] Meier-Hirmer C, Schumacher M (2007). Multi-state model for studying an intermediate event using time-dependent covariates: Application to breast cancer *BMC Medical Research Methodology*, **13**:80.

[28] Tassistro E, Bernasconi DP, Rebora P, Valsecchi MG, Antolini L (2020). Modeling the hazard of transition into the absorbing state in the illness-death model *Biometrical Journal*, **62**(3):836?851.

Appendix A

Appendix - R codes

In this Appendix the main lines of R code used to create the simulated data example and the simulation protocol are reported.

A.1 Simulated data example

```
set.seed(12345)
n <- 100
lambda00AE <- 1
lambda01AE <- 3
lambda10AE <- 3
lambda11AE <- 9
lambda00RL <- 2
lambda01RL <- 6
lambda10RL <- 5
lambda11RL <- 15
# generate a binary covariate X1
x1 <- rbinom(n = n, size = 1, prob = 0.3)
# generate another binary covariate
x2 <- rbinom(n = n, size = 1, prob = 0.4)
# generate the time of Relapse (competing event)
genFt  <- runif(n, min = 0, max = 1)
T_RL <- ifelse(x1 == 0 & x2 == 0, -1/lambda00RL * log(1-genFt),
        ifelse(x1 == 0 & x2 == 1, -1/lambda01RL * log(1-genFt),
        ifelse(x1 == 1 & x2 == 0, -1/lambda10RL * log(1-genFt),
               -1/lambda11RL * log(1-genFt))))
# generate the time of Adverse Event
genFt  <- runif(n, min = 0, max = 1)
```

```
T_AE <- ifelse(x1 == 0 & x2 == 0, -1/lambda00AE * log(1-genFt),
        ifelse(x1 == 0 & x2 == 1, -1/lambda01AE * log(1-genFt),
        ifelse(x1 == 1 & x2 == 0, -1/lambda10AE * log(1-genFt),
               -1/lambda11AE * log(1-genFt))))
# create the status indicator (1 = AE, 0 = RL)
status <- ifelse(pmin(T_RL, T_AE) == T_RL, 0, 1)
# create the time variable as minimum of T_RL and T_AE
time <- pmin(T_RL, T_AE)
id <- seq(1, n, 1)
# create the dataset
ds <- data.frame (id, x1, x2, T_RL, T_AE, time, status)
```

A.2 Study simulation

The code will be reported only for scenario 1. The other scenarios could be obtained by simply varying the initial parameters denoted as `lambdaRL`.

A.2.1 Theoretical survival probability

```
lambda00AE <- 1
lambda01AE <- 3
lambda10AE <- 3
lambda11AE <- 9
lambda00RL <- 2
lambda01RL <- 2
lambda10RL <- 2
lambda11RL <- 2
time_vector <- c(0.2, 0.3)
reference_surv <- 0.7*0.6*exp(-lambda00AE*time_vector) +
                  0.7*0.4*exp(-lambda01AE*time_vector) +
                  0.3*0.6*exp(-lambda10AE*time_vector) +
                  0.3*0.4*exp(-lambda11AE*time_vector)
```

A.2.2 Expected number of subjects at risk

```
n <- 300
P00 <- 0.7*0.6
P01 <- 0.7*0.4
P10 <- 0.3*0.6
```

```
P11 <- 0.3*0.4
exp_numb_0.2 <- n * ((P00*exp(-lambda00AE*time_vector[1])*
                          exp(-lambda00RL*time_vector[1])) +
                     (P01*exp(-lambda01AE*time_vector[1])*
                          exp(-lambda01RL*time_vector[1])) +
                     (P10*exp(-lambda10AE*time_vector[1])*
                          exp(-lambda10RL*time_vector[1])) +
                     (P11*exp(-lambda11AE*time_vector[1])*
                          exp(-lambda11RL*time_vector[1])))
exp_numb_0.3 <- n * ((P00*exp(-lambda00AE*time_vector[2])*
                          exp(-lambda00RL*time_vector[2])) +
                     (P01*exp(-lambda01AE*time_vector[2])*
                          exp(-lambda01RL*time_vector[2])) +
                     (P10*exp(-lambda10AE*time_vector[2])*
                          exp(-lambda10RL*time_vector[2])) +
                     (P11*exp(-lambda11AE*time_vector[2])*
                          exp(-lambda11RL*time_vector[2])))
```

A.2.3 Simulation

```
library(survival)
library(cmprsk)
library(dynpred)
library(prodlim)
library(rms)
nsim <- 1000
naive_surv <- c()
weighted_surv1_KM <- c()
weighted_surv2_KM <- c()
weighted_surv1_cox <- c()
weighted_surv2_cox <- c()
ipcw_surv1 <- c()
ipcw_surv2 <- c()

for (i in 1:nsim) {
x1 <- rbinom(n = n, size = 1, prob = 0.3)
x2 <- rbinom(n = n, size = 1, prob = 0.4)
genFt  <- runif(n, min = 0, max = 1)
T_RL <- ifelse(x1 == 0 & x2 == 0, -1/lambda00RL * log(1-genFt),
```

```
        ifelse(x1 == 0 & x2 == 1, -1/lambda01RL * log(1-genFt),
        ifelse(x1 == 1 & x2 == 0, -1/lambda10RL * log(1-genFt),
               -1/lambda11RL * log(1-genFt))))
genFt  <- runif(n, min = 0, max = 1)
T_AE <- ifelse(x1 == 0 & x2 == 0, -1/lambda00AE * log(1-genFt),
        ifelse(x1 == 0 & x2 == 1, -1/lambda01AE * log(1-genFt),
        ifelse(x1 == 1 & x2 == 0, -1/lambda10AE * log(1-genFt),
               -1/lambda11AE * log(1-genFt))))
status <- ifelse(pmin(T_RL, T_AE) == T_RL, 0, 1)
time <- pmin(T_RL, T_AE)
ds <- data.frame (id = seq(1, n, 1), x1, x2, T_RL, T_AE, time, status)

### NAIVE KAPLAN-MEIER ESTIMATOR
naive_est <- summary(prodlim(Hist(time, status == 1) ~ 1, data = ds),
                     times = time_vector)
wrong_surv <- c(wrong_surv, wrong_est$table[1,5], wrong_est$table[2,5])

### WEIGHTED AVERAGE KAPLAN-MEIER ESTIMATOR - X1
est_KM1 <- survfit(Surv(time, status == 1) ~ as.factor(x1), data = ds)
unique_time <- unique(ds$time)
unique_time <- c(unique_time, time_vector)
surv_x0 <- evalstep(time = est_KM1[treat=1]$time,
                    stepf = est_KM1[treat=1]$surv,
                    newtime = unique_time, subst = 1,
                    to.data.frame = T)
surv_x1 <- evalstep(time = est_KM1[treat=2]$time,
                    stepf = est_KM1[treat=2]$surv,
                    newtime = unique_time, subst = 1,
                    to.data.frame = T)
colnames(surv_x0)[2] <- "surv_x0"
colnames(surv_x1)[2] <- "surv_x1"
ds_surv1 <- merge(surv_x0, surv_x1, by = 'newtime')
ds_surv1$surv_tot <- length(which(ds$x1 == 0))/n * ds_surv1$surv_x0 +
                     length(which(ds$x1 == 1))/n * ds_surv1$surv_x1
weighted_surv1_KM <- c(weighted_surv1_KM,
                       ds_surv1$surv_tot[ds_surv1$newtime == 0.2],
                       ds_surv1$surv_tot[ds_surv1$newtime == 0.3])

### WEIGHTED AVERAGE KAPLAN-MEIER ESTIMATOR - X1 and X2
```

```
est_KM2 <- survfit(Surv(time, status == 1) ~ as.factor(x1) + as.factor(x2),
                   data = ds)
surv_x00 <- evalstep(time = est_KM2[treat=1]$time,
                     stepf = est_KM2[treat=1]$surv,
                     newtime = unique_time, subst = 1,
                     to.data.frame = T)
surv_x01 <- evalstep(time = est_KM2[treat=2]$time,
                     stepf = est_KM2[treat=2]$surv,
                     newtime = unique_time, subst = 1,
                     to.data.frame = T)
surv_x10 <- evalstep(time = est_KM2[treat=3]$time,
                     stepf = est_KM2[treat=3]$surv,
                     newtime = unique_time, subst = 1,
                     to.data.frame = T)
surv_x11 <- evalstep(time = est_KM2[treat=4]$time,
                     stepf = est_KM2[treat=4]$surv,
                     newtime = unique_time, subst = 1,
                     to.data.frame = T)
colnames(surv_x00)[2] <- "surv_x00"
colnames(surv_x01)[2] <- "surv_x01"
colnames(surv_x10)[2] <- "surv_x10"
colnames(surv_x11)[2] <- "surv_x11"
ds_surv2 <- merge(surv_x00, surv_x01, by = 'newtime')
ds_surv2 <- merge(ds_surv2, surv_x10, by = 'newtime')
ds_surv2 <- merge(ds_surv2, surv_x11, by = 'newtime')
ds_surv2$surv_tot <- length(which(ds$x1 == 0))/n *
                     length(which(ds$x2 == 0))/n * ds_surv2$surv_x00 +
                     length(which(ds$x1 == 0))/n *
                     length(which(ds$x2 == 1))/n * ds_surv2$surv_x01 +
                     length(which(ds$x1 == 1))/n *
                     length(which(ds$x2 == 0))/n * ds_surv2$surv_x10 +
                     length(which(ds$x1 == 1))/n *
                     length(which(ds$x2 == 1))/n * ds_surv2$surv_x11
weighted_surv2_KM <- c(weighted_surv2_KM,
                       ds_surv2$surv_tot[ds_surv2$newtime == 0.2],
                       ds_surv2$surv_tot[ds_surv2$newtime == 0.3])

### WEIGHTED AVERAGE COX MODEL - X1
est_cox1 <- cph(Surv(time, status == 1) ~ x1, data = ds, surv = T)
```

```
mean_surv1 <- colMeans(survest(est_cox1, newdata = ds,
                      times = time_vector, se.fit = F)$surv)
weighted_surv1_cox <- c(weighted_surv1_cox, mean_surv1[[1]], mean_surv1[[2]])

### WEIGHTED AVERAGE COX MODEL - X1 and X2
est_cox2 <- cph(Surv(time, status == 1) ~ x1 + x2, data = ds, surv = T)
mean_surv2 <- colMeans(survest(est_cox2, newdata = ds,
                      times = time_vector, se.fit = F)$surv)
weighted_surv2_cox <- c(weighted_surv2_cox, mean_surv2[[1]], mean_surv2[[2]])

### IPCW - X1 (mod_den1) - X1 and X2 (mod_den2)
ds$censored <- ifelse(ds$status == 0, 1, 0)
ds.long <- survSplit(ds, cut = ds$time, end = "time", start = "Tstart",
                    event = "status", id = "id2")
ds.long <- ds.long[order(ds.long$id, ds.long$time), ]
ds.long.cens <- survSplit(ds, cut = ds$time, end = "time", start = "Tstart",
                         event = "censored", id = "id2")
ds.long.cens <- ds.long.cens[order(ds.long.cens$id, ds.long.cens$time), ]
ds.long$censored <- ds.long.cens$censored
mod_den1 <- cph(Surv(Tstart, time, censored) ~ x1, data = ds.long,
                surv = T)
mod_den2 <- cph(Surv(Tstart, time, censored) ~ x1 + x2, data = ds.long,
                surv = T)
ds.long$prob_den1 <- NULL
ds.long$prob_den2 <- NULL

for(j in 1:nrow(ds.long)){
dataj <- ds.long[j, ]
ds.long$prob_den1[j] <- as.numeric(survest(mod_den1, newdata = dataj,
                        times = dataj$Tstart, se.fit = F)$surv)
ds.long$prob_den2[j] <- as.numeric(survest(mod_den2, newdata = dataj,
                        times = dataj$Tstart, se.fit = F)$surv)
}
ds.long$weights_unst1 <- 1/ds.long$prob_den1
ds.long$weights_unst2 <- 1/ds.long$prob_den2
res_unst1 <- survfit(Surv(Tstart, time, status) ~ 1, data = ds.long,
                    weights = weights_unst1, timefix = F)
ipcw_surv1 <- c(ipcw_surv1, evalstep(time = res_unst1$time,
                stepf = res_unst1$surv, newtime = time_vector,
```

```
                subst = 1, to.data.frame = F))
res_unst2 <- survfit(Surv(Tstart, time, status) ~ 1, data = ds.long,
                     weights = weights_unst2, timefix = F)
ipcw_surv2 <- c(ipcw_surv2, evalstep(time = res_unst2$time,
                stepf = res_unst2$surv, newtime = time_vector,
                subst = 1, to.data.frame = F))
}
```

Received: 30 August 2018 | Revised: 21 May 2019 | Accepted: 12 June 2019

DOI: 10.1002/bimj.201800267

Biometrical Journal

RESEARCH PAPER

Modeling the hazard of transition into the absorbing state in the illness-death model

Elena Tassistro | **Davide Paolo Bernasconi** | **Paola Rebora** | **Maria Grazia Valsecchi** | **Laura Antolini**

Center of Biostatistics for Clinical Epidemiology, School of Medicine and Surgery, University of Milano-Bicocca, Monza, Italy

Correspondence
Davide Paolo Bernasconi, Center of Biostatistics for Clinical Epidemiology, School of Medicine and Surgery, University of Milano-Bicocca, Via Cadore 48, 20900 Monza, Italy.
Email: davide.bernasconi@unimib.it

Funding information
Associazione Italiana per la Ricerca sul Cancro, Grant/Award Number: Fellowship FIRC 16 triennial id: 18150; Ministero dell'Istruzione, dell'Università e della Ricerca, Grant/Award Number: SIR RBSI14LOVD

Abstract

The illness-death model is the simplest multistate model where the transition from the initial state 0 to the absorbing state 2 may involve an intermediate state 1 (e.g., disease relapse). The impact of the transition into state 1 on the subsequent transition hazard to state 2 enables insight to be gained into the disease evolution. The standard approach of analysis is modeling the transition hazards from 0 to 2 and from 1 to 2, including time to illness as a time-varying covariate and measuring time from origin even after transition into state 1. The hazard from 1 to 2 can be also modeled separately using only patients in state 1, measuring time from illness and including time to illness as a fixed covariate. A recently proposed approach is a model where time after the transition into state 1 is measured in both scales and time to illness is included as a time-varying covariate. Another possibility is a model where time after transition into state 1 is measured only from illness and time to illness is included as a fixed covariate. Through theoretical reasoning and simulation protocols, we discuss the use of these models and we develop a practical strategy aiming to (a) validate the properties of the illness-death process, (b) estimate the impact of time to illness on the hazard from state 1 to 2, and (c) quantify the impact that the transition into state 1 has on the hazard of the absorbing state. The strategy is also applied to a literature dataset on diabetes.

KEYWORDS

illness-death, Markov model, survival, time scales, transition hazard

1 | INTRODUCTION

The illness-death model is the simplest multistate model where the transition from the initial state to the absorbing state may involve or not an intermediate state, such as illness (Beyersmann, Allignol, & Schumacher, 2012). The impact of the intermediate transition on the subsequent hazard of the absorbing state enables insights to be gained into the disease evolution and refine the prediction. The standard approaches of analysis rely on the choice of the time scale(s) to measure the follow-up time after the transition to illness (Eulenburg, Mahner, Woelber, & Wegscheider, 2015; Putter, Fiocco, & Geskus, 2007). The original (or "clock forward") time scale, commonly used to measure the follow-up time before the first transition, can be considered even after the transition to illness. The time to such transition can be included as a covariate in a model estimated on the entire set of subjects. The follow-up time after the transition can be also measured in the "clock reset" scale that uses the intermediate state as a new origin (Andersen, Esbjerg, & Sorensen, 2000; Putter et al., 2007). In this setting, models are commonly estimated only on the subsample of subjects who developed illness, and the time of the transition to illness can be included as a covariate

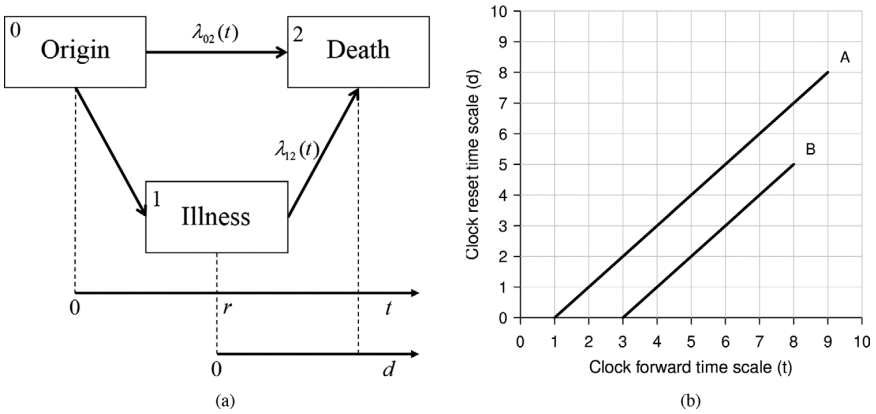

FIGURE 1 (a) Illness-death model with clock forward and clock reset time scales. Boxes and arrows represent the states and the possible transitions, respectively; (b) Lexis diagram of two patients with times to illness $r = 1$ for patient A and $r = 3$ for patient B

(*subsample models*). In a recently proposed approach (Iacobelli & Carstensen, 2013; Meier-Hirmer & Schumacher, 2013; Rebora, Galimberti, & Valsecchi, 2015), the follow-up time after the transition is measured both in the original scale and in the clock reset scale. The model is estimated on the entire sample of subjects and the time of the transition is included as covariate (*full-sample models*). A further possibility is to measure the follow-up time in the original scale before the transition and in the clock reset scale after the transition.

The paper aims to set up a strategy a statistician can follow to fit the most suitable full-sample model on the hazards of transition to the absorbing state. A feature of the strategy is that the scale to measure time after illness, for transition hazard to the absorbing state, is not a choice done a priori, but it depends only on the Markov, semi-Markov, and extended semi-Markov properties of the illness-death process (Bernasconi, Rebora, Iacobelli, Valsecchi, & Antolini, 2016). In case of non-Markov data, we developed a novel modeling approach that ensures the interpretability of the model coefficient of the time to illness.

In Section 2, we set up the notation and provide definitions of the clock forward and clock reset time scales. In addition, we review the meaning of the Markov, semi-Markov, and extended semi-Markov properties. In Section 3, we introduce models on the subsample of patients who developed illness to validate the properties of the illness-death process. In Section 4, we present full-sample models on the entire sample of patients. In Section 5, a simulation study is proposed. In Section 6, results of the simulation study are presented. In Section 7, the proposed methods are applied to a literature dataset on diabetes. The paper ends with a discussion where the strategy of analysis is summarized.

2 | NOTATION AND BACKGROUND

We consider an illness-death model, where the initial state, the intermediate state, and the absorbing state are called 0, 1, and 2, respectively. We assume that the transition to the intermediate state is due to the development of an intermediate event (*illness*) and the transition to the absorbing state is due to the development of *death*. At the beginning of follow-up, all patients are in state 0 and they can move directly to the absorbing state 2 or they can move first to state 1 and then to state 2. The hazard of transition to the absorbing state is denoted by $\lambda_{02}(t)$ or $\lambda_{12}(t)$, depending on whether the subject was in state 0 or 1. A graphical visualization of this process is provided in Figure 1a, where three time measurements are represented:

- t: It is time measured starting from the origin, which is the natural way to measure time for patients in state 0, but it can be considered also for patients in state 1, since time from origin can be calculated even after the intermediate event. t is called *clock forward* time scale emphasizing that time "keeps going forward" even after the intermediate event.
- r: It is time of development of the intermediate event, which is a value in the clock forward time scale that we will call in the remaining *time to illness*. It is defined only for patients in state 1.
- d: It is time measured from the intermediate state 1, thus from a new origin represented by the time r of occurrence of the intermediate event. This time scale is called *clock reset* time scale and can be indeed measured only for patients in state 1. It is the natural way to measure time after the intermediate event.

Let us consider a generic time point t^* in the clock forward scale and a patient in state 0. This patient experienced up to t^* the hazard of transition to the absorbing state 2 $\lambda_{02}(t)$, and also the hazard of transition to state 1. The latter hazard is not relevant for the aim of this paper. If the patient will enter in the intermediate state 1 at some time r in the clock forward scale, he/she will start experiencing the hazard $\lambda_{12}(t)$ for $t > r$, which may be different from the corresponding $\lambda_{02}(t)$ due to the development of the intermediate event and at the time r where the transition happened.

The relationship between the clock forward t and the clock reset d time scales is defined as follows. At the transition to the intermediate state, t is equal to r, the clock reset scale is defined as $d = 0$, and the time to illness r is set. The next time unit is $t = r + 1$ and the clock reset scale starts increasing with $d = 1$. In general, the relationship between t and d is $t = r + d$ or equivalently $d = t - r$.

The value of $\lambda_{12}(t)$ acting on patients who developed the intermediate event may depend on time t ($t > r$) through two factors:

(i) the time when the transition to the intermediate state occurred, i.e. the fixed covariate time to illness r,

(ii) the time past from the transition d.

Depending on these factors, three properties of $\lambda_{12}(t)$ for $t > r$ are defined: Markov (M), semi-Markov (SM), and extended semi-Markov (ESM).

Markov property

The process is Markovian if the value of $\lambda_{12}(t)$ for $t > r$ depends only on time t from origin. This means, in other words, that the value of $\lambda_{12}(t)$ depends on time d after illness and on the fixed covariate time to illness r only through their sum $d + r = t$ and not on the values d and r taken separately. The clock forward time scale alone can be conveniently used to measure time even after the intermediate event and the notation $\lambda_{12}(t)$ can be kept, meaning that in this case the hazard depends only on t.

To enlighten the meaning of this property, let us consider two patients in state 1 at time $t^* = 6$, as represented in the Lexis diagram, where on the x-axis there is the clock forward scale t and on the y-axis the clock reset scale d (Figure 1b): patient A developed the intermediate event at time to illness $r_A = 1 < t^*$ and patient B two time units later, at time to illness $r_B = r_A + 2 = 3$, $r_B < t^*$. The times after transition to illness d_A and d_B of these patients are different, being $d_A = t^* - r_A = 5$ for patient A and $d_B = t^* - r_B = t^* - r_A - 2 = d_A - 2 = 3$ for patient B. Although $r_A \neq r_B$ and thus $d_A \neq d_B$, at time t^* the two patients share the same hazard value $\lambda_{12}(t^*)$. The only condition that matters is that both patients are in state 1 at time t^*, regardless of when this transition happened and of the time after the transition.

Semi-Markov property

The process is semi-Markovian if the value $\lambda_{12}(t)$ for $t > r$ depends only on time d after illness. Thus, the clock reset time scale alone can be conveniently used to measure time after the intermediate event and the new notation $\lambda_{12}(d)$ for $d > 0$ is set, meaning that time is measured on the clock reset scale.

To enlighten the meaning of this property, let us consider again the two patients of the Lexis diagram in Figure 1b. Let us now consider for these patients a time point $d^* = 4$ in the clock reset scale: here, $d_A = d_B = d^*$, whereas $t_A = r_A + d^* = 5$ is different from $t_B = r_B + d^* = r_A + 2 + d^* = 7$. Although $r_A \neq r_B$ and thus $t_A \neq t_B$, at time d^* the two patients share the same hazard value $\lambda_{12}(d^*)$. In other words, the only condition that matters is that both patients are in state 1 after d time units from the transition to the intermediate event, regardless of when this transition happened and of the time from origin.

Extended semi-Markov property

The process is extended semi-Markovian if the value of $\lambda_{12}(t)$ for $t > r$ depends on time d after illness and on the fixed covariate time to illness r, taken separately. The time d after illness, and thus the clock reset time scale, becomes again the natural way to measure time after the intermediate event and one can set the new notation $\lambda_{12}(r, d)$ meaning that time is measured on the clock reset scale and r has an impact on the hazard $\lambda_{12}(r, d)$ for $d > 0$.

To enlighten the meaning of this property, let us consider again the two patients of the Lexis diagram in Figure 1b. Consider the time point $d^* = 4$ in the clock reset scale: here, $d_A = d_B = d^*$, whereas $t_A = r_A + d^* = 5$ is different from $t_B = r_B + d^* = r_A + 2 + d^* = 7$. Suppose that an earlier transition to state 1 implies a greater $\lambda_{12}(r, d)$: since patient A has an earlier time to illness than patient B, d^* time units after the transition to the intermediate event, patient A will have a greater hazard of transition to the absorbing state 2. Of note, the dependency of $\lambda_{12}(r, d)$ from the time t is fully captured by the dependency on both d and r.

TABLE 1 Subsample models

Model	t	d	r	$t \geq r$
CFr	✓	–	✓	$\lambda_{12}(t,r) = \lambda_0(t)e^{\beta \cdot r}$
CFd	✓	✓	–	$\lambda_{12}(t,d) = \lambda_0(t)e^{\beta \cdot d}$
CRr	–	✓	✓	$\lambda_{12}(d,r) = \lambda_0(d)e^{\beta \cdot r}$

$\lambda_0(\cdot)$ represents the baseline hazard, β is the coefficient of r in CFr and CRr models, and of d in CFd model.
CFr = clock forward model with r as covariate, CFd = clock forward model with d as covariate, CRr = clock reset model with r as covariate.

From the meaning of the three properties, we can thus recognize that the time to illness r has a role on the hazard after illness only in the extended semi-Markov scenario, that is, when Markov and semi-Markov properties are not satisfied. Therefore, the analysis of the possible role of r, and of its impact, should start from assessing if either the Markov or semi-Markov properties could be justified.

3 | SUBSAMPLE MODELS FOR TRANSITION 1 → 2

Let us consider the subsample of patients who developed the intermediate event. Table 1 collects the three subsample models on $\lambda_{12}(t)$ for $t > r$ one might consider:

- CFr model: The clock forward scale t is used to measure time even after the transition to state 1, and the time to illness r is included as a fixed covariate, since we do not consider time before the transition to the illness state.
- CFd model: The clock forward scale t is used to measure time even after the transition to state 1, and the time after illness d is included as covariate. This covariate is time varying since it is not defined before the transition at $t = r$ and increases deterministically in time with $d = t - r$ (Andersen et al., 2000). Of note, CFr and CFd models are algebraically equivalent, since $r = t - d$ for $t > r$ implies $\lambda_{12}(t,r) = \lambda_0(t) \cdot \exp(\beta r) = \lambda_0^*(t) \cdot \exp(-\beta d) = \lambda_{12}(t,d)$, where the baseline hazards satisfy the relationship $\lambda_0^*(t) = \lambda_0(t) \cdot \exp(\beta t)$.
- CRr model: The clock reset scale d is used to measure time after the transition to state 1, and the time to illness r is included as fixed covariate.

To check whether the Markov property can be assumed, the CFr (or the equivalent CFd) model is the natural choice since, under the Markov property, $\lambda_{12}(t)$ should depend only on the time since origin t and neither the time to illness r nor the linear transformation $d = t - r$ should have an impact on $\lambda_{12}(t)$. Thus, if the time to illness r in CFr model, or the time after illness d in CFd model, does not show an impact on the hazard $\lambda_{12}(t)$, the Markov property could be justified. Before relying on the coefficient estimate of r (or d as an alternative) in the CFr (or CFd) model to test the Markov property, one should check if the proportionality of the hazards and the log-linear effect of r are plausible. This can be performed by the analysis of the Schoenfeld and martingale residuals from the subsample model obtained according to the Cox model specification. The proportional hazards assumption and the log-linear effect of r can also be checked graphically by plotting the logarithm of the smoothed hazard function against time in subgroups of subjects with similar times to illness (e.g., according to deciles) (Rebora, Salim, & Reilly, 2014). If the curves tend to keep constant distances in time, the proportionality of the hazards may be plausible. This suggests a naive method to test directly the plausibility of the Markov property, which consists of plotting a smoothed estimate of the hazard function (or its logarithm) in subgroups of patients with similar times to illness against time measured in the clock forward scale (Bernasconi et al., 2016). If the obtained curves are overlapping, the Markov property could be reasonable.

If the Markov property cannot be justified, the validity of the semi-Markov property should be investigated using the clock reset scale as natural time scale by the CRr model. If the time to illness does not show an impact on the hazard $\lambda_{12}(d)$, the semi-Markov property could be justified. Before relying on this model, the proportionality of the hazards and the log-linear effect of r should be investigated. Again, a naive graphical method to test the semi-Markov property could also be used by plotting a smoothed estimate of the hazard function in subgroups of patients with similar times to illness against time in the clock reset scale. In case of overlapping curves, the semi-Markov property could be justified.

When neither the Markov nor the semi-Markov property hold, we are in the extended semi-Markov scenario and the CRr model provides the strategy for a suitable estimation of the impact of the time to illness r.

TABLE 2 Full-sample models on $\lambda_{02}(t)$ and $\lambda_{12}(t)$

Model	t	d	r	$t < r$	$t \geq r$
CFr	✓	–	✓	$\lambda_{02}(t)$	$\lambda_{12}(t,r) = \lambda_{02}(t)e^{\beta_{ill}}e^{\beta \cdot r}$
CFr*	✓	–	✓	$\lambda_{02}(t)$	$\lambda_{12}(t,r) = \lambda_{02}(t)e^{\beta_{ill}+\gamma(t)}e^{\beta \cdot r}$
CFrd	✓	✓	✓	$\lambda_{02}(t)$	$\lambda_{12}(t,r,d) = \lambda_{02}(t)\mu(d)e^{\beta_{ill}}e^{\beta \cdot r}$
CFrd*	✓	✓	✓	$\lambda_{02}(t)$	$\lambda_{12}(t,r,d) = \lambda_{02}(t)\mu(d)e^{\beta_{ill}+\gamma(t)}e^{\beta \cdot r}$
CRr	✓	✓	✓	$\lambda_{02}(t)$	$\lambda_{12}(r,d) = \lambda_{02}(d)e^{\beta_{ill}}e^{\beta \cdot r}$
CRr*	✓	✓	✓	$\lambda_{02}(t)$	$\lambda_{12}(r,d) = \lambda_{02}(d)e^{\beta_{ill}+\gamma(t)}e^{\beta \cdot r}$

The function $\gamma(t)$ is an explicit time-varying component in the hazard ratio $\lambda_{12}(t)/\lambda_{02}(t)$, and β is the coefficient of the time to illness r.
CFr = clock forward model with r as covariate, CFrd = clock forward model with r and d as covariates, CRr = clock reset model with r as covariate; models with * = models with a flexible effect of the intermediate event.

4 | FULL-SAMPLE MODELS FOR TRANSITIONS 0 → 2 AND 1 → 2

In a full-sample model, the transition hazard $\lambda_{02}(t)$ is used as a baseline hazard that could be modified into $\lambda_{12}(t)$ at $t > r$ if illness occurred at time to illness r. Table 2 collects six models one might consider. In all models, the transition hazard $\lambda_{02}(t)$ is implemented in the clock forward scale, as this is the only time scale available for subjects in state 0. The transition hazard $\lambda_{12}(t)$ can be implemented according to three specifications, depending on the time scale and on the covariates included, as follows:

- CFr model: Clock forward scale and time to illness r as covariate. This covariate is time varying since it is not defined before the transition at $t = r$. After the transition, r is fixed.
- CFrd model: Clock forward scale and time to illness r as time-varying covariate (not defined before the transition at $t = r$, fixed after the transition), and time after illness d as time-varying covariate, where $\mu(d)$ was used to indicate a non-log-linear effect of the time-varying covariate d. Substituting $d = t - r$, the hazard $\lambda_{12}(t) = \lambda_{02}(t) \cdot \exp(\beta_{ill}) \cdot \mu(t-r) \cdot \exp(\beta r)$ depends on t through $\mu(t-r)$. For this reason, the CFrd model is a clock forward model with two explanatory time-varying covariates: the time to illness r and the time after illness d, respectively, with log-linear and non-log-linear effects. Of note, the two covariates cannot be included both with log-linear effects since this would lead to the component $\exp(\beta_d d) \cdot \exp(\beta_r r) = \exp(\beta_d(t-r)) \cdot \exp(\beta_r r)$ with a doubled inclusion of r. Approaches similar to CFrd can be found in the works of Iacobelli and Carstensen (2013) and Meier-Hirmer and Schumacher (2013).
- CRr model: Clock reset scale d and time to illness r as fixed covariate.

The relationship between the transition hazards $\lambda_{02}(t)$ and $\lambda_{12}(t)$ for $t > r$ is modeled by a proportional hazards-type specification through the β_{ill} parameter in CFr, CFrd, and CRr models. Of note, this specification does not mean necessarily proportionality in the clock forward time scale t for $t > r$. Let us consider the hazard ratio $\lambda_{12}(t)/\lambda_{02}(t)$.

- CFr model: The hazard ratio is equal to $\exp(\beta_{ill}) \cdot \exp(\beta r)$ for $t > r$, thus in this case proportionality is assumed in the clock forward scale (given the fixed r).
- CFrd model: The hazard ratio is $\exp(\beta_{ill}) \cdot \mu(d) \cdot \exp(\beta r)$ that is nonproportional since owing to $d = t - r$ for $t > r$, it becomes $\exp(\beta_{ill}) \cdot \mu(t-r) \cdot \exp(\beta r)$, which depends on t through $\mu(t-r)$ given the fixed r, and thus proportionality is not assumed.
- CRr model: We need to distinguish the case with $r = 0$ from the case with $r > 0$. In the first case, we can write $d = t - 0 = t$ and $\lambda_{12}(d)$ can be written as $\lambda_{12}(t)$. Thus, the hazard ratio $\lambda_{12}(t)/\lambda_{02}(t)$ is equal to $\exp(\beta_{ill})$ for any $t > 0$, and proportionality is assumed in the clock forward scale. If $r > 0$, the hazard ratio is $\lambda_{02}(d) \cdot \exp(\beta_{ill}) \cdot \exp(\beta r)/\lambda_{02}(t)$ for $t > r$ and thus $\lambda_{02}(t-r) \cdot \exp(\beta_{ill}) \cdot \exp(\beta r)/\lambda_{02}(t)$ depends on t through $\lambda_{02}(t-r)/\lambda_{02}(t)$, given the fixed r.

All models can be extended including an explicit time-varying component (denoted by $\gamma(t)$) in the hazard ratio $\lambda_{12}(t)/\lambda_{02}(t)$ for $t > r$ to model with more flexibility the impact of the intermediate event (models identified by * in Tables 2 and 4-6).

If the Markov property can be justified, the clock forward scale is the natural choice even for $\lambda_{12}(t)$. The CFr (or CFr*) model can be applied and, as expected, the time to illness r will not have an impact.

If the semi-Markov property can be justified, the clock reset scale should be used for $\lambda_{12}(d)$, whereas the clock forward scale should be kept for $\lambda_{02}(t)$. The CRr (or CRr*) model can be applied and, as expected, the time to illness will not have an impact.

When neither the Markov nor the semi-Markov property holds (i.e., extended semi-Markov), the CRr (or CRr*) model provides the strategy for a suitable joint modeling of $\lambda_{02}(t)$ and $\lambda_{12}(r,d)$, enlightening the impact of the time to illness r on the

hazard of the absorbing event after illness. Of note, in the CFrd (and CFrd*) model the clock forward time scale is used even after the transition to the intermediate state through the common baseline $\lambda_{02}(t)$ and thus the impact of r solely on $\lambda_{12}(d)$ cannot be estimated.

5 | SIMULATION PROTOCOL

5.1 | Data generation

We generated 1,000 datasets of $n = 500$ observations, using the inversion method proposed by Bender, Augustin, and Blettner (2005), according to this strategy:

- Potential time to illness R, potential time of absorbing event without developing the intermediate event T_{02}, and potential time after the transition to the illness state D were simulated independently.
- The binary variable X indicating the development of the intermediate event was set equal to 1 if $R < T_{02}$ and 0 otherwise.
- The observed time since origin, denoted by T, was equal to $R + D$ if $X = 1$ and equal to T_{02} if $X = 0$.

The time to illness R was simulated according to an exponential distribution with parameter $\lambda = 0.5$. The times T_{02} and D were generated according to exponential and Weibull distributions representing constant or decreasing hazard behaviors of $\lambda_{02}(t)$ and $\lambda_{12}(t)$. More details on how the time after the transition D was generated are presented in the Appendix. At first, we did not consider the presence of censoring since we were only interested in the ability of the models to capture the effects.

Nine different scenarios were considered and summarized in Table A1: three under the Markov property, three under the semi-Markov property, and three under the extended semi-Markov property.

Markov data
Under the Markov property, hazards $\lambda_{02}(t)$ and $\lambda_{12}(t)$ are constant in M0, proportional and decreasing in M1, and $\lambda_{02}(t)$ is constant and $\lambda_{12}(t)$ is decreasing in M2, thus they are nonproportional.

Semi-Markov data
Under the semi-Markov property, hazards $\lambda_{02}(t)$ and $\lambda_{12}(d)$ are proportional and decreasing in SM1, and $\lambda_{02}(t)$ is constant and $\lambda_{12}(d)$ is decreasing in SM2, thus they are nonproportional; $\lambda_{02}(t)$ and $\lambda_{12}(d)$ are nonproportional and decreasing in SM3.

Extended semi-Markov data
In the extended semi-Markov scenarios for $r = 0$, $\lambda_{02}(t)$ and $\lambda_{12}(0, d)$ are identical to those in the corresponding semi-Markov scenarios, whereas for $r > 0$, $\lambda_{12}(r, d) = \lambda_{12}(0, d) \cdot \exp(\beta r)$, where β is fixed and equal to 0.5.

The hazard functions corresponding to the nine scenarios summarized in Table A1 are represented in Figure A1, where $\lambda_{12}(t)$ is displayed for three patients with times to illness $r = 0, 1, 2$. The ratio between $\lambda_{12}(t)$ and $\lambda_{02}(t)$ for $t \geq r$ is displayed in Figure A2.

The simulation protocol was extended to evaluate the performance of the models in the presence of the following additional issues:

1. independent censoring,
2. censoring dependent on an additional binary covariate,
3. nonmonotone hazard functions.

Simulations were carried out using the `R` software available at: http://cran.r-project.org/ (source code to reproduce data and results are available as Supporting Information).

5.2 | Model implementation

Subsample and full-sample models were estimated by the Poisson regression, including the baseline hazard as a parametric function and the length of the time interval as an offset (Iacobelli & Carstensen, 2013). The follow-up data were splitted in small intervals due to the fact that the Poisson regression assumes constant hazard rates in each time interval. Time scales were

TABLE 3 Simulation results of the subsample models in all scenarios

		$\beta = 0$			
Scenario	**Model**	**Est.**	**SE**	**% rejec.**	**Miss-spec.**
M0	CFr	0.001	0.035	5.4	
	CFd	−0.001	0.035	5.4	
M1	CFr	−0.024	0.114	4.4	
	CFd	0.024	0.114	4.4	
	CFr	0.565	0.119	97.7	✓
SM1	CFd	−0.565	0.119	97.7	✓
	CRr	−0.023	0.081	7.3	
		$\beta = 0.5$			
Scenario	**Model**	**Est.**	**SE**	**% rejec.**	**Miss-spec.**
	CFr	1.390	0.150	100.0	✓
ESM1	CFd	−1.390	0.150	100.0	✓
	CRr	0.489	0.125	99.0	

β is the coefficient of r in CFr and CRr models, and of d in CFd model.
Est. = mean of the estimates, SE = standard error of the estimates, % rejec. = percentage of rejected null hypotheses, Miss-spec. = if ✓ the model is miss-specified; M0 = Markov 0, M1 = Markov 1, SM1 = semi-Markov 1, ESM1 = extended semi-Markov 1; CFr = clock forward model with r as covariate, CFd = clock forward model with d as covariate, CRr = clock reset model with r as covariate.

TABLE 4 Simulation results of the full-sample models in the Markov scenarios (β is the coefficient of the time to illness r)

		$\beta = 0$			β_{ill}					
Scenario	**Model**	**Est.**	**SE**	**% rejec.**	**Est.**	**Bias**	**SE**	**% rejec.**	**% rejec.***	**Miss-spec.**
M0	CFr	0.001	0.034	5.5	1.103	0.004	0.145	100.0		
	CFr*	0.001	0.035	5.4	1.082	−0.017	0.455	64.9	6.2	
M1	CFr	−0.004	0.086	4.4	1.100	0.001	0.129	100.0		
	CFr*	−0.025	0.113	4.7	1.001	−0.098	0.615	44.1	6.0	
M2	CFr	−0.119	0.088	32.6	1.408	–	0.133	100.0		✓
	CFr*	−0.018	0.100	4.9	–	–	–	–	56.0	

*Likelihood ratio test.
Est. = mean of the estimates, SE = standard error of the estimates, % rejec. = percentage of rejected null hypotheses, Bias = difference between the mean of the estimates and the true value ($\beta_{ill} = 1.099$), Miss-spec. = if ✓ the model is miss-specified; M0 = Markov 0, M1 = Markov 1, M2 = Markov 2; CFr = clock forward model with r as covariate; models with * = models with a flexible effect of the intermediate event.

included with a flexible effect through natural splines, while covariates were included with log-linear or non-log-linear effects (d in CFrd model) through natural splines.

Since CFr (both in subsample and full-sample approaches), CFd, and CFrd models are based on the clock forward scale even for the transition hazard $\lambda_{12}(t)$, the time to illness r has to be included also as left truncation time for time measured after the intermediate event. In CRr models, the time after the intermediate event is directly measured in the clock reset scale d and there is no need of setting a left truncation time. In all full-sample models where modeling of $\lambda_{02}(t)$ is involved, an artificial right censoring is included at the time where the intermediate event occurs.

6 | RESULTS

Simulation results are presented in Tables 3-6, where some summary indicators are calculated. All the tables show the mean and the standard error of the coefficients estimates obtained in the 1,000 simulated samples and the percentage of simulated samples in which the coefficient of the time to illness is statistically significant ($p < .05$). The tables of the simulation results obtained in the full-sample models (Tables 4-6) also show the bias, calculated as the difference between the mean of the 1,000 coefficients estimates and the true values ($\beta_{ill} = 1.099$ when it has a linear effect, $\beta = 0.5$ in the extended semi-Markov scenarios) and the results of the likelihood ratio test to test the presence of a time-varying β_{ill}. In each table, we indicate whether the model is

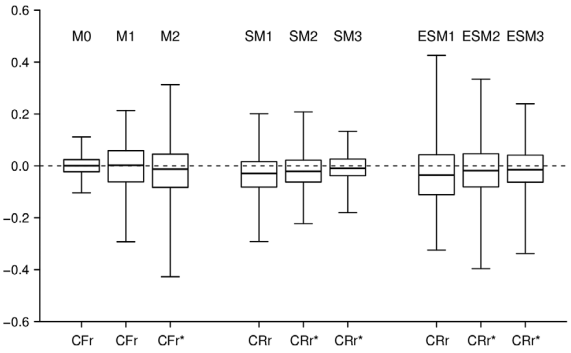

FIGURE 2 Bias of the estimates of the β coefficient of the time to illness obtained from the correctly specified models according to the scenario. M0 = Markov 0, M1 = Markov 1, M2 = Markov 2, SM1 = semi-Markov 1, SM2 = semi-Markov 2, SM3 = semi-Markov 3, ESM1 = extended semi-Markov 1, ESM2 = extended semi-Markov 2, ESM3 = extended semi-Markov 3; CFr = clock forward model with r as covariate, CRr = clock reset model with r as covariate, models with * = models with a flexible effect of the intermediate event

miss-specified according to the considered scenario. The results dealing with the extension of the simulation protocol (censoring and nonmonotone hazards) are available as Supporting Information.

6.1 | Subsample models

Simulation results presented in Table 3 show the application of the sub-sample models to check the properties of the transition hazard $\lambda_{12}(t)$ to assess the role of the time to illness r. Only four scenarios (M0, M1, SM1, and ESM1) were taken into account because the subsample models approach considers only $\lambda_{12}(t)$, which is exactly the same within M1 and M2, SM1 and SM2, ESM1 and ESM2, and it has a similar behavior (i.e., monotone decreasing) within SM2 and SM3, and ESM2 and ESM3.

We applied first the CFr and CFd clock forward models of Table 1 to check the Markov property. In case of violation of this property, we subsequently applied the CRr clock reset model to check the semi-Markov property and to estimate the impact of the time to illness r when the property was not fulfilled (extended semi-Markov).

CFr or CFd model to check Markov property

In Table 3, one can notice that CFr and CFd models lead to the same absolute value of the estimate of β, as explained in Section 3.

In M0 and M1, none of the two models showed a significant impact of the time to illness r, with a rejection fraction of $H_0 : \beta = 0$ around the nominal 5%. This was expected since data were generated under the Markov property.

In SM1, a significant impact of the time to illness r is shown with a high rejection fraction of $H_0 : \beta = 0$ in both models. This was expected since data were generated under the semi-Markov property and has to be interpreted as violation from the Markov assumption only and not as an impact of the time to illness r on $\lambda_{12}(t)$. This impact is an artifact due to the time after illness d and not to a different shape of $\lambda_{12}(d)$ and can be thought as a spurious effect of r. This spurious effect can be noticed in Figure A1 (panel C), considering that at time t the value of $\lambda_{12}(t)$ becomes higher as the time to illness r increases. In ESM1, CFr and CFd models showed again a significant impact of r (or d) with a high rejection fraction of $H_0 : \beta = 0$, indicating again the violation of the Markov assumption only. In fact, even in this case a spurious effect of r is still present, being the corresponding β clearly overestimated.

CRr model to check semi-Markov property

In SM1 and ESM1 scenarios, where the violation of the Markov property was present, the CRr clock reset model was applied to check the semi-Markov property. In Table 3, one can notice in SM1 a nonsignificant impact of the time to illness r with a rejection fraction of $H_0 : \beta = 0$ around the nominal value, indicating the validity of the property. By contrast, in ESM1, departure from the semi-Markov property is demonstrated by the high rejection fraction of $H_0 : \beta = 0$. Of note, in both scenarios the model provided an unbiased estimate of β addressing properly the role of r.

6.2 | Full-sample models

Simulation results presented in Tables 4-6 show the application of the full-sample models on all the nine scenarios described in the simulation protocol. Figures 2 and 3 report the distribution of the simulation estimates of β and β_{ill} coefficients, respectively, of the correctly specified models in each scenario.

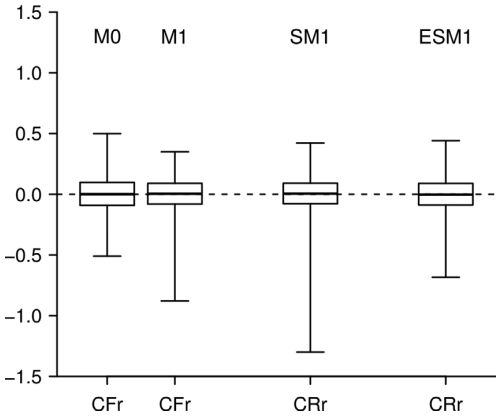

FIGURE 3 Bias of the estimates of the β_{ill} coefficient of illness obtained from the correctly specified proportional hazards models according to the scenario. M0 = Markov 0, M1 = Markov 1, SM1 = semi-Markov 1, ESM1 = extended semi-Markov 1; CFr = clock forward model with r as covariate, CRr = clock reset model with r as covariate

TABLE 5 Simulation results of the full-sample models in the semi-Markov scenarios (β is the coefficient of the time to illness r)

		$\beta = 0$			β_{ill}					
Scenario	**Model**	**Est.**	**SE**	**% rejec.**	**Est.**	**Bias**	**SE**	**% rejec.**	**% rejec.***	**Miss-spec.**
SM1	CFr	0.317	0.101	90.2	1.331	0.232	0.143	100.0		✓
	CFr*	0.565	0.118	97.4	1.365	0.266	0.595	68.7	49.6	✓
	CFrd	0.230	0.110	57.4	1.966	0.867	0.252	100.0		✓
	CFrd*	0.186	0.229	13.5	0.979	−0.120	1.113	31.0	24.5	✓
	CRr	−0.029	0.076	7.7	1.101	0.002	0.130	99.9		
	CRr*	−0.029	0.078	8.4	1.088	−0.011	0.178	100.0	6.7	
SM2	CFr	0.175	0.099	55.8	1.677	–	0.157	100.0		✓
	CFr*	0.572	0.103	99.7	–	–	–	–	99.1	✓
	CFrd	−0.022	0.096	5.7	2.601	–	0.233	100.0		✓
	CFrd*	−0.009	0.196	5.1	–	–	–	–	4.8	✓
	CRr	−0.022	0.077	13.0	1.625	–	0.142	100.0		✓
	CRr*	−0.023	0.067	8.0	–	–	–	–	90.2	
SM3	CFrd	0.325	0.123	84.0	0.837	–	0.278	85.6		✓
	CFrd*	−0.051	0.628	9.1	–	–	–	–	40.5	✓
	CRr	−0.007	0.044	3.2	0.183	–	0.128	39.7		✓
	CRr*	−0.007	0.048	5.6	–	–	–	–	97.3	

*Likelihood ratio test.

Est. = mean of the estimates, SE = standard error of the estimates, % rejec. = percentage of rejected null hypotheses, Bias = difference between the mean of the estimates and the true value ($\beta_{ill} = 1.099$), Miss-spec. = if ✓ the model is miss-specified; SM1 = semi-Markov 1, SM2 = semi-Markov 2, SM3 = semi-Markov 3; CFr = clock forward model with r as covariate, CFrd = clock forward model with r and d as covariates, CRr = clock reset model with r as covariate; models with * = models with a flexible effect of the intermediate event.

CFr and CFr models*

In Markov scenarios (Table 4), we assumed the Markov property previously verified through the subsample models strategy and thus we applied the CFr and CFr* clock forward models. Nevertheless, the covariate r was included to check whether even in a full-sample model its null coefficient was maintained. In M0 and M1, the estimates of the coefficient β of r obtained by the CFr model are close to 0 with a rejection fraction of $H_0 : \beta = 0$ around the nominal value. In M2, the rejection fraction of $H_0 : \beta = 0$ is 32.6%. This is due to the fact that in M2 the hazard ratio $\lambda_{12}(t)/\lambda_{02}(t)$ is decreasing (see Figure A2, panel B), whereas the model assumes proportional hazards by the β_{ill} parameter. This generates a spurious effect of r only due to nonproportionality, which however disappears when the time-varying nature of the hazard ratio is taken into account by the CFr* model. The presence of censoring (independent or dependent from a covariate) and nonmonotone hazards does not affect the validity of these considerations (see the results in the Supporting Information).

CFr model was also applied in SM1 and SM2 (Table 5) to corroborate the idea that a clock forward model in a non-Markov scenario generates a spurious effect of r due to the fact that on the clock forward time scale subjects with different r have also

TABLE 6 Simulation results of the full-sample models in the extended semi-Markov scenarios (β is the coefficient of the time to illness r)

		$\beta = 0.5$				β_{ill}					
Scenario	**Model**	**Est.**	**Bias**	**SE**	**% rejec.**	**Est.**	**Bias**	**SE**	**% rejec.**	**% rejec.***	**Miss-spec.**
ESM1	CFrd	0.813	0.313	0.147	100.0	1.983	0.884	0.263	100.0		✓
	CFrd*	0.775	0.275	0.434	55.0	1.011	−0.088	1.100	31.5	20.5	✓
	CRr	0.468	−0.032	0.112	99.4	1.098	−0.001	0.134	100.0		
	CRr*	0.477	−0.023	0.116	99.3	1.033	−0.066	0.270	98.4	7.0	
ESM2	CFrd	0.484	−0.016	0.127	98.2	2.663	–	0.251	100.0		✓
	CFrd*	0.525	0.025	0.379	36.2	–	–	–	–	5.2	✓
	CRr	0.551	0.051	0.109	99.9	1.660	–	0.154	100.0		✓
	CRr*	0.487	−0.013	0.099	99.9	–	–	–	–	84.7	
ESM3	CFrd	0.879	0.379	0.119	100.0	0.895	–	0.264	92.2		✓
	CFrd*	0.755	0.255	0.890	50.4	–	–	–	–	34.9	✓
	CRr	0.404	−0.096	0.062	100.0	0.211	–	0.130	44.8		✓
	CRr*	0.488	−0.012	0.075	100.0	–	–	–	–	97.9	

*Likelihood ratio test.

Est. = mean of the estimates, SE = standard error of the estimates, % rejec. = percentage of rejected null hypotheses, Bias = difference between the mean of the estimates and the true value ($\beta = 0.5$, $\beta_{ill} = 1.099$), Miss-spec. = if ✓ the model is miss-specified; ESM1 = extended semi-Markov 1, ESM2 = extended semi-Markov 2, ESM3 = extended semi-Markov 3; CFrd = clock forward model with r and d as covariates, CRr = clock reset model with r as covariate; models with * = models with a flexible effect of the intermediate event.

different times after illness. This is evident also in the flexible version of the model, CFr*. In SM1, the constant hazard ratio $\lambda_{12}(d)/\lambda_{02}(t)$ for $r = 0$, defined by the β_{ill} parameter, is clearly overestimated by CFr model. Results of this model in SM3 and ESM1, ESM2, and ESM3 scenarios are omitted since they lead to similar conclusions.

CFrd and CFrd models*

CFrd and CFrd* models were used in all semi-Markov (Table 5) and extended semi-Markov (Table 6) scenarios to investigate whether these approaches would enable to remove the spurious effect of r due to the time after the transition d, since they include d as covariate on top of r.

When applying CFrd model in SM1, a spurious effect of r is still present and the constant hazard ratio $\lambda_{12}(d)/\lambda_{02}(t)$ for $r = 0$, defined by the β_{ill} parameter, is overestimated. The spurious effect of r is observed also in the CFrd* model, where the β_{ill} parameter is underestimated. In SM2, the time-varying nature of $\lambda_{12}(d)/\lambda_{02}(t)$, in time t for any r, is not captured even by the CFrd* model since the rejection fraction is only 5.8%. Of note, the spurious effect of r disappears in both CFrd and CFrd* and the rejection fractions of $H_0 : \beta = 0$ are around the nominal value. This is due to the peculiarity of SM2 scenario, where $\lambda_{02}(t)$ is constant. In fact in SM3, where $\lambda_{02}(t)$ is decreasing, the spurious effect of r is again observed. In this scenario, the time-varying nature of $\lambda_{12}(d)/\lambda_{02}(t)$, in time t for any r, is captured by the CFrd* model but with a rejection fraction of only 40.5%.

In ESM1, the effect of r and the constant hazard ratio $\lambda_{12}(r, d)/\lambda_{02}(t) = \exp(\beta_{ill})$ for $r = 0$ are still overestimated by CFrd model. In ESM2, the time-varying nature of $\lambda_{12}(r, d)/\lambda_{02}(t)$ for any r is not captured even by the CFrd* model where the rejection fraction is only 5.2%. Of note, the effect of r is adequately estimated in both CFrd and CFrd*, again owing to the peculiarity of ESM2 where $\lambda_{02}(t)$ is constant. In fact, in ESM3, where $\lambda_{02}(t)$ is decreasing, the spurious effect of r is again observed. The time-varying nature of $\lambda_{12}(r, d)/\lambda_{02}(t)$, in time t for any r, is captured by the CFrd* model but with a rejection fraction of only 34.9%.

CRr and CRr models*

CRr and CRr* models were used in all semi-Markov (Table 5) and extended semi-Markov (Table 6) scenarios to investigate whether these approaches would enable to remove the spurious effect of r, since they include the clock forward time scale only for patients in state 0 and the clock reset time scale for patients in state 1 and r as covariate. In all scenarios, the coefficient of r is properly estimated: in the semi-Markov scenarios, the coefficient of r is close to 0 with a rejection fraction for $H_0 : \beta = 0$ between 3.2% and 13%, while in the extended semi-Markov scenarios the effect of r is recognized and estimated with bias lower than 0.096 in absolute value and very high rejection fractions. In SM1 and ESM1, the constant hazard ratio $\lambda_{12}(d)/\lambda_{02}(t) = \exp(\beta_{ill})$ is correctly estimated with very high rejection fractions, whereas in the remaining scenarios, the time-dependent nature

TABLE 7 Results of the subsample Poisson models to test the Markov and the semi-Markov assumptions on the data of the Danish National Diabetes Register

Model	Variable	Est.	SE	z	p-value
CFr	Intercept	−2.448	0.130	−18.769	<.001
	$f(t)1$	−0.626	0.245	−2.553	.011
	$f(t)2$	−2.691	0.414	−6.501	<.001
	$f(t)3$	−1.774	0.352	−5.034	<.001
	Time to Ins. prescr.	0.184	0.022	8.514	<.001
CRr	Intercept	−2.269	0.103	−22.011	<.001
	$f(d)1$	0.146	0.318	0.459	.646
	$f(d)2$	−2.942	0.399	−7.371	<.001
	$f(d)3$	−2.123	0.624	−3.402	<.001
	Time to Ins. prescr.	0.060	0.014	4.163	<.001

$f(\cdot)$ is a natural spline with 3 degrees of freedom. Ins. prescr. = second insulin prescription, Time to Ins. prescr. = time to the second insulin prescription (in years); Est. = estimate of the coefficient, SE = standard error.

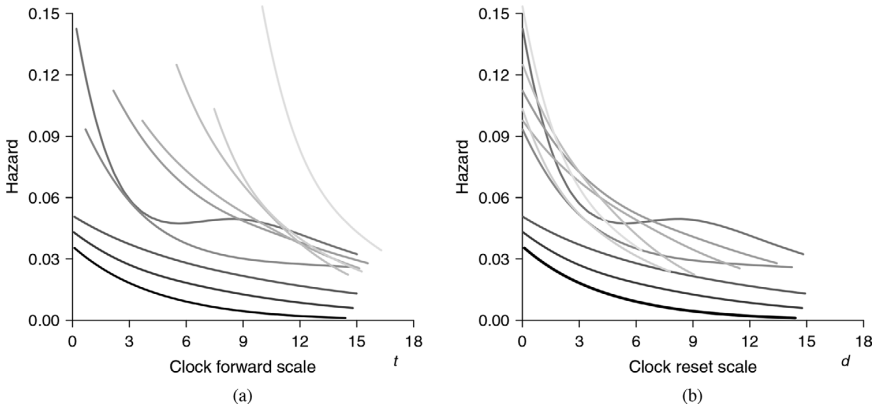

FIGURE 4 Plot of the smoothed estimate of the hazard function in subgroups of patients with similar times to insuline prescription (according to deciles) against time measured in the clock forward scale (a) and in the clock reset scale (b). Black curves correspond to patients with smaller times to insuline prescription, whereas gray curves correspond to patients with larger times to insuline prescription

of $\lambda_{12}(d)/\lambda_{02}(t)$ for any r is captured. The presence of censoring (independent or dependent from a covariate) and non-monotone hazards does not affect the validity of these considerations (see the results in the Supporting Information).

7 | APPLICATION TO THE DANISH NATIONAL DIABETES REGISTER

We analyzed survival data of a random sample of 10,000 patients among those with date of diagnosis of diabetes after 1995 from the Danish National Diabetes Register, available in the `Epi` package of the `R` software (Carstensen, Kristensen, Ottosen, & Borch-Johnsen, 2008). We considered an illness-death model with the inclusion in the register as the initial state, the second prescription of insulin as the intermediate state and death as the final state.

First, we applied the CFr and CRr models on the subsample of patients with the second prescription of insulin ($n = 1,791$) to the test the properties of Markov and semi-Markov, respectively. Results are presented in Table 7. Since in both models the coefficient of the time to the second prescription of insulin is statistically different from 0, we cannot assume either the Markov or the semi-Markov property, so we can conclude that data satisfy the extended semi-Markov property. The coefficient is positive, meaning that a delayed second insulin prescription increases the subsequent hazard of death. The Markov and semi-Markov properties are checked also graphically by plotting a smoothed estimate of the hazard function in subgroups of patients with similar times to insuline prescription against time in the clock forward (Figure 4a) and clock reset (Figure 4b) scales, respectively. The curves in Figure 4a are not overlapping, so the data cannot satisfy the Markov property. Since also in Figure 4b

TABLE 8 Results of the full-sample Poisson models on the data of the Danish National Diabetes Register

Model	Variable	Est.	SE	z	p-value
CRr	Intercept	−2.949	0.078	−37.689	<.001
	$f(u)1$	0.157	0.113	1.393	.164
	$f(u)2$	−0.378	0.198	−1.904	.057
	$f(u)3$	0.057	0.178	0.321	.748
	Ins. prescr.	−0.120	0.073	−1.654	.098
	Time to Ins. prescr.	0.089	0.016	5.575	<.001
	Sex (female vs. male)	−0.100	0.044	−2.288	.022
CRr*	Intercept	−3.078	0.091	−33.998	<.001
	$f(u)1$	0.178	0.121	1.467	.142
	$f(u)2$	0.004	0.226	0.017	.986
	$f(u)3$	0.322	0.186	1.734	.083
	Ins. prescr.	0.435	0.174	2.501	.012
	$f(u)1$*Ins. prescr.	−0.060	0.345	−0.173	.862
	$f(u)2$*Ins. prescr.	−2.055	0.538	−3.819	<.001
	$f(u)3$*Ins. prescr.	−2.114	0.677	−3.121	.002
	Time to Ins. prescr.	0.065	0.017	3.903	<.001
	Sex (female vs. male)	−0.098	0.044	−2.236	.026

$f(\cdot)$ is a natural spline with 3 degrees of freedom; u is equal to the time from inclusion in the register (clock forward scale) before the transition to the second insulin prescription and equal to the time from the second insulin prescription (clock reset scale) after that transition, Ins. prescr. = second insulin prescription, Time to Ins. prescr. = time to the second insulin prescription (in years); Est. = estimate of the coefficient, SE = standard error; models with * = models with a flexible effect of the intermediate event.

the curve are far from being overlapping, an effect of the time to insuline prescription can be hypothesized, therefore the extended semi-Markov property is reasonable. From the analysis of the Schoenfeld residuals in the clock reset time scale, the proportional hazards assumption is reasonable ($p = .125$). Slight deviations from the log-linearity are also apparent in the martingale residuals (not shown). This would suggest the inclusion of a flexible effect of r instead of the simple linear effect we included.

Following our strategy, we then applied the CRr full-sample model to investigate simultaneously the impact of the second prescription of insulin and again the time to this intermediate event on the incidence of mortality. Results are reported in Table 8. In the CRr model, the coefficient of insulin prescription is negative and marginally significant ($p < .1$), suggesting an overall protective effect of the treatment. Confirming the results of the CRr subsample model, the effect of the time to second insulin prescription is significantly greater than zero. In order to obtain a better model fitting, one can implement the analogous CRr* model with a time-varying effect of insulin prescription. This was achieved by applying a spline function of time to the coefficient of the intermediate event. Although the interpretability of the effect of the intermediate event becomes difficult, we have a better estimate of the impact of the time to second insulin prescription, which is again significantly positive and very similar to the subsample CRr model estimate. In both full-sample models, sex has an estimated similar impact on the hazard of death.

8 | DISCUSSION

In this work, we proposed a strategy (described in the flowchart of Figure 5) to model hazard functions for illness-death data with the aim of quantifying the impact of time to illness, and the transition to illness, on the subsequent hazard of the absorbing event. The strategy involves two phases: (1) investigate whether the data will reflect the Markov, semi-Markov, or extended semi-Markov properties using subsample models; (2) quantify the impact of the transition to illness on top of the impact of time to illness in a full-sample model. We discuss in detail the two phases.

About phase 1, it can be noticed that often in multistate literature, for sake of simplicity, the Markov property is commonly assumed (Andersen & Perme, 2008; Lauseker et al., 2015; Putter et al., 2007). From the clinical point of view, however, it is difficult to justify that the hazard a patient experiences after illness is neither influenced by the time to illness (semi-Markov) nor by the time after illness (extended semi-Markov) (Andersen et al., 2000; Eulenburg et al., 2015; Meira-Machado, de Una-Alvarez, Cadarso-Suarez, & Andersen, 2009). From the methodological point of view, it is clearer to start from the Markov

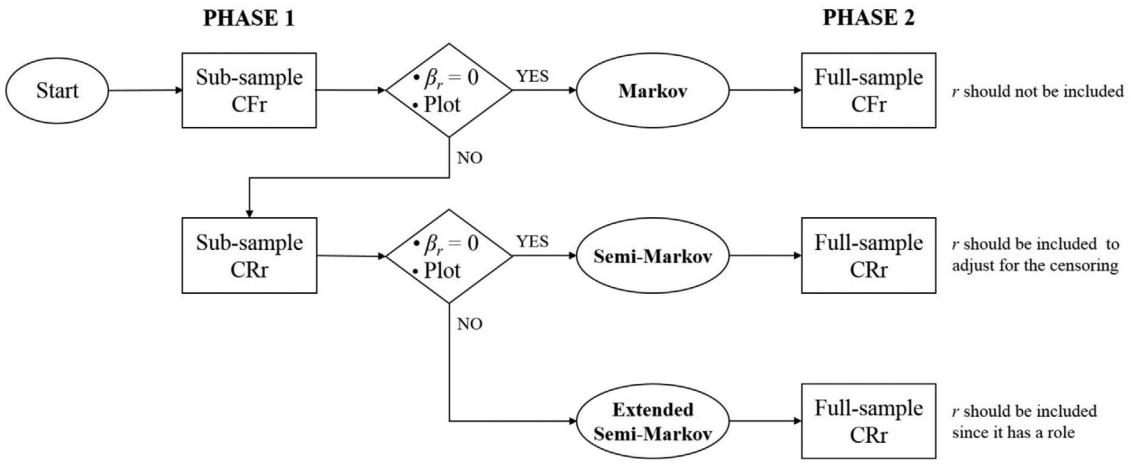

FIGURE 5 Graphical representation of the strategy to use when analyzing illness-death data. Plot = plot of the smoothed estimate of the hazard function in subgroups of patients with similar times to illness against time in the clock forward scale (to check the Markov property) or in the clock reset scale (to check the semi-Markov property)

assumption as a reference model. If the assumption is violated, we then suggest to check if either the semi-Markov or the extended semi-Markov property holds.

Since the Markov, semi-Markov, and extended semi-Markov properties concern the behavior of the hazard function after illness, it is natural to investigate them by analyzing the subsample of patients who experienced illness with the sequential use of the CFr (or CFd) and CRr models we presented. The use of a model-based approach, however, relies on the proportionality of the hazards and on a possible log-linear effect of r. These assumptions could be checked both through residuals analysis and naive graphical methods, as presented in Section 3. Violations of these assumptions could lower the power on testing the model coefficients, where the null hypothesis of the coefficient equal to zero represents the presence of Markov or semi-Markov data. The graphical check we described enables also to verify directly if the Markov property is tenable and, in negative case, if the semi-Markov property is tenable.

In the case of Markov data, the use of the clock forward time scale is the natural way to measure the follow-up time. The clock reset scale should be considered in case of non-Markov data, since forcing to use the clock forward scale will result in a spurious effect of the time to illness, due to the time after illness and not to a different shape of the hazard function after illness. One may note that it is only in an extended semi-Markov scenario that time to illness has an influence on the subsequent hazard, which is kept by the use of the clock reset time scale in the CRr model. The motivating data on the Danish National Diabetes Register were consistent with an extended semi-Markov scenario with proportionality of the hazards but deviated from the assumption of log-linear effect. These deviations were not accounted for the sake of simplicity.

In phase 2, the choice of the full-sample model is guided by the results of phase 1. The model in the clock forward scale with the time-varying covariates representing the transition to illness and the time to illness is suitable to be used in Markov data and, as we expected, it will not reveal an impact of time to illness. Nonproportionality between the hazards may be handled through a time-varying effect of the transition to illness. The use of a model involving the clock forward scale after the transition to illness in non-Markov data will result again in a spurious effect of the time to illness due to the time after illness. In addition, the estimate of the impact of the transition to illness is biased even in simple scenarios (e.g., SM1 in Table 5). One may note that this issue holds even for the more complex CFrd model, where the time after illness is used as a time-varying covariate in a model involving the clock forward time scale as background. In fact, a model including the clock reset time scale for the follow-up time after illness on top of the original time scale, which will be still considered as baseline, will necessarily imply that the time to illness is included as a time-varying covariate. By contrast, if the original time scale is not considered in the model formulation to measure the follow-up time after illness, the time to illness can be included as a fixed covariate. Given its flexibility, the CFrd model may be used for prediction purposes, such as in Rebora et al. (2015).

In the case of non-Markov data, when the goal is to estimate the impact of time to illness on the hazard from state 1 to state 2, only the CRr model, which uses the clock forward scale to measure time before illness and the clock reset scale for time after the intermediate event, should be considered. This was done in our motivating data. This approach captures the real effect of the time to illness, which is null if data satisfy the semi-Markov property, on the subsequent hazard. It is worth to note that, even in

this scenario, the inclusion in the model of the fixed time to illness covariate could be useful to relax the censoring assumption (i.e., the model is valid even under censoring dependent on covariates). In addition, the impact of the transition to illness is correctly quantified, also in complex scenarios (e.g., ESM3 in Table 6).

Finally, we point out that even under totally independent censoring (e.g., administrative), the CRr model induces a more strict censoring assumption. In fact, adopting the clock reset scale to measure time after illness implies that censoring is dependent from time to illness. However, the problem is tackled by simply including in the model the fixed time to illness covariate, even when it has no effect (semi-Markov scenario).

ACKNOWLEDGMENTS

DPB is supported by a fellowship of the Associazione Italiana per la Ricerca sul Cancro (AIRC). PR was supported by the grant SIR RBSI14LOVD of the Italian Ministry of Education, University and Research. This research was partially supported by AIRC (2013-14634).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Elena Tassistro https://orcid.org/0000-0001-5166-7305
Davide Paolo Bernasconi https://orcid.org/0000-0001-8771-5428
Paola Rebora https://orcid.org/0000-0003-0606-5852
Maria Grazia Valsecchi https://orcid.org/0000-0001-5574-3504
Laura Antolini https://orcid.org/0000-0003-1040-4191

REFERENCES

Andersen, P. K., Esbjerg, S., & Sorensen, T. I. (2000). Multi-state models for bleeding episodes and mortality in liver cirrhosis. *Statistics in Medicine*, *19*, 587–599.

Andersen, P. K., & Perme, M. P. (2008). Inference for outcome probabilities in multi-state models. *Lifetime Data Analysis*, *14*, 405–431.

Bender, R., Augustin, T., & Blettner, M. (2005). Generating survival times to simulate Cox proportional hazards models. *Statistics in Medicine*, *24*, 1713–1723.

Bernasconi, D. P., Rebora, P., Iacobelli, S., Valsecchi, M. G., & Antolini, L. (2016). Survival probabilities with time-dependent treatment indicator: Quantities and non-parametric estimators. *Statistics in Medicine*, *35*, 1032–1048.

Beyersmann, J., Allignol, A., & Schumacher, M. (2012). *Competing risks and multistate models with R*. New York, NY: Springer.

Carstensen, B., Kristensen, J. K., Ottosen, P., & Borch-Johnsen, K. (2008). The Danish National Diabetes Register: Trends in incidence, prevalence and mortality. *Diabetologia*, *51*, 2187–2196.

Eulenburg, C., Mahner, S., Woelber, L., & Wegscheider, K. (2015). A systematic model specification procedure for an illness-death model without recovery. *PLoS ONE*, *10*, e0123489.

Iacobelli, S., & Carstensen, B. (2013). Multiple time scales in multi-state models. *Statistics in Medicine*, *32*, 5315–5327.

Lauseker, M., Hasford, J., Hoffmann, V. S., Müller, M. C., Hehlmann, R., & Pfirrmann, M. (2015). A multi-state model approach for prediction in chronic myeloid leukaemia. *Annals of Hematology*, *94*, 919–927.

Meier-Hirmer, C., & Schumacher, M. (2013). Multi-state model for studying an intermediate event using time-dependent covariates: Application to breast cancer. *BMC Medical Research Methodology*, *13*, 80.

Meira-Machado, L., de Una-Alvarez, J., Cadarso-Suarez, C., & Andersen, P. K. (2009). Multistate models for the analysis of time-to-event data. *Statistical Methods Medical Research*, *18*, 195–222.

Putter, H., Fiocco, M., & Geskus, R. B., (2007). Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine*, *26*, 2389–2430.

Rebora, P., Galimberti, S., & Valsecchi, M. G., (2015). Using multiple timescale models for the evaluation of a time-dependent treatment. *Statistics in Medicine*, *34*, 3648–3660.

Rebora, P., Salim, A., & Reilly, M. (2014). bshazard: A flexible tool for nonparametric smoothing of the hazard function. *R Journal*, *6/2*, 114–122.

SUPPORTING INFORMATION

Additional Supporting Information including source code to reproduce the results may be found online in the supporting information tab for this article.

How to cite this article: Tassistro E, Bernasconi DP, Rebora P, Valsecchi MG, Antolini L. Modeling the hazard of transition into the absorbing state in the illness-death model. *Biometrical Journal*. 2020;62:836–851. https://doi.org/10.1002/bimj.201800267

APPENDIX

In M1 and M2 scenarios, we simulated the time after illness D as follows.

TABLE A1 Details on the simulation protocol of the Markov, semi-Markov and extended semi-Markov scenarios ($\beta_{ill} = \log 3$, $\beta = 0.5$)

Scenario	T_{02}	D
M0	$Exp(0.1)$	$Exp(0.1e^{\beta_{ill}})$
M1	$Weib(0.6, 0.7)$	$Weib(0.6e^{\beta_{ill}/0.7}, 0.7)$*
M2	$Exp(0.4)$	$Weib(0.6e^{\log 3/0.7}, 0.7)$*
SM1	$Weib(0.6, 0.7)$	$Weib(0.6e^{\beta_{ill}/0.7}, 0.7)$
SM2	$Exp(0.4)$	$Weib(0.6e^{\log 3/0.7}, 0.7)$
SM3	$Weib(0.4, 0.5)$	$Weib(0.6, 0.7)$
ESM1	$Weib(0.6, 0.7)$	$Weib(0.6e^{\beta_{ill}/0.7}e^{\beta \cdot r/0.7}, 0.7)$
ESM2	$Exp(0.4)$	$Weib(0.6e^{\log 3/0.7}e^{\beta \cdot r/0.7}, 0.7)$
ESM3	$Weib(0.4, 0.5)$	$Weib(0.6e^{\beta \cdot r/0.7}, 0.7)$

*The time after illness D is simulated from a Weibull distribution as described below.
M0 = Markov 0, M1 = Markov 1, M2 = Markov 2; SM1 = semi-Markov 1, SM2 = semi-Markov 2, SM3 = semi-Markov 3; ESM1 = extended semi-Markov 1, ESM2 = extended semi-Markov 2, ESM3 = extended semi-Markov 3.

Since patients can die because of the development of illness only after the development of illness, the survival function was calculated, accounting for the delayed entry at time r, integrating the hazard function on the time interval $[r, t]$, where t is the time at which the risk is evaluated. The hazard at time t, given that the patient visited the intermediate event, follows a Weibull distribution. The survival function of a generic time $T \sim \text{Weibull}(\lambda, p)$, given an initial time $R = r$, is

$$\begin{aligned} S(t) &= P(T > t | R = r) = P(T_{\cdot|R} > t) \\ &= \exp\left(-\int_r^t \lambda(u)du\right) \\ &= \exp\left(-\int_r^t p\lambda^p u^{p-1}du\right) \\ &= \exp(-(\lambda^p t^p - \lambda^p r^p)) \\ &= \exp(-\lambda^p(t^p - r^p)). \end{aligned}$$

From $S(t)$, we can calculate the cumulative distribution function $F(t)$ and solve for t

$$\begin{aligned} F(t) = 1 - S(t) &= 1 - \exp(-\lambda^p(t^p - r^p)) \\ log(1 - F(t)) &= -\lambda^p(t^p - r^p) \\ \left(-\frac{log(1 - F(t))}{\lambda^p} + r^p\right)^{1/p} &= t. \end{aligned} \tag{A1}$$

With the inversion method proposed by Bender et al. (2005), we simulated the $F(t)$ from a uniform distribution [0,1] and substituted the obtained value in the formula (A1). We found a value for the time of interest $T_{\cdot|R}$. Finally, the time after the transition D is calculated as a difference between the time t and the time to illness R previously generated in order to account for the delayed entry and the dependence of the time after illness from the time to illness.

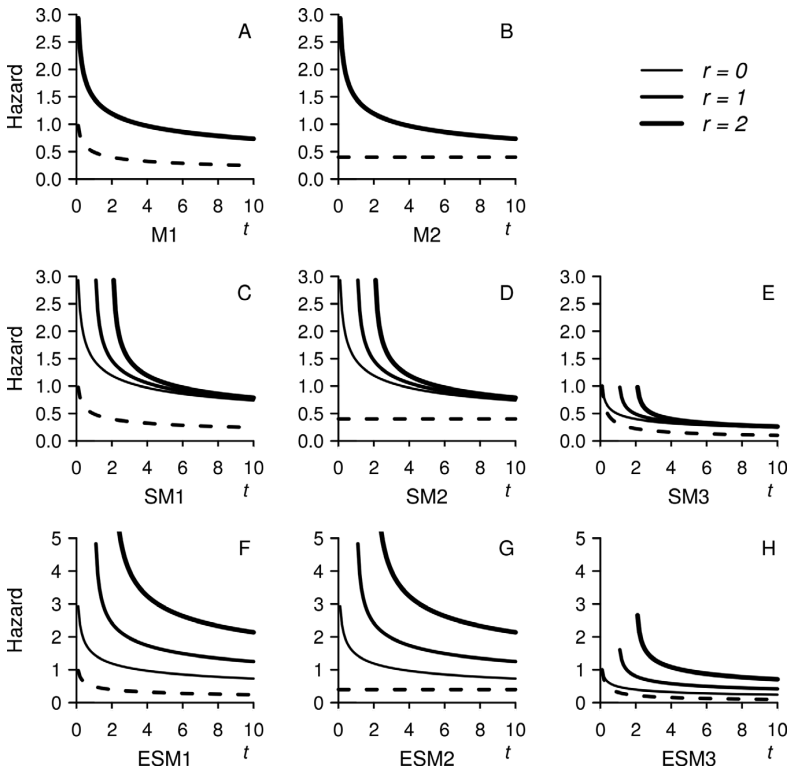

FIGURE A1 Hazards $\lambda_{02}(t)$ (dashed line) and $\lambda_{12}(t)$ (solid line) in all scenarios for three patients with times to illness $r = 0, 1, 2$. On the x-axis is represented the time from origin t (*clock forward time scale*), and on the y-axis the transition hazard to the absorbing state. M1 = Markov 1, M2 = Markov 2; SM1 = semi-Markov 1, SM2 = semi-Markov 2, SM3 = semi-Markov 3; ESM1 = extended semi-Markov 1, ESM2 = extended semi-Markov 2, ESM3 = extended semi-Markov 3

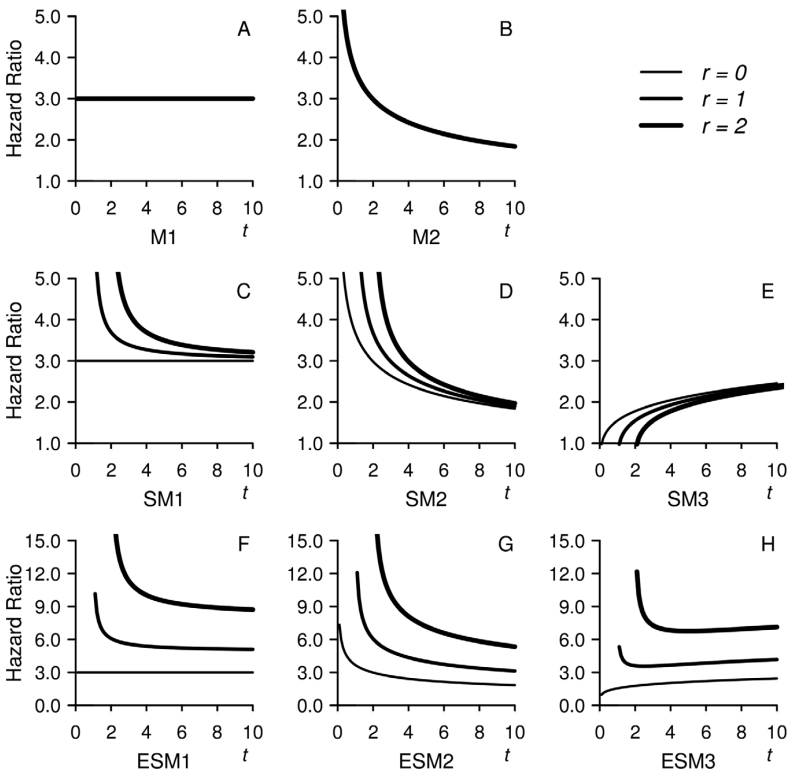

FIGURE A2 Hazard ratio between $\lambda_{12}(t)$ and $\lambda_{02}(t)$ for $t \geq r$ in all scenarios for three patients with times to illness $r = 0, 1, 2$. On the x-axis is represented the time from origin (*clock forward time scale*), and on the y-axis the hazard ratio. M1 = Markov 1, M2 = Markov 2; SM1 = semi-Markov 1, SM2 = semi-Markov 2, SM3 = semi-Markov 3; ESM1 = extended semi-Markov 1, ESM2 = extended semi-Markov 2, ESM3 = extended semi-Markov 3